

**Lexikogrammatische Besonderheiten des Englischen im
World Wide Web: eine formale und funktionale Analyse
von Mehrworteinheiten in verschiedenen Web-Genres**

**Inaugural-Dissertation zur Erlangung des Doktorgrades
der Philosophie der Justus-Liebig-Universität Gießen**

**vorgelegt von
Sven Saage
Liebigstr. 55
35392 Gießen**

Gießen 2013

Dekan: Prof. Dr. Magnus Huber

1. Berichterstatter: Prof. Dr. Joybrato Mukherjee
2. Berichterstatter: Prof. Dr. Magnus Huber

Tag der Disputation: 17. Juni 2013

Inhaltsverzeichnis

1. Einleitung	1
2. World Englishes und Lingua Franca-Sprachgebrauch	6
2.1. <i>World Englishes</i>	6
2.1.1. Die Entwicklung des Englischen zur Weltsprache	6
2.1.2. <i>World English</i> -Modelle	10
2.1.2.1. Wie lässt sich <i>ELF</i> in <i>World English</i> -Modellen verorten?	17
2.1.2.2. Ein neues Modell von <i>English as an International Language</i>	19
2.2. Englisch als Lingua Franca-Konzeption	21
2.2.1. <i>English as a Lingua Franca</i> vs. <i>English as an International Language</i>	23
2.2.2. Befindet sich <i>ELF</i> auf dem Weg, stabile sprachliche Merkmale auszubilden?	25
2.3. Zusammenfassung	30
3. Das World Wide Web und Sprache	32
3.1. Die Entwicklung des World Wide Web	32
3.2. Das <i>Netspeak</i> -Konzept	35
3.3. Der Begriff des Mediums	40
3.4. Mündlichkeit und Schriftlichkeit	41
3.4.1. Das Nähe-Distanz-Modell von Koch / Oesterreicher	44
3.4.1.1. Textsorte, Kommunikationsform, Register und Genre	47
3.4.1.2. Das Nähe-Distanz-Modell und Websprache	49
3.4.2. Konzeptionelle Mündlichkeit / Schriftlichkeit von Mehrworteinheiten	50
3.5. Mediale und sprachliche Merkmale des Webs.....	51
3.5.1 Mediale Merkmale	51
3.5.2. Sprachliche Merkmale.....	54
3.5.3. Merkmale der analysierten Web-Genres	57
3.5.3.1. Foren	57
3.5.3.2. Blogs	61
3.5.3.3. Homepages	64
3.5.3.4. Wikis	65
3.5.3.5. Eine Modellierung der Web-Genres	68
3.6. Zusammenfassung	72
4. Mehrworteinheiten	74
4.1. Die Bewertung von Mehrworteinheiten	74
4.2. Das Konzept der <i>Formulaic Language</i>	76
4.3. Kriterien zur Unterscheidung von Mehrworteinheiten	80
4.3.1. Kompositionalität und holistische Verarbeitung	80
4.3.2. Fixiertheit	81
4.3.3. Häufigkeit.....	81
4.3.4. Kollokationsstärke.....	82
4.4. Funktionale Eigenschaften von Mehrworteinheiten.....	83
4.4.1. Verminderter kognitiver Verarbeitungsaufwand	83

4.4.2. Sozio-interaktionelle Funktion	86
4.4.3. Diskurs-organisierende Funktion	88
4.5. Eingrenzung des Konzepts der <i>Formulaic Language</i>	88
4.6. Kognitive Verarbeitung von <i>Formulaic Language</i>	90
4.7. Eingrenzung des Mehrworteinheiten-Konzepts für die vorliegende Arbeit	93
4.8. Entwicklungen in der <i>n-gram</i> -Forschung	93
4.9. Fokussierung auf <i>4-grams</i>	96
4.10. Zusammenfassung und Hypothesen der Arbeit	97
5. Methodik der Arbeit	99
5.1. Korpora.....	99
5.1.1. Das Web-ELF-Korpus.....	99
5.1.1.1. Hintergründe zur Erstellung des Web-ELF-Korpus	99
5.1.1.2. Zusammensetzung des Web-ELF-Korpus	100
5.1.2. Referenzkorpora	105
5.2. Linguistische und statistische Analysetools	106
5.3. Methodische Konzeption der quantitativen Studien	109
5.3.1. Quantitative Studie – Formale Klassifizierung von Mehrworteinheiten im Websprachgebrauch.....	109
5.3.2. Quantitative Studie – Funktionale Klassifizierung von Mehrworteinheiten im Websprachgebrauch.....	110
5.4. Methodische Konzeption der qualitativen Analysen	111
5.5. Zusammenfassung	112
6. Quantitative Analyse.....	113
6.1. Formale <i>4-gram</i> Analyse.....	113
6.1.1. Konzeptioneller Hintergrund.....	113
6.1.2. Definition der Kategorien.....	115
6.1.3. Adaption der Kategorisierung auf Web-Genres.....	119
6.1.4. Ergebnisse der formalen Klassifikation – <i>4-gram-Types</i>	120
6.1.5. Ergebnisse der formalen Klassifikation – Distribution der Kategorien	122
6.1.6. Probleme.....	130
6.1.7. Zusammenfassung	132
6.2. Funktionale <i>4-gram</i> Analyse	132
6.2.1. Grundkonzeption.....	132
6.2.2. Identifikation von Mehrworteinheiten mit dem <i>Automatic n-gram Analyser</i>	133
6.2.3. Die funktionale Klassifikation von <i>4-grams</i>	134
6.2.3.1. Das Erstellen einer Klassifikation	134
6.2.3.2. Klassifikation der <i>4-grams</i> der vier Web-Genres	137
6.2.3.3. Ergebnisse der Klassifikation.....	139
6.2.3.4. Eine funktionale Klassifikation gemeinsamer <i>4-grams</i> der vier Web-Genres	141
6.2.3.5. Verbindungen von Form und Funktion.....	149
6.2.4. Zusammenfassung	152

7. Qualitative Analyse	154
7.1. Häufigkeitsliste <i>4-grams</i>	154
7.2. Analyse von <i>at the end of</i>	158
7.3. Analyse des <i>Phrase-Frames at the X of</i>	164
7.4. Analyse von <i>one of the most</i>	165
7.5. Analyse von <i>to be able to</i>	173
7.6. Zusammenfassung	178
8. Diskussion der Ergebnisse und Resümee	180
8.1. Zusammenfassung der beobachteten Ergebnisse	180
8.2. Erklärungsansätze für die sprachliche Heterogenität der Web-Genres im Hinblick auf die Verwendung von <i>4-grams</i>	182
8.3. Überprüfung der Hypothesen	185
8.4. Resümee und Ausblick	186
Literaturverzeichnis	191
Appendix	207

Abbildungsverzeichnis

Abbildung 1: <i>Three Circle model</i> (in Crystal 1995: 107, basierend auf Kachru 1985, 1992).	11
Abbildung 2: <i>The circle of World English</i> (McArthur 1998: 97).....	13
Abbildung 3: <i>The Centripetal Circle of International English</i> (Modiano 1999).....	16
Abbildung 4: <i>Three-dimensional Model of English Use</i> (Yano 2007: 38)	19
Abbildung 5: Das Nähe-Distanz-Modell (Koch / Oesterreicher 1985, 1994).....	45
Abbildung 6: Web-Genres im Nähe-Distanz-Modell	72
Abbildung 7: Van Lanckers Kontinuum (Van Lancker 1987: 56).....	79
Abbildung 8: Formelhafte Sequenzen als Kompensation für kognitive Beschränktheit (aus Wray 2000: 16).....	85
Abbildung 9: Funktionen routinierter Sprache (aus Wray 2002: 97).....	87
Abbildung 10: Einstellungsmöglichkeiten von ANGA (aus Heller 2013: 49).	108
Abbildung 11: Foren im prozentualen Vergleich zu den Registern in Biber et al. (2004)	123
Abbildung 12: Blogs im prozentualen Vergleich zu den Registern in Biber et al. (2004)	124
Abbildung 13: Kommerzielle Homepages im prozentualen Vergleich zu den Registern in Biber et al. (2004).....	125
Abbildung 14: Wikipedia-Artikel im prozentualen Vergleich zu den Registern in Biber et al. (2004)	126
Abbildung 15: Prozentualer Vergleich der Web-Genres untereinander	127
Abbildung 16: Prozentualer Vergleich aller mündlichen und schriftsprachlichen Kategorien der Web-Genres.....	129
Abbildung 17: Verteilung von <i>Lexical Bundles</i> in funktionalen Kategorien (Biber et al. 2004: 397).....	137
Abbildung 18: Zusammenfassung der funktionalen Klassifikation.....	140
Abbildung 19: Interaktion struktureller und funktionaler Kategorien (Biber et al. 2004: 398)	150
Abbildung 20: Ergebnisse der Analyse von Form und Funktion.....	151
Abbildung 21: Übersicht relevanter Kommunikationsbedingungen für die Web-Genres.....	184

Tabellenverzeichnis

Tabelle 1: Merkmale mündlicher und schriftlicher Sprachrealisierung (vgl. Crystal 2003b)..	36
Tabelle 2: Kriterien zur Unterscheidung von gesprochener und geschriebener Sprache (Baron 2008: 47).....	43
Tabelle 3: Vergleich der Kommunikationsbedingungen der Web-Genres	69
Tabelle 4: Größe der individuellen Komponenten des Web-ELF-Korpus	103
Tabelle 5: Prozentuale Verteilung der <i>4-grams</i> in <i>Structural Patterns</i>	114
Tabelle 6: Normalisierte Gesamtzahl (<i>pmw</i>) aller <i>4-gram-Types</i> in den verschiedenen Korpora	121
Tabelle 7: Auszug aus der <i>4-gram</i> Liste des Web-Genres kommerzielle Homepages (normalisiert <i>pmw</i>).....	133
Tabelle 8: Typische <i>4-grams</i> aus verschiedenen Web-Genres des Web-ELF-Korpus für die einzelnen funktionalen Kategorien (vgl. Saage et al. 2012: 146).....	136
Tabelle 9: Auszug aus der <i>4-gram</i> Liste des Web-Genres Foren (normalisiert <i>pmw</i>).....	138
Tabelle 10: Prozentuale Verteilung der <i>4-grams</i> entlang der funktionalen Kategorien im Web-ELF-Korpus	139
Tabelle 11: Prozentuale Verteilung der <i>4-grams</i> entlang der funktionalen Kategorien in Biber et al. (2004).....	139
Tabelle 12: Distribution gemeinsamer <i>4-grams</i> in verschiedenen Funktionen (Biber et al. 2004: 396).....	141
Tabelle 13: Distribution gemeinsamer <i>4-grams</i> der Web-Genres	143
Tabelle 14: Gemeinsame <i>4-grams</i> für die Kategorie <i>Stance expressions</i>	144
Tabelle 15: Gemeinsame <i>4-grams</i> für die Kategorie <i>Discourse organizers</i>	145
Tabelle 16: Gemeinsame <i>4-grams</i> für die Kategorie <i>Referential expressions</i>	147
Tabelle 17: Statistische Auswertung der funktionalen Ähnlichkeit der Web-Genres.....	148
Tabelle 18: Die zwanzig häufigsten <i>4-gram-Types</i> des Web-Genres Foren.....	154
Tabelle 19: Die zwanzig häufigsten <i>4-gram-Types</i> des Web-Genre Blogs	156
Tabelle 20: Die zwanzig häufigsten <i>4-gram-Types</i> des Web-Genres kommerzielle Homepages	157
Tabelle 21: Die zwanzig häufigsten <i>4-gram-Types</i> des Web-Genres Wikipedia-Artikel.....	158
Tabelle 22: Ergebnis der semantischen Analyse von <i>at the end of</i>	160
Tabelle 23: Kollokate von <i>at the end of</i> im Web-Genre Foren	162
Tabelle 24: Kollokate von <i>at the end of</i> im Web-Genre Blogs	163

Tabelle 25: Kollokate von <i>at the end of</i> im Web-Genre kommerzielle Homepages	163
Tabelle 26: Kollokate von <i>at the end of</i> im Web-Genre Wikipedia-Artikel	163
Tabelle 27: Prozentuale Verteilung des <i>Phrase-Frames at the X of</i> in die semantischen Kategorien	165
Tabelle 28: Kollokate von <i>one of the most</i> im Web-Genre Foren	166
Tabelle 29: Kollokate von <i>one of the most</i> im Web-Genre Blogs	167
Tabelle 30: Kollokate von <i>one of the most</i> im Web-Genre kommerzielle Homepages	169
Tabelle 31: Kollokate von <i>one of the most</i> im Web-Genre Wikipedia-Artikel	171
Tabelle 32: Übersicht der prozentualen Verteilung der konnotierten <i>4-grams</i>	172
Tabelle 33: Kollokate von <i>to be able to</i> im Web-Genre Foren	174
Tabelle 34: Kollokate von <i>to be able to</i> im Web-Genre Blogs	176
Tabelle 35: Kollokate von <i>to be able to</i> im Web-Genre kommerzielle Homepages	176
Tabelle 36: Kollokate von <i>to be able to</i> im Web-Genre Wikipedia-Artikel	177
Tabelle 37: Kollokate aller Web-Genres links vom <i>Node-Word</i>	178
Tabelle 38: Zusammenfassung der Hauptkenntnisse der einzelnen Studien	181
Tabelle 39: Vergleich der Kommunikationsbedingungen der Web-Genres des Web-ELF- Korpus	183

1. Einleitung

Seit einigen Jahren findet ein öffentlicher Diskurs über den Einfluss des World Wide Web (WWW oder Web) auf den Sprachgebrauch statt. Dem Web wird dabei eine entscheidende Rolle am Sprachwandel und damit auch am Niedergang der Sprachkultur zugeschrieben.

Baron (2008: 163) zitiert beispielhaft zwei Zeitungsartikel, die dies auf polemische Art und Weise thematisieren. In einem Artikel wird kritisch angemerkt "[t]ext chats are starting to bleed into other aspects of life" (*National Post*, January 4, 2005), während in dem anderen Artikel resümiert wird: "[T]he changes we see taking place today in the language will be a prelude to the dying use of good English" (*Sun*, April 24, 2001).

Zweifelsohne hat das Web seit seiner öffentlichen Nutzung in den frühen 1990er Jahren die menschliche Kommunikation weltweit revolutioniert. Das Web unterscheidet sich von anderen Medien im Hinblick auf seine Multimedialität, Virtualität, Interaktivität und Nicht-Linearität. Um mit anderen Teilnehmern zu kommunizieren, können Webnutzer auf eine große Vielfalt an unidirektionalen, dialogischen und multidirektionalen Kommunikationsformen (*vulgo* Web-Genres) zurückgreifen. Trotz einer technischen und stilistischen Heterogenität wurde in der Sprachforschung wiederholt die Existenz sprachlicher Universalien für die Gesamtheit aller Kommunikationsformen im Medium Internet postuliert. Es wird dabei beispielsweise von einem dritten Medium, dem sogenannten *Netspeak* (Crystal: 2001) gesprochen, das sich zwischen den beiden etablierten medialen Polen "gesprochene Sprache" und "geschriebene Sprache" verorten ließe. *Netspeak* zeichne sich dadurch aus, dass es Eigenschaften der ersten beiden Medien in sich vereinige und damit als Hybridwesen aus mündlicher und schriftlicher Sprache ein eigenständiges neues Medium bilde. Es müsse allerdings näher am Pol der Schriftlichkeit positioniert werden, wie Crystal (2001: 51) feststellt: "Netspeak is on the whole better seen as written language which has been pulled in the direction of speech than spoken language which has been written down". Entgegen der ursprünglich positiven Rezeption wurde in den letzten Jahren das *Netspeak*-Konzept jedoch zunehmend kritischer diskutiert (vgl. Dürscheid 2003a). Dies führte zu einer Polarisierung innerhalb der Internetsprachforschung. Dabei haben bereits einige Studien (vgl. Crystal 2006, Baron 2008) gezeigt, dass das Web durch die große Bandbreite an verschiedenen Web-Genres ein stilistisch vielfältiges Medium darstellt. So unterscheidet sich beispielsweise das Web-Genre *Chat* sowohl in seiner Form als auch in seiner Funktion in hohem Maße vom Web-Genre *Wiki*. Es muss vor diesem Hintergrund angezweifelt werden, ob sich trotz aller sprachlicher Heteroge-

nität ein Kern aus universellen Sprachmerkmalen identifizieren lässt, der die Existenz einer homogenen Websprache belegt.

Nachdem in den 1990er Jahren nur peripher Forschung betrieben wurde, die sich der Beschreibung der sprachlichen Unterscheidung der verschiedenen Web-Genres widmete, wurde diesem Thema in jüngeren Jahren zunehmend mehr Beachtung geschenkt (vgl. Giltrow / Stein 2009). Allerdings wurde dabei der Fokus der Untersuchungsebene vor allem auf lexikalische Phänomene gelegt (vgl. Baron 2008, Crystal 2006). Eine Beschreibung und Unterscheidung verschiedener Web-Genres im Hinblick auf phraseologische Besonderheiten wurde bisher nicht annähernd so umfassend geleistet. Bisher existieren lediglich vereinzelte, sehr spezifische Studien, in denen beispielhaft einzelne lexikogrammatische Muster analysiert werden. So hat Fritz (2011) zum Beispiel eine funktionale Analyse von bestimmten Textmustern in dem Blog *Language Log* durchgeführt. Die vorliegende Arbeit fokussiert sich auf ein Teilgebiet der Phraseologie, der formalen und funktionalen Analyse von Mehrworteinheiten.

Das Web ist ein sprachlich höchst heterogenes Medium, in dem eine Vielzahl von Sprachen und Dialekten genutzt wird. Als klare Tendenz zeichnet sich jedoch ab, dass eine geringe Zahl von Sprachen einen Großteil der Sprachverwendung ausmacht. Das Englische hat dabei in diesem Medium eine ähnlich übergeordnete Rolle wie im nicht-elektronischen Kontext (Miniwatts Marketing Research 2011). Weltweit beherrschen etwa zwei Milliarden Menschen die englische Sprache. Der Anteil der Muttersprachler ist jedoch mit circa 350 Millionen Sprechern verhältnismäßig gering (Crystal 2006). Die Tatsache, dass seit einigen Jahrzehnten die Anzahl der Nichtmuttersprachler kontinuierlich ansteigt und inzwischen bei Weitem die Gesamtheit der Muttersprachler übertrifft, manifestiert sich in einer Ausdifferenzierung der Varietäten des Englischen. Insgesamt ist ohne Zweifel eine Tendenz zur stärkeren sprachlichen Heterogenität der englischen Sprache zu erkennen.

Im Jahr 2011 haben weltweit etwa 2,1 Milliarden Menschen das Web genutzt (Miniwatts Marketing Research 2011). Obwohl sich im Web neben den Sprachen mit hoher Sprecheranzahl zunehmend kleinere Sprachen etablieren, zeigt sich in der Gesamtverteilung aller Sprachen eine eindeutige Präferenz für die Nutzung von Sprachen, die von vielen Sprechern weltweit beherrscht werden. In allererster Linie trifft dies auf das Englische zu, das bereits seit jeher die Sprache des Webs darstellt. Obwohl der prozentuale Anteil an englischsprachigen Webseiten relativ stark geschrumpft ist, von etwa 80% Mitte der 1990er Jahre (Warschauer 2000) auf etwa 56% im Jahr 2002 (Netz-Tipp 2002), ist das Englische weiterhin eine der

Sprachen, die im Web einen großen Einfluss ausübt.¹ So sind neun der zehn am häufigsten besuchten Webseiten weltweit in englischer Sprache verfasst (Alexa 2012). Darüber hinaus ist das Englische im Vergleich zu anderen Sprachen die am häufigsten verwendete Sprache des Webs; so kommunizieren 26,8% aller Webnutzer auf Englisch (Miniwatts Marketing Research 2011).

Für die vorliegende Arbeit werden mehrere Studien zur lexikogrammatistischen Beschreibung von verschiedenen Web-Kommunikationsformen durchgeführt. Dazu wurden vier Web-Genres (Foren, Blogs, kommerzielle Homepages und Wikipedia-Artikel) gewählt, die anhand von formalen und funktionalen Analysen von phraseologischen Einheiten (Mehrworteinheiten) voneinander unterschieden werden sollen. Als Vergleichsparameter dient dabei traditioneller mündlicher und schriftlicher Sprachgebrauch. Das Ziel der Analysen ist es, einerseits die vier untersuchten Web-Genres voneinander abzugrenzen, andererseits zu überprüfen, ob sich trotz ihrer Unterschiede ein Kern an Web-Genre übergreifenden Gemeinsamkeiten finden lässt.

Auf der lexikogrammatistischen Ebene sind Genreunterschiede in der Regel gradueller Natur und können in Form von Häufigkeitsunterschieden im Gebrauch sprachlicher Formen und Strukturen erfasst werden. Daher wird ein großes und repräsentatives Textkorpus benötigt, um auf dieser empirischen Datengrundlage verlässlich Frequenzunterschiede zu beschreiben. Für die vorliegende Arbeit wird das Web English as a Lingua Franca (Web-ELF) Korpus verwendet, das aus insgesamt 2,7 Millionen Wörtern besteht und Texte aus den Web-Genres Foren, Blogs, kommerzielle Homepages und Wikipedia-Artikel beinhaltet, die allesamt im thematischen Kontext *Medical Surgery / Medical Tourism* entstanden sind.²

Im Fokus der Studien stehen formale und funktionale Analysen von *4-grams*, das heißt Mehrworteinheiten, die aus vier Elementen bestehen, wie zum Beispiel *on the other hand*. Ausgangspunkt waren die Studien von Biber et al. (1999) und Hyland (2008), die gezeigt haben, dass Mehrworteinheiten in verschiedenen Kommunikationssituationen (beziehungsweise Registern), in unterschiedlichem Maße genutzt werden. Sie wiesen dabei nach, dass man zwischen "typisch mündlichen" *4-grams* (z.B. *I don't know what*) und "typisch schriftlichen" (z.B. *on the other hand*) *4-grams* unterscheiden kann.

¹ Seit 2002 wurden keine Studien mehr durchgeführt, die den prozentualen Anteil englischsprachiger Webseiten im Verhältnis zu anderssprachigen Webseiten untersucht haben.

² Die Anzahl der Wörter aus den einzelnen Korpuskomponenten beträgt: Foren: 1.200.000 Wörter, Blogs 215.000 Wörter, kommerzielle Homepages: 570.000 Wörter, Wikipedia-Artikel: 740.000 Wörter.

In einer ersten Studie werden alle *4-grams* der verschiedenen Web-Genres auf formaler Ebene untersucht. Biber et al. (1999) fanden bereits in früheren Studien heraus, dass in traditionell mündlichen und schriftlichen Kommunikationssituationen unterschiedliche *4-gram*-Muster (*patterns*) verwendet werden. So wird in mündlicher Konversation beispielsweise das Muster *personal pronoun + lexical VP* sehr häufig verwendet, wohingegen in schriftlicher Wissenschaftssprache das Muster *NP + of-phrase fragment* verwendet wird. Für ihre Studie identifizierten sie manuell alle Mehrworteinheiten (*lexical bundles*) des COBUILD Korpus³ und ordneten diese in eigens gebildete Kategorien ein. Sie untersuchten dabei alle Muster der beiden Register *Conversation* und *Academic prose*. Für die vorliegende Studie werden die Kategorien von Biber et al. (1999) übernommen, um alle *4-grams* der vier Web-Genres automatisiert zu untersuchen. Die Ergebnisse der Analyse werden schließlich in einer grafischen Form aufbereitet, die es ermöglicht, Gemeinsamkeiten und Unterschiede besser zu visualisieren.

In einer weiteren Studie werden die am häufigsten auftretenden *4-grams* identifiziert und funktionalen Kategorien zugeordnet, die auf Biber et al. (2004) basieren. Das Ziel der Studie ist es, im ersten Schritt die typischsten *4-grams* für vier verschiedene Kommunikationssituationen des Webs (Foren, Blogs, kommerzielle Homepages, Wikis) zu identifizieren und diese dann im zweiten Schritt funktional zu klassifizieren, um die vier untersuchten Web-Genres im Hinblick auf den Grad ihrer konzeptionellen Mündlichkeit / Schriftlichkeit auf der lexikogrammatischen Ebene bewerten zu können.

Neben diesen quantitativen Studien werden ebenfalls mehrere qualitative Analysen von einzelnen *4-gram*-Mustern durchgeführt. Dabei werden auffällige *4-gram*-Muster in ihrem weiteren Kontext betrachtet und ihre Verwendung in den verschiedenen Web-Genres des Web-ELF-Korpus überprüft. Die Ergebnisse zielen darauf ab, die Resultate der quantitativen Analyse zu ergänzen und weitere Belege dafür zu finden, dass die einzelnen Web-Genres sich auf vielfältiger Ebene voneinander unterscheiden. Im Anschluss an die Präsentation der Ergebnisse werden zunächst Erklärungsansätze für die Ergebnisse der einzelnen Studien formuliert werden, bevor ein Resümee gezogen wird. Die vorliegende Arbeit untergliedert sich wie folgt:

In dem Kapitel 2 wird die Entwicklung des Englischen zur globalen Sprache kurz skizziert. Im Fokus steht dabei besonders die Verwendung des Englischen als Lingua Franca im Kon-

³ Collins Birmingham University International Language Database Korpus.

text der *World Englishes*. Dabei wird die Frage aufgeworfen, inwieweit der Sprachgebrauch im Web eine neue Form des Lingua Franca-Sprachgebrauchs darstellt.

In dem Kapitel 3 wird zunächst die historische Entwicklung des Internets nachgezeichnet. Im Anschluss wird das *Netspeak*-Konzept beschrieben und kritisch diskutiert. Es werden zudem sowohl die medialen als auch die sprachlichen Besonderheiten des Mediums Internet herausgestellt. Das Nähe-Distanz-Modell von Koch / Oesterreicher (1985, 1994) erlaubt es verschiedene Register voneinander abzugrenzen. Das Modell wird in dieser Studie verwendet werden, um auch die vier untersuchten Web-Genres bezüglich ihrer konzeptionellen Mündlichkeit / Schriftlichkeit voneinander zu unterscheiden.

In dem Kapitel 4 wird die Bedeutung der "formelhaften Sprache" (*Formulaic Language*) mit all ihren unterschiedlichen Ausprägungsformen für den alltäglichen Sprachgebrauch im Detail erläutert. Dabei wird aufgezeigt, dass völlig unterschiedliche Definitionen des Mehrworteinheiten-Konzepts existieren. Das Konzept, das dieser Arbeit zu Grunde liegt, wird ausführlich beschrieben werden. Im Anschluss folgt ein Überblick über aktuelle Studien zum Thema *n-grams*. Am Ende des Kapitels werden schließlich die Forschungshypothesen, die der vorliegenden Arbeit zu Grunde liegen, vorgestellt.

In dem Kapitel 5 wird die Methodik beschrieben, die für die vorliegenden Studien gewählt wurde. In dem Kapitel 6 werden die quantitativen formalen und funktionalen Studien, die für die Arbeit durchgeführt wurden, präsentiert und die Ergebnisse der Analysen im Detail aufgeführt. Das Kapitel 7 widmet sich komplementär dazu der Darstellung der qualitativen Analysen von mehreren auffälligen *4-gram*-Mustern. In dem Kapitel 8 werden die Ergebnisse aller Studien quantitativen und qualitativen Studien zusammengefasst. Dabei werden auch die Gründe für die Unterschiedlichkeit der Web-Genres erörtert. Das Kapitel schließt mit einer Zusammenfassung der Arbeit, bevor ein Ausblick auf potenzielle Nachfolgestudien gegeben wird.

2. World Englishes und Lingua Franca-Sprachgebrauch

The new language which is rapidly ousting the language of Shakespeare as the world's lingua franca is English itself – English in its new global form.

(Graddol 1997: 11)

2.1. World Englishes

2.1.1. Die Entwicklung des Englischen zur Weltsprache

Die zunehmende globale Verbreitung des Englischen als kulturübergreifendes Kommunikationsmittel ist ein Phänomen, das in den letzten Jahrzehnten ausführlich beschrieben wurde (vgl. Crystal 2003a, Graddol 1997). Dabei wird wiederholt betont, dass die Rolle und der Status des Englischen im Laufe des 20. Jahrhunderts aus wirtschaftlichen, kulturellen und politischen Gründen auf globaler Ebene enorm expandiert sind. Als ein weiterer bestimmender Faktor ist das World Wide Web zu nennen, das die Vormachtstellung des Englischen als *de facto* globale Lingua Franca weiter gefestigt hat. Aus einer jahrhundertlang relativ unbedeutenden Sprache, die beinahe exklusiv auf den britischen Inseln verwendet wurde, ist somit eine Weltsprache geworden, die zu Beginn des neuen Millenniums etwa zwei Milliarden Menschen beherrschen (Crystal 1997: 61).

Es muss dennoch betont werden, dass die Vormachtstellung des Englischen das Ergebnis eines mehrere Jahrhunderte dauernden kontinuierlichen Prozesses ist. Der Aufstieg des Englischen begann mit der Etablierung des Vereinigten Königreichs als Weltmacht und der damit einhergehenden Kolonialpolitik unter Königin Elizabeth I. im 16. Jahrhundert. Bis dahin beschränkte sich der Gebrauch des Englischen fast ausschließlich auf die britischen Inseln (vgl. Trudgill / Hannah 2008: 8). Durch die imperialistische Politik des Vereinigten Königreichs, die von militärischer Stärke und internationalem Handel geprägt war, wurde das Englische schließlich in alle Teile der Welt getragen. Als Meilenstein auf dem Weg zur politischen Weltmacht erwiesen sich vorrangig die Kolonisierung und die wirtschaftliche Ausbeutung von Teilen Afrikas, Asiens und Nordamerika.

Jenkins (2003, 2009) und Mesthrie / Bhatt (2008) beschreiben die Expansion des Englischen als Resultat zweier Diaspora. Als erste Diaspora gilt die Emigration von mehreren tausend Briten in die neu gegründeten Kolonien nach Nordamerika, Australien und Südafrika. Die Herkunft der Emigranten hatte einen entscheidenden Einfluss auf die Herausbildung der örtlichen Varietäten. Diese sollten sich schließlich in neuen normativen Standards, wie zum Bei-

spiel das amerikanische Englisch oder das australische Englisch, niederschlagen. Als zweite Diaspora wird die Kolonialisierung von Teilen Afrikas und Asiens verstanden. Im Gegensatz zur ersten Diaspora emigrierte nur eine geringe Zahl an britischen Muttersprachlern in diese ohnehin schon dicht bevölkerten Kontinente. Der Einfluss der englischen Minorität auf die vorherrschenden Lokalsprachen, die von den Einheimischen weiterhin als Erstsprache verwendet wurden, war somit letztlich stark limitiert. Das Englische diente in vielen dieser Länder nur als Zweitsprache für administrative Zwecke. Mit zunehmender Dauer wurde die Nutzung des Englischen jedoch auch in andere Bereiche expandiert. Schließlich haben sich vielerorts norm-entwickelnde (*norm-developing*, vgl. Kachru 1985) lokale Englischvarietäten herausgebildet, die sich zunehmend von der amerikanischen und britischen Norm entfernten, wie beispielsweise das *Indian English* (Jenkins 2009: 5).

Ein bedeutender Faktor für die globale Expansion des Englischen stellt die Unabhängigkeit der Vereinigten Staaten von der britischen Krone dar. Dieses Ereignis führte zu einer Emanzipation der amerikanischen Varietät von der britischen Varietät, so dass fortan *American English* als eigenständige Varietät akzeptiert wurde. Während das Vereinigte Königreich im Laufe des 20. Jahrhunderts zunehmend an weltpolitischer und wirtschaftlicher Bedeutung einbüßte, begannen die Vereinigten Staaten von Amerika in diesem Zeitraum die weltpolitische Bühne zunehmend zu beherrschen. Im Zuge dessen expandierte die englische Sprache in Form des *American English* auf globaler Ebene. Als entscheidend dafür erwies sich ebenfalls der kulturelle Einfluss der Vereinigten Staaten, in Form der amerikanischen Film- und Musikindustrie.

Auf wissenschaftlicher Ebene ist im 20. Jahrhundert ebenfalls ein Paradigmenwechsel vollzogen worden. Das Englische hat im 20. Jahrhundert andere bis dato führende Sprachen (Französisch, Deutsch, Lateinisch) als Lingua Franca der Wissenschaft abgelöst (Knapp 1991). Auf den meisten internationalen Konferenzen wird mittlerweile auf Englisch kommuniziert und Forschungsergebnisse werden zumeist in englischer Sprache veröffentlicht. Laut Knapp (1991: 20) werden etwa drei Viertel aller wissenschaftlichen Veröffentlichungen weltweit auf Englisch publiziert. Weitere Studien (Ammon 1998, Kaplan 2001) haben gezeigt, dass in den letzten Jahren die Anzahl an Publikationen auf Englisch beständig gewachsen ist, vor allem im naturwissenschaftlichen Bereich, wo andere ehemals wichtige Sprachen, wie zum Beispiel das Deutsche, klar an Bedeutung verloren haben. So wurden Ende der 1990er Jahre 91% aller naturwissenschaftlichen Publikationen auf Englisch veröffentlicht (Ammon 1998: 152). In einer aktuelleren Studie stellen Clarke et al. (2007) fest, dass 96,5% aller wis-

senschaftlichen Beiträge zum Thema Gesundheitswesen (*public health*) auf Englisch veröffentlicht wurden.

Mit der zunehmenden Verwendung verschiedener Ausprägungen elektronischer Kommunikation in allen Lebensbereichen ist in den letzten Jahrzehnten ein weiterer Faktor für die globale Verbreitung des Englischen hervorgetreten. McLuhan hat dies bereits vor über fünfzig Jahren in seinem einflussreichen medientheoretischen Werk *The Gutenberg Galaxy* (McLuhan 1962) vorausgesehen. Er wies darauf hin, dass die Welt durch elektronische Vernetzungen zu einem *global village* (globalen Dorf) zusammenwachsen würde. Das Web wird heutzutage oft als Sinnbild für seine Kernaussage herangezogen, da dieses seinen Nutzern ermöglicht, unabhängig vom physischen Standort global miteinander zu kommunizieren. McLuhan (1962) prognostizierte zudem bereits dreißig Jahre vor der öffentlichen Zugänglichkeit des World Wide Web, dass elektronische Kommunikation die traditionelle Printkultur ablösen würde und dass dies beträchtliche Auswirkungen auf das gesellschaftliche Selbstverständnis haben würde. In diesem globalen Dorf würde die Individualität des Einzelnen zugunsten einer kollektiven Identität aufgegeben werden. McLuhans gezeichnetes Szenario des globalen Dorfes ist inzwischen zur Realität geworden. In einer globalisierten Welt ist die Nutzung einer gemeinsamen Verkehrssprache, die es Teilnehmern unterschiedlicher Sprach- und Kulturkreise ermöglicht, miteinander zu kommunizieren, von großer Bedeutung. Diese Rolle füllte in vielen Fällen das Englische aus. Die Erfindung des World Wide Web hat schließlich seinen Teil dazu beigetragen, dass das Englische seine Rolle als globale Lingua Franca weiter festigen konnte. Das Web hat den Alltag revolutioniert und füllt inzwischen verschiedenste Funktionen sowohl im beruflichen als auch im privaten Bereich aus. Es hat sich zu einem globalen Interaktionsmedium entwickelt, in dem sich Nutzer aus verschiedenen Ländern und Kulturkreisen in diversen Kommunikationssituationen, wie zum Beispiel Foren oder Chatrooms, verständigen können.

Graddol (2006) prognostiziert, dass das Englische in den Bereichen Wirtschaft und Forschung im Laufe des 21. Jahrhunderts weiter an Bedeutung gewinnen wird. Daher erwartet er, dass zunehmend mehr Menschen Englisch lernen werden. In Bezug auf das Web sieht Graddol jedoch einen gegenläufigen Trend voraus. Graddol (2006: 44) verweist dabei auf die stark rückläufige Quote englischer Muttersprachler im Web: sie sank von 51,3% im Jahr 2000 auf 32% im Jahr 2005. Für den starken prozentualen Rückgang des Englischen innerhalb dieses Zeitraums zeichnet er in erster Linie die zunehmende Web-Nutzung chinesischer Sprachnutzer verantwortlich.

Das Web ist durch die fortschreitende globale Zugänglichkeit innerhalb der letzten zwei Dekaden zu einem multilingualen Versammlungsort für Nutzer unterschiedlicher Muttersprachen geworden. Die heutige Gemeinschaft von Webnutzern definiert sich als multikulturell und multiethnisch. So tauschen sich verhältnismäßig immer mehr chinesisch und arabisch sprechende Nutzer im Web aus. Im selben Maße sinkt die Anzahl englischer Muttersprachler proportional zur Gesamtzahl der Webnutzer. Dies ist ein Trend, der die realen globalen Bevölkerungsverhältnisse widerspiegelt. Englisch wird mittlerweile von weit mehr Nichtmuttersprachlern als von Muttersprachlern verwendet. Laut *Ethnologue* (vgl. Graddol 2006) verwendeten im Jahr 2005 328 Millionen Menschen weltweit das Englische als Muttersprache und 508 Millionen Englisch als Zweitsprache. Graddol (2006: 62) vermutet, dass die tatsächliche Zahl der *English as a Second Language (ESL)*-Sprachverwender mit rund einer Milliarde Sprechern wesentlich höher liegt. Zudem rechnet er zukünftig mit einer steigenden Anzahl an Sprechern, vor allem durch die stark wachsende Zahl an Fremdsprachenlernern des Englischen. Graddol (2006: 66) ist davon überzeugt, dass diese Entwicklung einen Sprachwandel nach sich ziehen wird. Aus diesem Wandel, in dessen Prozess sich das Englische bereits befindet, werde demnach letztendlich eine neue Form des Englischen hervorgehen: das *Global English*. In dieser Varietät üben Muttersprachler, aufgrund ihrer marginalen Rolle im Vergleich zu den Nichtmuttersprachlern, keinen entscheidenden normativen Einfluss mehr auf die Sprachentwicklung des Englischen aus. Statistiken zur globalen Nutzung des Englischen bestätigen den von Graddol prognostizierten Trend. Beneke (1991) berichtet, dass englische Muttersprachler an 80% aller globalen Interaktionen, die auf Englisch geführt werden, nicht teilnehmen. Seidlhofer (2005) stellt fest, dass nur etwa jeder vierte Nutzer des Englischen Muttersprachler ist.

Obwohl das Englische bereits seit einigen Jahrzehnten als die meistgenutzte internationale Kommunikationssprache dient, wird es in seiner Funktion als *Lingua Franca* von der Forschung erst seit relativ kurzer Zeit intensiv beleuchtet. Im Fokus der Betrachtung steht seit den 1990er Jahren insbesondere, inwieweit *English as a Lingua Franca (ELF)*-Sprachgebrauch eine eigenständige Varietät des Englischen darstellt (vgl. Firth 1996, Meierkord 1996, Mollin 2006a, Maley 2009). Seidlhofer (2001) bemerkt, dass sich die linguistische Forschung vorrangig auf muttersprachliche Varietäten konzentriert, obwohl *ELF*-Sprachgebrauch zu Beginn des 21. Jahrhunderts den Großteil der Kommunikation in englischer Sprache ausmacht. Sie fordert daher, dass der *ELF*-Sprachgebrauch besser beschrieben werden muss (Seidlhofer 2001: 133ff.). Diesem Postulat wurde Folge geleistet, so dass zu Beginn der 2000er Jahre schließlich die ersten großangelegten *ELF*-Korpusprojekte, in Form des *Vienna-Oxford*

International Corpus of English (VOICE) (vgl. Seidlhofer 2001) und des *English as a Lingua Franca in Academic Settings (ELFA)*-Korpus (vgl. Mauranen 2003), begonnen wurden. Diese beiden Korpora ermöglichten erste deskriptive Beschreibungen von *ELF*-Sprachgebrauch.

Zusammenfassend muss festgehalten werden, dass die Gründe für den Aufstieg des Englischen zur Weltsprache vielfältiger Natur sind. Politische, wirtschaftliche und kulturelle Ursachen haben dazu geführt, dass das Englische Einzug in alle Teile der Welt gefunden hat und dass das Erlernen der englischen Sprache vielerorts als gewinnbringend empfunden wird. Diesbezüglich stellt Dröschel (2011: 28) fest: "English has an image of prosperity and many people are learning it in order to climb the socio-economic ladder". Das Beherrschen der englischen Sprache wird vor allem in Schwellen- und Entwicklungsländern mit sozialem Aufstieg gleichgesetzt (Brutt-Griffler 2008, Verma 2008). Heutzutage wird in vielen Nationen weltweit Englisch als erste Fremdsprache an Schulen gelehrt. Kritiker (u.a. Philipson 1992) führen die Dominanz der englischen Sprache vor allem auf den politischen und kulturellen Einfluss der beiden Supermächte, der Vereinigten Staaten von Amerika (USA) und Großbritannien, zurück. Unabhängig davon, welcher Standpunkt in dieser Debatte vertreten wird, fungiert das Englische aktuell weiterhin als globale Lingua Franca, trotz der stetig wachsenden globalen Bedeutung des Chinesischen.

2.1.2. World English-Modelle

Seit den frühen 1980er Jahren wird in der Varietätenforschung versucht, alle möglichen Ausprägungsformen des Englischen in Form eines *World English*-Modells zusammenzutragen. Innerhalb eines *World English*-Modells wird häufig zwischen *English as a Native Language (ENL)*-, *English as a Second Language (ESL)*- und *English as a Foreign Language (EFL)*-Varietäten unterschieden.

Als eines der einflussreichsten Modelle gilt das *Three Circle Model* von Kachru (1985, 1988) (vgl. Abbildung 1). Das Modell besteht aus drei konzentrischen Kreisen, dem *Inner Circle*, dem *Outer Circle* und dem *Expanding Circle*. Kachru sortiert alle Varietäten des Englischen in diese drei Kreise ein. Die Eingruppierung geschieht ausschließlich nach historischen und geografischen Gesichtspunkten. Im *Inner Circle* befinden sich alle Varietäten in denen Englisch als Muttersprache genutzt wird. In den aufgeführten Nationen wird das Englische sowohl in formellen als auch informellen Kommunikationssituationen verwendet. Zu diesem Kreis zählen die USA, Großbritannien, Kanada, Irland, Australien und Neuseeland (vgl. Kachru 1992: 230). Im *Outer Circle* befinden sich alle Varietäten in denen Englisch als

Zweitsprache verwendet wird. Das Englische wird in diesen Nationen unter anderem für administrative Zwecke genutzt. Die Nationen, die diesem Kreis angehören, weisen im Normalfall eine koloniale Vergangenheit auf. Zu dieser Gruppe gehören unter anderem Indien, Singapur und Sri Lanka. Im *Expanding Circle* befinden sich schließlich alle Varietäten in denen Englisch als Fremdsprache genutzt wird. Zu dieser Gruppe gehören unter anderem China, Russland oder Deutschland. Die Varietäten, die sich im *Inner Circle* befinden, sind *norm-providing*, das heißt in diesen Varietäten werden endonormative Standards für das Englische geschaffen, an denen sich Englischlerner orientieren. Dahingegen sind die Varietäten im *Outer Circle norm-developing*. Dies bedeutet, dass sie im Begriff sind, eigene Normen zu entwickeln und sich damit von den Varietäten der Muttersprachler im *Inner Circle* zu emanzipieren.

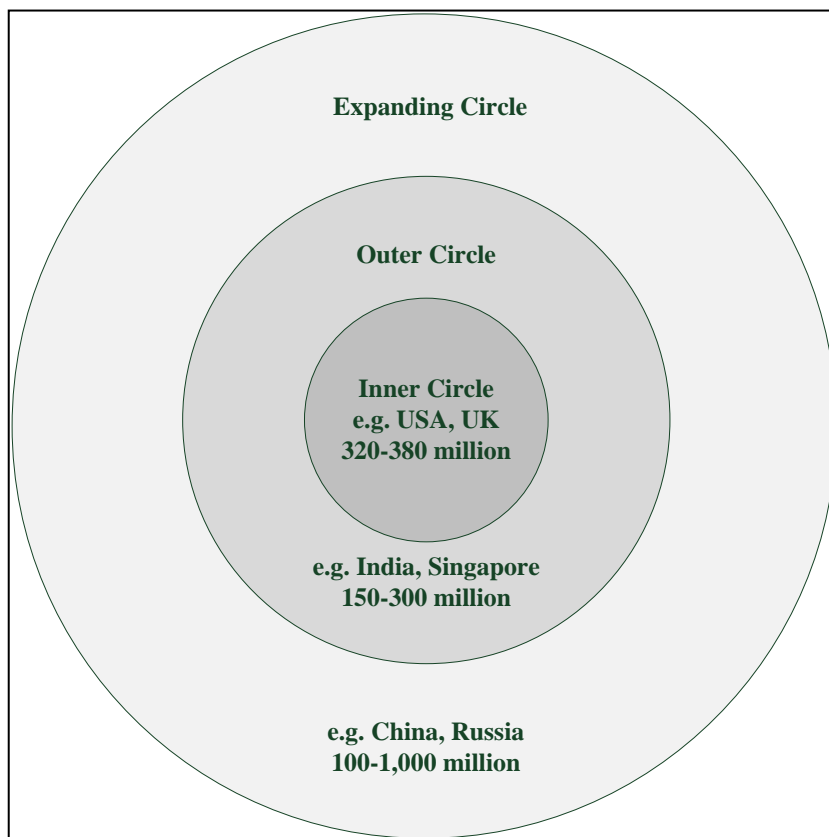


Abbildung 1: *Three Circle model* (in Crystal 1995: 107, basierend auf Kachru 1985, 1992)

Die Varietäten im *Expanding Circle* sind schließlich *norm-dependant*, das heißt sie folgen exonormativen muttersprachlichen Standards. So wurde beispielsweise in weiten Teilen Deutschlands viele Jahre lang das britische Englisch als Standardnorm betrachtet. Die Beherrschung dieses Standards auf muttersprachlichem Niveau (sowohl auf phonetischer, lexika-

lischer, grammatischer als auch pragmatischer Ebene) dient Lernern im *Expanding Circle* als anzustrebende Zielnorm.

Das *Three Circle Model* von Kachru ist seit seiner Erscheinung unterschiedlich rezipiert worden. Als positiv ist zu erachten, dass das Modell einen Überblick über die Vielfalt der *World Englishes* gibt. Trotz der Einfachheit des Modells erweist es sich als sehr hilfreiche visuelle Unterstützung zur Abgrenzung der diversen Varietäten des Englischen. Der Erfolg des Modells beruht darauf, dass es in den 1980ern der Emanzipation der *New Englishes* diene. Kachrus Modell schuf nach seinem Erscheinen für viele Jahre den Rahmen für etliche weiterführende Arbeiten im Bereich der *World Englishes* (vgl. Bamgbose 1998, Crystal 2003a, Jenkins 2007).

Trotz der weitreichenden Verwendung des Modells in der Varietätenforschung ist es wegen zahlreicher Mängel dennoch vielfach kritisiert worden (vgl. Bruthiaux 2003). Mollin (2006a: 1) stellt dazu fest: "Its most serious drawback is that it does not spell out clearly what it wants to categorize, since nations, types of speakers, functions of English as well as types of variety are all referred to". Beispielsweise würden Muttersprachler im *Inner Circle* von Nichtmuttersprachlern im *Outer* und *Expanding Circle* unterschieden werden. Gleichzeitig würde der Eindruck erweckt, dass Englisch in Nationen des *Inner Circle* in allen Domänen und für alle kommunikativen Zwecke verwendet werde, ebenso wie für alle wichtigen intranationalen Funktionen im *Outer Circle* (vgl. Mollin 2006a: 1). Im *Expanding Circle* hingegen würde Englisch für die Kommunikation mit Mitgliedern des *Inner* und *Outer Circle* erlernt werden. Jenkins (2003: 17) führt weitere Kritikpunkte auf, unter anderem, dass das Modell ausschließlich auf historisch gewachsene regionale Unterscheidungen der Varietäten fokussiert ist. Die Grenzen werden dabei als absolut aufgefasst. So kann beispielsweise nicht im Modell abgebildet werden, dass englische Muttersprachler in Indien leben oder dass in den Vereinigten Staaten von Amerika eine große Gruppe von Immigranten lebt, für die Englisch nur als Zweitsprache dient. Zudem können Kompetenzunterschiede innerhalb einer Sprechergemeinschaft nicht anhand des Modells sichtbar gemacht werden. Stattdessen wird eine sprachliche Homogenität der Sprachnutzer vorgetäuscht, die der Realität in keinster Weise entspricht (Jenkins 2003: 17). Ein weiterer schwerwiegender Nachteil des Modells ist, dass der interzirkuläre Gebrauch des Englischen als Lingua Franca nicht abgebildet wird. Dies bedeutet, dass die *ELF*-Kommunikation innerhalb und zwischen Kachrus Kreisen vollends unbeachtet bleibt (vgl. Seidlhofer 2009). Zudem kritisiert Jenkins (2003: 17), dass eine Hierarchisierung des Englischen vorgenommen wird, indem durch die Anordnung der Kreise der Eindruck erweckt wird, dass

ENL-Varietäten eine zentralere Rolle einnehmen im Vergleich zu *ESL*- und *EFL*-Varietäten. Kurzum, das Modell von Kachru wird nicht der komplexen Realität der englischen Sprache gerecht und kann diese daher nicht in zufriedenstellendem Maße abbilden.

Neben Kachru haben weitere Varietätenforscher versucht, ein Modell zu entwickeln, das das Verhältnis der Varietäten des Englischen zueinander veranschaulicht. Es fällt auf, dass in vielen dieser Modelle eine geografische Abgrenzung der Varietäten anhand von Kreisen vorgenommen wird. Eines dieser Modelle ist das *The circle of World English*-Modell von McArthur (1987) (vgl. Abbildung 2).

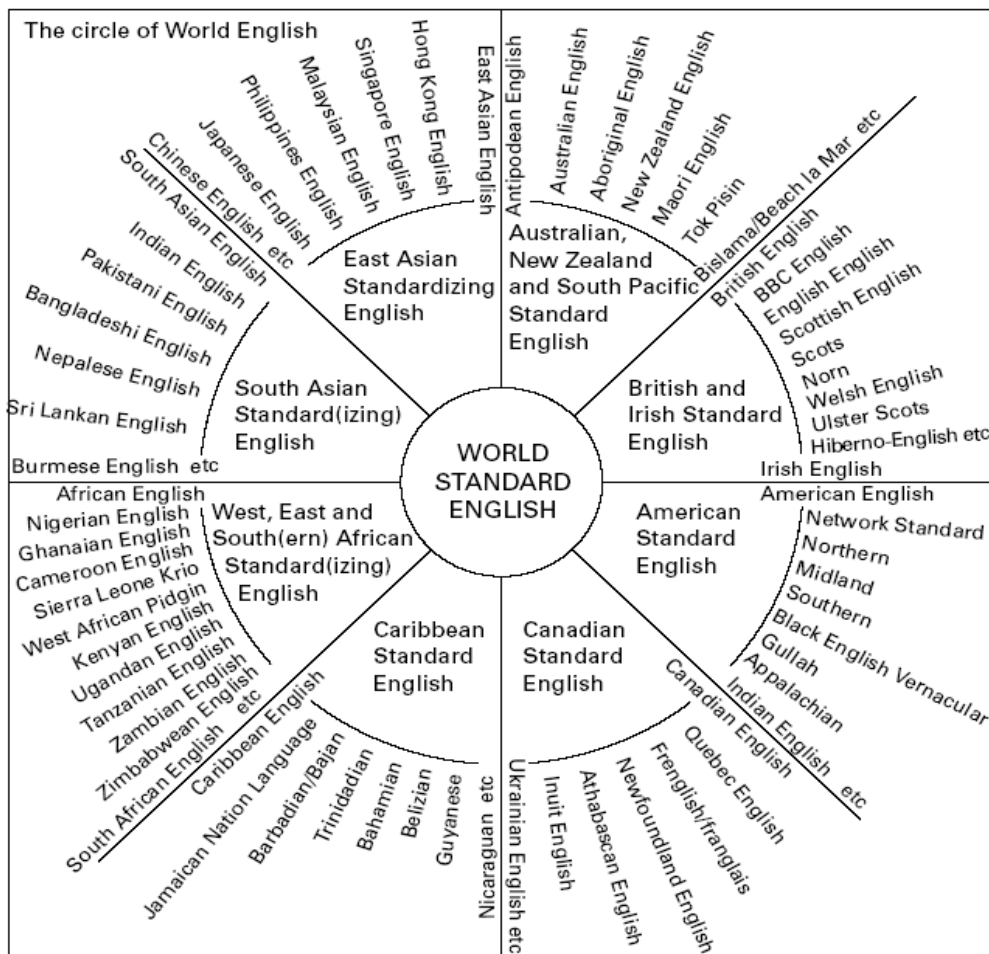


Abbildung 2: *The circle of World English* (McArthur 1998: 97)

Den Kern des Modells bildet das sogenannte *World Standard English*, das die schriftliche, international anerkannte Varietät des Englischen darstellt. Dieser Kern ist umringt von einem Kreis, der acht geografische Regionen des Englischen abbildet, die sowohl *Standard Englishes* als auch *Standardizing Englishes* beinhalten. Die Regionen Großbritannien / Irland,

Amerika, Kanada, Australien / Neuseeland / Südpazifik und Karibik bilden eigenständige Standardvarietäten des Englischen, während die Regionen in Asien und Afrika lediglich *Standardizing* Englishes darstellen. Dies sind Varietäten, die sich im Prozess der Standardisierung befinden. Um diesen Kreis formt sich schließlich ein weiterer Kreis, in dem sich die Subvarietäten der jeweiligen Varietät befinden, so zum Beispiel *Trinidadian* und *Belizian*, die dem *Caribbean Standard English* unterstehen.

Analog zu Kachrus Modell veranschaulicht McArthurs Modell sehr gut die verschiedenen regionalen Varietäten des Englischen. McArthur bezieht sogar Subvarietäten wie beispielsweise einige Pidgin- und Kreolsprachen und Ethnolekte (zum Beispiel das *African American Vernacular English*) mit ein. Zudem weist er auf die verschiedenen Entwicklungsstadien der *World Englishes* hin, indem er zwischen bereits standardisierten *Englishes* und sich noch standardisierenden *Englishes* unterscheidet.

McArthurs Modell muss dennoch gleichermaßen aus verschiedenen Gründen kritisiert werden. Die Existenz eines *World Standard English* (WSE) wird zwar bescheinigt, es wird jedoch nicht explizit erläutert, was diese Varietät sprachlich charakterisiert. Aus diesem Grund bleibt WSE ein abstraktes Konstrukt. Als weiterer Kritikpunkt muss erwähnt werden, dass das Modell sich ausschließlich auf die Darstellung von *ENL*- und *ESL*-Varietäten beschränkt und dass *EFL*-Sprachgebrauch nicht berücksichtigt wird. Ebenso wenig findet *ELF* Eingang in das Modell, da *ELF*-Sprachgebrauch nicht mit der Nutzung des *World Standard English* gleichgesetzt werden kann. Zudem werden zwar Pidgin- und Kreolsprachen aufgeführt, jedoch werden diese nicht von *ENL*- und *ESL*-Varietäten abgegrenzt. Zum Teil werden diese sogar den Standardvarietäten zugeordnet, obwohl ihr Standardisierungsprozess noch nicht abgeschlossen ist.

Ein in jüngerer Zeit sehr einflussreiches Modell stellt das *Dynamic Model of the Evolution of Postcolonial Englishes* von Schneider (2003, 2007) dar. Im Unterschied zu den eingangs beschriebenen Modellen wird in Schneiders Modell ein sprachhistorischer Prozess der *variety-formation* modelliert, der auf die Entstehung aller Varietäten des Englischen angewendet werden kann. Demnach müssen postkoloniale Varietäten verschiedene Phasen durchlaufen, bis sie eigenständige endonormative Standards herausbilden. Das Modell ist unterteilt in die Phasen *foundation*, *exonormative stabilization*, *nativization*, *endonormative stabilization* und *differentiation*.

Während der *foundation*-Phase emigriert eine signifikante Anzahl englischer Siedler in eine neugegründete Kolonie. Sie bringen die englische Sprache damit in ein Land, das bereits von einer indigenen Bevölkerung besiedelt ist und zu dem vorher kein Sprachkontakt bestand.

In der Folgezeit etabliert sich die Sprache in ihrem neuen Umfeld (Schneider 2007: 33ff.). In der zweiten Phase (*exonormative stabilization*) beginnt die politische, soziale und sprachliche Stabilisierung der neuen Gemeinschaft. Eine dialektale Konvergenz wird zwischen dem Englischen und den lokalen Sprachen erreicht. Die Siedler entlehnen Elemente der lokalen Sprachen in das Englische und die lokale Bevölkerung akzeptiert die Sprache der Kolonialmacht. Das Englische wird fortan für administrative Zwecke verwendet und die Kolonialiserten werden bilingual erzogen (Schneider 2007: 36ff.). In der dritten Phase (*nativization*) ereignen sich kulturelle und sprachliche Umwälzungen, die dazu führen, dass das Englische als Zweitsprache akzeptiert wird. Gleichzeitig streben die kolonialisierten Länder nach politischer Unabhängigkeit von den ehemaligen Kolonialherren (Schneider 2007: 40ff.). Die vierte Phase (*endonormative stabilization*) ist gekennzeichnet von der graduellen Akzeptanz einer eigenen etablierten Norm. In dieser Phase bilden politische und psychologische Freiheit notwendige Voraussetzungen um ein Selbstverständnis von einem eigenen endonormativen Standard zu schaffen (Schneider 2007: 48ff.). Die fünfte Phase (*differentiation*) schließt die endonormative Stabilisierung ab. Von diesem Zeitpunkt an können wiederum neue Prozesse in Gang gesetzt werden, die zu einer weiteren internen Ausdifferenzierung der Varietät führen (Schneider 2007: 52ff.).

Das *Dynamic Model* ermöglicht es, den sprachhistorischen Prozess der Entwicklung von Varietäten nachzuzeichnen. Eine Modellierung von *ELF*-Sprachgebrauch liegt hingegen nicht im Erkenntnisinteresse von Schneider, da er *ELF* nicht den Status einer Varietät zugesteht (Schneider 2012: 60).

Ein *World English*-Modell, das sich nicht auf eine historische und geografische Abbildung beschränkt, ist das überarbeitete *The Centripetal Circles of International English*-Modell von Modiano (1999) (vgl. Abbildung 3). Den Kern des Englischen bilden kompetente Nutzer des *English as an International Language (EIL)*, also Sprecher, die das Englische für kulturübergreifende Kommunikation nutzen. In diesem Kern befinden sich sowohl Muttersprachler als auch Nichtmuttersprachler des Englischen, die varietätenübergreifend einen gemeinsamen Kern an sprachlichen Merkmalen verwenden. Der Kreis, der den Kern umfasst, enthält weitere sprachliche Merkmale. Diese werden im Laufe der Zeit entweder von der *EIL*-Gemeinschaft genutzt und somit Teil des sprachlichen Kerns oder sie werden nicht genutzt und bleiben damit außerhalb des *EIL*-Kerns. Die äußeren Kreise können in fünf Gruppen unterteilt werden: *American English, British English, Foreign Language Speakers, Major varieties,*

Other varieties. Die aufgelisteten Varietäten besitzen eigene, varietätenspezifische sprachliche Merkmale, die nur von Sprechern der jeweiligen Varietät verwendet werden.

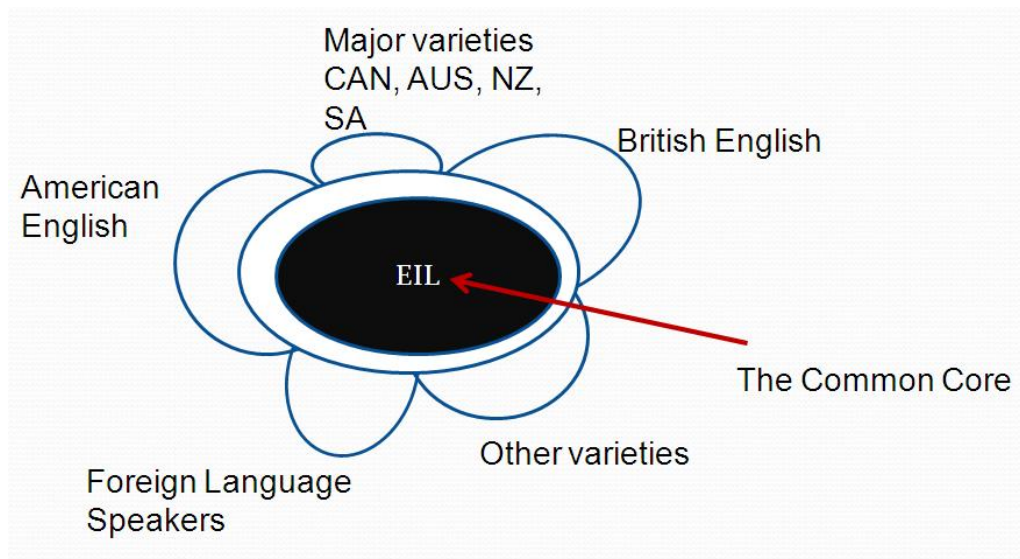


Abbildung 3: *The Centripetal Circle of International English* (Modiano 1999)

Ebenso wie andere *World English*-Modelle beschreibt Modiano zwar einen *EIL*-Sprachkern, jedoch kann auch er nicht im Detail erläutern, welche Charakteristika diesen Kern ausmachen. Diese Problematik ist im Zuge der Lingua Franca-Debatten der letzten Jahrzehnte immer wieder evident geworden. Trotz zahlreicher Versuche (vgl. Seidlhofer 2001, Jenkins 2000, 2007) konnte bisher noch kein gemeinsamer *EIL*-Kern an sprachlichen Merkmalen identifiziert werden.

Crystal (2003a) hat ein weiteres Modell entwickelt, das die englische Sprache in seiner Gesamtheit zu erfassen versucht. Crystal (2003a: 177f.) prognostiziert, dass zukünftig eine zunehmende Vielfalt an Varietäten existieren werde und somit eine Notwendigkeit für globale Verständigung bestehe. Aus diesem Grund müsse jeder Englischnutzer zwei Normen (ein lokales und ein globales Englisch) erlernen. Je nach Kontext müsse der Englischnutzer dann entscheiden, welches dieser beiden Varietäten angewendet werde. Die globale Varietät bezeichnet Crystal als *World Standard Spoken English (WSSE)*. Diese enthalte viele Elemente des geschriebenen Standard-Englisch und werde charakterisiert durch "careful pronunciation, conventional grammar, and standard vocabulary" (Crystal 2003a: 185). Crystal nennt explizit das Amerikanische Englisch als sehr einflussreich für die Entwicklung des *WSSE*.

Im Gegensatz zu den anderen beschriebenen *World English*-Modellen verschließt sich Crystals Modell nicht der Tatsache, dass Englisch vielfach in einem Lingua Franca-Kontext

eingesetzt wird. Den Gebrauch des Englischen als Lingua Franca fasst Crystal konkret unter dem *WSSE*-Konzept zusammen. *WSSE* beschreibt er als eine im Vergleich mit den lokalen Varietäten strukturell reduzierte Varietät des Englischen. Ein Kritikpunkt, der bei den anderen Modellen formuliert wurde, greift auch hier. Crystal liefert zum einen weder eine exhaustive Beschreibung der sprachlichen Merkmale des *WSSE*, noch erklärt er im Detail, auf welche Weise sich eine *WSSE*-Varietät in der Realität etabliert.

2.1.2.1. Wie lässt sich *ELF* in *World English*-Modellen verorten?

Obwohl sich die oben beschriebenen Modelle mehr oder weniger eignen, um *ENL*- und *ESL*-Varietäten oder *EFL*-Sprachgebrauch zu beschreiben, existiert bisher kein Modell, das *ELF*-Sprachgebrauch auf angemessene Weise abbildet. Als erhebliches Problem erweist sich, dass sich *ELF*-Sprachgebrauch, im Gegensatz zu den anderen Varietäten, nicht räumlich festlegen lässt, wie Mollin (2006b: 42) feststellt: "[...] we do not yet seem to have a consensus as to the location of such conversations in theoretical, conceptual space.". Aus diesem Grund wird *ELF*-Sprachgebrauch heutzutage oft losgelöst von den etablierten *ENL*-, *ESL*- und *EFL*-Varietäten betrachtet. So beschreibt Jenkins (2007: 2) *ELF* als "[...] an emerging English that exists in its *own right* and which is being described in its *own terms* rather than by comparison with *ENL*". *ELF*-Sprachgebrauch kann nur schwerlich mit dem bestehenden dreigliedrigen Konzept von *ENL*-, *ESL*- und *EFL*-Varietäten in Einklang gebracht werden, da sowohl varietätenübergreifend als auch innerhalb der Varietäten ein Lingua Franca-Sprachgebrauch stattfindet. Es stellt sich somit die Frage, wie *ELF*-Sprachgebrauch innerhalb dieser Taxonomie zu verorten ist.

Das Verhältnis zu bestehenden Standardnormen ist ein zentrales Charakteristikum für die Bewertung der Entwicklung von *ESL*- und *EFL*-Varietäten. Im *ESL*-Sprachgebrauch wird Abweichung von der *Native Speaker*-Norm häufig als Innovation auf dem Weg zur Herausbildung einer eigenen Norm betrachtet (vgl. Schneider 2003). Vor diesem Hintergrund stellt sich im Fall von *ELF*-Sprachgebrauch die Frage, ob Variation als Ausdruck von Innovation oder als Fehler zu bewerten ist. Nach Ansicht von Seidlhofer / Berns (2009: 190) wird Variation im *ELF*-Sprachgebrauch oftmals mit Fehlerhaftigkeit gleichgesetzt, da Englischlernern im Gegensatz zu Sprechern des *Inner Circles* und des *Outer Circles* nicht das Recht eingeräumt wird, Innovationen hervorzubringen.

Es zeigt sich, dass in der *ELF*-Forschung keine Einigkeit darüber herrscht, wie sich *ELF*-Sprachgebrauch zukünftig entwickeln wird. Während die Fraktion um Seidlhofer (2001),

Jenkins (2000) und Mauranen (2003) propagiert, dass *ELF*-Sprachgebrauch einen Kern aus gemeinsamen sprachlichen Merkmalen herausbilden werde, postulieren Yano (2001) und Mollin (2006b), dass *ELF* sich in verschiedene lokale Varietäten aufspalten werde. Yano (2001) ist davon überzeugt, dass das Englische sich in die sechs verschiedenen Lingua Franca-Standards *Euro-English*, *Asian English*, *Latin English*, *Arab English*, *African English* und *Anglo-American English* ausdifferenzieren wird. Jede dieser Varietäten werde beeinflusst durch lokale kulturelle Einflüsse und die Identität der Sprachteilnehmer. Es würden sich dadurch eigenständige Standardvarietäten mit inhärenten sprachlichen Charakteristika herausbilden, die sich durch die vergleichsweise häufige Verwendung im regionalen Kontext verfestigen würden (Yano 2001: 127ff.).

Ein weiterer Diskussionspunkt, der in den letzten Jahren aufgeworfen wurde, ist die Frage, inwieweit das unterschiedliche Sprachniveau der *ELF*-Nutzer einen Einfluss auf die Entwicklung einer möglichen *ELF*-Varietät hat. Jenkins (2009: 23) betont dabei, dass das Sprachniveau ein relevanterer Faktor für die Sprachentwicklung sei als die regionale Herkunft: "Degree of proficiency or expertise is an eminently (and possibly the most) useful way to approach the English of its entirety of speakers nowadays, regardless of where they come from and what other language(s) they speak.". Im Zuge dessen rückt auch die Rolle des Englischen als globale Interaktionssprache in einem Lingua Franca-Kontext zunehmend ins Zentrum des Erkenntnisinteresses, wie Pennycook (2011: 521) feststellt:

A plausible case can in fact be made that the ELF focus is trying to address precisely that gap left by the holes in the WE model: how to come to grips with a non-centrist understanding of English as an international language that is dependant neither on hegemonic versions of central English nor on nationally defined new Englishes, but rather attempts to account for the ever-changing negotiated spaces of current language use.

Dieser Paradigmenwechsel hat schließlich Eingang gefunden in ein neues, dreidimensionales *World English*-Modell (Yano 2009), das im Folgenden präsentiert wird.

2.1.2.2. Ein neues Modell von *English as an International Language*

Yano (2009) beschreibt mit seinem *Three-dimensional Model of English Use* ein neues Modell, das sowohl die Herkunft als auch die verschiedenen Sprachniveaus der Englischnutzer berücksichtigt (vgl. Abbildung 4).

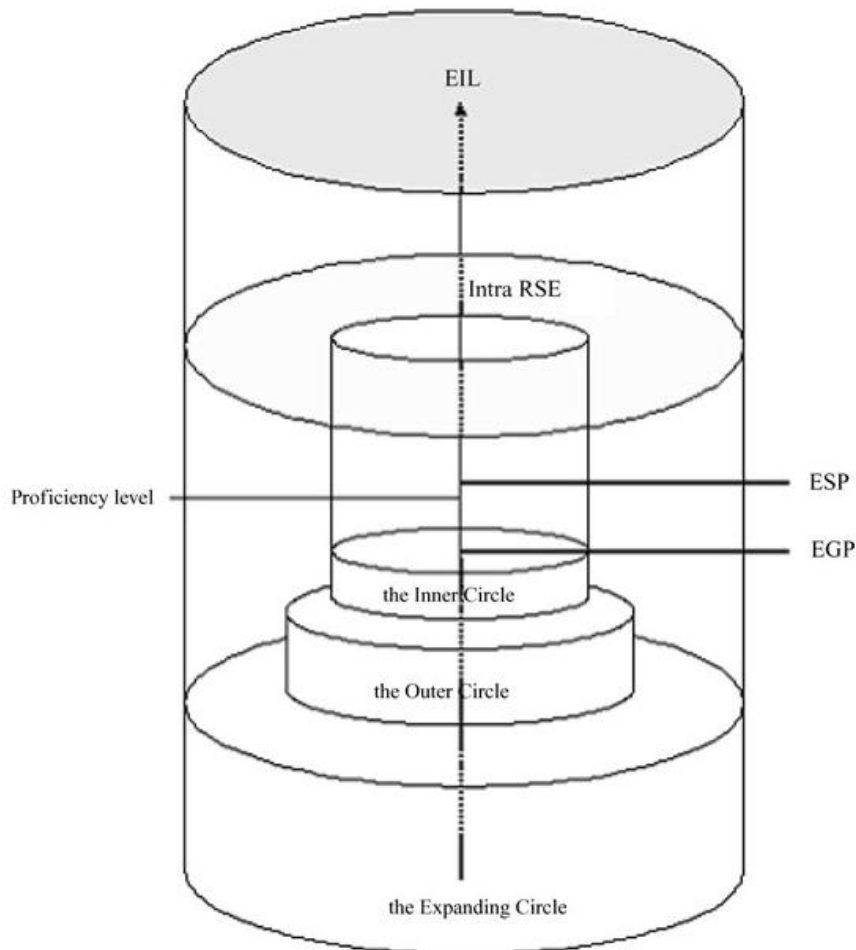


Abbildung 4: *Three-dimensional Model of English Use* (Yano 2007: 38)

Mit Hilfe des Modells kann grundsätzlich die Sprachkompetenz von Englischnutzern abgebildet werden. Der Pfeil in der Mitte des Modells weist das individuelle Sprachniveau des jeweiligen Sprachnutzers aus. Je weiter oben sich ein Englischnutzer auf diesem Pfeil befindet, desto größer ist seine Sprachkompetenz. Das Ziel eines nichtmuttersprachlichen Fremdsprachlerner aus dem *Expanding Circle* ist prinzipiell das Erreichen des Sprachniveaus eines erwachsenen Muttersprachlers aus dem *Inner Circle*. Dieses Sprachniveau bezeichnet Yano als *English for General Purposes (EGP)*. Lerner, die dieses Sprachniveau erreicht haben, be-

sitzen eine grundsätzliche Kompetenz auf Englisch kommunizieren zu können (Yano 2009: 250). Zum Erreichen des *English for Specific Purposes (ESP)*-Niveaus ist es nötig, dass Lerner professionelle und berufliche Fachausdrücke erwerben. Für ein Erreichen des *English for Specific Cultures (ESC)*-Niveaus, das sich auf derselben Ebene befindet, müssen Lerner kulturspezifische Ausdrücke erwerben. Gleichwohl schränkt Yano (2009: 251) ein, dass Lerner ein hohes Maß an *ESP*-Wissen haben können, ohne ein hohes Maß an *EGP*-Wissen zu besitzen. Yano (2009: 251f.) betont dabei explizit, dass alle Englischnutzer (auch *ENL*- und *ESL*-Sprecher) *ESP*- und *ESC*-Wissen erst erwerben müssen.

Das *Intra-Regional Standard English (Intra RSE)* weist Yano als nächsthöhere Stufe aus. Die Sprachnutzer, die diese Stufe erreicht haben, sind befähigt zur Kommunikation mit Sprachnutzern regionaler Englischformen (z.B. Englisch in Asien). Die Spitze des Sprachniveaus bildet schließlich *English as an International Language (EIL)*. Ein idealisierter Sprachnutzer auf diesem Sprachlevel ist zu interregionaler globaler Kommunikation befähigt und weist dabei ein hohes Level an *ESP* und *ESC* Kompetenz auf (Yano 2009: 251). Yano betont dabei ausdrücklich, dass *EIL* nicht als eine eigenständige Varietät des Englischen zu verstehen ist, sondern als ein loses plurizentrisches Konglomerat von verschiedenen nationalen *Englishes* (z.B. *German English*) begriffen werden muss.

Yanos Modell zeichnet ein idealisiertes Bild des internationalen Gebrauchs des Englischen. Es illustriert sehr gut die verschiedenen Lernschritte, die im Laufe des Spracherwerbs erreicht werden müssen, um ein Sprachniveau zu erreichen, das es Sprachnutzern ermöglicht, in einem globalen Kontext erfolgreich auf Englisch kommunizieren zu können. Das Besondere an Yanos Modell ist, dass keine explizite Trennung zwischen muttersprachlichen und nicht-muttersprachlichen Englischlernern vorgenommen wird. Vielmehr zielt es auf die Abbildung des erreichten Sprachkompetenzniveaus von Englischlernern ab. Das Ziel jedes Lerners ist dabei die Erreichung eines Sprachniveaus, das es ihnen ermöglicht sich in einem internationalen Kontext austauschen zu können. Aus diesem Grund bevorzugt Yano die Verwendung des Begriffs *EIL* an Stelle von *ELF*.

Yano distanziert sich von den traditionellen geografischen Unterscheidungen, die in anderen *World English*-Modellen gemacht werden und fokussiert sich stattdessen auf das Sprachniveau von Englischnutzern, das für ihn von viel zentralerer Bedeutung ist. Aus diesem Grund eignet sich das Modell besser als die vorher genannten Modelle für die Abbildung von Websprache, da im Medium Internet auch die physischen Grenzen der Webnutzer verschwimmen.

Nichtsdestotrotz ist auch dieses Modell nur bedingt auf Websprache anzuwenden, da die Anonymität des Mediums es erschwert, das Sprachniveau einzelner Webnutzer aufzuzeigen.

Insgesamt zeigt sich, dass kein einziges *World English*-Modell die sprachliche Realität des Webs vollständig abbildet. Von allen Modellen kann Yanos Modell noch am ehesten den Sprachgebrauch im Web widerspiegeln. Im Web treten kulturelle Unterschiede zwischen Webnutzern in den Hintergrund. Der Fokus wird stattdessen auf die erfolgreiche Kommunikation untereinander gelegt, unabhängig vom Kompetenzniveau der Webnutzer. Vor diesem Hintergrund wird ersichtlich, dass bestehende Modelle es nicht vermögen, Webnutzer voneinander zu unterscheiden. Weder der kulturelle Hintergrund noch die sprachliche Kompetenz von Webnutzern kann in jedem Fall zweifelsfrei festgestellt werden. Dies ist ein Kernmerkmal des Webs.

Eine Unterscheidung der einzelnen Webnutzer liegt nicht im Erkenntnisinteresse der vorliegenden Arbeit. Im Fokus steht vielmehr die Beschreibung von einzelnen Web-Genres in ihrer Gesamtheit, unabhängig von einzelnen Webnutzern. Die vorliegende Arbeit basiert auf der Prämisse, dass alle Formen des hier untersuchten Sprachgebrauchs im Web einen *Lingua Franca*-Sprachgebrauch darstellen. Da es diverse Definitionen des *Lingua Franca*-Begriffs gibt, soll zunächst eine Eingrenzung des Begriffs geleistet werden.

2.2. Englisch als *Lingua Franca*-Konzeption

Der Terminus *Lingua Franca* ist in den letzten Jahren in diversen Kontexten, oftmals in undifferenzierter Weise, verwendet worden. Um konzeptionelle Missverständnisse zu vermeiden, muss daher eine genauere Eingrenzung des Konzeptes geleistet werden. Zunächst wird dazu der etymologische Ursprung des Begriffs skizziert. Im Anschluss wird der *Lingua Franca*-Begriff, der dieser Arbeit zu Grunde liegt, explizit erörtert werden.

Ursprünglich bezeichnete *Lingua Franca* eine Pidginsprache, die auf dem Italienischen fußte und die Einflüsse von einer Vielzahl von Sprachen (Griechisch, Arabisch, Spanisch, Französisch und Türkisch) aufwies. Sie diente den sprachlich und kulturell unterschiedlichen Mittelmeernationen vom 14. Jahrhundert bis ins 19. Jahrhundert als gemeinsame Handels- und Verkehrssprache (Schuchardt 1909, Knapp 2002). Die ursprüngliche Bedeutung des *Lingua Franca*-Begriffs wurde im Laufe der Zeit erweitert, so dass *Lingua Franca* schließlich als Bezeichnung für eine Verkehrssprache zwischen Sprechern unterschiedlicher Muttersprachen diente (Jenkins 2007).

Die Verwendung einer Lingua Franca als internationale Verkehrssprache ist kein neuzeitliches Konzept. So diente bereits in der Antike Latein als Lingua Franca für den westlichen Teil des römischen Imperiums, Griechisch füllte diese Rolle für dessen östlichen Teil aus. Das Französische fungierte vom 17. Jahrhundert bis ins 20. Jahrhundert hinein als Sprache der Diplomatie in Europa (Barotchi 1994). In allen Epochen in allen Teilen der Welt wurden stets regionale und überregionale Lingua Francae verwendet. Jedoch haben die neuen Medien und die globale Vernetzung in den letzten Jahrzehnten zu einer zunehmenden Nutzung von Lingua Francae geführt. Die politischen und wirtschaftlichen Entwicklungen auf globaler Ebene in den letzten Jahrhunderten haben schließlich dazu geführt, dass heutzutage diverse Lingua Francae weltweit existieren, die in verschiedenen Regionen unterschiedliche Funktionen ausfüllen. *ELF* füllt dabei im globalen Kontext eine besondere Rolle aus.

Gleichzeitig muss konstatiert werden, dass innerhalb der Forschung eine Uneinigkeit darüber herrscht, wie genau *ELF*-Sprachnutzer zu definieren sind. So gibt es eine allgemeine inklusive Definition, die Kommunikation zwischen Sprechern unterschiedlicher Muttersprachen als *ELF*-Sprachgebrauch definiert (vgl. Jenkins 2007). Den Gegenpol bildet eine exklusive Definition, in der Muttersprachler ausdrücklich ausgeschlossen werden, so dass nur Konversationen zwischen Nichtmuttersprachlern als *ELF*-Sprachgebrauch gewertet werden (vgl. Baumgarten 2011). Eine sehr allgemeine inklusive Definition liefert Jenkins (2007: 1), die Lingua Franca als "[...] a contact language used among people who do not share a first language, and is commonly understood to mean a second (or subsequent) language of its speakers" definiert. Diese Definition stellt den Kommunikationsaspekt in den Mittelpunkt. Die Bedeutung von Muttersprachlern wird dabei marginalisiert. Firth (1996) liefert eine weitere inklusive Definition, die sich von der Definition Jenkins (2007) dadurch unterscheidet, dass sie um das Konzept der Kultur ergänzt wird. Firth (1996: 240) definiert *ELF* als "[...] a contact language between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen foreign language of communication".

Vor dem Hintergrund der vorliegenden Arbeit, die sich mit Websprache beschäftigt, erweist sich eine exklusive Lingua Franca-Definition als ungeeignet. Für das Web muss eine inklusive Definition des Lingua Franca-Konzepts verwendet werden, da durch die Anonymität des Mediums Internet oftmals nicht nachgewiesen werden kann, ob ein Text von Muttersprachlern oder von Nichtmuttersprachlern verfasst wurde.

2.2.1. *English as a Lingua Franca vs. English as an International Language*

Ein wesentliches Problem in der Lingua Franca-Debatte stellt nicht nur die genaue Definition des Lingua Franca-Begriffs dar, sondern auch dessen Abgrenzung von anderen Konzepten, die zum Teil mit *ELF*-Sprachgebrauch gleichgesetzt werden. So wird eine Vielzahl an alternativen Termini verwendet, um dem *ELF*-Konzept Ausdruck zu verleihen, so zum Beispiel *English as an International Language (EIL)* (Modiano 1999, Kirkpatrick 2007), *English as a Global Language (EGL)* (Crystal 2003a) oder *World English* (Brutt-Griffler 2002). In einigen dieser Konzepte wird Lingua Franca-Sprachgebrauch gleichgesetzt mit globalem Sprachaustausch, während er in anderen Fällen gleichgesetzt wird mit der Herausbildung von lokalen *Englishes* (vgl. Canagarajah 2006). Aus diesem Grund muss im Folgenden eine genauere Eingrenzung des *ELF*-Konzeptes geleistet werden. Dazu wird zunächst ein allgemeiner Überblick über die aktuelle Debatte gegeben.

Bruthiaux (2003) bewertet Englisch, das in einem Lingua Franca-Kontext entsteht, als ein restriktives *English for Specific Purposes (ESP)*-Register, das von einer kleinen Gruppe von Sprechern in einem sehr speziellen Themenschwerpunkt verwendet wird. Eine stabile *ELF*-Varietät könne dabei nicht entstehen. Brutt-Griffler (1998: 382) glaubt, dass sich weltweit verschiedene globale Register (wie z.B. *Legal English*) entwickeln, und somit eine große Bandbreite an unterschiedlichen *EILs* existiert. Widdowson (1997) stimmt dieser Betrachtung zu, merkt dabei jedoch kritisch an, dass Lingua Franca-Sprachgebrauch oftmals nicht hinreichend differenziert wird. Er fordert daher eine klare Trennung von Register (*ESP*)-Sprachgebrauch und dialektalem (*ENL, ESL, EFL*) Sprachgebrauch, da sich *ESP*-Sprachgebrauch nicht einer bestimmten geografischen Region zuordnen lasse. Widdowson (1998: 400) sieht zudem eine unterschiedliche Entwicklung für diese beiden Formen des Sprachgebrauchs voraus. Er prognostiziert, dass sich Varietäten im dialektalen Sprachgebrauch sprachlich zunehmend voneinander unterscheiden werden, während der Sprachgebrauch in verschiedenen Registern sich zunehmend aneinander anpassen werde.

Das Entstehen einer *ESP*-Sprachsituation wird von Hüllen (1987) näher beschrieben. Sprachnutzer finden sich demnach in einer *Intersociety* zusammen. Dies ist eine Sprechergemeinschaft, die über keine gemeinsame Kommunikationsgeschichte, begrenzte Erfahrungen und eine eigene spezifische Kulturcharakterisierung verfügt. Diese *Intersociety* beinhaltet Sprecher unterschiedlicher Nationalitäten und Muttersprachen. In ihr wird eine gemeinsame Sprache zur Verständigung genutzt. Die kulturellen Merkmale der einzelnen Mitglieder werden in dieser Sprechergemeinschaft bewusst außen vor gelassen. Die Kommunikation ist zum

einen gekennzeichnet durch Interferenz der jeweiligen Muttersprachen, zum anderen durch die Aushandlung von Bedeutungen. Das Ziel der Kommunikation in einer *ESP*-Sprachsituation ist es, dass sich die Mitglieder schnell, effizient und verständlich über ein spezifisches Thema austauschen können. House (1999) meint, dass kulturspezifischer Diskurs interaktional und kommunikativ irrelevant für eine *Intersociety* ist. Sie folgert daraus, dass somit in einer *Intersociety* auch die muttersprachlichen Normen der Mitglieder unwichtig sind. Wie Swales (1990) dabei richtigerweise feststellt, hat die Bildung von Sprechergemeinschaften auch einen Einfluss auf die schematische Struktur des Diskurses sowie auf Inhalt und Stil der Kommunikation. Diese Kategorisierung von Kommunikationsteilnehmern ist auch für Linguisten eminent wichtig, da die Einteilung in bestimmte Sprechergemeinschaften (*Communities*) hilft, ihr Sprachverhalten zu interpretieren und die Sprachvariation zwischen verschiedenen Sprechergemeinschaften zu erklären (vgl. Hyland 2006). Das Konzept der *Intersociety* ähnelt insofern in seiner Konzeption dem Konzept der *Community of Practice* (vgl. Lave / Wagner 1991, Meyerhoff 2004). Eckert / McConnell-Ginet (1992: 464) definieren diese als "[...] an aggregate of people who come together around mutual engagement in an endeavour... practices emerge in the course of this mutual endeavour".

Unabhängig vom sprachlichen Hintergrund der Teilnehmer, kann es somit in den einzelnen *Web Communities* zur Bildung eines *ESP*-Registers kommen, in dem sich eigene Identitätsmerkmale entwickeln können, wie Crystal (2001: 60) hervorhebt:

If real Internet communities are relatively small-scale, they will demonstrate their solidarity by evolving (consciously or unconsciously) measures of identity, some of which will be nonlinguistic (e.g. shared knowledge, a particular morality) and some linguistic in character. The linguistic features will take time to evolve, but eventually they will provide the community with an occupational dialect which newcomers will have to learn if they wish to join it.

Das Konzept der *Intersociety* weist Parallelen auf zu Seidlhofers Verständnis von *ELF*-Sprachgebrauch. Seidlhofer (2001: 140f.) weist ausdrücklich darauf hin, dass *ELF*-Sprachnutzer üblicherweise einen anderen Zweck verfolgen als Lerner des Englischen:

ELF speakers are usually not [...] concerned with emulating the way native speakers use their mother tongue within their own communities, nor with

socio-psychological and ideological meta-level discussions. Instead, the central concerns for this domain are efficiency, relevance and economy in language learning and language use.

In den letzten Jahren wurden in der Varietätenforschung die Begriffe *ELF* und *EIL* nicht ausreichend differenziert verwendet. Dabei bestehen durchaus konzeptionelle Unterschiede zwischen *EIL* und *ELF*. Dröschel (2011: 52) unterscheidet zwischen *EIL* als Lingua Franca-Kommunikation im globalen Kontext und *ELF* als Lingua Franca-Kommunikation in lokalen Kontexten. Sie plädiert für eine strikte Trennung der Begriffe *ELF* und *EIL*. *ELF* versteht Dröschel (2011) als dialektale intraregionale Varietät und *EIL* als globales Register. Widdowson (1997: 143) definiert *EIL* als "[...] a range of self-regulating registers for international use". Modiano (2009: 209) ergänzt, dass *EIL* interkulturelle Kommunikation in einer multi-kulturellen Umgebung darstellt.

Erste Ansätze zur Beschreibung der verschiedenen *EIL*-Register sind bereits unternommen worden. So identifiziert Baumgarten (2011) idiosynkratische sprachliche Merkmale für verschiedene *EIL*-Register. Anhand dieser Merkmale können schließlich Register voneinander unterschieden werden (beispielsweise das Register Wirtschaftssprache von dem Register medizinische Fachsprache). Allerdings weist Baumgarten (2011) explizit darauf hin, dass die Verwendung des jeweiligen Registers abhängig ist von der Kommunikationssituation jedes einzelnen Sprechers. Dies setzt voraus, dass die Identität der Kommunikationsteilnehmer bekannt sein muss. Diese Forderung lässt sich im Hinblick auf Websprache nicht erfüllen, da medienbedingt die Autorenschaft der Webnutzer in den meisten Fällen nicht bekannt ist. Das Web stellt in dieser Hinsicht somit ein Sonderfall dar, den es genauer zu beleuchten gilt.

2.2.2. Befindet sich *ELF* auf dem Weg, stabile sprachliche Merkmale auszubilden?

Die Frage nach der Stabilität eines *ELF*-Sprachgebrauchs hat in jüngerer Vergangenheit hitzige Diskussionen innerhalb der Varietätenforschung entfacht. Im Hinblick auf diesen Aspekt haben sich zwei Lager herausgebildet. Zum einen gibt es Verfechter für die These, dass *ELF* im Prozess ist genuine sprachliche Merkmalen herauszubilden (vgl. Seidlhofer 2001, Jenkins 2000, 2007, Mauranen 2005). Darüber hinaus propagiert diese Gruppe, dass *ELF*-Sprachgebrauch Einzug in bestehende Lehrwerke an Schulen finden müsse. Anhand von speziellen Lingua Franca-Korpora (*VOICE*, *ELFA*) versuchen Seidlhofer (2001) und Mauranen (2005) sprachliche Merkmale von *ELF*-Sprachgebrauch quantitativ zu beschreiben und daraus Uni-

versalien abzuleiten. Im Kern haben Jenkins (2000), Seidlhofer (2001) und Mauranen (2005) dabei eine Liste von inhärenten phonetischen, lexikalischen, grammatischen und pragmatischen Merkmalen erstellt, die *ELF* als eigenständigen Sprachgebrauch markieren sollen. Dieser unterscheidet sich von *ENL*-Varietäten insofern, als *ELF*-Nutzer sprachliche und kulturelle Eigenheiten ihrer eigenen Muttersprachen in ihren Sprachgebrauch mit einfließen lassen. Aus diesem Konglomerat an sprachlichen und kulturellen Einflüssen entstehe schließlich ein neuer Sprachgebrauch. Als ein vielzitiertes Beispiel fungiert der Wegfall des 3. *Prs. Sg. -s* bei Verben (Seidlhofer 2001). Breiteneder (2009) hat in ihrer Studie dieses Phänomen im *VOICE*-Korpus untersucht. Ihrem Ergebnis nach lassen einige untersuchte *ELF*-Sprecher demnach das 3. *Prs. Sg. -s* in manchen Fällen wegfällen. Breiteneder (2009: 256) vergleicht diese Tatsache mit Tendenzen, die sich auch in anderen Varietäten des Englischen nachweisen lassen: "the occurrence of this feature follows general principles of language usage that have been observed in numerous varieties of English around the globe and indicate affinities between ELF and various world Englishes (WEs)". Die beobachteten Parallelen belegen ihrer Meinung nach, dass *ELF* sich in einem Indigenisierungsprozess befinde (vgl. Schneider 2003). Gleichzeitig gibt sie einschränkend zu, dass die Lingua Franca-Nutzer in ihrer Untersuchung in 80% aller Fälle das 3. *Prs. Sg. -s* Morphem regelkonform zur muttersprachlichen Standardnorm verwendet haben. Daher räumt sie ein: "[...] ELF usage in fact very largely corresponds to standard English norms" (Breiteneder 2005: 10).

In der Vergangenheit wurden auf unterschiedlichen Ebenen bereits einige Versuche unternommen, einen Lingua Franca-Sprachkern herauszuarbeiten. Peyawary (1999) hat beispielsweise aus verschiedenen Korpora einen gemeinsamen *EIL*-Vokabelkern herauskristallisiert.⁴ Jenkins (2000) hat einen phonetischen *ELF*-Kern identifiziert, der ausschließlich Phoneme beinhaltet, die essentiell sind für das sprachliche Verständnis. Der dentale Frikativ, der als /θ/ oder /ð/ realisiert werden kann, bereitet beispielsweise vielen nichtmuttersprachlichen Lernern Schwierigkeiten. Aus diesem Grund ist es nicht Teil dieses Kerns, da alternative sprachliche Realisierungen des Phonems, beispielsweise als /f/, in der Regel nicht zu Bedeutungsveränderungen und damit nicht zu Missverständnissen führen. Auch auf der pragmatischen Ebene wurden Eigenheiten des *ELF*-Sprachgebrauchs beschrieben. So führt Jenkins (2000) aus, dass es einige sprachliche Konstruktionen gibt, die im Gegensatz zum muttersprachlichen Sprach-

⁴ Es zeigt sich, dass in verschiedenen Studien die Begriffe *EIL* und *ELF* immer wieder synonym verwendet werden. Konzeptionell entspricht Peyawarys (1999) *EIL*-Begriff dem Lingua-Franca-Konzept, so wie es in der vorliegenden Arbeit verwendet wird.

gebrauch irrelevant sind für *ELF*-Sprachgebrauch. Dazu zählt sie "idiomatic usage, slang, phrasal verbs, puns, proverbs, cultural allusions and the like" (Jenkins 2000: 220).

Um die Ergebnisse der einflussreichen Studien von Seidlhofer (2001, 2004), Mauranen (2003, 2005) und Jenkins (2000, 2007) richtig einordnen zu können, muss nochmals auf ihr zugrundeliegendes Verständnis von *ELF*-Sprachgebrauch verwiesen werden. In all diesen Studien werden englische Muttersprachler explizit vom *ELF*-Sprachkontakt ausgeschlossen. Daher kritisiert Prodromou (2007) diese Studien, da durch die Exklusion von englischen Muttersprachlern ein gewichtiger Teil des weltweiten *ELF*-Sprachgebrauchs vollends ignoriert werde. Für diese Art des *ELF*-Sprachgebrauchs würden dementsprechend die sprachlichen Merkmale eines *ELF*-Kerns nicht greifen. Zudem weist Prodromou (2008: 32) auf die widersprüchlichen Aussagen von Seidlhofer und Jenkins hin, da sie einerseits die Existenz einer monolithischen *ELF*-Varietät abstreiten würden, andererseits fortwährend einen universellen *ELF*-Kern mit eigenen präskriptiven Normen beschreiben würden.

Neben den Befürwortern der Idee, dass *ELF* stabile sprachliche Merkmale entwickelt, gibt es eine Vielzahl von Varietätenforschern, die dieser Betrachtungsweise vehement widersprechen. Gnutzmann / Intemann (2008) stimmen Seidlhofer zwar insofern zu, dass *ELF*-Sprachnutzer sich in ihrem Sprachgebrauch einander anpassen (*speech accommodation*), allerdings sind sie davon überzeugt, dass dies nur bis zu einem gewissen Grad geschieht; Sprachnormen würden weiterhin eingehalten werden. So folgern sie: "[...] native speakers and the majority of proficient non-native speakers will not stop inflecting 3rd person singular verbs" (Gnutzmann / Intemann 2008: 15). Einschränkend räumen Gnutzmann / Intemann (2008: 15) jedoch ein:

However, it can be envisaged that native speakers will make allowance on the pragmatic level of communication by adjusting their grammar and lexis as well as refraining from idioms and colloquial expressions, i.e., reducing and simplifying its complexity in ELF communication, in order to ensure communicative success.

Es sei demnach möglich, dass Muttersprachler zumindest ihre Grammatik und Lexik an den *ELF*-Sprachgebrauch anpassen würden. Beispielsweise würden Muttersprachler davon absehen, Idiome und kolloquiale Ausdrücke zu verwenden, um die Komplexität in der *ELF*-Sprachsituation zu reduzieren. Haegeman (2002) bestätigt, dass sich Sprecher mit höherer

Kompetenz dem Sprachniveau ihrer Kommunikationspartner anpassen, um Kommunikationsprobleme zu vermeiden.

Es stellt sich allgemein die Frage, inwieweit im *ELF*-Sprachgebrauch eine sprachliche Reduktion stattfindet und auf welchen unterschiedlichen Ebenen diese beobachtet werden kann. Meißner (2005: 138) sieht das Englische in Europa auf bestem Wege "von einer Fremdsprache zu einer *Sekundärsprache* (sekundäre Vehikularsprache) mit reduzierter kommunikativer Funktion zu werden". Meierkord (2005) untersucht interkulturelle Interaktionen zwischen Sprechern unterschiedlicher Muttersprachen mit divergierenden Sprachniveaus. Dabei stellt sie fest, dass das gemeinsam verwendete Vokabular nur gering von den indigenisierten *Englishes* der an der Kommunikation beteiligten Sprecher beeinflusst wird (Meierkord 2005: 101). Als weitere sprachliche Besonderheit nennt sie die geringe Nutzung von Idiomen und *Phrasal verbs*. Zudem wird ein insgesamt reduzierter Wortschatz verwendet. Dies schlägt sich unter anderem in einem niedrigen *Type-Token*-Verhältnis und einer geringen lexikalischen Dichte nieder. Zusammenfassend beschreibt Meierkord (2005) interkulturellen Sprach-austausch als lexikalisch reduziert und kulturell neutral.

Mukherjee (2005) prägt den Terminus *English as a Global Pidgin (EGP)*. *EGP* stelle eine von der Kultur der Nutzer losgelöste Sprachgebrauchssituation des Englischen dar, die formal und funktional reduziert ist. Als Beispiel führt er den Gebrauch des Englischen als Wissenschaftssprache auf. Crystal (2003: 58) sieht bei der Betrachtung von *World Standard Spoken English (WSSE)* ebenfalls Reduktionstendenzen: "[...] sentence construction is simplified, idioms and slang vocabulary are avoided, linguistic tempo is reduced and assimilations, elisions and weak forms are avoided". Trotz der Heterogenität der Termini (*ELF*, interkulturelle Kommunikation, *EGP*, *WSSE*) zeichnet sich in einer interkulturellen Sprachsituation insgesamt eine gemeinsame Tendenz zur Sprachreduktion ab.

Ein entscheidender Faktor für die Herausbildung einer *ELF*-Varietät ist die Beständigkeit einer *ELF*-Sprechergemeinschaft. Durch konstanten Sprachgebrauch innerhalb einer Sprechergemeinschaft bilden sich üblicherweise Sprachmerkmale heraus, die charakteristisch für diese Gemeinschaft sind. Im Fall von *ELF*-Sprechergemeinschaften zeichnet sich jedoch ein anderes Bild ab, da diese nur sehr lose Sprechergemeinschaften darstellen, die sich für einen bestimmten Zweck ad hoc zusammenfinden. Meierkord (2004: 115) zweifelt an der Stabilität von *ELF*-Sprechergemeinschaften:

[...] unlike dialect contact, ELF cannot be conceived as a permanent form of English, since its users do not constitute a stable community. Rather, it is a

variety in constant flux, involving different constellations of speakers of diverse individual Englishes in every single interaction.

Ferguson (2009: 121) unterstreicht ebenfalls den unstabilen Charakter von *ELF*-Sprecher-gemeinschaften und glaubt nicht, dass diese sich in einem Nativisierungsprozess im Sinne von Schneider (2003) befinden. *ELF*-Kommunikation weise eine zu große Variation auf, als dass sich daraus zukünftig eine stabile *ELF*-Varietät entwickeln könne. Ferguson (2009: 129f.) glaubt, dass selbst im Falle, dass *ELF* im Laufe der Zeit eine kodifizierte Form entwickeln sollte, diese dennoch nicht zwangsläufig von den Sprachverwendern als Varietät akzeptiert werden würde. House (2006: 88) ergänzt in diesem Zusammenhang: "In ELF there is no consistency of form that goes beyond participant level". James (2008: 141) spricht einer potenziellen *ELF*-Varietät die Existenzberechtigung ab, indem er *ELF*-Sprachgebrauch als "[...] fragmented, contingent, marginal, transitional indeterminate, ambivalent and hybrid in various ways" beschreibt. Maley (2010: 31) hinterfragt die Sinnhaftigkeit der Suche nach globalen *ELF*-Universalien: "The aggregate of NNS-NNS interactions globally does not add up to a speech community. It is no more than an inchoate and disconnected agglomeration of instances of use". Prodromou (2007: 49) verweist darauf, dass die Ergebnisse aus den bisherigen Studien, die die Existenz von *ELF*-Normen proklamieren, auf einer dünnen Datenbasis fußen und daher nicht verallgemeinert werden können:

What has been published (Seidlhofer, 2004) is based on a small sample of users of ELF which does not amount to evidence that there are common norms in 'Euro-English', and even less so in the huge range of contexts in which English is used globally.

Die Kritiker einer *ELF*-Norm wehren sich gegen ein präskriptives *ELF*-Varietätenmodell, da sich ihrer Ansicht nach das Englische in *ELF*-Sprachsituationen durch enorme sprachliche Unterschiede der Sprachkompetenz auszeichnet. Daher meinen sie, dass Lingua Franca-Sprachgebrauch nicht zwangsläufig Ausdruck von simplifiziertem, reduziertem Englisch sein muss.

2.3. Zusammenfassung

Der Fokus der vorliegenden Arbeit liegt auf der Analyse von Sprachgebrauch im Web. Das Web erlaubt es jedem Sprachverwender mit Internetzugang, virtuell zu kommunizieren. Innerhalb des Webs verschwimmen jedoch die physischen Grenzen regionaler Englischvarietäten. Außerdem ist im Web die Identität der Webnutzer oft ungeklärt, so dass häufig die Herkunft der Webnutzer nicht ersichtlich ist. Aus diesem Grund können bestehende regionale *World English*-Modelle, wie das von Kachru (1985) oder das von McArthur (1998), nicht auf den Sprachgebrauch im Web angewendet werden. Im Gegensatz zu diesen Modellen konzentriert sich Yanos Modell auf die Abbildung des Sprachniveaus von Sprechern des Englischen, unabhängig davon, ob diese *ENL*-, *ESL*- oder *EFL*-Sprecher darstellen und welchen kulturellen Hintergrund sie haben. In dieser Arbeit stehen jedoch nicht einzelne Webnutzer im Erkenntnisinteresse, sondern typische Webnutzer in ihrer Gesamtheit.

Der vorliegenden Arbeit liegt ein Lingua Franca-Konzept im Sinne von Dröschel (2011) zu Grunde. Demnach befürworte ich eine klare Trennung zwischen lokalem Lingua Franca-Sprachgebrauch (*ELF*) und globalem Lingua Franca-Sprachgebrauch (*EIL*). Dabei erscheint mir vor dem Hintergrund der Untersuchung von Websprache die Verwendung des *EIL*-Begriffs sinnvoll. *EIL*-Sprachgebrauch ist in dieser Arbeit dabei keineswegs gleichzusetzen mit dem Konzept eines globalen Lingua Franca-Sprachgebrauchs mit stabilen sprachlichen Merkmalen im Sinne von Seidlhofer (2001) und Jenkins (2000). Vielmehr verstehe ich den untersuchten Sprachgebrauch ausdrücklich als ein Konglomerat von *ESP*-Registern im Medium Internet. Der kulturelle Hintergrund der untersuchten Sprachnutzer spielt im Web eine untergeordnete Rolle, ebenso wie die Frage, ob diese Muttersprachler oder Nichtmuttersprachler des Englischen sind. Die Anonymität des Mediums lässt nur schwer auf die Sprachkompetenz von Webnutzern schließen. Da Webnutzer oft weder wissen, mit wem sie kommunizieren noch wie hoch die Sprachkompetenz der anderen Nutzer ist, ist es durchaus denkbar, dass sie ihre eigene Sprachproduktion vorsorglich adaptieren. So könnten sie lexikalisch und syntaktisch reduzierte Texte produzieren, um sicherzustellen, dass die Kommunikation erfolgreich sein wird.

Nicht die Herkunft der Sprachnutzer oder ihr kultureller Hintergrund, sondern der thematische Kontext ist ausschlaggebend für die Textproduktion. Abhängig vom gewählten Thema lassen sich unterschiedliche sprachliche Besonderheiten finden. Es ist demnach davon auszugehen, dass sich die Sprachverwendung in einem medizinischen Fachforum von der Sprachverwendung in einem Fussballfanforum unterscheidet. Die Texte, die für die vorliegende

Arbeit analysiert wurden, sind allesamt im Themenbereich *Medical Surgery / Medical Tourism* entstanden. Diese Tatsache hat vermutlich einen entscheidenden Einfluss auf die Sprachverwendung in den untersuchten Web-Genres.

Das vorliegende Kapitel behandelte im Detail die Frage nach der Konzeption von Lingua Franca-Sprachgebrauch im Web. Dabei zeigte sich, dass sich der Sprachgebrauch im Web anhand bestehender Modelle nur schlecht abbilden lässt. Insbesondere die Eigenheiten des Mediums Internet unterscheiden den Sprachgebrauch im Web von traditionellem mündlichem und schriftlichem Sprachgebrauch. Aus diesem Grund werde ich im nächsten Kapitel genauer auf die medialen und sprachlichen Besonderheiten des Webs eingehen und die Gemeinsamkeiten und Unterschiede zu typisch mündlichen und schriftlichen Texten aufzeigen.

3. Das World Wide Web und Sprache

If we are to understand truly how the Internet might shape our language, then it is essential that we seek to understand how different varieties of language are used on the Internet. (Paolillo 1999)

3.1. Die Entwicklung des World Wide Web

Allgemein wird die erste Phase des Internetzeitalters vom Anfang der 1990er Jahre bis zum Ende des Millenniums als Web 1.0 bezeichnet (vgl. O'Reilly 2007). In den Anfangsjahren wurde das Web in erster Linie dazu verwendet nach Informationen zu suchen und mit anderen Webnutzern zu kommunizieren (beispielsweise über E-Mails oder *Instant Relay Chat*). Webseiten wurden üblicherweise von professionellen Programmierern angefertigt. Für den durchschnittlichen Webnutzer war es zu dieser Zeit mit viel Aufwand verbunden eigene Webpräsenzen zu erstellen.

Seit dem Ende der 1990er Jahre wurden die Funktionen des Webs auf vielfache Weise erweitert, so dass diese Zeit den Beginn des Web 2.0 markiert (vgl. O'Reilly 2007). Das Web 2.0 setzt den Fokus auf Kollaboration, Partizipation und Kreativität seitens der Webnutzer (Warschauer / Grimes 2007). So entstand in dieser Zeit beispielsweise das Web-Genre *Web Logs (Blogs)*. Diese fungieren als öffentliche Tagebücher, in denen Webnutzer sich zu allen möglichen Themen äußern können. Auch die Entstehung von Wikipedia ist eng mit dem Web 2.0 verknüpft. Wikipedia ist eine öffentliche, frei zugängliche Enzyklopädie, deren Beiträge in einem Kollaborationsprozess von einer Vielzahl von Webnutzern entstehen. Im Kontext des Web 2.0 entstanden des Weiteren *Social Network Sites*. Diese stellen ein global bahnbrechendes Phänomen dar, das gravierende Auswirkungen auf das Sozialverhalten von Menschen in aller Welt hat (vgl. Baron 2008). So verzeichnet der Marktführer Facebook mittlerweile 800 Millionen aktive Mitglieder, von denen sich mehr als die Hälfte mindestens einmal täglich in das soziale Netzwerk einloggen.⁵

Der Drang, permanent Zugriff auf das Web haben zu wollen, stellt inzwischen ein vielbeobachtbares Massenphänomen dar (Baron 2008: 4f.). Diese Entwicklung wurde maßgeblich von sozialen Netzwerken wie Facebook oder Google+ befördert. Diese Entwicklung hat jedoch auch ihre Schattenseiten. So zeigt eine aktuelle Studie im Auftrag der Bundesdrogenbe-

⁵ Facebook Statistik: <http://www.facebook.com/press/info.php?statistics>, 30.12.2012.

auftragten, dass 560.000 Deutsche internetsüchtig sind (Berliner Kurier 2011). Die Tendenz ist dabei, insbesondere bei Kindern und Jugendlichen, steigend.

Ein weiteres Phänomen, das im Kontext des Web 2.0 in den letzten Jahren für mediales Aufsehen gesorgt hat, ist das Sammeln von persönlichen Daten der Webnutzer zu kommerziellen Zwecken. Der soziale Netzwerkbetreiber Facebook ist in den letzten Jahren wiederholt ins Kreuzfeuer der Kritik geraten, da nachweislich Daten über ihre Nutzer gesammelt und gegen Bezahlung an dritte Parteien weitergegeben wurden. Diese Tatsache ist ein Grund dafür, dass Webnutzer mittlerweile vorsichtiger mit persönlichen Daten im Web umgehen.

Seit dem Übergang vom Web 1.0 zum Web 2.0 sind bereits einige Jahre verstrichen. Dabei zeigt sich, dass sich das Web weiterhin in einem ständigen Wandel befindet. So hat in den letzten Jahren eine neuerliche Revolution des Webs stattgefunden, die als Web 3.0 oder *semantisches Web* bezeichnet wird (Berners-Lee et al. 2001). Durch die Ergänzung von Metadaten werden im *semantischen Web* Strukturen geschaffen, die es erlauben, die Flut an Informationen im Web effektiver zu strukturieren. Die Annotationen erlauben Informationen schneller zu finden und zu ordnen. Beispielsweise versucht die Suchmaschine *Wolfram Alpha*⁶ inhaltliche Antworten auf Suchanfragen zu geben. Anhand einer großen Datenbasis versucht *Wolfram Alpha* die Ergebnisse der Suchanfragen zu berechnen und dabei die Antworten bereitzustellen, nach denen der Webnutzer gesucht hat.

Die rapide Entwicklung des Webs in den letzten Jahren hat einen großen Einfluss auf die sprachwissenschaftliche Forschung genommen. Das Web ermöglicht es, relativ leicht auf eine erhebliche Menge an Text zuzugreifen. Somit ist es prinzipiell einfach geworden, große maschinenlesbare Korpora zu kompilieren, um beispielsweise den Sprachgebrauch in bestimmten Web-Genres zu analysieren. Die Vielfalt an verschiedenen Kommunikationssituationen und die thematische Breite erlaubt es zudem, sehr spezielle Korpora zusammenzustellen. Das Web wird daher zunehmend als Steinbruch genutzt, von dem große Mengen an Daten heruntergeladen werden können. Gleichzeitig ist in vielen Studien bei der Kompilation eines Webkorpus die konzeptionelle Ebene oftmals vernachlässigt worden. Der Einfluss der verschiedenen Kommunikationssituationen im Web auf die Sprachproduktion wurde in der Vergangenheit bei der Bewertung der Analysen oftmals außer Acht gelassen.

Generell muss konstatiert werden, dass es an einflussreichen linguistischen Veröffentlichungen zum Thema Websprache und neuen Medien mangelt. Thurlow / Mroczek (2011) bemängeln, dass seit 1996 lediglich vier Sammelbände erschienen sind, die sich genuin mit der

⁶ <http://www.wolframalpha.com>, 30.12.2012.

Beschreibung von Sprache in neuen Medien befassen: *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (Herring 1996), *The Multilingual Internet: Language, Culture, and Communication Online* (Danet / Herring 2007), *Language and New Media: Linguistic, Cultural, and Technological Evolutions* (Rowe / Wyss 2009) und *Digital Discourse: Language in the New Media* (Thurlow / Mroczek 2011). Die Arbeiten von Baron (2001, 2008) und Crystal (2006) stellen einflussreiche Monographien zum Thema Neue Medien dar. Es herrscht insbesondere ein Mangel an Arbeiten, die sich mit Gemeinsamkeiten und Unterschieden von Web-Genres befassen; meines Wissens haben einzig Giltrow / Stein (2009) einen Sammelband zu diesem Thema veröffentlicht. Eine systematische Gegenüberstellung von Web-Genres ist jedoch bis zum jetzigen Zeitpunkt noch nicht durchgeführt worden.

Das Web zählt zu den Ausprägungen computervermittelter Kommunikation (*computer-mediated communication = CMC*), über dessen Einfluss auf den Sprachgebrauch erstmals in den 1980er Jahren intensiv debattiert wurde (Baron 2008: 13). Die Erfindung von weiteren elektronischen Gerätschaften, wie zum Beispiel dem Mobiltelefon, hat zu einer terminologischen Erweiterung des *CMC*-Konzepts geführt, so dass inzwischen auch der Terminus *electronically-mediated communication (EMC)* verbreitet ist. Baron (2008: 15ff.) unterscheidet Ausprägungen von *EMC* anhand der beiden Dimensionen *synchronicity* und *audience scope*. Die Dimension *synchronicity* kann dichotomisch in die beiden Zustände synchron und asynchron unterteilt werden. Synchrone Kommunikation findet in Echtzeit statt, während es in asynchroner Kommunikation eine zeitliche Verzögerung gibt. Die Dimension *audience scope* bezieht sich auf die Hörerschaft der Kommunikation. So kann entweder eine Person (*one-to-one*) oder eine Gruppe von Personen (*one-to-many*) adressiert werden (Baron 2008: 15ff.). Mit Hilfe dieser beiden Dimensionen kann schließlich eine Matrix errichtet werden, anhand derer man Web-Genres voneinander abgrenzen kann. Ein Beispiel für asynchrone *one-to-one* Kommunikation stellt eine berufliche E-Mail an einen Kollegen dar, während ein Blogbeitrag als Beispiel für asynchrone *one-to-many* Kommunikation dient. Während *Instant Messaging* als Beispiel für synchrone *one-to-one* Kommunikation fungiert, stellt ein Sprachchat in einem öffentlichen Chatraum synchrone *one-to-many* Kommunikation dar. Es existieren auch Web-Genres, die Mischformen darstellen. So können Nutzer in *Massive Multiplayer Online Role-Playing Games* (MMORPGs) sowohl medial schriftlich (per Texteingabe) als auch medial mündlich (per Mikrofon) miteinander kommunizieren.

Aus sprachlicher Perspektive stellt das Web ein äußerst vielschichtiges Medium dar. Es besteht aus einer Vielzahl unterschiedlicher Web-Genres, die sich in ihren medialen und sprachlichen Merkmalen in unterschiedlichem Maße voneinander unterscheiden. Medial betrachtet konstituiert ein Großteil der Web-Genres schriftliche Sprache, da die meisten Texte im Web schriftlich produziert werden. Auf konzeptioneller Ebene bestehen jedoch große Unterschiede zwischen den einzelnen Web-Genres bezüglich ihrer Mündlichkeit / Schriftlichkeit.

Es hat sich in der Vergangenheit wiederholt gezeigt, dass es schwierig ist, einen gemeinsamen Kern an universellen Sprachmerkmalen zu identifizieren, der sich gleichermaßen für eine Vielzahl unterschiedlicher Web-Genres nachweisen lässt. Der Sprachgebrauch in den jeweiligen Web-Genres muss differenzierter betrachtet werden, da ihnen oftmals verschiedene Kommunikationsziele zu Grunde liegen. Allerdings zeigt sich, dass pauschale Aussagen zu Web-Genres schwierig sind, da auch innerhalb der Web-Genres große Unterschiede vorherrschen können. Chesher (2005) zeigt beispielsweise, dass sich Newsblogs und persönliche Blogs sowohl stilistisch als auch sprachlich enorm voneinander unterscheiden können. Es sind in der Vergangenheit Versuche unternommen worden, diese Unterscheidungen systematisch darzulegen. Crystal (2001) formuliert beispielsweise mehrere Parameter und Merkmale, mit deren Hilfe seiner Meinung nach Web-Genres voneinander abgegrenzt werden können. Gleichzeitig versucht Crystal (2006) Universalien zu identifizieren, die für alle Ausprägungen des Sprachgebrauchs im Web gültig sind. Diese subsumiert er unter dem Konzept des *Netspeak*.

3.2. Das *Netspeak*-Konzept

Über die Existenz eines *Netspeak*-Konzepts äußert sich Crystal (2006: 20) folgendermaßen:

There is a widely held intuition that some sort of *Netspeak* exists – a type of language displaying features that are unique to the Internet, and encountered in all the above situations, arising out of its character as a medium which is electronic, global, and interactive.

Demnach würden Web-Genres einen gemeinsamen Kern an sprachlichen Merkmalen aufweisen, der Websprache somit als drittes Medium zwischen gesprochener und geschriebener Sprache etablieren würde. *Netspeak* beinhalte dabei sowohl Elemente mündlicher als auch schriftlicher Sprachrealisierungen. Um die sprachlichen Charakteristika beschreiben zu können, bemüht sich Crystal (2003b) um eine Taxonomie von Merkmalen, die er als typisch für

mündliche und schriftliche Sprache erachtet (vgl. Tabelle 1). Es muss hierbei betont werden, dass Crystal die aufgeführten Merkmale keineswegs als allgemeingültig versteht, sondern, dass er sie lediglich als typisch für die meisten mündlichen und schriftlichen Kommunikationssituationen erachtet.

Speech:	Writing:
<ul style="list-style-type: none"> - Time-bound - Spontaneous - Face-to-face - Socially interactive - Loosely structured - Immediately revisable - Prosodically rich 	<ul style="list-style-type: none"> - Space-bound - Contrived - Visually decontextualized - Factually communicative - Elaborately structured - Repeatedly revisable - Graphically rich

Tabelle 1: Merkmale mündlicher und schriftlicher Sprachrealisierung (vgl. Crystal 2003b)

Laut Crystal (2003b) ist gesprochene Sprache kurzlebig und vergänglich. Es gibt keine Zeitverzögerung zwischen Produktion und Rezeption gesprochener Sprache. Satzgrenzen sind oftmals nicht identifizierbar. Sie ist darüber hinaus geprägt von *face-to-face*-Kommunikation, in der non-verbale Kommunikation (zum Beispiel Gestik, Mimik oder Tonalität) wichtig ist. Der Prosodie wird dabei eine zentrale Rolle zugewiesen; so kann beispielsweise durch Lautstärke, Tempo oder Betonung Bedeutung kommuniziert werden. Lexikalisch ist gesprochene Sprache geprägt durch die Nutzung von Kontraktionen, *Slang*, und dem Potenzial zur Obszönität im Ausdruck. Auffällig ist die Produktion längerer komplexer syntaktischer Konstruktionen. Gesprochene Sprache erfüllt eine phatische Funktion, das heißt, sie ist wichtig um soziale Beziehungen aufrecht zu erhalten. Gesprochene Sprache kann während des Formulierens zwar verändert werden, allerdings können einmal gemachte Äußerungen nicht mehr rückgängig gemacht werden. Unterbrechungen und Überlappungen sind üblich.

Schriftsprache ist langlebig und mehr oder weniger unveränderlich. Es existiert eine Zeitverzögerung zwischen Produktion und Rezeption, ein mehrmaliges Rezipieren des Geschriebenen ist möglich. Satzgrenzen sind durch Interpunktion und Layout eindeutig erkennbar. Non-verbale Signale können nicht vermittelt werden, ebenso wenig wie prosodische Merkmale. Elaborierte, verschachtelte Satzkonstruktionen sind durchaus üblich. Mit Hilfe von Tabellen und Grafiken können bestimmte Sachverhalte besser thematisiert und memoriert werden. Fehler und andere sprachliche Missstände können im Nachhinein noch editiert

werden, ohne dass der Rezipient diesen Korrekturprozess wahrnimmt. Unterbrechungen und Überlappungen von Sprache sind unüblich.

Crystal (2006) behauptet, dass *Netspeak* Merkmale sowohl von mündlicher als auch von schriftlicher Sprache aufweise und somit als eigenständiges Medium konstituiert werden könne. Für Crystal bilden Mündlichkeit und Schriftlichkeit zwei Pole eines Kontinuums, auf dem sich die verschiedenen Web-Genres verorten lassen. So setzt Crystal (2006: 31) Homepages beispielsweise sehr nah an den Pol der Schriftlichkeit, da sie viele Elemente von Schriftsprache aufweisen würden, während er Blogs nah am mündlichen Pol platziert, da sie viele Elemente von gesprochener Sprache beinhalten würden. Einschränkend weist Crystal (2006: 32f.) darauf hin, dass *Netspeak* von gesprochener Sprache unterschieden werden müsse, da es elementare Bedingungen für gesprochene Sprache nicht erfülle. So herrsche beispielsweise, im Gegensatz zu gesprochener Sprache, in Chats immer eine unidirektionale Kommunikationssituation vor. Zudem seien Unterbrechungen und Sprachüberlappungen, die für phonisch realisierte Sprache üblich sind, nicht möglich. Auch das Fehlen von non-verbalen Signalen könne nur zum Teil durch orthografische Zeichen, wie zum Beispiel *Emoticons*, kompensiert werden. Die Sprachgeschwindigkeit könne selbst in getippter Sprache niemals die von gesprochener Sprache erreichen. Darüber hinaus existiere immer eine zeitliche Verzögerung zwischen Produktion eines Textes und dessen Rezeption, die teilweise selbst beim Chatten bis zu mehreren Sekunden dauern kann.

Der Sprecherwechsel (*turn-taking*) ist einer der fundamentalsten Merkmale gesprochener Sprache. Dieser lässt sich jedoch nur schwer beziehungsweise nur anders im Web realisieren. Die prinzipiell chaotischen Diskursstrukturen in einigen Kommunikationssituationen, wie zum Beispiel dem Gruppenchat, erlauben keinen effektiven Sprecherwechsel, so wie er in gesprochener Sprache mit *turn-constitutional units* und *transition-relevance points* (vgl. Sacks et al. 1974) vorgenommen werden kann.

Die aufgezählten Argumente führen Crystal (2006: 44) zur Beobachtung: "[...] in my estimation the actual amount that Netspeak has in common with speech is very limited". Zusammenfassend stellt er fest, dass *Netspeak* einerseits wenige Elemente gesprochener Sprache aufweist, andererseits auch nicht mit konventioneller Schriftsprache gleichgesetzt werden kann.

Im Gegensatz zu geschriebener Sprache ist *Netspeak* im Nachhinein veränderbar, da die Möglichkeit besteht, Texte jederzeit zu editieren, ohne dass in vielen Fällen für den Rezipienten eine Veränderung sichtbar wird. Zudem lässt sich beobachten, dass sich innerhalb der

Web-Genres der Aufbau und die Struktur der Satzkonstruktionen erheblich voneinander unterscheiden. So existieren Web-Genres, in denen unvollständige Sätze produziert werden. Auch in Bezug auf die Textkohärenz kann sich *Netspeak* von geschriebener Sprache unterscheiden. Beispielsweise kann es Rezipienten in langen Foren-*Threads* mitunter schwerfallen, dem Diskurs inhaltlich zu folgen, da dieser typischerweise nicht kohärent strukturiert ist.

Trotz der aufgeführten Inkongruenzen betont Crystal (2006: 50), dass *Netspeak* wesentlich mehr Elemente geschriebener als gesprochener Sprache aufweise: "[...] it is plain that Netspeak has far more properties linking it to writing than speech". Dies führt Crystal (2006: 51) zu seinem Fazit, dass *Netspeak* geschriebene Sprache sei, die Elemente von gesprochener Sprache enthielte: "But on the whole, Netspeak is better seen as written language which has been pulled some way in the direction of speech than as spoken language which has been written down". Crystal (2006: 52) bezeichnet *Netspeak* als "a genuine 'third medium'".

Seiner Ansicht nach haben sich, im Zuge der zunehmenden globalen Nutzung des Webs, idiosynkratische sprachliche Merkmale von *Netspeak* herausgebildet und sich sogar auf sprachliche Kontexte außerhalb computervermittelter Sprache ausgeweitet, insbesondere auf den Ebenen der Lexikologie und der Graphologie (Crystal 2006: 20f.). Demnach würden Ausdrücke, die exklusiv im Kontext des Webs entstanden sind auch zunehmend in nicht-elektronischer Kommunikation verwendet. Als Beispiel nennt Crystal (2006: 21) die Phrase "It's my turn to download now".

Die Validität und Plausibilität von Crystals *Netspeak*-Theorie wurde in den letzten Jahren zunehmend kritisch diskutiert. Mittlerweile existiert eine große Schar von Kritikern, die sich eindeutig davon distanzieren. Dürscheid (2003a: 2) bemängelt, dass Crystal in einem Medium Universalien zu finden sucht, das sich als heterogener darstellt als alle Medien zuvor.⁷ Zudem kritisiert sie, dass Crystal vollends die einflussreichen medienlinguistischen Arbeiten von Koch / Oesterreicher (1985, 1994) zur Unterscheidung medialer und konzeptioneller Mündlichkeit / Schriftlichkeit ignoriere. Zudem differenziere Crystal nicht zwischen dem Sprachgebrauch im Web und der Verwendung fachsprachlicher Bezeichnungen, die im Web entstanden seien. Im Erkenntnisinteresse sollte aber eindeutig der Sprachgebrauch im Web stehen. Die von Crystal beschriebenen sprachlichen Innovationen, die vom Web auf die nicht-virtuelle Welt übertragen werden, könnten somit als irrelevant für die Beschreibung von Sprachgebrauch im Web betrachtet werden und somit außer Acht gelassen werden. Dürscheid (2003a: 4) sieht es zudem als nicht zulässig an, E-Mail- und Chatkommunikation und den

⁷ Dürscheid (2003a) bezieht sich mit ihrer Kritik auf Crystal (2001).

Sprachgebrauch im Web unter dem Terminus *Netspeak* zu subsumieren. Da beim Sprachgebrauch im Web die dialogische Komponente in den Hintergrund tritt, herrscht ihrer Meinung nach ein anderer Sprachgebrauch vor, der nicht mit dialogischem Sprachgebrauch verglichen werden kann. Als maßgeblich für die Ablehnung des *Netspeak*-Begriffs sieht Dürscheid (2003a: 4) zudem an, dass sich viele der von Crystal postulierten mediumbedingten sprachlichen Merkmale auch in anderen Verwendungskontexten, die bereits vor der Erfindung des Internets existierten, belegen lassen. So finden sich einige charakteristische Merkmale der Webkommunikation bereits in der privaten Korrespondenz des ausgehenden 19. Jahrhunderts (vgl. Elspaß 2002). Abschließend stellt Dürscheid (2003a: 8) fest: "Ohnehin kann es bei der Vielzahl von Texten, die heute im Internet geschrieben werden, gar keine verallgemeinernden Aussagen zum Auftreten bestimmter sprachlicher Mittel mehr geben". Auch Schlobinski (2000: 77) streitet vehement die Existenz einer "sondersprachlichen Varietät" in Form eines *Netspeaks* ab. Anstatt nach Universalien zu suchen, erachtet Dürscheid (2003a: 14) es daher als sinniger, den Fokus "von der Makroebene auf die Mikroebene zu richten".

Zusammenfassend muss konstatiert werden, dass Crystals *Netspeak*-Konzept zu vereinfachend erscheint, um die Komplexität des Sprachgebrauchs im Web zufriedenstellend abzubilden. Als größtes Problem erweist sich dabei, dass Crystal (2006) bemüht ist, Universalien für den Websprachgebrauch zu finden, die sich weder eindeutig belegen lassen, noch durch vorherige Studien belegt werden konnten. Crystal (2011: 2) selbst hat sich inzwischen weitestgehend von der Idee distanziert, Web-Genre-übergreifende Universalien zu finden, da er einräumen muss, dass sogar innerhalb eines Web-Genres die stilistische Breite enorm variieren kann.

Cyperspeak, Netspeak, and other -speak coinages were often used in accounts aimed at a general public, but their weakness was that they placed undue emphasis on the potential linguistic idiosyncrasies of the medium and suggested that the medium was more homogeneous than it actually is.

Crystal (2011: 10) führt weiterhin aus: "It is difficult to find linguistic generalizations that apply comfortably to Internet language as a whole". Trotzdem beharrt er weiterhin auf der Position, dass das Web als ein drittes Medium neben gesprochener und geschriebener Sprache verstanden werden müsse. Dabei erläutert er jedoch nicht explizit, was genau ein 'Medium'

charakterisiert. Dass dieser Begriff durchaus konzeptionell umstritten ist, zeigt das folgende Kapitel.

3.3. Der Begriff des Mediums

Esser (2002) sieht den Begriff des 'Mediums' generell als problematisch an, da er durchaus auf unterschiedliche Konzepte verweisen kann. So kann 'Medium' im Sinne von Lyons (1981: 18) folgendermaßen verstanden werden:

When we use the term 'medium', rather than 'channel', we are not concerned with the actual transmission of signals on particular occasions, but with the systematic functional and structural differences between what is characteristically written or characteristically spoken.

Für Lyons (1981) steht die Unterscheidung von typisch mündlicher / schriftlicher Sprache anhand von strukturellen und funktionalen Merkmalen im Fokus, welche jedoch stets gradueller Natur ist. Svartvik / Ekedahl (1995) haben diesbezüglich gezeigt, dass sich öffentliche (gesprochene) Reden im Hinblick auf die Nutzung von Verben mehr von persönlicher gesprochener Sprache unterscheiden als von geschriebener Sprache. Im Gegensatz zu diesem Verständnis des 'Medium'-Konzepts weitet Crystal (2001) den Begriff auf die 'Channel'-Ebene aus, indem er "computer, the internet, writing and e-mail" ebenfalls als Medien versteht (Esser 2002: 84). Eine undifferenzierte Vermengung dieser beiden Konzepte, wie Crystal (2001) sie macht, ist in jedem Fall als problematisch zu bewerten (vgl. Esser 2002). Auch Mukherjee (2008) betont, dass zwischen Mündlichkeit / Schriftlichkeit als Medium und als Stil differenziert werden müsse. Aus medialer Perspektive sei Sprachgebrauch im Web (mit Ausnahme von Audio- und Videoformaten) demnach an die grafische Substanz gebunden, aus stilistischer Perspektive könnte er sowohl mündliche als auch schriftliche Ausprägungen finden (Mukherjee 2008: 2). Als prototypisch mündliche Charakteristika nennt Mukherjee (2008: 2) *contractions*, *discourse markers* und die Formulierung von unvollständigen Sätzen; als prototypisch schriftliche Charakteristika führt er das Fehlen von *contractions* und *discourse markers*, komplexere Nominalphrasen und die Formulierung von vollständigen Sätzen auf.

3.4. Mündlichkeit und Schriftlichkeit

Die Nutzung computervermittelter Kommunikation hat die bestehende Unterscheidung zwischen mündlichem und schriftlichem Sprachgebrauch verkompliziert. Im Zuge der zunehmenden Bedeutung der neuen Medien in den letzten Jahrzehnten wurde dabei immer wieder die Frage diskutiert, inwieweit sich diese tendenziell gesprochener oder geschriebener Sprache zuordnen lassen. Dabei muss erwähnt werden, dass sich insbesondere im Laufe des 20. Jahrhunderts gesprochene und geschriebene Sprache zunehmend einander angenähert haben (vgl. Sieber 1998: 192). Mair (1997: 1542) bezeichnet diese Entwicklung als "colloquialisation of written English".

Mündlichkeit und Schriftlichkeit bilden die beiden Realisierungsmöglichkeiten menschlicher Sprache. Gesprochene Sprache wird phonisch realisiert und akustisch wahrgenommen. Schriftliche Sprache wird grafisch realisiert und optisch wahrgenommen. Gesprochene Sprache ist primär, das heißt, sie wird wesentlich früher erlernt als geschriebene Sprache (Halliday 1989: 2f.). Sprechen ist ein Prozess, der automatisch ohne formelle Instruktionen durch soziale Kontakte erlernt wird (Akinnaso 1982: 111). Geschriebene Sprache ist somit sekundär und das Schreiben ist ein Prozess, der gezielt in einem formellen Umfeld, wie zum Beispiel der Schule, erlernt werden muss. Die Ausprägungen geschriebener Sprache, wie sie heutzutage bekannt sind, sind das Resultat einer evolutionären Transition von einer oralen Sprachkultur zu einer Schriftkultur. Allerdings sind, vor allem durch die Nutzung neuer Medien, durchaus gegenläufige Tendenzen zu erkennen.

Sowohl gesprochene Sprache als auch geschriebene Sprache zeichnen sich durch charakteristische mediumabhängige Merkmale aus, die typisch für die jeweilige Kommunikationssituation sind. In Anlehnung an Lyons' (1981) Konzept der *medium-transferability* unterscheidet Esser (2006) zwischen mediumunabhängigen Strukturen und mediumabhängigen Strukturen. So sind Betonung und Intonation Merkmale gesprochener Sprache, während Interpunktion Ausdruck geschriebener Sprache ist. Es ist unter normalen Umständen unmöglich, diese mediumabhängigen Eigenheiten auf das jeweilige andere Medium zu übertragen. Zu den mediumunabhängigen Charakteristika zählen sowohl grammatikalische als auch lexikalische Formen. Diese können, unabhängig von dem Medium in dem sie realisiert werden, in das jeweilige andere Medium transferiert werden (vgl. Esser 2000).

Geschriebene Sprache unterscheidet sich auf fundamentale Weise von gesprochener Sprache. Im Gegensatz zu geschriebener Sprache ist gesprochene Sprache nicht editierbar. Gesprochene Sprache entsteht durch zwischenmenschlichen Kontakt. Durch den Kontext, der

Betonung des Gesagten, aber auch durch Mimik und Gestik kann Bedeutung geschaffen werden (vgl. Esser 2000). Geschriebene Sprache hingegen kann ein funktionaler Ersatz für gesprochene Sprache sein, wie Lyons (1981: 14) feststellt: "[Writing] serves as a functional substitute for speech only in situations which make vocal-auditory communication impossible, unreliable or inefficient". Geschriebene Sprache wurde ursprünglich dafür verwendet, Kommunikation über Distanzen zu ermöglichen und juristische, religiöse und kommerzielle Sachverhalte zu dokumentieren (Lyons 1981: 14). Daraus erklären sich schließlich die größere Formalität und das größere Prestige geschriebener Sprache. Im formellen Umfeld, in dem schriftsprachliche Normen erlernt werden, wird Lernern beigebracht, auf die Wahl der Worte und deren Anordnung zu achten. Dabei kommt den Aspekten 'Grammatikalität' und 'Korrektheit' bei der Textproduktion eine entscheidende Rolle zu (Akinnaso 1982: 111).

Gesprochene und geschriebene Sprache unterscheiden sich ebenfalls bezüglich ihrer Komplexität voneinander. Allerdings ist keines der Beiden *per se* komplexer als das Andere. Im Gegenteil: Halliday (1989, 2001) hat festgestellt, dass sie sich durch unterschiedliche Komplexitätsgrade auf der Ebene des Satzes und der Ebene der Phrase auszeichnen. So ist gesprochene Sprache komplexer auf der Ebene des Satzes, während geschriebene Sprache komplexer auf der Ebene der Phrase ist. Schäpers (2009: 119) bestätigt, dass geschriebene Sprache komplexer auf der Ebene der Phrase ist. Gleichzeitig belegen ihre Ergebnisse allerdings, dass geschriebene Sprache ebenfalls komplexer auf der Ebene des Satzes ist (Schäpers 2009: 137).

Baron (2008) führt einen Katalog von Kriterien auf, anhand derer man systematisch gesprochene Sprache und Schriftsprache voneinander abgrenzen kann (vgl. Tabelle 2).

Structural Properties	Speech	Writing
Number of participants	dialogue	monologue
Durability	ephemeral (real-time)	durable
Level of specificity	more vague	more precise
Structural accoutrements	prosodic and kinetic cues	document formatting
Sentence Characteristics	Speech	Writing
Sentence length	shorter units of expression	longer units of expression
One-word sentences	very common	very few
Initial coordinate conjunctions	frequent	generally avoided
Structural complexity	simpler	more complex
Verb tense	present tense	varied (esp. past and future)
Vocabulary Characteristics	Speech	Writing
Use of contractions	common	less common
Abbreviations, Acronyms	infrequent	common
Scope of vocabulary	more concrete more colloquial narrower lexical choices more slang and obscenity	more abstract more literary wider lexical choices less slang or obscenity
Pronouns	many 1 st and 2 nd person	fewer 1 st or 2 nd person (except in letters)
Deictics (e.g., <i>here, now</i>)	use (since have situational context)	avoid (since have no situational context)

Tabelle 2: Kriterien zur Unterscheidung von gesprochener und geschriebener Sprache (Baron 2008: 47)

Dabei vergleicht sie mündliche und schriftliche Sprache anhand der drei Parameter *structural properties*, *sentence characteristics* und *vocabulary characteristics*, die jeweils mehrere Unterkategorien beinhalten. Die Gegenüberstellung von gesprochener Sprache und Schriftsprache belegt, dass sich diese beiden Realisierungsmöglichkeiten von Sprache erheblich voneinander unterscheiden, insbesondere im Hinblick auf ihre strukturellen, syntaktischen und lexikalischen Merkmale.

In den vergangenen Jahrzehnten wurde eine Vielfalt von Studien durchgeführt, die die Unterschiede zwischen gesprochener und geschriebener Sprache untersuchten. Drieman (1962: 39) veranschaulicht, dass in gesprochener Sprache eine wesentlich höhere Anzahl von Wörtern benutzt wird. Schäpers (2009: 15) zeigt auf, dass in gesprochener Sprache mehr Wörter benötigt werden, um dieselbe Menge an Information auszudrücken. Halliday (1989: 61) illustriert, dass geschriebene Sprache von einer wesentlich höheren lexikalischen Dichte geprägt ist als gesprochene Sprache. Zugleich werden in gesprochener Sprache wesentlich mehr

grammatikalische Einheiten verwendet. Halliday (1989: 79) resümiert, dass gesprochene Sprache keineswegs unstrukturierter ist als geschriebene Sprache. Poole / Field (1976: 309) stellen fest, dass gesprochene Sprache syntaktisch komplexer ist als geschriebene Sprache. Häcki Buhofer (2000: 256) betont, dass Schreiben einen langsameren und geplanteren Prozess darstellt als Sprechen. Das Ergebnis sind komplexer strukturierte und kohärentere Texte, die mehrmals korrigiert und revidiert werden können.

Bei genauerer Betrachtung der Studien zu gesprochener Sprache und Schriftsprache offenbart sich in vielen Fällen das Problem, dass häufig nicht ausreichend zwischen einer medialen und einer konzeptionellen Ebene von Mündlichkeit / Schriftlichkeit unterschieden wird. Koch / Oesterreicher (1985, 1994) haben diese Unterscheidung in einem Modell festgehalten, das systematisch die Unterschiede zwischen medialer und konzeptioneller Mündlichkeit / Schriftlichkeit beschreibt.

3.4.1. Das Nähe-Distanz-Modell von Koch / Oesterreicher

Koch / Oesterreicher (1985, 1994) haben mit ihrem "Nähe-Distanz-Modell" ein sehr einflussreiches Modell hervorgebracht, das einen Ansatzpunkt für die Modellierung des Verhältnisses von gesprochener und geschriebener Sprache bietet (vgl. Abbildung 5). Sie unterscheiden dabei zwischen "medialer Mündlichkeit / Schriftlichkeit" und "konzeptioneller Mündlichkeit / Schriftlichkeit". Medial bezieht sich dabei auf die phonische / grafische Realisierung von Sprache, während sich konzeptionell auf die in der Äußerung gewählte Ausdrucksform beruft.

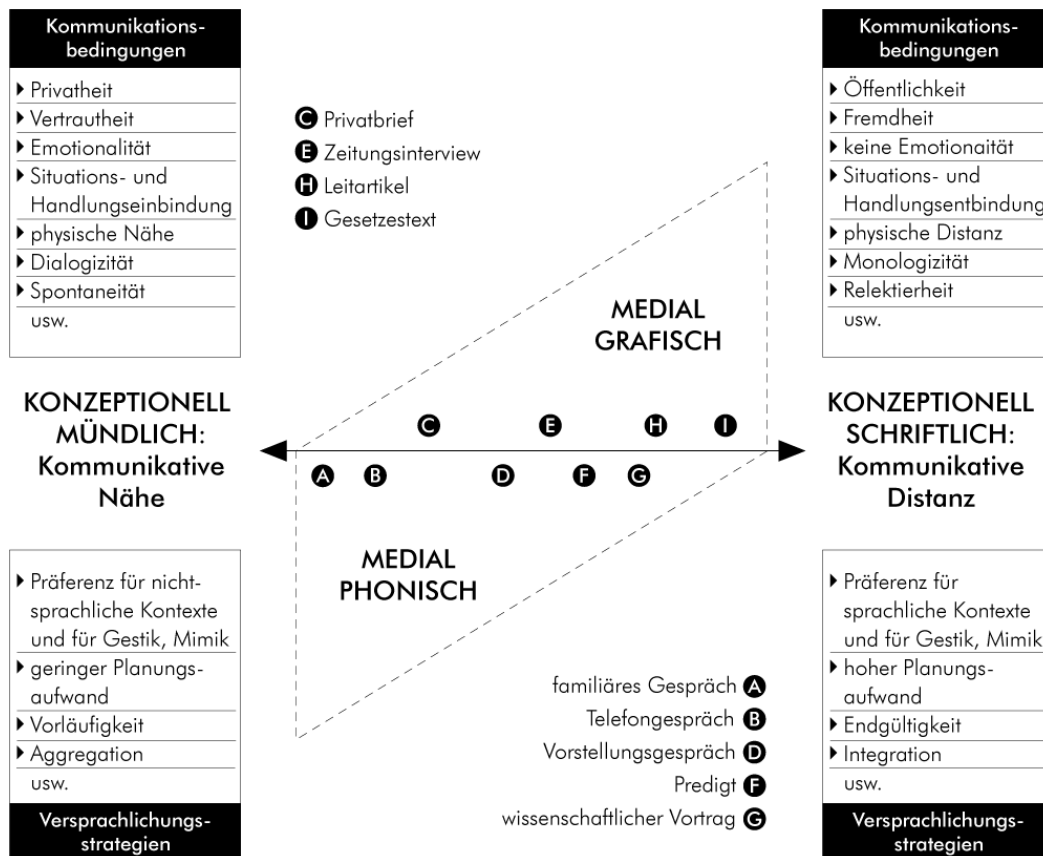


Abbildung 5: Das Nähe-Distanz-Modell (Koch / Oesterreicher 1985, 1994)

Während die mediale Ebene in die beiden Pole "grafisch" und "phonisch" unterteilt werden kann, kann die konzeptionelle Ebene in die beiden Pole "mündlich" und "schriftlich" aufgespalten werden. Während die Unterscheidung auf der medialen Ebene strikt dichotomisch ist, stellt sich diese auf der konzeptionellen Ebene als graduell dar. Die beiden Pole auf der konzeptionellen Ebene bezeichnen Koch / Oesterreicher als "Sprache der Nähe" und "Sprache der Distanz". Ihnen liegt eine Reihe skalierbarer Parameter zu Grunde, die als Kommunikationsbedingungen und Versprachlichungsstrategien aufgelistet werden. Die Kommunikationsbedingungen stellen außersprachliche Gegebenheiten dar, die einen Einfluss auf die Versprachlichungsstrategien haben. Konzeptionelle Mündlichkeit wird charakterisiert durch raum-zeitliche Nähe, emotionale und soziale Nähe, und die Vertrautheit der Gesprächspartner miteinander. Die Kommunikation zeichnet sich durch Emotionalität, Spontaneität und eine prinzipiell dialogische, offene Gesprächsentwicklung aus. Als Versprachlichungsstrategien sind unter anderem eine Präferenz für nichtsprachliche Kontexte und für Gestik und Mimik und ein geringer Planungsaufwand zu nennen.

Konzeptionelle Schriftlichkeit wird charakterisiert durch raum-zeitliche, emotionale und soziale Distanz. Zudem gibt es eine zeitliche Verschiebung zwischen Produktion und Rezeption. Dem Produktionskontext wird keine entscheidende Bedeutung zugemessen. Die Kommunikation zeichnet sich durch eine monologische Themenorganisation seitens des Produzenten aus, die nicht von dem Rezipienten beeinflusst werden kann. Als Versprachlichungsstrategien sind eine Präferenz für sprachliche Kontexte und für Gestik und Mimik, und ein hoher Planungsaufwand zu nennen.

Es können graduelle Unterscheidungen zwischen konzeptioneller Mündlichkeit / Schriftlichkeit gemacht werden. Nähe und Distanz bilden zwei oppositionelle Endpunkte eines Kontinuums, auf dem sowohl medial mündliche Kommunikationsformen als auch medial schriftliche Textsorten verortet werden können. Während also ein wissenschaftlicher Vortrag medial phonisch ist, da er mündlich vorgetragen wird, ist er konzeptionell eher als schriftlich zu bewerten, da viele Elemente von Schriftsprache zur Geltung kommen. Im Gegenzug ist ein informeller Brief an einen Bekannten zwar medial grafisch, er kann aber typische Elemente konzeptioneller Mündlichkeit aufweisen. Ein Herausstellungsmerkmal des Modells ist, dass graduelle Unterschiede zwischen einzelnen Kommunikationssituationen aufgezeigt werden können. Prinzipiell sind Transpositionen aller beschriebenen Äußerungsformen in das jeweilige andere Realisierungsmedium immer möglich, so kann beispielsweise ein Privatbrief vorgelesen werden oder ein wissenschaftlicher Vortrag abgedruckt werden (Koch / Oesterreicher 1985: 4).

Das Modell erweist sich insgesamt als nützlich, um Kommunikationsbedingungen und Versprachlichungsstrategien gesprochener und geschriebener Sprache abzubilden und die beiden Pole der sprachlichen Nähe und der sprachlichen Distanz zu definieren. Trotzdem offenbart das Modell auch einige Schwächen, die in den letzten Jahren zu einer zunehmend kritischen Rezeption geführt haben. Ägel / Hennig (2006: 13) weisen auf die Problematik der logisch-heterogenen Bezüge innerhalb des Modells hin. Sie kritisieren die Auswahl einzelner Parameter innerhalb der Kommunikationsbedingungen und Versprachlichungsstrategien und bemängeln, dass diese oftmals nicht präzisiert werden. Als fundamentales Problem fassen sie auf, dass eine Gleichrangigkeit zwischen den verschiedenen Kommunikationsbedingungen und Versprachlichungsstrategien nicht suggeriert wird und somit weder Abhängigkeiten aufgezeigt noch Gewichtungen vorgenommen werden können (Ägel / Hennig 2006: 14). Zudem erweise sich auch der Begriff der Versprachlichungsstrategie als unpassend, da er oft Aspekte aufgreift, die auch als Kommunikationsbedingungen verstanden werden können.

Thaler (2007: 153) kritisiert, dass Koch / Oesterreicher (1985: 17) mediale Mündlichkeit / Schriftlichkeit strikt dichotomisch voneinander abgrenzen. Diese Abgrenzung ist laut Thaler jedoch für computervermittelte Kommunikation nicht mehr haltbar, da es inzwischen technisch möglich ist, mündliche in schriftliche Sprache und schriftliche in mündliche Sprache zu konvertieren.

Thaler (2007: 151f.) kritisiert des Weiteren, dass Koch / Oesterreicher (1985, 1994) selbst nicht konkretisieren, inwieweit das Nähe-Distanz-Modell angewendet werden kann, um gesprochene und geschriebene Sprache anhand quantifizierbarer Parameter voneinander abzugrenzen. Die von Koch / Oesterreicher genannten Beispiele deuten darauf hin, dass ihr Modell für die Analyse von Textsorten (beziehungsweise im Fall dialogischer Kommunikation für die Analyse von Diskursarten) ausgelegt ist.⁸ Es bleibt offen, inwieweit es beispielsweise auch auf Kommunikationsformen angewendet werden kann.

3.4.1.1. Textsorte, Kommunikationsform, Register und Genre

In der medienlinguistischen Forschung wurde bisher keine einheitliche Terminologie verwendet, so dass üblicherweise die Termini "Textsorte" und "Kommunikationsform" als Synonyme Anwendung finden (Dürscheid 2003b: 40). Für die vorliegende Arbeit muss jedoch eine Differenzierung dieser beiden Konzepte im Sinne von Dürscheid (2003b) und Ziegler (2003) geleistet werden. Als wesentliches Kennzeichen von Kommunikationsformen hebt Dürscheid (2003b: 40) hervor, dass diese ausschließlich durch textexterne Merkmale bestimmt werden. Diese manifestieren sich in den Parametern Kommunikationsrichtung (monologisch / dialogisch), Anzahl der Kommunikationspartner und zeitliche Dimension der Kommunikation (vgl. Androusoopoulos / Schmidt 2001). Textsorten hingegen können sowohl über textexterne als auch interne Merkmale definiert werden (vgl. Ziegler 2003: 24). Im Gegensatz zu Kommunikationsformen haben Textsorten eine einheitliche thematische Funktion (vgl. Brinker 2001). Kommunikationsformen hingegen zeichnen sich durch ihre Multifunktionalität aus (Dürscheid 2003b: 40). Demnach kann das Web-Genre E-Mail unterschiedliche Funktionen ausfüllen. So unterscheidet sich beispielsweise eine Werbe-Mail in ihrer Funktion von einer Einladungsmail. Aufgrund der funktionellen Heterogenität innerhalb der einzelnen Kommunikationsformen bezweifelt Dürscheid (2003b: 47), dass sich diese im Kontinuum von Sprache

⁸ Unter Textsorten verstehen Koch / Oesterreicher (1985, 1994) Texte / Diskurse, die gleiche situative und sprachlich-strukturelle Charakteristika innehaben und eine gemeinsame kommunikative Funktion erfüllen.

der Nähe / Distanz positionieren lassen. Einzig die Verortung der einzelnen Textsorten sei möglich (so zum Beispiel Werbemails und Einladungsmails). Thaler (2007: 164f.) bestätigt, dass innerhalb einer Kommunikationsform eine große Vielfalt von Textsorten vorherrscht. Die einzelnen Textsorten können somit prinzipiell in die Taxonomie des Nähe- / Distanzpols eingeordnet werden.

Im Gegensatz zur deutschen Sprachwissenschaft, in der vorrangig die Termini Textsorte und Kommunikationsform Anwendung finden, wird in der anglistischen Forschung häufig zwischen den Begriffen Register und Genre unterschieden (vgl. Biber / Conrad 2009, Esser 1993). Biber / Conrad (2009: 6) definieren Register dabei wie folgt:

In general terms, a **register** is a variety associated with a particular situation of use (including particular communicative purposes). The description of a register covers three major components: the situational context, the linguistic features, and the functional relationships between the first two components [...].

Unter *situational context* lassen sich mehrere Aspekte zusammenfassen. Zum einen zeigt der *situational context*, ob Sprache mündlich oder schriftlich produziert wird, ob eine Interaktion stattfindet, und was die Kommunikationsziele der Interaktion sind. Als Beispiel nennen Biber / Conrad (2009) *face-to-face*-Konversation. Diese erfordert eine direkte Interaktion von mindestens zwei Personen, die sich am selben Ort zur selben Zeit befinden und beide an der Kommunikation teilnehmen müssen. Es werden üblicherweise Meinungen und Gedanken ausgetauscht, die mit dem persönlichen Leben der Teilnehmer oder mit dem unmittelbaren Kontext zusammenhängen. In Registern findet man sich wiederholende linguistische Merkmale, die typisch für das Register sind.⁹ Der Registerbegriff kann sowohl eng als auch weit aufgefasst werden. So kann *Written academic register* im weitesten Sinn als ein Register verstanden werden. In einem engeren Verständnis des Begriffs könnte man *Written academic register* weiter in die Register *Textbooks* und *Academic research books* unterteilen (Biber / Conrad 2009: 33).

Der Begriff Genre unterscheidet sich vom Begriff des Registers dadurch, dass linguistische Merkmale eines Genres nicht funktional, sondern konventionell sind. Sie sind dabei nicht hochfrequent, sondern kommen oftmals nur einmal in einem Text vor, oft am Anfang oder am

⁹ Beispielsweise finden sich im Register Konversation viele Personalpronomen der ersten Person (*I, we*) und der zweiten Person (*you*) und Fragen.

Ende eines Textes (Biber / Conrad 2009: 7). So gibt es beispielsweise für das Genre Geschäftsbrief formale Konventionen: Ein Geschäftsbrief beginnt mit Datum, Name und Anschrift des Schreibers, gefolgt von einer Begrüßung und Titel des Adressaten. Es folgt der Hauptteil des Briefes, der schließlich mit einer höflichen Formulierung und der Signatur des Schreibers schließt (Biber / Conrad 2009: 17). Biber (1988) postuliert, dass Genres auf externen Kriterien basieren (u.a. *audience, purpose, domain*). Genres werden somit anhand ihrer formalen und nicht anhand ihrer funktionalen Eigenschaften der Texte bewertet. Bibers Genre-Konzept korrespondiert somit grundsätzlich mit Dürscheids Kommunikationsform-Konzept.

Insgesamt eignet sich keiner der genannten Termini problemlos, um auf Sprachgebrauchssituationen im Web zu verweisen. Crystal (2011: 10) schlägt daher die Nutzung des neutraleren Terminus *outputs* vor. Gleichzeitig leistet er keine genaue Eingrenzung, welche Merkmale diese auszeichnen würden. Aus diesem Grund ist auch dieser Terminus nur bedingt geeignet. Generell muss für eine Eingrenzung des Sprachgebrauchs im Web entschieden werden, ob textexterne oder textinterne Merkmale als Parameter für dessen Bewertung genutzt werden sollen. Für die vorliegende Arbeit sollen verschiedene Web-Genres miteinander verglichen werden. Bei der Auswahl der Texte für das Korpus waren jedoch textexterne Merkmale (*audience, purpose, domain*) entscheidend. Daher verstehe ich den analysierten Sprachgebrauch als Genre im Sinne von Biber (1988) beziehungsweise als Kommunikationsform im Sinne von Dürscheid (2003b).

3.4.1.2. Das Nähe-Distanz-Modell und Websprache

Trotz der beschriebenen Kritikpunkte wurde das Nähe-Distanz-Modell von Koch / Oesterreicher (1985, 1994) in der Vergangenheit herangezogen, um auch Web-Genres bezüglich ihrer konzeptionellen Mündlichkeit / Schriftlichkeit zu bewerten, auch wenn es sich nur bedingt eignet, um den Sprachgebrauch im Web abzubilden, unter anderem da der Grad der Situations- und Handlungseinbindung medial beschränkt ist und da non-verbale und paraverbale Kommunikation nicht möglich ist (Thaler 2007: 161). Einschränkend muss erwähnt werden, dass Telepräsenz-Szenarien, wie zum Beispiel der Videochat, dabei zumindest einen Teil dieser non-verbale Kommunikation ermöglichen. Storrer (2001) nutzt beispielsweise das Modell um aufzuzeigen, dass Chatsprache viele Elemente konzeptioneller Mündlichkeit aufweist.

Dürscheid (2003b: 48f.) bemängelt jedoch, dass das Modell in der bestehenden Form nicht die Realität von Web-Genres abbilden kann, da diese in vielen Fällen weder synchronen noch asynchronen Sprachgebrauch darstellen und erweitert daher das Modell konsequenterweise um die Ebene der "Quasi-Synchronizität". Als prototypisches Beispiel nennt sie Chatkommunikation, in der immer ein *Time-lag* (eine zeitliche Verzögerung) zwischen Produktion und Rezeption vorliegt, so dass die Kommunikation als "quasi-synchron" bezeichnet werden muss.

Dürscheid (2003a: 13) betont, dass die Synchronie der Kommunikation einen Einfluss darauf hat, welche Ausdrucksmittel genutzt werden. Dies entscheidet wiederum über die Positionierung einer Textsorte innerhalb der Taxonomie von konzeptioneller Mündlichkeit / Schriftlichkeit: "Je 'synchroner' die Kommunikation, desto eher weist die Äußerung Merkmale auf, die Koch / Oesterreicher (1994) der konzeptionellen Mündlichkeit zuschreiben (Gebrauch von Gesprächspartikeln, Satzabbrüche etc.)" (Dürscheid 2003a: 13). In Anlehnung an Meise-Kuhn (1998) sieht Dürscheid (2003a: 14) somit die Interaktivität der Kommunikationsform und nicht das Medium als elementar für einen bestimmten Sprachgebrauch an.

Storrer (2001) untersucht Kommunikationsformen im Web bezüglich ihrer konzeptionellen mündlichen und schriftlichen Merkmale. Sie beschränkt sich jedoch auf die Beschreibung lexikalischer, syntaktischer und pragmatischer Besonderheiten in einem einzelnen Web-Genre. Im Erkenntnisinteresse der vorliegenden Arbeit liegt jedoch die Beschreibung phraseologischer Besonderheiten in mehreren Web-Genres, um diese klar voneinander abzugrenzen. Diese Besonderheiten werden anhand einer formalen und einer funktionalen Analyse von Mehrworteinheiten herausgestellt. Das Nähe-Distanz-Modell erweist sich dabei als sehr nützlich, um die Unterschiede zwischen den einzelnen Web-Genres aufzuzeigen.

3.4.2. Konzeptionelle Mündlichkeit / Schriftlichkeit von Mehrworteinheiten

In verschiedenen Studien wurden Mehrworteinheiten bereits im Hinblick auf Aspekte konzeptioneller Schriftlichkeit / Mündlichkeit untersucht. Unter anderem haben Biber et al. (1999) die Nutzung von unterschiedlichen *Lexical Bundles* in "Konversation" (*Conversation*) und in "akademischer Prosa" (*Academic prose*) erforscht. Sie haben gezeigt, dass, abhängig vom jeweiligen Register, andere Typen von *lexical bundles* bevorzugt werden.¹⁰ *Lexical Bundles*

¹⁰ So lässt sich in Konversation in 44% aller Fälle die Form *personal pronoun + lexical verb phrase (+ complement clause)* finden, während in akademischer Prosa in 33% aller Fälle die Form *preposition + noun phrase fragment* verwendet wird.

sind Mehrworteinheiten, die häufiger zusammen auftreten als statistisch erwartet werden kann. Biber et al. (1999: 989) beschreiben *Lexical Bundles* folgendermaßen:

[...] lexical bundles are the sequences of words that most commonly co-occur in a register. Usually they are not fixed expressions, and it is not possible to substitute a single word for the sequence; in fact, most lexical bundles are not structurally complete at all [...].

Lexical Bundles werden dadurch charakterisiert, dass die einzelnen Elemente innerhalb der Mehrworteinheiten variabel sind (im Gegensatz zu den Konstituenten von Idiomen, die invariabel sind).

In einer Nachfolgestudie haben Biber et al. (2004) *Lexical Bundles* systematischer nach ihren vorwiegenden funktionalen Eigenschaften unterschieden. Dazu haben sie die häufigsten *Lexical Bundles* für vier verschiedene Register (*Conversation, Classroom teaching, Textbooks, Academic prose*) identifiziert und unter drei funktionale Oberkategorien subsumiert. Das Ergebnis bestätigte, dass in gesprochener Sprache zum Teil andere *Lexical Bundles* präferiert werden als in geschriebener Sprache. Aufbauend auf der Methodik von Biber et al. (2004) haben Simpson-Vlach / Ellis (2010) in einer weiteren Studie schließlich eine feiner gegliederte Kategorisierung von *Lexical Bundles* erstellt. Anhand dieser Kategorisierung haben sie eine Liste von typisch mündlichen und typisch schriftlichen *Lexical Bundles* im akademischen Kontext generiert.

Insgesamt belegen die aufgeführten Studien, dass sich gesprochene und geschriebene Register bezüglich der formalen und funktionalen Nutzung von Mehrworteinheiten voneinander unterscheiden. Vor diesem Hintergrund steht zu erwarten, dass sich auch verschiedene Web-Genres diesbezüglich unterscheiden. Dies wird anhand einer formalen und funktionalen Analyse im praktischen Teil dieser Arbeit näher untersucht werden. Zunächst müssen jedoch die medialen und die sprachlichen Merkmale des Webs genauer beschrieben werden.

3.5. Mediale und sprachliche Merkmale des Webs

3.5.1 Mediale Merkmale

Der Sprachgebrauch im Web unterscheidet sich von typisch mündlichem und schriftlichem Sprachgebrauch im nichtvirtuellen Kontext. Die medialen Bedingungen im Internet haben dabei einen entscheidenden Einfluss auf den Sprachgebrauch. Im Folgenden werden die media-

len Besonderheiten des Webs, die für einen Großteil aller Web-Genres gültig sind, zunächst aufgelistet und im Folgenden beschrieben:¹¹

- a) Hybridcharakter des Mediums
- b) Infinität und Intransparenz
- c) Anonymität
- d) Flüchtigkeit und funktionale Erweiterung
- e) Formatierung
- f) Veränderbarkeit

a) Hybridcharakter des Mediums

Das Medium Internet zeichnet sich durch seinen Hybridcharakter aus. Computervermittelte Kommunikation ermöglicht Nutzern ehemals nur getrennt voneinander verfügbare Übertragungskanäle (Bild, Ton, Schrift) zu kombinieren und stellt damit eine eigenständige Form der Kommunikation dar (Thaler 2007: 147). Vor allem der vernetzte Computer kann dabei als "Hybrid-Kommunikationsmedium" (Höflich 1997) aufgefasst werden, der sich im Hinblick auf seine Interaktivität entscheidend von anderen klassischen Massenmedien unterscheidet. Er ermöglicht interpersonale (*one-to-one*), Gruppen- (*one-to-few* oder *few-to-few*) und Massenkommunikation (*one-to-many* oder *many-to-many*) (Thaler 2007: 147).

b) Infinität und Intransparenz

Ein weiteres elementares Merkmal des Webs ist seine scheinbare Infinität. Täglich werden Tausende von neuen Webseiten hinzugefügt, während viele andere Webseiten entweder verändert oder gelöscht werden. Seit Jahren steigt die Anzahl von Webseiten kontinuierlich an. Allein zwischen Dezember 2011 und Januar 2012 ist die Zahl der Webseiten um 4,9% gewachsen. Dies entspricht etwa 27,2 Millionen Webseiten (vgl. Netcraft 2012). Insgesamt besteht das Web inzwischen aus 582.716.657 Webseiten (vgl. Netcraft 2012).¹² Bei vielen Webseiten ist häufig nicht ersichtlich, ob diese noch aktiv genutzt werden. Zudem ist es oftmals schwierig nachzuvollziehen, wann genau die Texte einer Webseite produziert wurden. In vielen Fällen wird lediglich angezeigt, wann die Webseite das letzte Mal aktualisiert wurde. Eine Ausnahme bilden Online-Zeitungsartikel, auf denen üblicherweise das Datum der Publikation

¹¹ Die Kategorien wurden aus Bittner (2003), Thaler (2007) und Baron (2010) zusammengetragen und ergänzt.

¹² Zum Zeitpunkt 05.02.2012.

des Textes angegeben wird. Diese Tatsache erschwert es, diachrone Studien durchzuführen, um beispielsweise eine Veränderung des Sprachgebrauchs im Web nachzuweisen.

c) Anonymität

Das Web erlaubt es seinen Nutzern prinzipiell anonym zu bleiben. Es besteht zwar die Möglichkeit Webnutzer zu identifizieren, dies ist jedoch häufig mit viel Aufwand verbunden. Die Anonymität des Mediums kann einen Einfluss auf den Sprachgebrauch in ihm haben, da Webnutzer sich nicht für ihren Sprachgebrauch verantworten müssen. So findet sich beispielsweise in vielen Kommentaren im Videoportal Youtube ein markierter Sprachgebrauch, der insbesondere durch Obszönität und Vulgärsprache geprägt ist. Auf kommerziellen Homepages oder auf *Social Network Sites* findet man hingegen ein anderes sprachliches Verhalten, da die Nutzer prinzipiell identifizierbar sind und sprachliches Fehlverhalten Konsequenzen haben kann. Durch die fehlenden soziobiografischen Angaben ist es oftmals nicht ersichtlich, ob der Produzent eines Textes *ENL*-, *ESL*- oder *EFL*-Sprecher ist. Ebenso wenig lassen sich andere soziobiografische Angaben (Geschlecht, Alter, Ethnizität etc.) finden. Die Anonymität des Webs erschwert somit erheblich die Durchführung soziolinguistischer Untersuchungen. Zudem erschwert sie die Kompilation eines linguistischen Korpus, das repräsentativ für die Web-Gemeinschaft in ihrer Gesamtheit sein soll.

d) Flüchtigkeit und funktionale Erweiterung

In den letzten zwei Dekaden ist eine funktionale Erweiterung des Webs von einem reinen Informationsbeschaffungsmedium (Web 1.0) zu einem interaktiven multimedialen Medium (Web 2.0) vollzogen worden. Durch den medialen Wandel existiert mittlerweile eine Vielzahl verschiedener Web-Genres, die sich mehr oder weniger voneinander unterscheiden.

Das Web unterliegt ständigen Veränderungen. Einerseits entstehen ständig neue Web-Genres, andererseits verändern sich bestehende Web-Genres in ihrer Form. In den letzten Jahren entstand eine große Anzahl neuer Web-Genres, von denen sich nicht alle dauerhaft etablieren konnten. Beispielsweise erlangte das *Massive Multiplayer Online Role-Playing Game* (*MMORPG*) *Second Life* innerhalb kürzester Zeit eine große Popularität. Aufgrund zahlreicher Mängel und problematischer Inhalte verlor es jedoch wieder einen Großteil seiner Nutzer und steht nun nicht mehr so stark im medialen Interesse (vgl. Fittkau / Maaß Consulting 2008). Der soziale Netzwerkbetreiber Facebook hingegen hat es verstanden, die ursprünglichen Funktionen der Seite auf vielfältige Weise zu erweitern. So können inzwischen *Tweets*, *Instant Messages*, aber auch E-Mails verfasst werden, die mit externen Quellen, wie

zum Beispiel Videos, verlinkt werden können. Wandelbarkeit erweist sich als ein zentrales Kriterium, um in der schnelllebigen virtuellen Welt bestehen zu können; Web-Genres müssen sich ständig an die kommunikativen Anforderungen der Webnutzer anpassen.

e) Formatierung

Web-Genres unterscheiden sich im Hinblick auf ihr Layout und ihre Interaktivität erheblich voneinander. Dabei kann die Formatierung einer Webseite einen entscheidenden Einfluss auf den Inhalt der Texte haben. Die Microblogging-Plattform Twitter charakterisiert beispielsweise, dass nur Texte mit maximal 140 verfasst werden können (vgl. Schmidt et al. 2010). Dies führt dazu, dass dort die Sprachverwendung oftmals telegrammartig ist.

f) Veränderbarkeit

Generell unterscheiden sich Texte, die im Web produziert wurden von typisch schriftlichen Texten in Bezug auf ihre Editierbarkeit. Texte, die im Web produziert wurden, sind üblicherweise nachträglich editierbar. Als Beispiel fungiert das Web-Genre Wiki, in dem Texte von unterschiedlichen Nutzern jederzeit inhaltlich und sprachlich editiert werden können. Einen Ausnahmefall stellt das Web-Genre Chat dar. In ihm entsteht ein ad hoc Sprachgebrauch, der im Nachhinein nicht mehr verändert werden kann. Typischerweise ergibt sich bei der Veränderung von Texten im Web das Problem, dass die einzelnen Editierungsprozesse nicht mehr nachzuvollziehen sind. Wikitexte stellen diesbezüglich eine Ausnahme dar, da in Wikis auch ältere Textversionen abgespeichert werden und so auch im Nachhinein jede Änderung des Ursprungstexts nachvollzogen werden kann.

3.5.2. Sprachliche Merkmale

Sprachgebrauch kann durch das Medium entscheidend beeinflusst werden. Aus den medialen Bedingungen des Internets erwachsen somit sprachliche Besonderheiten, die für einen Großteil der Web-Genres gültig sind und die nun im Folgenden zunächst aufgelistet und dann näher beschrieben werden:¹³

- a) Fehlende sprachliche Normen
- b) Orthografische Abweichungen
- c) Fehlende Textkohärenz
- d) Die Nutzung non-verbaler Kommunikation

¹³ Die einzelnen sprachlichen Merkmale wurden aus Baron (2010), Crystal (2011) übernommen und ergänzt.

a) Fehlende sprachliche Normen

Die Existenz von Netiquette-Verhaltensregeln, die in Foren und Sprachchats bestehen, zeigt, dass sich noch keine einheitlichen sprachlichen und stilistischen Verhaltensregeln im Web etabliert haben. Webnutzer müssen explizit darauf hingewiesen werden, welches sprachliche Verhalten als problematisch bewertet wird und zu welchen Konsequenzen ein Fehlverhalten führen kann. Die Anonymität des Mediums erlaubt es Webnutzern jedoch, Netiquette-Regeln zu verletzen, da Zuwiderhandlungen in den seltensten Fällen sanktioniert werden können. Diese Form von markiertem sprachlichem Verhalten ist vermutlich einer der Gründe dafür, dass sich im öffentlichen Sprachbewusstsein fest verankert hat, dass Websprache eine minderwertige Form von Sprachverwendung darstellt (vgl. Baron 2010). Darüber hinaus wird dem Web oftmals attestiert, einen negativen Einfluss auf den Standardsprachgebrauch zu haben. Bisher existieren keine Studien, die eindeutig belegen können, dass das Web einen Einfluss auf Grammatik, Vokabular oder Orthografie genommen hat (Crystal 2011: 75). Zugleich weist Crystal (2011: 140) ausdrücklich darauf hin, dass der Mangel an Forschung in diesem Bereich die Verbreitung des negativen Einflusses des Websprachgebrauchs auf Standard-sprachgebrauch erst ermöglicht hat: "The absence of good up-to-date descriptions is the main reason for the growth in popular mythology about Internet language, such as the view that it is causing increased illiteracy".

Trotz des Fehlens von sprachlichen Normen zeigt sich, dass Webnutzer ein Bewusstsein dafür ausgeprägt haben, dass je nach Web-Genre ein anderes sprachliches Verhalten erforderlich ist. Dies führt letztendlich zu einer Adaption des Sprachgebrauchs in verschiedenen Web-Genres. So beobachtet Tagliamonte (2011) einen formelleren Sprachgebrauch in Teenager-E-Mail-Kommunikation im Vergleich zu Kommunikation in sozialen Netzwerken. Im Hinblick auf die aktuelle Einstellung von Sprachnutzern gegenüber Orthografie und Grammatik, prägt Baron (2008: 169) den Begriff *The Whatever Generation*. Demnach stehen heutige Sprachnutzer sprachlichen Neuerungen wesentlich offener gegenüber als frühere Generationen.

b) Orthografische Abweichungen

Im Web lassen sich typischerweise orthografische Abweichungen von der schriftsprachlichen Norm finden (vgl. Li 2010). Allerdings gibt es dabei große Unterschiede zwischen verschiedenen Web-Genres. In der Öffentlichkeit werden diese Unterschiede zwischen den einzelnen Web-Genres üblicherweise nicht wahrgenommen und stattdessen generalisierende Aussagen getroffen. So wird im Kontext von Webkommunikation von "elektronischen Kaffeekränzchen" oder "Geschwätz im globalen Dorf" gesprochen (Günther / Wyss 1996). Paton

(2008) beklagt die fehlenden orthografischen Kenntnisse von britischen Universitätsstudenten. In einem Kommentar der Frankfurter Allgemeinen Zeitung zeigt sich der Autor davon überrascht, dass sich die Leserkommentare zu einem Artikel über Keilschriften regelkonform zu den deutschen Standards verhalten:

Auffällig ist, dass die Kommentare weit über das übliche Online-Gebrabbel hinausgehen. Das zeigt schon die ungewöhnliche korrekte Orthografie und Interpunktion der Kommentare. Hier schreibt nicht der durchschnittliche Youtube-Nutzer. (Oehmen 2011)

Websprache wird als minderwertige Variante von Schriftsprache betrachtet, in der das Befolgen von orthografischen Regeln optional ist. Thurlow (2006) fasst den Metadiskurs zu Websprache in den Printmedien zusammen. Demnach verursacht die Verwendung von Akronymen, lexikalischen Kürzungen und Nichtstandard-Schreibungen von jungen Menschen eine "moralische Panik" (*moral panic*), die ständig in den Printmedien aufgegriffen wird und verallgemeinernd für alle Formen des Sprachgebrauchs im Web verwendet wird.

c) Fehlende Textkohärenz

Der Sprachgebrauch im Web unterscheidet sich im Vergleich zu traditioneller Schriftsprache im Hinblick auf seine Linearität. Webseiten sind häufig nicht darauf ausgelegt, linear rezipiert zu werden. Mit Hilfe von Hypertextlinks wird die Linearität aufgebrochen. Durch Hypertextlinks kann auf weitere Informationen oder andere Medien (z.B. Fotos, Videos, Audiodateien) zugegriffen werden. Die Rezipienten können dann entscheiden, ob sie den Verlinkungen folgen wollen oder beim rezipierten Text bleiben. Häufig bewegen sich Webnutzer auf diese Weise nichtlinear durch die vorhandenen Medieninhalte. Diese Nichtlinearität hat einen Einfluss auf den Inhalt und die Struktur der im Web produzierten Texte.

d) Die Nutzung non-verbaler Kommunikation

Das Fehlen von non-verbaler Kommunikation ist ein typisches Merkmal von Schriftsprache. Im Web hingegen kann ihr Fehlen durch Alternativen kompensiert werden. Dies geschieht beispielsweise durch Emoticons oder durch Interpunktion. Je nach Web-Genre kann es dabei große Unterschiede in der Nutzung non-verbaler Kommunikation geben.

Aus der Beschreibung der medialen und sprachlichen Gemeinsamkeiten von Websprache wird deutlich, dass im Vergleich zu gesprochener Sprache und Schriftsprache nur wenige

Universalien existieren, die für viele oder fast alle Web-Genres gültig sind. Es scheint jedoch wesentlich interessanter zu sein, die Unterschiede zwischen den einzelnen Web-Genres herauszustellen.

3.5.3. Merkmale der analysierten Web-Genres

Für die vorliegende Arbeit wurden vier unterschiedliche Web-Genres (kommerzielle Homepages, Wikipedia-Artikel, Blogs und Foren) ausgewählt und einer formalen und funktionalen Analyse von Mehrworteinheiten unterzogen. Dass es nur wenige sprachliche Gemeinsamkeiten zwischen den Web-Genres gibt, wurde bereits ausführlich dargelegt. Aus diesem Grund muss im Folgenden eine detailliertere Beschreibung der individuellen Besonderheiten der einzelnen Web-Genres geleistet werden. Um die einzelnen Web-Genres miteinander vergleichen zu können, werden gemeinsame Parameter der Beschreibung benötigt. Dafür wird im Folgenden auf das Nähe-Distanz-Modell von Koch / Oesterreicher (1984, 1994) zurückgegriffen werden, das bereits Bittner (2003) als Grundlage für die Beschreibung verschiedener Web-Genres diente. Dabei muss betont werden, dass die hier geleisteten Beschreibungen als typisch für die jeweilige Ausprägungsform des Web-Genres gelten, diese jedoch nicht universell für alle Ausprägungen der beschriebenen Web-Genres gelten können. Die vier Web-Genres werden mit Hilfe dieser Parameter miteinander verglichen werden. Zunächst wird jedoch eine Beschreibung der einzelnen Web-Genres geleistet.

3.5.3.1. Foren

Elektronische Diskussionsgruppen existieren bereits seit Mitte der 1980er Jahre. Die älteste Online Community der Welt, *The WELL (Whole Earth 'Lectronic Link)*, wurde 1985 gegründet. *The WELL* beinhaltet über 260 Untergruppen, die *Conferences* genannt werden. *Usenet* ist ein weiteres sehr bekanntes öffentliches Netzwerk, in der sich Webnutzer in sogenannten *Newsgroups* austauschen können, die thematisch einen weiten Bogen spannen, von wirtschaftlichen über wissenschaftliche Themen bis hin zu Freizeitthemen (Hoffmann 2007).

Unter elektronischen Foren (verkürzt: Foren) versteht man Webseiten, auf denen Nutzer Einträge zu einem bestimmten Thema *posten* (verfassen und öffentlich machen) können. Dazu treffen sich Webnutzer auf einer Forumsseite, wo sie sich dann in thematisch unterschiedlichen *Threads* (thematische Unterseiten) austauschen können. Foren zeichnet ein

hohes Maß an Interaktion aus, da jeder Nutzer prinzipiell auf andere *Posts* antworten kann. In gewisser Hinsicht spiegeln Foren damit konzeptionell dialogische Konversation wider.

Dennoch unterscheidet sich der Sprachgebrauch in Foren gravierend von konzeptionell mündlichem Sprachgebrauch, vor allem durch die Asynchronität der Kommunikation. Zudem besitzen Foren einen sprachlichen Filter. Administratoren und Moderatoren sorgen dafür, dass sprachliche Grundregeln eingehalten werden. Zu diesem Zweck existieren üblicherweise FAQs und Netiquette-Regeln, auf denen genauer erläutert wird, wie sich die Nutzer sprachlich zu verhalten haben. Ein Nichtbefolgen der Regeln führt im Normalfall zu einer Sanktionierung in Form einer Editierung oder der Löschung des *Posts*.¹⁴

Auf einer typischen Forenseite wird auf der Startseite das Thema des Forums umrissen und auf einzelne Unterforen verwiesen, in denen die Nutzer *Threads* zu speziellen Themenbereichen eröffnen können. In den Unterforen werden die einzelnen *Threads* angezeigt; üblicherweise werden die neuesten Einträge oben platziert, ältere Beiträge rutschen weiter auf der Seite herunter, so dass die Rezipienten weiter nach unten scrollen müssen, um diese zu finden.

Bedingt durch ihren strukturellen Aufbau haben Foren oftmals das Problem der fehlenden inhaltlichen Kohärenz. Es ist durchaus üblich, dass ein *Thread* zu einem bestimmten Thema initiiert wird, sich im Laufe des Diskurses jedoch immer weiter von diesem entfernt. In extremen Fällen ist schließlich ein thematischer Bezug zum ursprünglichen Thema nicht mehr erkennbar. Aus diesem Grund füllen Moderatoren eine wichtige Funktion in Foren aus, da sie dafür sorgen, dass die inhaltliche Kohärenz aufrecht erhalten bleibt, indem sie die Produzenten in den *Threads* immer wieder darauf hinweisen einen thematischen Bezug zur Ausgangsdiskussion herzustellen. Es erweist sich für Forenrezipienten dennoch oftmals schwierig in längeren *Threads* dem Diskurs zu folgen. Während in einigen Fällen referentielle Bezüge zu vorherigen *Posts* hergestellt werden, indem diese aufgegriffen werden, fehlen in anderen Fällen referentielle Bezüge völlig, so dass der Rezipient gezwungen wird für sich eine eigene kohärente Struktur zu finden. Je höher die Anzahl der *Posts* in einem *Thread*, desto schwieriger wird es, eine strukturelle Kohärenz beizubehalten.

Die Bekanntheit eines Forums und die Relevanz der Themen für die Rezipienten haben einen enormen Einfluss auf den Grad der Involviertheit der Rezipienten. Dieser lässt sich messen durch die statistischen Angaben, die sich üblicherweise in Foren finden lassen. In

¹⁴ So ist zum Beispiel das *Flamen*, also das Beleidigen anderer Teilnehmer, ebenso wie der Gebrauch vulgärer Ausdrücke, verboten. *Flamen* kann das Löschen des Beitrags zur Folge haben. Darüber hinaus kann es im Extremfall zu einem Ausschluss des Schreibers führen.

Foren wird beispielsweise immer die Anzahl der Antworten für die jeweiligen *Threads* angezeigt.

Im Gegensatz zu vielen Blogs und kommerziellen Homepages sind die Produzenten von Texten in Foren grundsätzlich anonym. Die einzige diesbezügliche Einschränkung ist, dass sich Forennutzer mit einer funktionierenden E-Mail-Adresse registrieren lassen müssen. In vielen Fällen müssen die Nutzer zwar ein Profil anlegen, in dem sie auch ihren Namen, und optional ihr Alter und ihren Wohnort angeben können, allerdings können sie an dieser Stelle falsche Angaben machen, ohne mit Konsequenzen rechnen zu müssen. Anhand der Profile der Nutzer (beispielsweise in den Foren des Web-ELF-Korpus) zeigt sich, dass es die meisten Nutzer vorziehen, anonym zu bleiben. Es kann nur mit einigem Aufwand auf die wahre Identität der Nutzer geschlossen werden.

In einem Forum werden üblicherweise einige soziobiografische Angaben der Nutzer angezeigt. So finden sich typischerweise zumindest der Nutzernamen, das Registrierungsdatum und die Gesamtzahl der *Posts*, die die Nutzer in dem jeweiligen Forum gemacht haben. Letzteres zeigt den Grad der Involviertheit der Schreiber im Forum. In einigen Foren wird auch der Aufenthaltsort der Nutzer angegeben. Neben den Diskussionsteilnehmern gibt es in jedem Forum eine hohe Zahl an Rezipienten, die sich nicht aktiv am Diskurs beteiligen, sondern lediglich *Threads* rezipieren. Diese Nutzergruppe kann in den meisten Forensystemen nicht statistisch erfasst werden. Forennutzer können in verschiedene Typen unterschieden werden, Claridge (2007: 92) differenziert in diesem Sinne in *simple readers*, *casual senders* und *hosts*. Jeder dieser Typen ist in unterschiedlichem Maße in einem Forum eingebunden. *Hosts* tragen beispielsweise viel zu Diskussionen bei und können die Rollen von Moderatoren ausfüllen. *Casual Senders* engagieren sich unregelmäßig, während *Simple Readers* ausschließlich Texte rezipieren.

Der Grad der Involviertheit und die allgemeine Rezeption der einzelnen *Threads* können je nach Forum, abhängig vom Themengebiet, enorm variieren. Während auf manchen Forenseiten eher ein breiter thematischer Fokus liegt, wird der Fokus in anderen Foren auf wesentlich spezialisiertere Themen gelegt. Dabei gibt es prinzipiell keine thematischen Einschränkungen, so dass Foren zu allen möglichen Themen existieren. Die Ernsthaftigkeit des Diskurses kann dabei von Fall zu Fall enorm variieren. Die Wahl des Themas hat dabei einen entscheidenden Einfluss auf die Art des Diskurses, der letztendlich in einem Forum stattfindet.

Bis zum heutigen Zeitpunkt sind lediglich einige wenige Studien erschienen, die sich explizit mit sprachlichen Merkmalen des Web-Genres Foren beschäftigt haben. Laut Crystal

(2006: 147) ist Sprachakkomodation ein typisches Merkmal von Forensprache. Crystal betont in diesem Kontext speziell die Ausbildung von *Jargon*, *Slang* und speziellen Abkürzungen, die sich in einer *Web Community* etablieren, da sie zunächst von einzelnen Nutzern verwendet werden und nach und nach von den restlichen Mitgliedern eines Forums übernommen werden. Davis / Brewer (1997: 85ff.) finden unter anderem, dass das Personalpronomen 1. Person Singular (*I*) und dass privative Verben (*think, feel, know*) sehr häufig in Foren verwendet werden. Claridge (2007: 97) stellt ebenfalls fest, dass in Foren viele Personalpronomen verwendet werden, allerdings sei ihre Anzahl geringer als in gesprochener Sprache. So wird *I* 23,3 Mal (*ptw*)¹⁵ gebraucht, während *you* 10,7 Mal (*ptw*) verwendet wird. Ein Vergleich mit dem schriftsprachlichen *FLOB*-Kopus¹⁶ (*I*: 4,9 *ptw* / *you*: 2,9 *ptw*), dem gesprochenen *WSC*-Kopus¹⁷ (*I*: 25,1 *ptw* / *you*: 19,3 *ptw*) und dem gesprochenen *BNC Spoken Demographic*-Kopus (*I*: 39 *ptw* / *you*: 31,3 *ptw*) zeigt, dass Forensprache in Bezug auf Pronominalnutzung zwischen gesprochener Sprache und Schriftsprache verortet werden muss. Die verhältnismäßig häufige Nutzung von Personalpronomen (vor allem der ersten Person Singular) zeigt, dass in Foren wesentlich mehr selbstreferentielle Bezüge hergestellt werden als in geschriebener Sprache. Im Wesentlichen veranschaulicht dies, dass Forennutzer häufig ihre Meinung gegenüber bestimmten Sachverhalten äußern. Ein charakteristisches sprachliches Merkmal von Foren ist zudem sprachliche Redundanz. Aussagen, die von anderen Nutzern getroffen wurden, werden teilweise von mehreren Nutzern wiederholt, um vorangegangene Diskussionen wiederaufzugreifen und zu vertiefen. Laut Claridge (2007: 100) gibt es in Foren eine hohe Zahl an epistemischen Einstellungsmarkern.¹⁸ Dies könnte damit in Verbindung stehen, dass Informationen aus dem Web grundsätzlich mit Misstrauen begegnet wird (Claridge 2007: 100).

Foren müssen von anderen Formen computervermittelter Kommunikation in vielerlei Hinsicht unterschieden werden. Sie sind dialogisch / polylogisch und stilistisch konversationell. Damit weisen sie prinzipiell dieselben Merkmale auf wie die Web-Genres Chat und E-Mail. Im Gegensatz zu diesen sind Foren allerdings öffentlich rezipierbar (Claridge 2007: 87). Im Gegensatz zu Chats sind Foreneinträge nicht flüchtig, sondern können auch noch im Nach-

¹⁵ Pro 1000 Wörtern.

¹⁶ The Freiburg - LOB Corpus of British English (FLOB).

¹⁷ The Wellington Corpus of Spoken New Zealand English (WSC).

¹⁸ Claridge (2007) vergleicht die Zahlen mit den Ergebnissen von Biber (1999) und Conrad / Biber (2009). Sie ist dabei zu den folgenden Ergebnissen gekommen: *Epistemic stance markers*: 12 instances *ptw*. *Attitudinal stance markers*: 2,2 instances *ptw*. *Style stance markers*: 0,5 instances *ptw*.

hinein im Web gefunden werden. Thematisch sind sie offener als Chats und behandeln ernsthaftere und kontroversere Themen (Claridge 2007: 88).

3.5.3.2. Blogs

Weblogs (verkürzt Blogs) gehören nicht zur ersten Generation von Web-Genres (wie beispielsweise E-Mail, Chat oder Foren) und sind erst im Laufe der evolutionären Entwicklung des Webs entstanden. In Blogs liegt eine klassische *one-to-many*-Kommunikationsituation vor. Die ersten Blogs, die im Jahr 1997 gestartet wurden, waren öffentliche persönliche Tagebücher. Blogs wurden von den Produzenten entweder genutzt, um sich selbst darzustellen oder um über ein spezifisches Thema (wie zum Beispiel der politischen Situation in den Vereinigten Staaten von Amerika nach dem 11. September) zu berichten. Walker (2005) definiert das Web-Genre Blog als "a frequently updated Web site consisting of dated entries arranged in reverse chronological order so the most recent post appears first". Charakteristisch für das Web-Genre Blog sind demnach eine häufige Aktualisierung des Inhalts und die umgekehrte chronologische Reihenfolge von veröffentlichten *Posts*.

Das Web-Genre Blog hat in den vergangenen Jahren einen enormen Popularitätsschub erhalten. Während zum Jahresende 2003 100.000 Blogs im Web auffindbar waren, steigerte sich diese Zahl von 3 Millionen im Juli 2004 auf 57 Millionen zum Jahresende 2006 (vgl. Technorati 2012). Laut Schätzungen von BlogPulse existierten im Januar 2012 schließlich etwa 180 Millionen Blogs. Dabei gestaltet es sich schwierig, die genaue Anzahl an aktiven Blogs zu determinieren, da nicht nachvollzogen werden kann, wie viele dieser Blogs inzwischen nicht mehr aktiv sind.

Grundsätzlich kann jeder Webnutzer mit einem Internetzugang einen Blog anlegen. Aus diesem Grund gibt es eine enorme stilistische Bandbreite innerhalb der Blogs, die je nach Produzent und gewähltem Thema stark voneinander divergieren können. Auch die Anzahl der Rezipienten kann enorm variieren, so haben manche Blogs nur einige wenige *Follower* (Leser), während Top-Blogs wie *The Daily Post* von mehreren Millionen Lesern täglich rezipiert werden (vgl. BlogPulse 2012).

Im Laufe der Zeit hat jedoch eine funktionelle Erweiterung innerhalb des Web-Genres stattgefunden, so dass in zunehmenden Maße auch Unternehmen Blogs nutzen, einerseits um Werbung für ihre Produkte zu machen, andererseits um ihre Unternehmenspolitik effektiver zu kommunizieren. Blogs eignen sich dabei in besonderem Maße, da sie Rezipienten auf einer persönlichen Ebene adressieren.

Um Bloggen zu können, nutzen Webnutzer üblicherweise eigens dafür vorgesehene Blogging-Seiten wie Wordpress.com oder blogger.com, auf denen ihnen *Webspace* zur Verfügung gestellt wird. Nach erfolgreicher Anmeldung müssen potenzielle Blogger ein *Template* (eine Vorlage) wählen, das zum Teil an die persönlichen Wünschen der Nutzer angepasst werden kann. Rettberg (2008) glaubt, dass viele Blogger das *Template* ihrer Blogs bewusst auswählen, da es entscheidend sein kann, um die intendierte Botschaft des Blogs zu übermitteln. Rettberg (2008: 4) stellt weiterhin fest: "A blog, however, cannot be read simply for its writing, but will always be seen as the sum of writing, layout, connections and links, and tempo". Die Form eines Blogs hat immer auch einen Einfluss auf die Rezeption von dessen Inhalt. Nach Auswahl des *Templates* kann schließlich die Seite mit Inhalt gefüllt werden. Üblicherweise findet man in Blogs soziobiografische Angaben über die Autoren. Diese können in ihrer Ausführlichkeit variieren. So beinhalten einige Blogs detaillierte soziobiografische Angaben und Bilder der Autoren, während andere Blogs entweder teilweise oder vollständig anonym sind. Blogs werden sowohl von Privatpersonen als auch von Unternehmungen geführt. In beiden Fällen können einzelne oder mehrere Produzenten an der Entwicklung der Inhalte des Blogs beteiligt sein. In Unternehmensblogs werden die *Posts* häufig von einer kleinen Gruppe von Individuen geschrieben (Rettberg: 2008: 21).

Die strukturellen Merkmale von Blogseiten sind typischerweise identisch. Die *Posts* von Bloggern werden gewöhnlich zentral in der Bildmitte angezeigt, in chronologischer Umkehrung, so dass die neuesten Beiträge oben auf der Seite gepostet werden und ältere Beiträge automatisch weiter nach unten rutschen. In vielen Fällen existiert ein Archiv, das am Rand der Seite angezeigt wird und in dem ältere Beiträge, oftmals in segmentierte Zeitabschnitte unterteilt, gespeichert werden. Dies erleichtert es den Rezipienten, ältere *Posts* zu finden. Die Verwendung von Hyperlinks ist ebenfalls ein charakteristisches Merkmal von Blogs. Hyperlinks sind zum einen wichtig um die strukturelle Kohärenz auf der Blogseite aufrecht zu erhalten, zum anderen helfen sie die inhaltliche Stringenz von *Posts* sicherzustellen, da in einigen Blogbeiträgen auf andere Webseiten verwiesen wird, die konsultiert werden müssen, um den Inhalt der *Posts* nachvollziehen zu können.

Ein Blogbeitrag gleicht in seiner Struktur in vielen Fällen einem Zeitungsartikel. Er hat üblicherweise eine Überschrift, die einen Hinweis auf dessen Inhalt gibt oder diesen kurz zusammenfasst. Zudem wird auch der Zeitpunkt des *Posts* immer angezeigt. Einige Blogs erlauben den Rezipienten Kommentare zu den Beiträgen zu machen, dies gilt jedoch nicht für alle Blogs. Blogs haben eine fragmentierte narrative Struktur, die sich in einer Form des episodischen

schen Erzählens widerspiegelt. Durch die vorgegebene Strukturierung (Überschriften, Modularisierung in verschiedene Unterseiten) weisen Blogs eine hohe Kohärenz auf. Die einzelnen *Posts* können sowohl als einzelne episodische Einträge gelesen werden, als auch in einen Zusammenhang mit den anderen vorhandenen Beiträgen gebracht werden. Ein hervorstechendes Merkmal von Blogs ist, dass die Blogeinträge typischerweise relativ kurz gehalten sind, dafür jedoch häufig neue Einträge gemacht werden. Aufgrund der Fülle an Text im Web und der mangelnden Zeit, die Webnutzern zur Verfügung steht, werden längere Texte im Web oftmals nicht in ihrer Gänze, sondern nur flüchtig rezipiert (Nielsen 1997). Aus diesem Grund werden Texte im Web oftmals nur auf das Wichtigste reduziert.

Aus stilistischer Sicht weisen Blogs im Vergleich zu kommerziellen Homepages einen eher informellen Charakter aus (Walker 2005). Jedoch hängt dies auch stark davon ab, um welchen Typ von Blog es sich handelt, da die Funktion eines Blogs dessen Form determinieren kann. Rettberg (2008: 9) unterscheidet zwischen drei Arten von Blogging: *Personal Blogging*, *Filterblogging* und *Topic-driven Blogging*.

Personal Blogging charakterisiert, dass Autoren Persönliches über sich selbst berichten. Die Autoren geben dabei üblicherweise persönliche soziobiografische Informationen preis und nennen Inhalt und Ziel des Blogs in einem einleitenden Teil. Der Blog kann schließlich um persönliche Fotos, oder auch Videos ergänzt werden. Im Gegensatz zu *Topic-driven Blogging* wird der Fokus nicht auf ein spezielles Thema gelegt. Ein typisches Beispiel für ein *Personal Blog* ist der Blog von Willy Sudiarto Raharjo.¹⁹

Filterblogging dient in erster Linie dazu, die Flut an Homepages im Web zu filtern, um daraus die relevantesten Webseiten herauszukristallisieren. Auf *Filterblogs* finden sich von anderen Webseiten zusammengetragene Informationen, die kondensiert wiedergegeben werden und Verlinkungen zu anderen themenrelevanten Webseiten. Ein typisches Beispiel für ein *Filterblog* ist die Webseite *Boing Boing*, auf der Neuigkeiten über bizarre Funde des Webs zusammengetragen werden.²⁰

Topic-driven Blogging sind themenbezogene Blogs, die zwar oftmals auch Elemente persönlicher Blogs aufweisen können, jedoch im Gegensatz zu diesen nicht den Produzenten des Blogs, sondern ein spezielles Thema in den Vordergrund rücken. Das Ziel von *Topic-driven Blogging* ist es, die Rezipienten zu informieren oder kritische Fragen aufzuwerfen, die die Rezipienten zum Denken anregen sollen. *Coping with Childhood Cancer* ist ein Beispiel für ein

¹⁹ Vgl. <http://willysr.blogspot.com> (30.12.2012).

²⁰ Vgl. <http://boingboing.net> (30.12.2012).

Topic-driven Blog, in dem Eltern Informationen zur Behandlung von krebskranken Kindern erhalten können.²¹

Herring et al. (2005) haben ermittelt, dass 70,4% aller Blogs *Personal Blogs* darstellen, während 12,6% aller Blogs *Filterblogs* sind. *Topic-driven Blogs*, die Herring et al. (2005) als *k-logs (knowledge logs)* bezeichnen, bilden mit 4,5% nur einen kleinen Teil aller Blogs. 90,8% der von Herring et al. (2005) untersuchten Blogs sind von nur einem Autor erstellt worden. Sie wurden durchschnittlich alle 2,2 Tage aktualisiert und beinhalteten 0,65 Links pro *Post*. 43% der Blogs erlaubten Kommentare, 9,2% von ihnen enthielten Bilder (vgl. Herring et al. 2005).

Blogs unterscheiden sich häufig in erheblicher Weise im Hinblick auf Autorenschaft, Hyperlinknutzung, Kommentarfunktion und Aktualität der Einträge (Warschauer et al. 2010: 495). Eine starke *authorial voice* animiert Rezipienten dazu, den Blog weiter zu verfolgen (Warschauer et al. 2010: 497). Die meisten Blogs werden sehr selten rezipiert und nur eine verschwindend geringe Zahl von ihnen erfährt einen großen Zuspruch seitens der Rezipienten (vgl. Shirky 2003, Technorati 2012).

3.5.3.3. Homepages

Homepages dienen als "Ausgangs-, Übersichts- und Ankerpunkt in einem Hypertextnetz wie dem World Wide Web [...]" (Bittner 2003: 71). Prinzipiell können Homepages dichotomisch in die beiden Kategorien persönliche Homepages und kommerzielle Homepages differenziert werden. Persönliche Homepages werden von einzelnen Personen betrieben und dienen dazu, individuelle Interessen, Hobbies und auch berufliche Themen zu thematisieren (Bittner 2003: 100). Im Gegensatz dazu werden kommerzielle Homepages von Unternehmen mit finanziellem Interesse betrieben, die Informationen zu einem Produkt anbieten, um die Rezipienten der Webseite zu einer Inanspruchnahme der Dienstleistung zu bewegen.

In der vorliegenden Arbeit werden ausschließlich kommerzielle Homepages untersucht. Das Ziel jeder kommerziellen Homepage ist es, eine Dienstleistung zu verkaufen. Durch eine transparente Informationspolitik und eine professionelle Außendarstellung soll Vertrauen geschaffen werden. Aus diesem Grund ist es für die Betreiber der Seiten von essentieller Bedeutung eine Webseite zu erstellen, die sowohl inhaltlich als auch strukturell fehlerfrei ist, aber auch ständig aktualisiert wird. Üblicherweise werden kommerzielle Homepages professionell editiert, um sie regelmäßig auf ihre inhaltliche und strukturelle Kohärenz zu überprüfen.

²¹ Vgl. <http://www.copingwithchildhoodcancer.blogspot.com> (30.12.2012).

Grundsätzlich gleichen sich kommerzielle Homepages relativ stark in ihrer Struktur. Die Navigationsseite ist klar strukturiert und verlinkt auf mehrere Unterpunkte. Unter anderem wird in diesen Unterabschnitten der Zweck der Homepage beschrieben und der Betreiber der Dienstleistung vorgestellt. In einem weiteren Unterabschnitt werden die einzelnen Dienstleistungen aufgeführt. Oftmals gibt es einen Unterabschnitt, in dem ein Überblick über alle Neuigkeiten geliefert wird. Zudem haben kommerzielle Homepages immer einen Unterabschnitt, in dem Kontaktinformationen aufgeführt werden.

3.5.3.4. Wikis

Der Softwareautor Ward Cunningham konzipierte 1995 das erste im Web gehostete *Wiki* (hawaiianisch für "schnell"): *WikiWikiWeb* (Baron 2008: 116). Wikis ermöglichen Webnutzern in einem kollaborativen Prozess Texte zu produzieren. Jeder beteiligte Autor hat dabei die Möglichkeit, beliebig oft einen Text zu verändern (Richardson 2006). Für eine erfolgreiche Zusammenarbeit ist es dabei nicht erforderlich, dass sich die Autoren persönlich kennen oder physisch zusammenkommen müssen, um einen Text zu produzieren. Eine anonyme Zusammenarbeit ist durchaus üblich.

Im Gegensatz zu anderen Web-Genres ist es wesentlich schwieriger die exakte Anzahl an existierenden Wikis im Web zu bestimmen. Einerseits existiert die Suchmaschine www.wiki.com, mit deren Hilfe man nach Wikis suchen kann. Andererseits sind viele Wikis nicht öffentlich nutzbar, da sie passwortgeschützt nur für eine kleine Gruppe von Webnutzern zugänglich sind. Wikis werden vor allem für berufliche Kontexte genutzt, beispielsweise in Firmen, oder in Lehr- und Forschungskontexten. Sie sind von anderen Web-Genres zu unterscheiden, wie Warschauer et al. (2010: 500) feststellen: "Whereas mail and chat, the most traditional CMC genres, facilitate informal, author-centric, personal exchange, writing on a wiki facilitates more formal, topic-centric, depersonalized exchange". In Wikis steht ausdrücklich das Thema und nicht der Produzent des Textes im Vordergrund.

Wikis fanden zu Beginn ihrer Nutzung besonders bei Softwareentwicklern einen regen Zuspruch, da sie das schnelle Austauschen von Ideen und Konzeptentwürfen ermöglichten. Bis zum Jahr 2001 erfreuten sich Wikis allerdings jedoch nur in diesen ausgewählten Kreisen großer Popularität. Als bahnbrechend für die Entwicklung seitdem stellte sich Wikipedia heraus. Wikipedia ist eine frei zugängliche Online Enzyklopädie, an der sich prinzipiell jeder Webnutzer als Mitautor betätigen kann. Laut Baron (2008: 121) haben im September 2006 75.561 aktive Editoren Wikiartikel geschrieben. 45.001 dieser Editoren verfassten dabei englisch-

sprachige Artikel. Die Gesamtzahl der Autoren ist bis zum Jahr 2012 angestiegen, so dass sich inzwischen etwa 100.000 Autoren an der Erstellung von Artikeln beteiligen (Wikipedia 2012). Seit der Gründung von Wikipedia im Jahr 2001 wurden bis zum Jahr 2007 6,4 Millionen Artikel mit 236 Millionen Editierungen von 5,77 Millionen Webnutzern erstellt (Wilkinson / Huberman 2007). Dabei zeigt sich, dass die Anzahl der Kollaboratoren und der Editierungen der jeweiligen Artikel einen Einfluss auf deren Qualität hat (Wilkinson / Huberman 2007).

Um sich in den Schreibprozess als Autor einbringen zu können, ist es in jedem Fall erforderlich, sich offiziell auf der Plattform zu registrieren. Bedingt durch das Wiki-Prinzip kann jeder Artikel jederzeit verändert und ergänzt werden. Änderungen im Vergleich zu früheren Versionen können stets nachvollzogen werden, da diese weiterhin einsehbar bleiben. Alle beteiligten Autoren werden darauf hingewiesen, dass sie bestimmte Grundsätze der wissenschaftlichen Textproduktion beim Verfassen der Artikel beachten müssen. Dazu gehört unter anderem, dass diese "dem Grundsatz des neutralen Schreibens entsprechen" (Wikipedia 2012) und geltendes Recht, wie zum Beispiel das Urheberrecht, beachtet wird. Jeder Eintrag muss zudem wissenschaftlich belegt sein und genügend Referenzen vorweisen, um nicht wieder entfernt zu werden. Durch diese Vorgaben soll sichergestellt werden, dass die Artikel den Ansprüchen einer Enzyklopädie genügen. Für den Fall einer Nichtbeachtung dieser Grundsätze drohen Sanktionen. Im besten Fall führt dies zu einer Entfernung des Beitrags oder einer Zurücksetzung auf die vorige Version des Texts, im schlechtesten Fall endet dies mit einem Ausschluss der Autoren. Für eine Einhaltung dieser Grundsätze sorgt ein Kontrollgremium, das aus ernannten Moderatoren besteht, die regelmäßig die Beiträge im Hinblick auf die Vorgaben überprüfen.

Kritiker werfen Wikipedia vor, dass die Artikel allesamt von Laien erstellt worden sind und dementsprechend nicht den Gehalt einer von Experten geschriebenen wissenschaftlichen Enzyklopädie, wie sie beispielsweise die Encyclopaedia Britannica darstellt, haben kann. Die Anonymität der Autoren gepaart mit den zum Teil fehlenden referentiellen Verweisen führt insbesondere im akademischen Kontext dazu, dass Wikipedia als unwissenschaftliche Quelle betrachtet wird, die möglichst nicht zitiert werden sollte. Zudem wird der Plattform eine stilistische Heterogenität zugeschrieben, mit großen qualitativen Unterschieden zwischen den individuellen Beiträgen (Rosenzweig 2006). Es erscheint plausibel, dass Wikipedia-Artikel stilistisch heterogen sein müssen, da sie von einer Vielzahl von Autoren in einem gemeinschaft-

lichen Prozess produziert werden und jeder Kollaborant seinen eigenen idiosynkratischen Schreibstil miteinbringt. Dies ist jedoch nicht der Fall, wie Baron feststellt:

As Wikipedia has grown, it has become increasingly uniform in style. Through the urgings (sometimes not so gentle) of fellow writers, contributors progressively prepare entries using a standardized format, include references, and embed links to other Wikipedia articles. (Baron 2008: 121)

Baron beobachtet eine zunehmende stilistische Uniformität, die dem Standardisierungsbestreben der Mitautoren geschuldet ist. Es muss in diesem Kontext darauf hingewiesen werden, dass in Wikipedia eine klare hierarchische Ordnung vorherrscht. So gibt es unterschiedliche Positionen innerhalb von Wikipedia, die mit einer unterschiedlichen Fülle von Machtbefugnissen einhergehen. Analog zu der Hierarchisierungsstruktur in Foren gibt es einfache Schreiber (*stewards*), Moderatoren (*bureaucrats*) und Administratoren (*administrators*). Letztere haben die höchste Verfügungsgewalt und können sowohl bestimmte Artikel sperren als auch Autoren ausschließen. Eine Beförderung in der Hierarchieebene ist generell möglich (*community-based promotions*). Autoren haben zwar die Möglichkeit, sich unter ihrem richtigen Namen und unter Angabe soziobiografischer Daten anzumelden, können jedoch ebenso anonym bleiben. Innerhalb der Wikipedia-Community können sich Mitglieder Affinitätsgruppen anschließen und sich miteinander austauschen.

Die aufgeworfene These, dass der Schreibstil in Wikipedia im Vergleich zu professionellen Enzyklopädien laienhaft und unwissenschaftlich ist, wurde in den letzten Jahren auf ihre Richtigkeit überprüft. So hat Bell (2007) die Online-Präsenz der Encyclopaedia Britannica mit Wikipedia verglichen. Die zugrundeliegenden Parameter waren dabei Lesbarkeit, Syntax und der Gebrauch von faktischen Aussagen im Verhältnis zum Gebrauch von wertenden Aussagen. Den Ergebnissen zufolge sind demnach beide Enzyklopädien in allen drei Kriterien in stilistischer Hinsicht vergleichbar. Elia (2007) untersuchte die *lexical density*, die durchschnittliche Wortlänge und den Gebrauch von *formal nouns* und *impersonal pronouns*. Die Ergebnisse der Studie bestätigen dabei, dass sich Wikipedia und die Encyclopaedia Britannica sowohl auf formaler als auch auf stilistischer Ebene in vielen Punkten ähneln. Die Anzahl von *formal nouns* und *impersonal pronouns*, typisch für den formellen Diskurs (5,3 vs. 5,2), und die Durchschnittswortlänge (5,4 vs. 5,2) sind fast identisch (Elia 2007: 18). Dahingegen sind Britannica-Artikel im Durchschnitt halb so lang (1728 gegenüber 3510 Wörtern in Wikipedia), haben eine höhere lexikalische Dichte (44,9% gegenüber 31,4% in Wikipedia)

und ein höheres *Level of total formality* (50,2% gegenüber 36,6% in Wikipedia). Insgesamt schlussfolgert Elia (2007: 18), dass Wikipedia-Artikel einen formalen und standardisierten Stil innehaben, der dem Stil der Encyclopaedia Britannica, trotz der Unterschiede in der Artikelgröße und trotz der höheren Anzahl an Hyperlinks, sehr ähnelt.

Der große Erfolg von Wikipedia beruht zum einen auf der Tatsache, dass es frei zugänglich ist, zum anderen darauf, dass nicht ausschließlich anerkannten Experten vorbehalten bleibt, Beiträge zu produzieren. Zudem können selbst mangelhafte Texte durch mehrstufige Überarbeitungen zu nützlichen Texten werden. Aus diesem Grund konnte innerhalb kurzer Zeit eine Fülle von Texten zu unterschiedlichsten Themen produziert werden. So existierten im Dezember 2012 etwa 4,1 Millionen englischsprachige Einträge. Dies übertrifft die Anzahl der Beiträge der Encyclopaedia Britannica um ein Vielfaches. Die Beiträge auf Wikipedia sind jedoch nicht nur auf die englische Sprache beschränkt; es gibt ebenfalls eine riesige Anzahl an Beiträgen in vielen anderen Sprachen. Im Dezember 2012 lagen etwa 23 Millionen Beiträge in 284 Sprachen vor (Wikipedia 2012). Mittlerweile existiert sogar ein *Simple English Wikipedia*. In dieser Varietät werden zum einen nur ein Basisvokabular des Englischen, zum anderen einfachere grammatikalische Strukturen verwendet. Diese in ihrer Komplexität reduzierte Englischvarietät ermöglicht es insbesondere Menschen mit Lernschwierigkeiten, einen Zugang zu Wissen zu erlangen (Warschauer et al. 2010: 500).

3.5.3.5. Eine Modellierung der Web-Genres

Für eine genauere Beschreibung der einzelnen Web-Genres bietet es sich an, eine Auswahl von Parametern zu bestimmen, anhand derer Gemeinsamkeiten und Unterschiede bezüglich ihrer Kommunikationssituation und Kommunikationsbedingungen offenbart werden können. Die acht ausgewählten Parameter wurden von Koch / Oesterreicher (1985, 1990) übernommen und die Web-Genres Foren, Blogs, kommerzielle Homepages und Wikipedia-Artikel diesbezüglich miteinander verglichen. Es muss dabei ausdrücklich betont werden, dass alle Subkorpora aus Texten bestehen, die im Kontext des Themas *Medical Surgery / Medical Tourism* entstanden sind. Dieser thematische Fokus ist entscheidend für die Bewertung der Parameter in den einzelnen Web-Genres. Die Tabelle 3 bietet eine Zusammenfassung dieses Vergleichs.

Parameter	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
a) Grad der Vertrautheit der Partner	gering	gering	gering	sehr gering
b) Grad der emotionalen Beteiligung	hoch	mittel	gering	sehr gering
c) Grad der Situations- und Handlungseinbindung	sehr hoch	mittel	gering	sehr gering
d) Physische Nähe der Kommunikationspartner	nicht existent	nicht existent	nicht existent	nicht existent
e) Grad der kommunikativen Kooperation	hoch	gering	nicht existent	nicht existent
f) Grad der Dialogizität	sehr hoch	gering	sehr gering	nicht existent
g) Grad der Spontaneität	sehr hoch	gering	sehr gering	sehr gering
h) Grad der Themenfixierung	mittel	hoch	sehr hoch	sehr hoch

Tabelle 3: Vergleich der Kommunikationsbedingungen der Web-Genres

a) Grad der Vertrautheit der Partner

Eine Vertrautheit der Partner ist charakteristisch für Mündlichkeit, während eine Fremdheit der Partner eher ein Ausdruck von Schriftlichkeit ist. In den Web-Genres Blogs, kommerzielle Homepages und Wikipedia-Artikel herrscht zwischen Produzent und Rezipient ein geringes Maß an Vertrautheit. In Foren variiert dies grundsätzlich stark. Abhängig von der Art und Größe des Forums kann dabei ein fester Kern aus Nutzern entstehen, der sich über einen längeren Zeitraum an Diskussionen beteiligt. Für die untersuchten Foren zeigt sich jedoch, dass der Grad der Vertrautheit gering ist und sich nur wenige Nutzer in einem permanenten Austausch befinden.

b) Grad der emotionalen Beteiligung

Ein hohes Maß an Emotionalität ist ein typisches Merkmal von Mündlichkeit, geringe Emotionalität ein Merkmal von Schriftlichkeit. Emotionale Beteiligung wird erreicht durch die Nutzung von sprachlichen und orthografischen Mitteln (Imperative, Emoticons, Reihung von Satzzeichen), aber auch durch den mündlich-sprachlichen Stil des Diskurses (vgl. Claridge 2007). Es können zudem Videos und *Soundfiles* in den Text eingebunden werden, um emotionale Beteiligung herbeizuführen (Bittner 2003: 88). Die strukturelle Gestaltung der Webseite (Farbgebung, Layout) kann ebenfalls einen Teil dazu beitragen, den Rezipienten zu emotionalisieren. In Foren gibt es einen hohen Grad an emotionaler Beteiligung, in Wikipedia-Artikel

hingegen wird Emotionalität vermieden, um wissenschaftliche Objektivität zu gewährleisten. Blogs können diesbezüglich stark voneinander divergieren.

c) Grad der Situations- und Handlungseinbindung

Eine hohe Situations- und Handlungseinbindung ist ein Ausdruck von Mündlichkeit. Kommerzielle Homepages sind oftmals situations- und handlungsunabhängig. Trotzdem stellt Bittner (2003: 89) fest, dass es grundsätzlich möglich ist, eine Situations- und Handlungseinbindung künstlich zu emulieren, beispielsweise im Falle einer Homepage, die abhängig von der jeweiligen Tageszeit den Rezipienten auf unterschiedliche Weise begrüßt. In Foren ist der Grad der Situations- und Handlungseinbindung sehr hoch, in Blogs variiert er: so kann der Blogproduzent beispielsweise auf den Kommentar eines Rezipienten eingehen.

d) Physische Nähe der Kommunikationspartner

Physische Nähe ist Ausdruck von Mündlichkeit, physische Distanz Ausdruck von Schriftlichkeit. Das Medium Internet bedingt die fehlende physische Nähe der Kommunikationspartner. Webnutzer bewegen sich in einem virtuellen Raum, der grundsätzlich den Distanzcharakter von Websprache auszeichnet und der nur durch sprachliche Hilfsmittel überbrückt werden kann.

e) Grad der kommunikativen Kooperation

Je höher der Grad an kommunikativer Kooperation, desto höher ist der Grad an Mündlichkeit. Je nach Web-Genre gibt es ein höheres oder geringeres Maß an kommunikativer Kooperation. Kommunikative Kooperation bezieht sich ausdrücklich auf einen sprachlichen Austausch zweier oder mehrerer Sprachverwender. Auf kommerziellen Homepages ist beispielsweise, bedingt durch die Trennung von Produktion und Rezeption, keine Kooperation mit den Rezipienten möglich (Bittner 2003: 90). Im gleichen Maße kommunizieren Autoren in Wikis nicht miteinander, obwohl sie prinzipiell gemeinsam an einem Text arbeiten. In Foren gibt es, bedingt durch die dialogische Kommunikationssituation, ein hohes Maß an kommunikativer Kooperation. In Blogs ermöglicht die Kommentarfunktion grundsätzlich eine kommunikative Kooperation. Abhängig vom Blogtyp wird davon mehr oder weniger Gebrauch gemacht. In den für die vorliegende Arbeit untersuchten Blogs wird die Kommentarfunktion selten genutzt.

f) Grad der Dialogizität

Dialogizität ist Ausdruck von Mündlichkeit, während Monologizität Ausdruck von Schriftlichkeit ist. Kommerzielle Homepages haben einen unidirektionalen Informationsfluss. Die Besucher der Homepage nehmen ausschließlich die Rolle des Rezipienten ein. Aus diesem Grund ist ein dialogischer Austausch prinzipiell nicht möglich und kann nur durch andere auf der Homepage eingebundene Funktionen, wie zum Beispiel Kontaktformulare, geschehen. Forenseiten hingegen haben einen polydirektionalen Informationsfluss. Das Ziel von Forenseiten ist es, dass Nutzer sich treffen, um sich über festgelegte Themen miteinander auszutauschen.

g) Grad der Spontaneität

Ein hoher Grad an Spontaneität ist typisch für Mündlichkeit, während ein hoher Grad an Reflektiertheit typisch für Schriftlichkeit ist. Kommerzielle Homepages und Wikipedia-Artikel benötigen einen hohen Produktionsaufwand, da orthografische Fehler und inhaltliche Fehler mit mangelnder Professionalität gleichgesetzt werden. In Blogs ist dies stark abhängig vom Thema: je ernster das Thema, desto höher ist der Planungsaufwand für den Produzenten. Der thematische Schwerpunkt *Medical Surgery / Medical Tourism* bedingt ein geringes Maß an Spontaneität. In Foren hingegen werden Beiträge üblicherweise spontan verfasst und nicht wiederholt revidiert.

h) Grad der Themenfixierung

Ein geringes Maß an Themenfixierung ist typisch für Mündlichkeit, ein hohes Maß an Themenfixierung ist typisch für Schriftlichkeit. Der Produzent einer kommerziellen Homepage bestimmt über ihre thematische Fixierung, da er als Einziger mit Befugnissen ausgestattet ist, Inhalte zu verändern. Wikipedia-Artikel beschränken sich auf ein Themengebiet. Je nach Blogtyp variiert der Grad der Themenfixierung. Die thematische Fixierung ist jedoch bedingt durch das Thema *Medical Surgery / Medical Tourism* hoch. In Foren gibt es üblicherweise nur ein geringes Maß an Themenfixierung. Theoretisch wird durch die Eröffnung von neuen *Threads* versucht, eine klare Themenfixierung zu schaffen. In der Realität geht jedoch insbesondere in größeren *Threads* oftmals der thematische Fokus verloren, so dass eine Themenfixierung nicht sichergestellt werden kann. Die Foren, die im Web-ELF-Korpus untersucht werden, haben jedoch eine stärkere thematische Fixierung als andere Forenseiten. Dies liegt daran, dass in den *Threads* sehr spezielle und oftmals ernste Probleme thematisiert werden. Es

liegt im Interesse der beteiligten Autoren, dass auf diese Sachverhalte auf sensible und seriöse Art und Weise geantwortet wird.

Ein Vergleich der Web-Genres bezüglich ihrer Parameter erlaubt es, diese in das Nähe-Distanz-Modell von Koch / Oesterreicher (1985, 1994) einzuordnen (vgl. Abbildung 6).

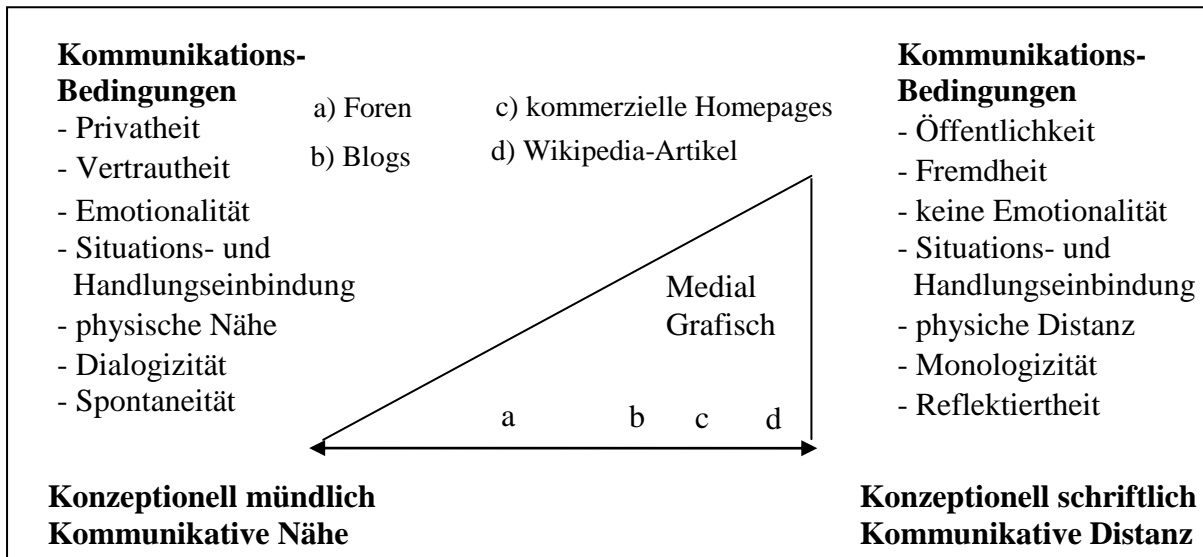


Abbildung 6: Web-Genres im Nähe-Distanz-Modell

Dabei zeigt sich, dass Foren viele Elemente konzeptioneller Mündlichkeit aufweisen und damit eher dem Pol der Mündlichkeit zugeordnet werden können. Die anderen drei Web-Genres müssen eher dem Pol der Schriftlichkeit zugeordnet werden. Dabei muss jedoch eine graduelle Unterscheidung vorgenommen werden. Während Blogs noch einige Merkmale konzeptioneller Mündlichkeit aufweisen, beinhalten kommerzielle Homepages wenige dieser Merkmale. Wikipedia-Artikel schließlich weisen ausschließlich Elemente konzeptioneller Schriftlichkeit auf.

3.6. Zusammenfassung

In dem vorliegenden Kapitel habe ich aufgezeigt, dass Besonderheiten des Sprachgebrauchs im Web existieren, die auf das Medium zurückzuführen sind. Gleichzeitig wurde deutlich herausgearbeitet, dass der Sprachgebrauch im Web nicht verallgemeinernd unter einem monolithischen *Netspeak*-Konzept zusammengefasst werden kann, da die einzelnen Web-Genres strukturell und sprachlich sehr unterschiedlich sind. Auch die in der vorliegenden Arbeit untersuchten vier Web-Genres erwiesen sich als sehr heterogen. Ein Vergleich ihrer Kommuni-

kationsbedingungen ermöglichte eine Verortung der Web-Genres in das Nähe-Distanz-Modell von Koch / Oesterreicher. Demnach können Foren eher dem Pol der konzeptionellen Mündlichkeit zugeordnet werden, während Blogs, kommerzielle Homepages und Wikipedia-Artikel näher am Pol der konzeptionellen Schriftlichkeit positioniert werden können.

Es liegt im Erkenntnisinteresse der vorliegenden Arbeit, phraseologische Unterschiede zwischen den vier untersuchten Web-Genres offenzulegen. Dabei steht zu erwarten, dass diese Unterschiede anhand einer formalen und funktionalen Analyse von Mehrworteinheiten konkret sichtbar gemacht werden können. Zunächst werde ich jedoch im nächsten Kapitel näher auf die theoretische Konzeption von Mehrworteinheiten und die Formelhaftigkeit von Sprache (*Formulaic Language*) eingehen. Dazu werde ich eine Eingrenzung des Konzepts leisten und dann die Bedeutung von formelhafter Sprache für den Sprachgebrauch herausarbeiten. Im Anschluss werde ich aktuelle Studien beschreiben, bevor schließlich das Mehrworteinheiten-Konzept definiert wird, das der vorliegenden Arbeit zu Grunde liegt. Das Kapitel schließt mit der Formulierung der Forschungsfragen, die im Fokus dieser Arbeit liegen, ab.

4. Mehrworteinheiten

Every genre has its own phraseology, and learning to be effective in the genre involves learning this. (Ellis / Simpson-Vlach 2009: 62)

4.1. Die Bewertung von Mehrworteinheiten

Die Korpuslinguistik hat seit den 1970er Jahren durch die Auswertung großer Sprachdatensammlungen aufgezeigt, dass im natürlichen Sprachgebrauch in erheblichem Maße sprachliche Routinen verwendet werden, die aus größeren lexikalischen Einheiten bestehen. Obwohl solchen Mehrworteinheiten seitdem zunehmend mehr Bedeutung beigemessen wurde, wurden zunächst nur wenige phraseologische Studien zur Bedeutung von Mehrworteinheiten für den Sprachgebrauch durchgeführt. Erst die Ausführungen von Sinclair (1991) zum Konzept des *Idiom Principle* rückten lexikogrammatische Routinen (beispielsweise Kollokationen und *Phrasal Verbs*) wieder vermehrt ins Blickfeld der Sprachforschung.

Laut Sinclair (1991) wirken im Sprachprozess jedes Menschen zwei unterschiedliche Prozesse zusammen. Diese beiden Prinzipien der Sprachverarbeitung bezeichnet er als *Idiom Principle* und *Open-Choice Principle*. Dem *Open-Choice Principle* nach ist Sprache das Resultat vieler komplexer Möglichkeiten: "[a]t each point where a unit is completed (a word, phrase, clause), a large range of choice opens up and the only restraint is grammaticalness" (Sinclair 1991: 109). Dieses Prinzip wird auch als *Slot and Filler*-Modell bezeichnet, da Sprachverwender einem permanenten analytischen Prozess unterstehen, der sie dazu bewegt, die *Slots* ihrer Äußerungen auszufüllen. Demnach steht es den Sprachverwendern frei, aus einer enormen Vielfalt von Möglichkeiten (die ihrem Sprachschatz entspricht) eine passende Äußerung auszuwählen, die sich als geeignet erweist, um den *Slot* in dem jeweiligen Kontext zu füllen. Dieses Prinzip stimmt grundsätzlich mit der Konzeption der generativen Grammatik überein. Sinclair ergänzt das *Open-Choice Principle* jedoch um das *Idiom Principle*, das nicht mit dem Verständnis der Generativisten vereinbar scheint. Sinclair (1991) behauptet, dass Sprachverwender nicht in jeder Situation in einem kreativen Akt aus einer unendlichen Menge an Alternativen eine Äußerung auswählen. In vielen Fällen nutzen sie stattdessen vorstrukturierte, sprachliche Routinen. Diese Idee fasst Sinclair (1991: 110) unter dem Begriff des *Idiom Principle* zusammen: "The principle of idiom is that a language user has available to him or her a large number of semipreconstructed phrases that constitute single choices, even though they might appear to be analysable into segments".

Das *Idiom Principle* tritt demnach immer an den Stellen ein, an denen das *Open-Choice Principle* nicht greift. Es besagt, dass in einer Sprache viele vorstrukturierte Elemente (*patterns*) existieren, so zum Beispiel Idiome und Sprichwörter, die kognitiv holistisch und nicht analytisch verarbeitet werden. Diese Muster dienen als Schablonen, die verwendet werden können, wann immer es eine kommunikative Situation erfordert. Die Existenz des *Idiom Principle* widerspricht damit dem generativistischen Grundgedanken, dass jede Sprachäußerung einen kreativen Akt darstelle. Anhand von Beispielen kann dies belegt werden. In einigen Kontexten steht einem englischen Sprachverwender beispielsweise keine Variationsmöglichkeit zur Verfügung (sogenannte *Cranberry Collocations*). So kann der Ausdruck "kith and..." ausschließlich mit dem Element "kin" ergänzt werden (cf. Moon 1998: 78).

Laut Sinclair (1991) ist das Zusammenspiel der beiden Prinzipien essentiell für die menschliche Sprachproduktion. Lexik und Grammatik interagieren demnach in einem ständigen Zusammenspiel und können nicht voneinander getrennt werden. Sinclair (1991:114) schlussfolgert daher: "a model of language which divides grammar and lexis, and which uses the grammar to provide a string of lexical choice points, is a secondary model". Die Sprachproduktion und das Sprachverständnis werden maßgeblich vom *Idiom Principle* bestimmt, wie Wray (2002: 18) feststellt:

In adult speakers [...], the relative balance of the two systems in operation appears to be in favour of the holistic, for we prefer a pragmatically plausible interpretation over a literal one, and we seem able to use with ease formulaic sequences whose internal form we have, apparently, never engaged with.

Hunston / Francis (2000) führen Sinclairs Konzept des *Idiom Principles* / *Open-Choice Principles* mit ihrem Konzept des *Pattern Flow* weiter. Demnach konstituieren Wörter eigene Muster und sind in kohärenten Texten auch Bestandteile anderer Muster. Renouf / Sinclair (1991: 140) führen weiter aus: "such a model, in which the linguistic unit for selection is a series of words rather than a single word, and the units blend into each other, is quite different from the 'slot-and-filler' models of old". Dem *Idiom Principle* nach sind sprachliche Routinen, die auch als formelhafte Sprache (*Formulaic Language*) bezeichnet werden, somit ein elementarer Bestandteil von Sprache.

Das Konzept der *Formulaic Language* wird in der Sprachwissenschaft auf verschiedene Weise verstanden und mit unterschiedlichen Termini belegt. Nattinger / DeCarrico (1992: 1) bezeichnen diese beispielsweise als *Lexical Phrases* und definieren sie als

[...] multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax, conventionalized form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time.

Neben dem Terminus *Lexical Phrases* existiert eine Vielzahl weiterer Begriffe, die näher erläutert werden müssen. Zunächst erfolgt jedoch eine chronologische Darlegung der Rezeption von *Formulaic Language* in der Sprachforschung.

4.2. Das Konzept der *Formulaic Language*

Die Nutzung formelhafter Sprache wurde bereits in der Mitte des 19. Jahrhunderts durch den Neurologen John Hughlings Jackson dokumentiert, der nachwies, dass aphasische Patienten ungewöhnlich viele sprachliche Routinen verwendeten (Wray 2002: 7). Die Tragweite der Bedeutung formelhafter Sprache für den Sprachgebrauch wurde jedoch erst zu Beginn des 20. Jahrhunderts allgemein sichtbar (vgl. de Saussure 1916 / 1966, Sapir 1921 und Bloomfield 1933). Jespersen (1924 / 1976) betont, dass es Sprachverwender prinzipiell kognitiv überfordern würde, sich jedes kleine Detail einer Sprache separat einzuprägen. Aus diesem Grund nutzen diese eine Vielzahl sogenannter *Formulas*. Eine *Formula* definiert Jespersen (1924 / 1976: 88) als "[...] a unit which cannot be further analyzed or decomposed in the way a free combination can". Diese Einheit könne dabei einen ganzen Satz, eine Gruppe von Wörtern, ein Wort oder sogar nur Teile eines Wortes darstellen. Hymes (1962 / 1968: 126f.) hebt die Bedeutung häufig auftretender Sprachmuster für die Sprache deutlich hervor:

A vast proportion of verbal behavior [...] consists of recurrent patterns, of linguistic routines [...] including] the full range of utterances that acquire conventional significance for an individual, group or whole culture.

Bolinger (1976: 1) stellt diesbezüglich fest: "our language does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs". Die Bedeutung von formelhafter Sprache für die Entwicklung des Sprachgebrauchs wird von Wray (2008) verdeutlicht. Wray (2008: 206) legt dar, dass während des evolutionären Übergangs von einer Protosprache zu einer voll entwickelten Sprache

die Notwendigkeit zur kognitiven Komplexitätsreduktion bestand, und sich dies letztendlich in der Entstehung sprachlicher Routinen niederschlug:

It [the modern language] introduced additional complexity and, at the point when working memory could no longer cope, a rescue strategy was needed to minimize the processing demands. This strategy entailed the processing of language in larger chunks.

Trotz der Erkenntnis, dass sprachliche Routinen ein elementarer Bestandteil von Sprache sind, dauerte es lange, bis diese Einsicht Beachtung in zeitgenössischen Grammatiken fand. Als Beispiel sei zu erwähnen, dass Quirk et al. (1972) in der bis dato umfassendsten Grammatik, die je veröffentlicht wurde, nur 3 von 1110 Seiten dem Thema *Formulaic Language* widmen. Lyons (1968) erwähnt zwar *Ready-Made Expressions*, misst ihnen jedoch nur eine untergeordnete Bedeutung im Sprachgebrauch bei.

Inzwischen haben sprachliche Routinen Eingang in zeitgenössische Grammatiken gefunden, zum Beispiel in der Konstruktionsgrammatik (Fillmore et al. 1988; Croft 2001; Goldberg 1995, 2006; Tomasello 2003; Wray 2002) und der kognitiven Grammatik (Langacker 1987, 1991, 2008; Taylor 2002).²² In der *Longman Grammar of Spoken and Written English* (Biber et al. 1999) werden dem Thema sprachlicher Routinen bereits 50 Seiten gewidmet.

In den letzten Jahren hat eine Fülle an Studien untermauert, dass *Formulaic Language* ein alldurchdringendes Merkmal menschlicher Sprache ist (vgl. Altenberg 1998, Spöttl / McCarthy 2004, Ellis / Simpson-Vlach 2009). Sie haben unter anderem aufgezeigt, dass Wörter häufig nicht zufällig zusammen auftreten, sondern einer komplexen Struktur kollokierender Muster (*Collocational Patterning*) folgen (vgl. Erman 2009). Dies widerspricht vehement der generativistischen Auffassung, dass jede Sprachverwendung einen kreativen Akt darstelle. Pawley / Syder (1983: 193) stellten bereits vor 30 Jahren die These auf, dass Muttersprachler oftmals nicht ihr kreatives Potenzial ausschöpfen, sondern häufig auftretende Sprachroutinen verwenden. Demnach speichern Sprachnutzer tausende komplette Teilsätze und Sätze kognitiv holistisch ab. Diese werden schließlich während des Prozesses der Sprachproduktion wieder aufgerufen. Pawley / Syder (2000: 195) unterstreichen die Bedeutung dieser Sprach-

²² So stellen Goldberg / Jackendoff (2004: 535) fest, dass es keine prinzipielle Trennung zwischen Lexikon und Regeln gibt und schlagen daher vor: "The proposal, then, is to expand the role of the traditional lexicon to include productive or semi-productive phrasal patterns that have previously been assumed to lie within the domain of syntax".

routinen für die Erreichung einer muttersprachlichen Sprachkompetenz: "[i]t is the store of memorized constructions and expressions, more than anything else, that is the key to native-like fluency". Pawley / Syder (1983: 213) postulieren, dass durchschnittliche Muttersprachler des Englischen einige tausend Lexeme kennen, ihnen hingegen mehrere hundert tausend *Sentence-Length Expressions* bekannt sind.

Ebenso wurde während der letzten Jahre schließlich in zahlreichen Studien der Versuch unternommen, exakt zu bestimmen, wie viel Prozent des Sprachgebrauchs aus sprachlichen Routinen bestehen. Die Resultate fallen dabei höchst unterschiedlich aus. Erman / Warren (2000: 55) meinen, dass präfabrizierte Sprache etwa 55% aller Äußerungen von Sprechern und Schreibern des Englischen ausmache. Altenberg (1990: 134, 1998: 102) behauptet, dass 70 bis 80% der Sprache präfabriziert seien. Foster (2001: 85) hingegen schätzt den Anteil auf 32,3%, während Moon (1998) mit 4 bis 5% einen wesentlich geringeren Wert aufführt.

Dass die Zahlen dermaßen weit auseinanderklaffen, muss dem unterschiedlichen Verständnis des *Formulaic Language*-Konzeptes geschuldet sein. Bei genauerer Betrachtung stellt sich dabei als zentrales Problem heraus, dass keine universelle Definition des Konzeptes existiert, die alle Aspekte der Formelhaftigkeit von Sprache berücksichtigen kann. Stattdessen gibt es eine Vielzahl an Definitionen, die sich vor allem bezüglich ihrer Inklusivität mehr oder minder voneinander unterscheiden. Es kann daher zwischen einem weiten und einem engen Verständnis des *Formulaic Language*-Konzepts unterschieden werden. Typischerweise werden je nach Verständnis andere Termini verwendet. Wray (2002: 261) wählt für ihre Zwecke eine eher weite Definition, in der sie sprachliche Routinen als *Formulaic Sequences* bezeichnet und dazu näher ausführt:

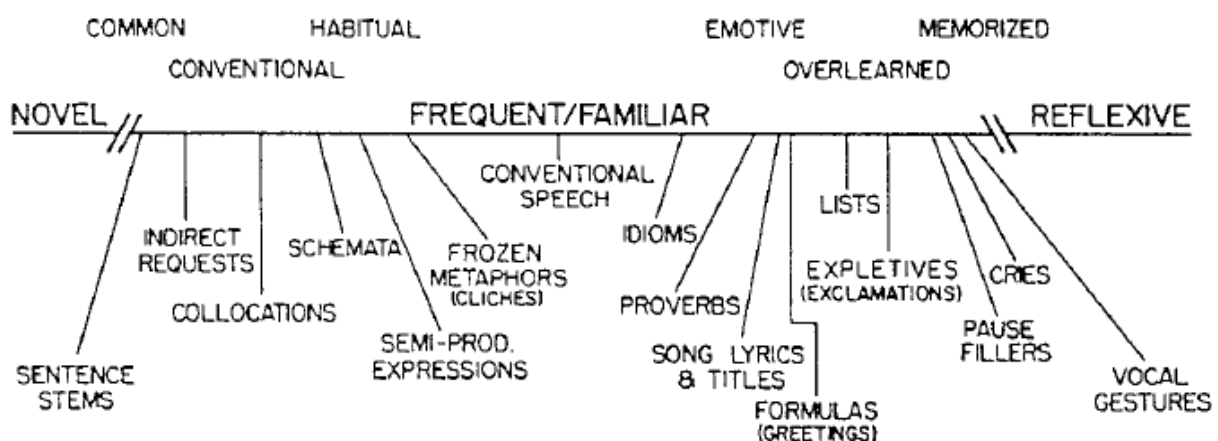
They can break phonological rules, by displaying fewer stresses than expected and by being articulated faster and less clearly. They can break syntactic rules, by resisting pluralization, passivization and so on, and by containing constituents which do not take on their normal grammatical function. They can break lexical rules, by containing items which are archaic or have no independent existence. And they can break semantic rules, by combining to mean something other than they ought to, and by being more idiomatic than an equivalent nonformulaic combination.

Wray (2002) betrachtet den Begriff *Formulaic Sequence* als Oberbegriff für eine ganze Reihe unterschiedlicher Ausprägungen formelhafter Sprache. Sie entwirft dabei ein Modell eines

Gradienten mit zwei oppositionellen Polen, auf dem sich die verschiedenen Formen formelhafter Sprache verorten lassen. Das eine Extrem bilden Idiome, die in ihrer Struktur fest fixiert sind und deren Konstituenten nicht verändert werden können. Das andere Extrem bilden *Phrase Frames*, die sich dadurch auszeichnen, dass alle Konstituenten prinzipiell frei austauschbar sind.²³ Zwischen diesen beiden Extremen lassen sich Mischformen formelhafter Sequenzen verorten, die sich bezüglich der Fixiertheit ihrer Konstituenten voneinander unterscheiden.

Eine Übersicht von Ausprägungen formelhafter Sprache in Form eines Kontinuums gibt auch Van Lancker (1987, siehe Abbildung 7). Auf dem Kontinuum befinden sich die beiden Pole *Novel* und *Reflexive*, welche die Extremformen darstellen. Zwischen ihnen lassen sich die verschiedenen Typen verorten.

PROPERTIES



Subsets of nonpropositional speech and their common properties, presented on a hypothetical continuum from most novel to reflexive.

Abbildung 7: Van Lanckers Kontinuum (Van Lancker 1987: 56).

Wray (2002: 63) betont jedoch, dass dieses Kontinuum problematisch ist, da die Termini sowohl Form (*sentence stems, collocation lists, etc.*) als auch Funktion (*indirect requests, song lyrics and titles, pause fillers, etc.*) und Bedeutung (*metaphors, idioms, proverbs, etc.*) beinhalten. Dies macht es extrem schwierig, die einzelnen Ausprägungen formelhafter Sprache

²³ Während das Idiom *to kick the bucket* in seiner Form fixiert ist und kein Element ausgetauscht werden kann, ohne dass die Bedeutung sich verändert, sind im *Phrase Frame I didn't know that* prinzipiell alle Elemente austauschbar.

miteinander zu vergleichen. Zudem beschränkt sich das Kontinuum lediglich auf den Grad der Fixiertheit der Elemente, andere Faktoren werden darin nicht berücksichtigt. Neben Van Lanckers Kontinuum wurden weitere Versuche unternommen, formelhafte Sprache in Form von Kontinuen darzustellen (u. a. Pawley / Syder 1983, Cowie 1988), jedoch schaffen es auch diese nicht, auf zufriedenstellende Weise alle Ausprägungen formelhafter Sprache vergleichbar zu machen. Sinclair (1987: 324) glaubt nicht, dass sich die Konzepte des *Idiom Principle* und des *Open-Choice Principle* so miteinander vereinbaren lassen, dass sich ein Kontinuum herausbilden kann. Sinclair (1987: 324) meint, die beiden *Principles* seien inkompatibel: "There is no shading of one into another; the switch from one model to the other will be sharp".

Im Gegensatz zu Wray (2002) wird in vielen Studien häufig nicht genau zwischen verschiedenen Ausprägungen von *Formulaic Language* unterschieden. In einigen Studien werden Mehrworteinheiten automatisiert analysiert (vgl. u.a. Biber et al. 1999, Hyland 2008), ohne dass eine Eingrenzung des *Formulaic Language*-Konzepts geleistet wird. Dies führt Biber et al. (2004: 372) zur Feststellung: "[...] despite the general consensus on the importance of multi-word units, there is surprisingly little agreement on their defining characteristics, the methodologies to identify them, or even what to call them". Bisher konnte selbst kein Konsens darüber erzielt werden, unter welchem Terminus sprachliche Routinen zu bezeichnen sind. Aus diesem Grund ist es sinnvoll, Kriterien aufzuzählen, anhand derer Mehrworteinheiten voneinander unterschieden werden können. Diese Kriterien werden nun im folgenden Kapitel genauer beschrieben.

4.3. Kriterien zur Unterscheidung von Mehrworteinheiten

4.3.1. Kompositionalität und holistische Verarbeitung

Mehrworteinheiten können sich in ihrem Grad der Kompositionalität voneinander unterscheiden. Die Grundannahme ist, dass eine formelhafte Sequenz von Wörtern nicht länger grammatikalisch regulär oder semantisch logisch sein muss (Wray 2002: 33). Mehrworteinheiten können fossilisiert werden und grammatikalische Formen beinhalten, die nicht mehr in der gegenwärtigen Sprache existieren. Als Beispiel dient der Ausdruck *If I were you*, in dem die Subjunktivform von *be* enthalten ist (Wray 2002: 33). Grundsätzlich gilt: je weniger kompositionell Mehrworteinheiten sind, desto höher ist die Wahrscheinlichkeit, dass diese im Gehirn als Ganzes abgespeichert werden. Diese "holistische Hypothese" (Jiang / Nekrasova 2007: 434) führt zu den Termini *Prefabricated* oder *Ready-Made Language Use* (präfabrizier-

ter, vorformulierter Sprachgebrauch). Als typisches Beispiel können Idiome genannt werden, da sie holistisch verarbeitet werden und somit als komplette Einheiten gelernt werden müssen, insbesondere wenn sie semantisch opak sind. Auf der phonologischen Ebene zeigt sich die holistische Verarbeitung von Mehrworteinheiten beispielsweise in der Flüssigkeit des Sprachgebrauchs, wie Nattinger / DeCarrico (1992: 32) feststellen:

It is our ability to use lexical phrases [...] that helps us speak with fluency. This prefabricated speech has both the advantage of more efficient retrieval and of permitting speakers (and hearers) to direct their attention to the larger structure of the discourse, rather than keeping it focused narrowly on individual words as they are produced.

4.3.2. Fixiertheit

Während einige Ausprägungen von Mehrworteinheiten in ihrer Form starr fixiert sind, können in anderen einzelne Elemente mehr oder weniger frei ausgetauscht werden. Idiome bilden dabei das eine Extrem, während *Lexical Bundles* das andere Extrem darstellen. Um ihre Bedeutung zu wahren, können Idiome kaum verändert werden. Es darf im Normalfall lediglich das Verb am grammatischen Tempus, Numerus und Aspekt ausgerichtet werden. Einige Studien belegen jedoch, dass selbst starre Mehrworteinheiten in ihrer Form variieren können. Moon (1998: 123) zeigt anhand von Daten aus dem *Oxford Hector Pilot Corpus* auf, dass Alternativen des Idioms *kick the bucket* in Form von *kick the pail* oder *kick the can* existieren. Allerdings ist das Idiom in seiner Variabilität beschränkt. Das Idiom könnte beispielsweise nicht zu *kick the fence* verändert werden, da es sonst die idiomatische Bedeutung (*to die*) verlieren würde. Als oppositioneller Gegenpol sind *Lexical Bundles* davon geprägt, dass ihre Konstituenten nicht fixiert sind, das heißt, dass sich ein oder mehrere Elemente frei austauschen lassen. Dies hat zur Folge, dass *Lexical Bundles* im Gegensatz zu Idiomen keine Bedeutungseinheiten ausmachen und somit nicht durch einzelne Wörter ersetzt werden können.

4.3.3. Häufigkeit

Mehrworteinheiten können sich bezüglich der Häufigkeit ihrer Verwendung voneinander unterscheiden. Während *Lexical Bundles* hochfrequent sind, sind Idiome typischerweise niedrigfrequent. Wray / Perkins (2000: 6) glauben, dass es unzweifelhaft einen Zusammenhang zwischen Formelhaftigkeit und Häufigkeit geben muss, da einige formelhafte Ausdrücke

sehr häufig verwendet werden. Sinclair (1991: 110) führt die Verwendung von formelhafter Sprache zurück auf "the recurrence of similar situations in human affairs [...] a natural tendency to economy of effort [and...] the exigencies of real-time conversation". Langacker 1991: 107) meint, dass sich formelhafter Sprachgebrauch durch einen Prozess der Schematisierung (schematization) herausbildet, "through the reinforcement and progressive entrenchment of recurring commonalities, as well as the 'cancellation' (non-reinforcement) of features that do not recur". Wray (2002: 25) führt dazu weiter aus: "[...] the more often a string is needed, the more likely it is to be stored in prefabricated form to save processing effort, and once it is so stored, the more likely it is to be the preferred choice when that message needs to be expressed". Eine häufige Verwendung einer Mehrworteinheit kann zu einer Konventionalisierung führen, so dass diese holistisch abgespeichert wird. Gleichzeitig ist eine große Häufigkeit allein kein Indikator für ihre holistische Speicherung. Während beispielsweise *Lexical Bundles* große Häufigkeiten aufweisen, ist dies bei Idiomen nicht der Fall. Howarth (1998: 27) schlussfolgert daher: "phraseological significance means something more complex and possibly less tangible than what any computer algorithm can reveal". Wray (2002: 31) betont, dass es noch weitere Faktoren gibt, die einen Ausdruck formelhaft werden lassen. Als Beispiel wählt sie den Ausdruck *keep your hair on*, der nicht formelhaft ist, wenn er im Sinne von *don't remove the wig* verwendet wird, jedoch formelhaft ist, wenn er im Sinne von *calm down* gebraucht wird. Das Beispiel zeigt, dass auch der pragmatische Kontext bei der Bewertung von Mehrworteinheiten miteinbezogen werden muss.

4.3.4. Kollokationsstärke

Biber et al. (1999: 988) definieren Kollokationen als "associations between lexical words, so that the words co-occur more frequently than expected by chance". Im Gegensatz zu Idiomen sind Kollokate eher statistische Assoziationen als relativ fixierte Ausdrücke. Die einzelnen Wörter innerhalb von Kollokationen behalten ihre eigene Bedeutung. Jedes Wort hat eine erweiterte Bedeutung, da es zusammen mit einer bestimmten Gruppe von Kollokaten auftritt. Einige scheinbar synonyme Wörter können sogar anhand ihrer bevorzugten Kollokate voneinander unterschieden werden. Als Beispiel führen Biber et al. (1999: 988) *little* auf, das häufig mit *baby*, *boy* oder *dog* kollokiert, während *small* häufig mit *amount*, *part* oder *world* auftritt.

Die Kollokationsstärke von Mehrworteinheiten kann anhand statistischer Verfahren gemessen werden, beispielsweise mit Hilfe des MI-Scores oder des T-Score Tests. Allerdings

sind diese Methoden in den letzten Jahren zunehmend kritisiert worden (vgl. auch Gries 2010).

Renouf / Sinclair (1991) haben das Konzept der Kollokation um die des *Collocational Frameworks* ergänzt. Sie definieren dieses als "a discontinuous sequence of two words, positioned at one word remove from each other; they are therefore not grammatically self-standing; their well-formedness is dependent on what intervenes" (Renouf / Sinclair 1991: 128). Ihrem Konzept zur Folge treten Lexeme nicht zufällig zusammen auf. Stattdessen existieren grammatikalische Kollokationen, die einen variablen lexikalischen *Slot* haben, der zwischen ihnen steht und der ausgefüllt werden kann. Im Grunde genommen sind *Collocational Frameworks* die konsequente Erweiterung des *Idiom Principle* von Sinclair (1991).

Collocational Frameworks bilden einen Rahmen, in dem Elemente ersetzt werden können. Ein Beispiel ist die Konstruktion *to + Inf. Verb + the + Nomen*, wie zum Beispiel *to drive the car*. Ein *Collocational Frame* kann beschrieben werden als ko-okkurrierende grammatikalische Wörter, die paradigmatische *Slots* beinhalten, die von lexikalischen Einheiten ausgefüllt werden können. Viele Mehrworteinheiten eines Korpus können diesen *Collocational Frameworks* zugerechnet werden. Beispielsweise finden sich im Web-ELF-Korpus zahlreiche Ausprägungen des *Collocational Frameworks a + * + of*, in dem * einen paradigmatischen *Slot* darstellt, der durch eine Auswahl an Wörtern ausgefüllt werden kann, so zum Beispiel *lot* oder *fraction*.

4.4. Funktionale Eigenschaften von Mehrworteinheiten

Neben formalen Parametern können auch funktionale Eigenschaften aufgeführt werden, die für Mehrworteinheiten typisch sind. Wray (1999: 215) schreibt Mehrworteinheiten zwei Hauptfunktionen zu, zum einen den verminderten kognitiven Verarbeitungsaufwand, zum anderen ihre sozio-interaktionelle Funktion. Zusätzlich zu diesen zwei Faktoren zählt sie noch eine Dritte auf, die Diskurs-organisierende Funktion.

4.4.1. Verminderter kognitiver Verarbeitungsaufwand

Mehrworteinheiten reduzieren den kognitiven Verarbeitungsaufwand, wie Wray (1999: 213) feststellt: "[...] formulaic language offers processing benefits to speakers and hearers, by providing a short cut to production and comprehension". Becker (1975) stellte bereits fest, dass es wenig Sinn ergeben würde, jedes Mal aufs Neue Formulierungen zu bilden, die regel-

mäßig verwendet werden. In den letzten Jahren wurden kognitiv-linguistische Studien (u.a. McCrone 1999 und Raichle 1998) durchgeführt, die zeigen, dass das Gehirn, sobald es mit einer linguistischen Aufgabe vertraut ist, die Prozesskette, die verwendet wurde, um diese zu lernen, überbrückt. Bestimmte Wortketten werden memorisiert und mit vereinbarten Bedeutungen assoziiert, die entweder transparent sein können, wie z.B. *I have known _for years in my capacity as _* oder indirekt sein können, wie z.B. *put the kettle on, will you?* (mit der Bedeutung: bitte bereite mir ein heißes Getränk zu) (Wray / Perkins 2002: 16). Somit bilden sich Kollokationen aus, die prinzipiell mehrere Bedeutungen haben können, wie zum Beispiel *bullet point*, für die es jedoch eine konventionelle Bedeutung gibt, die allgemein akzeptiert wird. Die häufige Verwendung derselben Wortsequenzen ist ein Indikator dafür, dass nicht jede sprachliche Äußerung einen kreativen Akt darstellt. Wray (2002: 17) führt dazu weiter aus:

The flexibility afforded by novel construction will be sacrificed both in routine interaction, where it is not needed, and also where processing pressures are abnormally high, such as when a person is trying to concentrate on something else while speaking, like listening to the radio [...].

Der Gebrauch routinierter wiederkehrender Mehrworteinheiten kompensiert die kognitive Speicherlimitierung des menschlichen Gehirns. Mehrworteinheiten erfüllen unterschiedliche Funktionen, um dieses Defizit auszugleichen. Wray / Perkins (2000: 16) haben eine Zusammenfassung aller Funktionen routinierter Mehrworteinheiten zusammengestellt (siehe Abbildung 8).

Table 3
Formulaic sequences as compensatory devices for memory limitations

Function	Effects	Type	Examples
Processing short-cuts	Increased production speed and/or fluency	<ul style="list-style-type: none"> • Standard phrases (with or without gaps) • Standard ideational labels with agreed meanings 	<ul style="list-style-type: none"> • Put the kettle on, will you?; I have known __ for __ years in my capacity as __ • Personal computer; bullet point; the current economic climate
Time-buyers	<ul style="list-style-type: none"> • Vehicles for fluency, rhythm and emphasis • Planning time without losing the turn 	<ul style="list-style-type: none"> • Standard phrases with simple meanings • Fillers • Turn-holders • Discourse shape markers • Repetitions of preceding input 	<ul style="list-style-type: none"> • Make a decision; draw a conclusion; a sea change; at the end of the day (<i>in the sense of 'really'</i>); one way and another • If the truth be told; if you want my opinion; if you like • And another thing; and let me just say.... • There are three points I want to make. Firstly... Secondly... Thirdly/Lastly... • (A: What's the capital of Peru?) B: What's the capital of Peru? (Lima isn't it?)
Manipulation of information	Gaining and retaining access to information otherwise unlikely to be remembered	<ul style="list-style-type: none"> • Mnemonics • Lengthy texts one is required to learn • Rehearsal 	<ul style="list-style-type: none"> • Thirty days hath September. . . ; Richard of York gave battle in vain • Shall I compare thee to a summer's day? • <i>Rehearsing a telephone number while looking for a pen</i>

Abbildung 8: Formelhafte Sequenzen als Kompensation für kognitive Beschränktheit (aus Wray 2000: 16)

Folglich führt die vermehrte Verwendung sprachlicher Routinen zu einer höheren Sprachproduktion und einem höheren Sprachfluss. Ebenso können diese genutzt werden, um Zeit bei der Sprachproduktion zu gewinnen, um einen Sprecherwechsel (*turn-taking*) zu verhindern. Zuletzt werden sprachliche Routinen ebenfalls verwendet, um Informationen zu memorisieren, die auf andere Weise nur sehr schwer abzurufen sind, beispielsweise in Form von memo-technischen Versen (*mnemonics*).

Die routinehafte Verwendung bekannter Wortketten kann dazu führen, dass Wörter, die nicht routinehaft verwendet werden, bewusster wahrgenommen werden. Wray (1999: 215) veranschaulicht dies am Beispiel der Mehrworteinheit *Happy birthday to you*. Wenn es gesungen wird, liegt der Fokus auf dem Namen der Person, die Geburtstag hat.

Die frequente Nutzung wiederkehrender Mehrworteinheiten ist ein alldurchdringendes Merkmal menschlichen Sprachgebrauchs. Insbesondere in spontan gesprochener Kommunikation werden viele sprachliche Routinen verwendet. Sie sind aus pragmatischer Sicht von hoher Bedeutung, da sie dem Sprecher Zeit geben, um die weitere Kommunikation zu planen. Als klassisches Beispiel gelten Sportkommentare (Kuiper 1996), bei denen die Sprecher häufig auf Sprachroutinen zurückgreifen müssen, da sie einerseits sehr schnell auf wechselnde Kontextbedingungen reagieren müssen, andererseits ihren Zuhörern nicht zu viel Information auf einmal zumuten können. So gibt Kuiper (1996: 17f.) einen typischen Kommentar aus einer Live-Berichterstattung zu einem Pferderennen wieder: "They're off and racing now; threading its way through; round the turn they come.". Weitere typische Register, in denen viele sprachliche Routinen verwendet werden, sind die Register Wettervorhersagen (Crystal 1995) und Auktionärssprache (Kuiper 1996).

4.4.2. Sozio-interaktionelle Funktion

Nattinger / DeCarrico (1992) beschreiben eine Vielzahl von Situationen, in denen Mehrworteinheiten verwendet werden, um ein interaktionelles Ziel zu erreichen. Demzufolge werden je nach Kontext bestimmte Mehrworteinheiten bevorzugt gebraucht, zum Beispiel in Begrüßungen (*good morning*), Danksagungen (*thanks a lot*) oder Anfragen (*would you mind*). In manchen Fällen werden sogar vorgeschriebene performative Sprechakte verwendet (*I name this ship*) (Wray 1999: 214). Dies geschieht vor allem in wiederkehrenden konventionalisierten Kontexten.

Sprachliche Routinen werden vor allem genutzt, um eine möglichst effiziente Verständlichkeit auf Seiten des Hörers sicherzustellen. Dennoch handelt der Sprecher nicht aus altruistischen Gründen. Wray (1999: 216) postuliert, dass diese hauptsächlich verwendet werden, um einen Hörer in seiner Wahrnehmung oder in seinem Handeln zu beeinflussen. Die Beeinflussung im Handeln geschieht durch Befehle, Anfragen und Verhandlungen. Die Beeinflussung in der Wahrnehmung wird durch verbale Marker durchgeführt, die die Hierarchiebeziehung von Sprecher und Hörer aufzeigen oder den Sprecher als Gruppenmitglied oder außenstehend markieren. Es ergibt Sinn, dass immer wieder dieselben sprachlichen Routinen verwendet werden, um eine erfolgreiche Interaktion sicherzustellen. Ein allgemein akzeptierter konventioneller Gebrauch sprachlicher Routinen erhöht die Wahrscheinlichkeit, dass der Hörer die Botschaft des Sprechers erfolgreich dekodiert. Als Beispiel führt Wray (1999: 216) eine überfüllte laute Bar auf, in der man jemand Unbekanntes darum bitten will, Platz zu

machen. Die Verwendung der sprachlichen Routinen *excuse me* oder *mind your backs* sollte in dieser Situation üblicherweise erkannt werden als nicht-konfrontative Aufforderung. Würde hingegen der weniger formelhafte Ausdruck *I'm just walking behind you* gebraucht werden, wäre nicht sichergestellt, dass damit die Botschaft effektiv vermittelt wird. Eine Übersicht über die Funktionen routinierter Sprache aus Sicht von Sprechern und Hörern wird in Abbildung 9 gegeben.

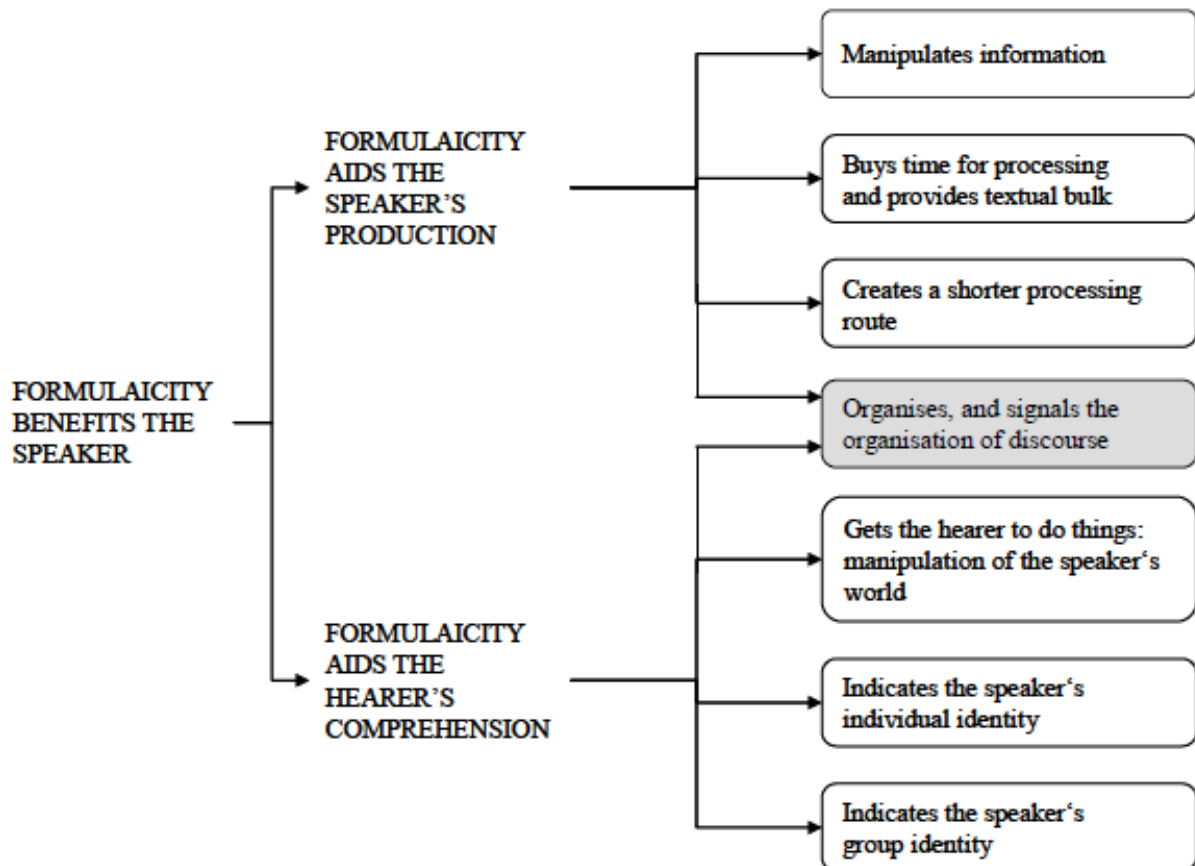


Abbildung 9: Funktionen routinierter Sprache (aus Wray 2002: 97)

Sprachliche Routinen sind vor allem für Sprecher vorteilhaft. Sie helfen ihnen bei der Sprachproduktion. Hörern hingegen erleichtern sie das Dekodieren der Botschaft des Sprechers. Durch den Gebrauch sprachlicher Routinen kann Gruppenzugehörigkeit ausgedrückt werden, so zum Beispiel, wenn bestimmte Ausdrücke ausschließlich im Freundeskreis verwendet werden und diese Routinen sich im Sprachgebrauch der Gruppe verfestigen.

4.4.3. Diskurs-organisierende Funktion

Eine weitere Funktion von Mehrworteinheiten besteht in der Diskurs-Organisation. Einige sprachliche Routinen dienen lediglich dazu, die Bedeutung der folgenden Proposition zu betonen (zum Beispiel *Listen to this now; You'll never believe this, but...*) (vgl. Wray 1999: 216). Andere Mehrworteinheiten, wie zum Beispiel *My first point is* oder *Most importantly*, schaffen eine Verbindung von dem zuvor Gesagten und zu dem, was folgt. Wray / Perkins (2000) argumentieren, dass diese Formen sprachlicher Routinen die beiden Funktionen (Verminderung des kognitiven Verarbeitungsaufwands und sozio-interaktionelle Funktion) in sich vereinen. Dem Sprecher helfen sie, den Text zu strukturieren und die Gedanken zu ordnen. Gleichzeitig beeinflussen sie den Hörer, da durch sie die Struktur des Textes vorgegeben wird und damit die Rezeption des Textes gelenkt wird. Zudem wird ihnen signalisiert, wann sie den Sprecher unterbrechen dürfen, um eine Frage zu stellen oder diesem zu widersprechen.

Zusammenfassend lässt sich feststellen, dass es schwierig ist, formale Charakteristika zu extrapolieren, die es ermöglichen, die Routinehaftigkeit von Mehrworteinheiten festzustellen. Die Fixiertheit der Elemente innerhalb einer Wortkette oder ihre Häufigkeit können Indikatoren dafür sein, dass diese holistisch verarbeitet werden, allgemeingültige Aussagen lassen sich anhand dieser Kriterien jedoch nicht machen. Im Gegensatz dazu können Aussagen zur Formelhaftigkeit von Wortketten anhand der beiden funktionalen Eigenschaften (Verminderung des kognitiven Verarbeitungsaufwands und sozio-interaktionelle Funktion) gemacht werden.

4.5. Eingrenzung des Konzepts der *Formulaic Language*

In den vergangenen Jahren wurde in der Forschung eine Vielzahl unterschiedlicher Termini verwendet, um dem Konzept der Formelhaftigkeit von Sprache Rechnung zu tragen. Die Begriffe *Lexicalised Sentence Stems* (Pawley and Syder 1983), *Lexical Phrases* (Nattinger and DeCarrico 1992), *Lexical Bundles* (Biber 1999), *Chunks* (Myles et al. 1999), *Prefabs* (Erman and Warren 2000), *Formulaic Sequences* (Wray 2002) oder *Morpheme Equivalent Units* (Wray 2008) stellen nur eine Auswahl von Termini dar, die sich allesamt auf das Konzept formelhafter Sprache beziehen. In ihrer nicht-exhaustiven Auflistung führt Wray (2002) nicht weniger als 58 verschiedene Termini auf.²⁴ Eine genauere Betrachtung verdeutlicht, dass die Charakteristiken und Funktionen von *Formulaic Language* in den einzelnen Definitionen so vielschichtig sind, dass aus ihnen keine zufriedenstellende universelle Definition destilliert

²⁴ Für eine vollständige Liste, vgl. Moon (1998).

werden kann. Oftmals wurden bestimmte Termini bewusst eingeführt, um sich mehr oder weniger von bereits bestehenden Termini abzugrenzen, in denen andere Schwerpunkte gesetzt werden. Gleichzeitig wurde der Terminus *Formulaic Language* in den vergangenen Jahren in der Forschung zu häufig verwendet, um als neutral gelten zu können (vgl. Wray 2002). Aus diesem Grund schlägt Wray (2008: 12) vor, den Begriff *Morpheme Equivalent Unit* (MEU) zu nutzen, den sie definiert als: "a word or word string, whether incomplete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any sub-parts it may have". Obwohl diese Definition sehr allgemein formuliert ist, kann auch sie nicht alle Ausprägungen von *Formulaic Sequences* abbilden.²⁵ Sie betont den Aspekt der holistischen Verarbeitung von Mehrworteinheiten, der ein bedeutendes Merkmal für viele Definitionen von *Formulaic Language* ist. Rekurrenz gilt üblicherweise als Indikator dafür, dass bestimmte Mehrworteinheiten holistisch abgespeichert werden können. Es muss dennoch hervorgehoben werden, dass Rekurrenz nicht das einzige Herausstellungsmerkmal für die Verarbeitung von Mehrworteinheiten sein kann. Im Englischen existiert eine Vielzahl häufig auftretender Mehrwortsequenzen, die Elemente aus zwei benachbarten Phrasen beinhalten. In großen Textkorpora können beispielsweise Mehrworteinheiten wie *to be on the* oder *and I have to* sehr häufig auftreten. Insofern muss kritisch hinterfragt werden, ob Sprachnutzer die genannten Beispiele holistisch als Sprachroutinen abspeichern.

Insgesamt lassen sich die Definitionen konzeptionell in zwei Positionen unterteilen. Ein Teil der Definitionen befasst sich mit *Formulaic Language* aus einer kognitivistischen Perspektive. Nur sprachliche Einheiten, die holistisch abgespeichert werden, werden dabei als formelhafte Sprache bewertet. Der andere Teil der Definitionen beschäftigt sich ausschließlich mit der statistischen Signifikanz von *Formulaic Language*. In diesen Definitionen ist es unerheblich, ob und inwiefern *Formulaic Language* holistisch verarbeitet wird. Stattdessen werden alle häufig auftretenden Sequenzen, die häufiger auftreten als statistisch gesehen erwartet werden dürfte, als formelhaft bewertet. Biber (2009) unterscheidet diesbezüglich zwischen den Begriffen *Multi-Word Lexical Collocations* und *Multi-Word Formulaic Sequences*. In seiner Studie untersuchte er die Fixiertheit (*fixedness*) der einzelnen Komponenten von Mehrworteinheiten. Dazu produzierte er eine Liste von 3- (*3-grams*) und 4-Worteinheiten (*4-grams*), das heißt Mehrworteinheiten, die aus drei respektive vier lexikalischen Elementen

²⁵ Wrays Definition nach würde die häufig auftretende Konstruktion *and I have to* keine *Morpheme Equivalent Unit* darstellen, weil sie keine strukturelle Einheit darstellt.

bestehen, für die beiden Genres *Conversation* und *Academic prose*.²⁶ Im Anschluss bestimmte er Frequenzen und MI-Scores²⁷ für ausgewählte Mehrworteinheiten dieser beiden Genres. Die Analyse illustrierte schließlich, dass Mehrworteinheiten mit hohen MI-Scores aus einer Sequenz von lexikalischen Wörtern bestehen. Diese bezeichnete Biber als *Multi-Word Collocations*. Hingegen belegte Biber hochfrequente Sequenzen, die sowohl funktionelle als auch lexikalische Wörter beinhalteten, mit dem Begriff der *Multi-Word Formulaic Sequences*.²⁸ Die Unterscheidung zwischen *Multi-Word Lexical Collocations* und *Multi-Word Formulaic Sequences* erlaubt es, sprachliche Routinen auf verschiedenen Betrachtungsebenen zu untersuchen.

Die Studien von Biber et al. (1999, 2004), Hyland (2008) und Simpson-Vlach / Ellis (2010) zeigen, dass die Verwendung von Sprachroutinen mit Hilfe größerer Korpora sehr gut untersucht werden kann. Konkordanzprogramme erlauben es, automatisiert sprachliche Muster zu identifizieren. Mit Hilfe quantitativer Analysen können anschließend allgemeingültige Aussagen über Sprachgemeinschaften getroffen werden. Analysen dieser Art sind insbesondere nützlich, um die Verwendung sprachlicher Routinen nachzuweisen, da sich diese durch ihre hohe Frequenz auszeichnen. Nur wenn ausreichend viele Belege gefunden werden, können allgemeingültige Aussagen über Sprachgemeinschaften getroffen werden. Aus diesem Grund wurde für die vorliegende Arbeit mit einem Textkorpus (Web-ELF) gearbeitet, das ausreichend groß ist (etwa 2,7 Millionen Wörter), um den Sprachgebrauch in verschiedenen Web-Genres zu beschreiben.

4.6. Kognitive Verarbeitung von *Formulaic Language*

Der Frage nach der kognitiven Verarbeitung von Sprachroutinen wurde in verschiedenen Spracherwerbsstudien nachgegangen. Dabei stellte sich heraus, dass der primäre Nutzen von formelhafter Sprache darin liegt, die Komplexität bei der kognitiven Verarbeitung von Sprache zu reduzieren. Vor allem Studien zum frühkindlichen Spracherwerb stellten dabei die Bedeutung von formelhafter Sprache für den Sprachgebrauch heraus. Peters (1977) betont, dass abhängig von der Spracherwerbsstufe, Kinder eine Sprache eher holistisch (*gestalt*) oder infe-

²⁶ Ein Beispiel für ein *3-gram* ist *the end of*, während *on the other hand* ein Beispiel für ein *4-gram* darstellt.

²⁷ Der MI-Score stellt einen statistischen Wert dar, der die Kollokationsstärke von zwei kokkurrierenden Wörtern feststellt.

²⁸ In früheren Studien (Biber et al. 1999: 990) nutzte Biber den Terminus *Lexical Bundles*. Die beiden Begriffe *Lexical Bundles* und *Multi-Word Formulaic Sequences* verweisen im Grunde genommen auf dasselbe Konzept.

rentiell (analytisch) lernen. Eine holistische Produktion von Sprache bedeutet, dass Äußerungen reproduziert werden, ohne dass diese analysiert werden und ohne dass die zugrundeliegenden grammatikalischen Regeln verstanden werden. Stattdessen werden holistisch memorisierte Äußerungen reproduziert, ohne erkennbare distinktive Pausen zwischen Wörtern und mit einer typischen phonetischen Ungenauigkeit, die sich in der Reduktion der phonemischen Artikulation äußert. Eine inferentielle (oder analytische) Produktion von Sprache bedeutet, dass der Sprachverwender dazu befähigt ist, die eigene Sprachproduktion zu analysieren und eigenständig neue Äußerungen zu formulieren. Hakuta (1974) und Wong-Fillmore (1976) haben den Zweitspracherwerb (L2) japanischer und spanischer Kinder untersucht und dabei herausgestellt, dass das Kennen und Nutzen von sprachlichen Routinen bedeutsam für die Sprachentwicklung sind. Peters (1983) untersuchte den Spracherwerb von Kleinkindern. Ihre Studie zeigt, dass Kinder eine große Anzahl an präfabrizierten, routinierten Sequenzen nutzen, die sie nicht in ihre einzelnen Komponenten segmentieren.²⁹ Erst in einer späteren Phase des Spracherwerbs beginnen Kinder, syntaktische Einheiten zu segmentieren und damit ein Sprachsystem zu verstehen und analytisch zu nutzen.

Hoey (2005) erklärt die Speicherung von *Formulaic Language* anhand des Konzepts des *Lexical Priming*. Beim lexikalischen Spracherwerb begegnet ein Sprachlerner Wörtern in bestimmten Kontexten. So lernen Sprachlerner unterschwellig, die erlernten Wörter ebenfalls in diesen Kontexten zu verwenden. Folglich hat, so Hoey (2005), jeder Mensch eine mentale Konkordanz zu allen Wörtern, die ihm begegnet sind. Hoey (2005: 11) führt weiter aus: "This mental concordance is accessible and can be processed in much the same way that a computer concordance is, so that all kinds of patterns, including collocational patterns, are available for use". Divjak / Gries (2008) liefern mit ihrer Studie quantitative Belege für die Existenz des *Lexical Priming*-Konzepts.

In weiterführenden Studien (u.a. Granger 1998, De Cock 2004) zum Zweitspracherwerb wurde untersucht, inwieweit sich L2-Sprecher von Muttersprachlern im Hinblick auf die Verwendung sprachlicher Routinen unterscheiden. Die Ergebnisse zeigten, dass formelhafte Sprache nichtmuttersprachlichen Lernern des Englischen grundsätzlich große Probleme bereitet. Sie vertrauen in frühen Phasen des Spracherwerbs häufig auf vorformulierte Sprache, zeigen allerdings Schwierigkeiten, sie idiomatisch anzuwenden. Durrant / Schmitt (2009) zeigen beispielsweise, dass fortgeschrittene Englischlerner bestimmte sprachliche Routinen zu

²⁹ Demnach verwenden Kinder die Phrase *what is that* als eine einzelne, unsegmentierte lexikalische Einheit.

häufig verwenden, während andere zu selten verwendet werden, im Vergleich zu Muttersprachlern.

Nicht nur auf der Ebene des kindlichen Spracherwerbs, sondern auch in Studien mit Probanden, die an aphasischen Störungen litten, finden sich Anzeichen für die holistische Speicherung von sprachlichen Routinen. Eine Vielzahl von Studien wies nach, dass Menschen mit eingeschränkten Gehirnfunktionen sehr viele sprachliche Routinen nutzen um zu kommunizieren.³⁰

In einigen Studien wurden *Eyetracker* verwendet, um die Blickbewegungen von muttersprachlichen und nichtmuttersprachlichen Probanden aufgezeichnet. Ihr Ziel war es, die holistische Verarbeitung sprachlicher Routinen nachzuweisen. Conklin / Schmitt (2008) stellen fest, dass sprachliche Routinen einen Verarbeitungsvorteil beim Lesen bieten. Jiang / Nekrasova (2007) fanden heraus, dass sprachliche Routinen schneller und weniger fehlerhaft verarbeitet werden als nicht-formelhafte Routinen, sowohl bei Muttersprachlern als auch bei Nichtmuttersprachlern des Englischen. Künstler et al. (in Vorb.) zeigen, dass Muttersprachler sprachliche Routinen wesentlich schneller verarbeiteten als Nichtmuttersprachler. Insgesamt belegen die Studien, dass sprachliche Routinen bei der Rezeption und Produktion von Sprache eine zentrale Rolle einnehmen.

Siyanova-Chanturia et al. (2011a) untersuchte die kognitive Verarbeitung von Idiomen in ihrer übertragenen und in ihrer wörtlichen Bedeutung. Ihren Resultaten zu Folge verarbeiten Muttersprachler beide Formen kognitiv schneller als neugebildete Phrasen, während Nichtmuttersprachler Idiome und kompositionelle Phrasen gleich schnell verarbeiteten. Zudem verarbeiteten sie Idiome mit wörtlicher Bedeutung schneller als Idiome mit übertragener Bedeutung. In einer weiteren Studie untersuchten Siyanova-Chanturia et al. (2011b) die Verarbeitung binomialer Phrasen bei Muttersprachlern und Nichtmuttersprachlern. Binomiale Phrasen bestehen aus zwei Inhaltswörtern aus derselben lexikalischen Klasse, wie beispielsweise *black* und *white* oder *day* und *night* und sind in ihrer Stellung zumeist austauschbar. Beiden Testgruppen wurden hochfrequente und niedrigfrequente Phrasen präsentiert. Die Ergebnisse zeigten, dass sowohl Muttersprachler als auch Nichtmuttersprachler hochfrequente binomiale Phrasen schneller verarbeiten als niedrigfrequente Formen. Muttersprachler und fortgeschrit-

³⁰ Einen Überblick über *Formulaic Language* und Aphasie geben unter anderem Code (1984) und Van Lancker (1988). Wray (2002) gibt einen sehr detaillierten Überblick über verschiedene Studien zum Erst- und Zweiterwerb von Sprache und den Einfluss formelhafter Sprache auf die Sprachentwicklung.

tene Lerner konnten zudem binomiale Phrasen in der üblichen Form schneller verarbeiten als die in umgekehrter Reihenfolge.

In der Studie von Ellis et al. (2008) wurden Muttersprachler und L2-Sprecher bezüglich der kognitiven Verarbeitung von akademischen *Lexical Bundles* untersucht, die anhand verschiedener Kriterien (Häufigkeit, Länge, MI-Score) zusammengestellt wurden. Die Ergebnisse zeigen, dass sich für Muttersprachler Verarbeitungsvorteile ergeben, wenn der MI-Score hoch ist, während für Nichtmuttersprachler eher die Häufigkeit entscheidend ist für eine schnellere Verarbeitung.

4.7. Eingrenzung des Mehrworteinheiten-Konzepts für die vorliegende Arbeit

Im Gegensatz zu Wray (2008), die Mehrworteinheiten mit ihrem Konzept der *MEU* auf kognitivistische Weise betrachtet, liegt der Fokus der vorliegenden Arbeit auf einer quantitativ statistischen Betrachtung von Mehrworteinheiten. Hierfür werden im Sinne von Biber (2009) *Multi-Word Formulaic Sequences* in einem Multimillionenwortkorpus bezüglich ihrer formalen und funktionalen Verwendung untersucht. Frequenz diente hierbei als alleiniges Kriterium, um *4-grams* automatisiert aus dem Korpus zu extrapolieren. Somit wurden auch viele Einheiten extrahiert, die nicht Wrays (2008) Mehrworteinheiten-Definition entsprechen. Faktisch stellt die Mehrheit der extrahierten *4-grams* keine semantisch kohärenten Einheiten dar.³¹ Bei vielen dieser Einheiten muss somit angezweifelt werden, ob sie holistisch verarbeitet werden. Diese Tatsache ist jedoch im Hinblick auf das Erkenntnisinteresse dieser Arbeit nicht weiter von Bedeutung.

4.8. Entwicklungen in der *n-gram*-Forschung

Seit Sinclair (1991) ist eine Vielzahl quantitativer Studien zu Mehrworteinheiten durchgeführt worden. Biber et al.s (1999) quantitative Studie zur Verwendung von *n-grams* in verschiedenen Registern war jedoch ein wichtiger Bezugspunkt vieler nachfolgender Studien. Ein *n-gram* ist eine zusammenhängende Sequenz einer beliebigen Anzahl von Wörtern (*n*) aus einem Text. Es steht an der Schnittstelle zwischen Lexik und Syntax.

N-gram-Analysen sind vor allem deshalb populär, da man mit Hilfe von Programmen relativ einfach Listen erstellen kann, die alle *n-grams* eines Korpus nach ihrer Häufigkeit sortiert

³¹ Die automatische Extrapolation von *4-grams* für die vorliegenden Studien produzierte eine Liste, in der unter anderem Elemente wie *on the musculoskeletal system* oder *kodiyeri balakrishnan minister of* auftauchten.

aufführen. In den vergangenen Jahren wurde eine Fülle von *n-gram*-Studien durchgeführt. Insbesondere in der kognitiven Forschung und der Fremdsprachenforschung sind dabei einige neue Erkenntnisse erlangt worden.

Underwood et al. (2004) illustrieren, dass formelhafte Mehrworteinheiten, die bis zu sechs Wörter lang sind, schneller gelesen werden als Sequenzen, die nicht formelhaft sind. Götz / Schilk (2011) vergleichen die Nutzung von *3-grams* durch *ENL*-, *ESL*- und *EFL*-Sprecher und weisen nach, dass *EFL*-Sprecher in gesprochener Sprache im Vergleich signifikant weniger *3-grams* verwenden. Zudem zeigen sie, dass *ESL*- und *EFL*-Sprecher einen kleineren Kern von *3-gram Types* wesentlich häufiger verwenden.

Neben kognitiven Studien wurden in jüngerer Zeit vermehrt Studien zu *n-grams* und Registern durchgeführt. Als Ausgangspunkt dienten die Arbeiten von Biber (1988) und Conrad / Biber (2001), in denen Multifaktoren-Analysen beschrieben wurden, um gesprochene und geschriebene Genres miteinander vergleichen zu können. Crossley / Louwse (2007) haben darauf basierend eine multidimensionale Register-Klassifikation anhand von *bigrams* durchgeführt und Unterschiede in der Verwendung von *bigrams* zwischen den verschiedenen Registern herausgestellt. Peng et al. (2003) haben anhand von *n-gram*-Sprachmodellanalysen in verschiedensprachigen Korpora (Griechisch, Japanisch, Chinesisch, Englisch) Texte klassifiziert in den Kategorien Sprachen, Autorenschaft, Genres und Themen. Anhand von *bigrams* konnten sie die einzelnen Texte mit einer hohen Genauigkeit klassifizieren. Statamatos et al. (2001) haben anhand von häufigen Unigrammen aus dem *British National Corpus* herausgefunden, dass diese Genres zuverlässiger markieren als lexikalische Einheiten aus ihrem Testkorpus, dem *Wall Street Journal*. Stefanowitsch / Gries (2003: 214) führen den Begriff der *collostructions* ein, die sie wie folgt beschreiben:

Lexemes that are attracted to a particular construction³² are referred to as *collexemes* of this construction; conversely, a construction associated with a particular lexeme may be referred to as a *collostruct*; the combination of a collexeme and a collostruct will be referred to as a *collostruction*.

(Stefanowitsch / Gries 2003: 215)

³² Eine Konstruktion definieren Stefanowitsch / Gries 2003: 212) als "any linguistic expression, no matter how concrete or abstract, that is directly associated with a particular meaning or function, and whose form or meaning cannot be compositionally derived."

Demnach ziehen sich Lexeme und grammatische Konstruktionen an. Eine *collostruction*-Analyse zeige demnach [...] "which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected) [...]". Sie schlussfolgern, dass die Abhängigkeiten zwischen Wörtern und grammatikalischen Strukturen Theorien stützen, die grammatikalische Strukturen als linguistische Zeichen verstehen:

If syntactic structures served as meaningless templates waiting for the insertion of lexical material, no significant associations between these templates and specific verbs would be expected in the first place (Stefanowitsch / Gries 2003: 237).

Gries (2010) erweitert das *Gravity Counts*-Konzept von Daudaravičius / Marcinkevičienė (2004) um das Konzept der *Lexical Gravity G*, das nicht die Anzahl der *Tokens* berücksichtigt (wie andere statistische Verfahren, wie beispielsweise der MI-Score), sondern die Anzahl der *Types*. Gries Ergebnisse belegen, dass sich mit diesem Konzept nicht nur gesprochene und geschriebene Sprache sehr zuverlässig unterscheiden lassen, sondern weist auch bessere Analyse von *bigrams* aus als der häufig verwendete T-Score-Test.

Mukherjee / Gries (2010) haben das *Lexical Gravity G*-Konzept übernommen, um zu überprüfen, inwiefern sich *Asian Englishes* bezüglich der Verwendung von *n-grams* unterscheiden. Gleichzeitig zeigen sie, dass die Berechnung von Kollokationsstärke anhand von MI-Scores problematisch ist und warnen davor, diesen unkritisch zu verwenden. Sie erweitern *Lexical Gravity* zudem um das Konzept der *Lexical Stickiness*. Gries et al. (2011) haben herausgefunden, dass mit Hilfe von identifizierten *n-grams* aussagekräftige Erklärungen über Register gemacht werden können. Biber (2006) hat gezeigt, dass in verschiedenen Genres eine unterschiedliche Anzahl von *Lexical Bundles* verwendet wird. In *Classroom teaching* werden im Vergleich zu *Conversation* und *Textbooks* doppelt so viele respektive viermal so viele *Lexical Bundles* verwendet. Hyland (2008: 59) vergleicht die Nutzung von *4-grams* in verschiedenen akademischen Registern und zeigt auf, dass innerhalb des Registers *Academic writing* Subregister existieren, in denen zum Teil andere *4-grams* bevorzugt werden. Es existieren viele weitere Studien, die sich mit der Identifikation von *n-grams* in akademischen Registern beschäftigt haben (vgl. u.a. Biber et al. 2004, Simpson 2004, Nesi / Basturkmen 2006, Simpson-Vlach / Ellis 2010).

4.9. Fokussierung auf 4-grams

Für die vorliegenden Studien wird der Fokus bewusst auf *4-grams* gelegt. Die Entscheidung, sich ausschließlich auf *4-grams* zu beschränken, basiert auf mehreren Gründen. Ein wichtiger Grund ist, dass in der vorliegenden Arbeit sowohl formale als auch funktionale Studien durchgeführt wurden. Insbesondere für die funktionalen Studien mussten ausreichend große Mehrworteinheiten untersucht werden, da grundsätzlich das Problem besteht, dass nicht allen automatisch identifizierten Mehrworteinheiten eine Funktion zugeschrieben werden kann. Den meisten *2-grams* und *3-grams* kann beispielsweise keine Funktion zugewiesen werden. Ab einer Größe von vier Elementen hingegen gibt es wesentlich mehr Mehrworteinheiten, denen man eine Funktion zugewiesen kann.

Ein zweiter wichtiger Grund für die Fokussierung auf *4-grams* liegt darin, dass die Ergebnisse der vorliegenden Studie mit denen anderer Studien verglichen werden sollen. Die Referenzstudien (Biber et al. 1999, 2004; Simpson-Vlach / Ellis 2010) haben sich allesamt auf *4-grams* konzentriert. Um eine Vergleichbarkeit der Ergebnisse zu ermöglichen, mussten daher ebenfalls *4-grams* untersucht werden.

Die Limitierung auf *4-grams* war auch für die automatisierte formale Analyse von Vorteil. Bei Mehrworteinheiten, die weniger als vier Elemente beinhalten, hat der *Automatic n-gram Analyser* Probleme, die einzelnen Elemente korrekt zu annotieren. Um ein fehlerfreies *Tagging* zu ermöglichen, musste die Software bereits für die *4-gram*-Analyse umfassend optimiert werden, ein fehlerfreies *Tagging* von *2-grams* oder *3-grams* war mit der aktuellen Version der Software nicht möglich. Diese hätte dazu noch umfangreicher optimiert werden müssen.

Des Weiteren habe ich bereits mit Hilfe einer Pilotstudie erste Analysen von *4-grams* durchgeführt, im Rahmen derer ich ihre formale Struktur untersucht habe und die Resultate mit den Ergebnissen von Biber et al. (1999) verglichen habe. Dafür habe ich ein Pilotkorpus erstellt, das ausschließlich Texte enthalten sollte, die im Kontext des Themas *Medical Surgery / Medical Tourism* entstanden sind. Die Ergebnisse der Studie belegten, dass es auf formaler Ebene Unterschiede zwischen den verschiedenen Web-Genres gibt. Foren entsprachen in der Verwendung von *4-grams* mündlichem Sprachgebrauch (*Conversation*), während Wikipedia-Artikel in ihrer Form akademischen Texten (*Academic prose*) glichen. Für Blogs und kommerzielle Homepages konnten keine klaren Aussagen getroffen werden. Sie wiesen jedoch tendenziell mehr Elemente des Registers *Academic prose* auf. Des Weiteren zeigte die Pilotstudie, dass der thematische Fokus auf *Medical Surgery / Medical Tourism* sicherstellte,

dass sowohl Nichtmuttersprachler als auch Muttersprachler des Englischen an der Textproduktion beteiligt waren. Die globale Relevanz des Themas schien einen entscheidenden Einfluss auf die Zusammensetzung der Webnutzer zu haben. Die Ergebnisse bestätigten, dass die Beschränkung auf das Thema *Medical Surgery / Medical Tourism* es ermöglicht, unterschiedliche Web-Genres miteinander zu vergleichen. Folglich wurde diese thematische Beschränkung auch für die Kompilation des Web-ELF-Korpus beibehalten.

Obwohl sich zahlreiche Studien (Biber et al. 1999, 2004; Hyland 2008) bereits mit der Analyse von *4-grams* in unterschiedlichen Registern befasst haben, gibt es meines Wissens keine Studie, die eine Untersuchung des Sprachgebrauchs im Web durchgeführt hat. Im Fokus der Studien für die vorliegende Arbeit steht der Vergleich verschiedener Web-Genres anhand mehrerer quantitativer und qualitativer *4-gram*-Analysen.

4.10. Zusammenfassung und Hypothesen der Arbeit

Die Sprachforschung hat verdeutlicht, dass Mehrworteinheiten eine wichtige Konstituente des menschlichen Sprachgebrauchs sind. Die Verwendung von Mehrworteinheiten bewirkt eine kognitive Komplexitätsreduktion (vgl. Wray 2008). Aus diesem Grund werden sowohl im mündlichen als auch im schriftlichen Sprachgebrauch viele formelhafte Einheiten verwendet.

Das Konzept von Mehrworteinheiten kann jedoch höchst unterschiedlich verstanden werden und muss aus diesem Grund genauestens definiert werden. Mehrworteinheiten können einerseits aus kognitivistischer Perspektive betrachtet werden. Demnach sind sie präfabrizierte Einheiten, die von Sprachnutzern holistisch abgespeichert werden. Mehrworteinheiten können andererseits auf rein statistischer Betrachtungsebene bewertet werden. Demnach werden alle häufig auftretenden Sequenzen, die öfter auftreten, als statistisch gesehen erwartet werden dürfte, als formelhaft bewertet. Für die vorliegende Studie untersuche ich Mehrworteinheiten im Sinne von Bibers Konzept von *Lexical Bundles* beziehungsweise *Multi-Word Formulaic Sequences*. Die kognitive Verarbeitung von Mehrworteinheiten steht ausdrücklich nicht im Fokus dieser Arbeit.

Im theoretischen Teil der vorliegenden Arbeit habe ich den Sprachgebrauch im Web und die aktuelle Lingua Franca-Debatte verknüpft. Im Web kommunizieren Sprachnutzer aus unterschiedlichen Teilen der Welt miteinander. Dabei gibt es keinen einheitlichen Standard (u.a. *American English, British English*), dem alle Nutzer gemeinsam folgen. Das Englische im Web stellt somit eine neue Form des Englischen dar, die erst näher beschrieben werden muss. Dabei zeigt sich, dass das Medium einen entscheidenden Einfluss auf den Sprachgebrauch im

Web hat, der sich jedoch je nach Web-Genre stark voneinander unterscheiden kann. Das Nähe-Distanz-Modell von Koch / Oesterreicher (1985, 1994) ermöglicht es prinzipiell, die einzelnen Web-Genres anhand einer Reihe von Merkmalen zu vergleichen und voneinander abzugrenzen. Es zeigt sich dabei, dass sich die einzelnen Web-Genres bezüglich ihrer konzeptionellen Mündlichkeit / Schriftlichkeit graduell unterscheiden. Foren weisen viele Merkmale konzeptioneller Mündlichkeit auf, Blogs, kommerzielle Homepages und Wikipedia-Artikel weisen im graduell zunehmenden Maße viele Merkmale konzeptioneller Schriftlichkeit auf.

Das Ziel der Arbeit ist es, Web-Genres anhand ihrer formalen und funktionalen Eigenschaften voneinander abzugrenzen. Als Ausgangsbasis dafür dienen Biber et al. (1999) und Biber et al. (2004), die gezeigt haben, dass unterschiedliche Muster von Mehrworteinheiten in traditionell mündlichen und schriftlichen Registern gefunden werden können. Die Ergebnisse dieser Analyse sollen dabei auch im Hinblick auf den Einfluss des Mediums Internet auf die Entwicklung des Lingua Franca-Sprachgebrauchs diskutiert werden. Im Folgenden werden nun die Forschungshypothesen, die der vorliegenden Arbeit zu Grunde liegen, formuliert:

1. Web-Genres unterscheiden sich in Bezug auf ihre formalen und funktionalen Verwendungen von Mehrworteinheiten entscheidend voneinander. Während Foren und Blogs mehr Elemente konzeptioneller Mündlichkeit enthalten, enthalten kommerzielle Homepages und Wikipedia-Artikel mehr Elemente konzeptioneller Schriftlichkeit.
2. Der Sprachgebrauch im Web unterscheidet sich entscheidend von traditionellem konzeptionell mündlichem und konzeptionell schriftlichem Sprachgebrauch. Obwohl einige Web-Genres auf der formalen und funktionalen Ebene typisch mündlichen und schriftlichen Registern ähneln, können sie dennoch nicht mit ihnen gleichgesetzt werden, da die vier Web-Genres Hybride aus mündlichem und schriftlichem Sprachgebrauch darstellen.
3. Obwohl sich die Web-Genres strukturell stark voneinander unterscheiden, erlaubt es der thematische Fokus auf *Medical Surgery / Medical Tourism*, die einzelnen Web-Genres miteinander zu vergleichen.

In Kapitel 8 sollen diese Hypothesen getestet und schließlich die Forschungsfragen beantwortet werden. Doch zunächst wird im folgenden Kapitel die Methodik präsentiert, die den quantitativen und qualitativen Studien zu Grunde liegt. Dazu werden das Web-ELF-Korpus und die linguistischen *Tools*, die für die beiden Analysen verwendet wurden, im Detail vorgestellt. Im Anschluss werde ich näher auf die Konzeptionen der einzelnen Studien eingehen.

5. Methodik der Arbeit

In der vorliegenden Arbeit werden mehrere korpusgestützte Studien durchgeführt, die sich mit der formalen und funktionalen Untersuchung phraseologischer Einheiten in verschiedenen Web-Genres befassen. Aus diesem Grund wird einerseits ein Korpus benötigt, das eine solche Untersuchung ermöglicht, andererseits passende Referenzkorpora, mit denen das Web-Korpus verglichen werden kann. Obgleich eine Vielzahl von Korpora existiert, die alle genuin aus Webtexten bestehen (beispielsweise das *Financial Times Corpus*³³ oder das *SAVE Korpus*³⁴), sind diese jedoch entweder nicht frei zugänglich oder erweisen sich als ungeeignet für eine vergleichende phraseologische Untersuchung verschiedener Web-Genres, da sie sich zumeist auf ein Web-Genre beschränken. Aus diesem Grund wurde eigens ein Korpus verwendet, mit dessen Hilfe die eingangs aufgeworfenen Fragestellungen untersucht werden konnten.

5.1. Korpora

5.1.1. Das Web-ELF-Korpus

5.1.1.1. Hintergründe zur Erstellung des Web-ELF-Korpus

Das Web-ELF-Korpus beinhaltet Texte, die aus verschiedenen Web-Genres aus dem Web extrahiert wurden. Es besteht aus etwa 2,7 Millionen Wörtern aus vier verschiedenen Web-Genres (Foren, Blogs, kommerzielle Homepages und Wikipedia-Artikel). Die Texte sind alle samt im Kontext des Themas *Medical Surgery / Medical Tourism* in einem Lingua Franca-Kontext entstanden (vgl. Künstler et al. 2012). Das Korpus beinhaltet sowohl kommerzielle Webseiten als auch Webseiten privater Betreiber. Im Gegensatz zu anderen bekannten *ELF*-Korpora (z.B. dem *Vienna-Oxford International Corpus of English (VOICE)*, vgl. Seidlhofer 2010), liegt der Fokus des Web-ELF-Korpus nicht auf dem nichtmuttersprachlichen lerner-sprachlichen Gebrauch des Englischen als Fremdsprache. Stattdessen wird das Web allgemein als ein virtueller Kommunikationsort verstanden, in dem die Herkunft der Sprecher und ihr kultureller Hintergrund keine entscheidende Rolle spielen. Im Web dient das Englische einer Vielzahl von muttersprachlichen und nichtmuttersprachlichen Sprachverwendern gleichermaßen als Lingua Franca. Somit wird im Web-ELF-Korpus sowohl muttersprachlicher als auch nichtmuttersprachlicher Sprachgebrauch abgebildet.

³³ Vgl. Forchini / Murphy (2008)

³⁴ *South Asian Varieties of English (SAVE) Korpus* (vgl. Bernaisch et al. 2011)

Die Grundvoraussetzung für die Erstellung eines repräsentativen Korpus ist, dass die Texte sehr sorgfältig selektiert werden. Biber / Kurjian (2007: 119) haben gezeigt, dass bei der Erstellung von Web-Korpora klare Kriterien festgelegt werden müssen, nach denen die Texte für das Korpus ausgewählt werden. Die Ergebnisse einer Untersuchung sind nur dann aussagekräftig, wenn bekannt ist, welche Ausprägungsform von Sprache die ausgewählten Texte repräsentieren. Die Intention bei der Erstellung des Web-ELF-Korpus war es, vier Web-Genres abzubilden und diese miteinander zu vergleichen. Vor diesem Hintergrund wurden Kriterien formuliert, die die Texte erfüllen mussten, um ins Web-ELF-Korpus aufgenommen zu werden. Die thematischen und strukturellen Einschränkungen sollten sicherstellen, dass die Subkorpora miteinander verglichen werden konnten.

Das Korpus zielt darauf ab, den globalen englischen Sprachgebrauch in verschiedenen Web-Genres abzubilden. Da das Web jedoch prinzipiell ein anonymes Medium darstellt, innerhalb dessen die Identität der Nutzer üblicherweise nicht eindeutig festgestellt werden kann, erweist es sich als grundsätzlich schwierig, passende Texte zu finden, die diese Bedingung erfüllen. Aus diesem Grund musste ein Kriterium gewählt werden, das sicherstellte, dass das Korpus sowohl muttersprachlichen als auch nichtmuttersprachlichen Sprachgebrauch abbildet. Die Fokussierung auf ein global diskutiertes Thema erlaubte es sicherzustellen, dass im Web-ELF-Korpus sowohl der Sprachgebrauch von Muttersprachlern als auch von Nichtmuttersprachlern abgebildet werden konnte. Die Pilotstudie zeigte, dass sich das Thema *Medical Surgery / Medical Tourism* hervorragend dafür eignete.

5.1.1.2. Zusammensetzung des Web-ELF-Korpus

Das Web-ELF-Korpus besteht aus vier Subkorpora, die ausschließlich Texte enthalten, die sowohl von muttersprachlichen als auch von nichtmuttersprachlichen Webnutzern im Kontext des Themas *Medical Surgery / Medical Tourism* entstanden sind. Neben der generellen thematischen Beschränkung wurden bei der Suche nach passenden Texten für jedes der Subkorpora weitere Kriterien definiert, die erfüllt sein mussten, um eine Aufnahme ins Korpus zu ermöglichen.

Für das Subkorpus Foren wurden ausschließlich Forenseiten ausgewählt, auf denen sich sowohl Muttersprachler als auch Nichtmuttersprachler miteinander austauschten. In den Foren diskutierten sowohl Laien als auch Experten über chirurgische Eingriffe und Operationsmöglichkeiten in verschiedenen Ländern der Welt. Für das Korpus konnten jedoch nicht alle *Threads* einer Forenseite verwendet werden. So wurden für das Korpus ausschließlich

Threads extrahiert, die zumindest fünf Beiträge vorweisen konnten. Zudem wurde sichergestellt, dass sich auf den Forenseiten eine Vielzahl unterschiedlicher Nutzer am Austausch beteiligte, so dass für das Korpus eine überproportionale Abbildung idiosynkratischen Sprachgebrauchs ausgeschlossen werden konnte. In den einzelnen *Threads* wurde zumeist über medizinische Eingriffe diskutiert. Einerseits haben sich sowohl Laien als auch Experten über die Chancen und Risiken von Operationen ausgetauscht, andererseits haben sie über die Möglichkeit chirurgischer Eingriffe im Ausland diskutiert. Obwohl Nutzer in Foren üblicherweise dazu neigen, thematisch abzudriften (und durch Moderatoren häufig darauf hingewiesen werden müssen, zum eigentlichen Thema zurückzukehren), zeigte sich für die untersuchten Foren, dass die Nutzer auch ohne das Eingreifen von Moderatoren das Thema beibehielten. Dies ist vermutlich der Ernsthaftigkeit und Relevanz der Themen für die Webnutzer geschuldet. Insgesamt wurden die Texte für dieses Subkorpus aus einer Vielzahl von *Threads* aus 18 Forenseiten extrahiert. Direkte Zitationen vorhergehender Posts wurden herausgefiltert, so dass es diesbezüglich keine Dopplungen gab.

Das Subkorpus Blogs besteht ausschließlich aus privaten Webseiten von Bloggern aus verschiedenen Ländern, die sich zum Thema *Medical Surgery / Medical Tourism* äußerten. Die Blogs wurden allesamt von einzelnen Autoren erstellt; in der Mehrzahl der Fälle gaben diese jedoch nur sehr wenig oder gar keine soziobiografischen Informationen preis. Durch die inhaltlichen Bezüge der Beiträge auf den jeweiligen Blogseiten konnte jedoch auf die Herkunft der Blogger geschlossen werden. Diese stammten aus verschiedenen Regionen der Welt. Als ein wichtiges Kriterium zur Aufnahme potenzieller Blogs in das Korpus wurde festgelegt, dass der Fokus auf das ursprüngliche Thema beibehalten werden musste. Die Blogs zielten darauf ab, medizinische Eingriffe zu erklären, Behandlungsmethoden zu diskutieren und über finanzielle Unterschiede bei der Durchführung medizinischer Eingriffe im Ausland aufzuklären. Insgesamt stellte es eine Herausforderung dar, im Web Blogs ausfindig zu machen, die sich diesem speziellen Thema widmeten. Die Schwierigkeit, genügend Texte zu finden, die alle geforderten Kriterien erfüllten, erklärt die vergleichsweise geringe Anzahl an Texten, die in diesem Subkorpus enthalten sind. Insgesamt wurden Texte aus 16 Blogseiten für das Subkorpus extrahiert.

Um sicherzustellen, dass das Subkorpus kommerzielle Homepages für eine globale Leserschaft produziert wurde, wurden ausschließlich englischsprachige Webseiten in das Korpus aufgenommen, die für chirurgische Operationen in Ländern warben, in denen das Englische nicht als Erstsprache fungiert. Alle für das Korpus extrahierten kommerziellen Homepages

zielten darauf ab, potenzielle Kunden über chirurgische Eingriffe zu informieren und sie zu beraten, um sie von einem chirurgischen Eingriff in einer Spezialklinik zu überzeugen. Die Autorenschaft ist im Fall kommerzieller Homepages unklar, da auf den Webseiten darüber keine genauen Angaben gemacht werden. Im Gegensatz zum Subkorpus Blogs ließ sich für dieses Web-Genre eine Vielzahl von Texten finden. Für das Subkorpus wurden Texte aus 48 kommerziellen Homepages extrahiert.

Das Subkorpus Wikipedia-Artikel beinhaltet 898 Beiträge, die allesamt von der Online-Enzyklopädie Wikipedia extrahiert wurden und sich thematisch mit medizinischen Eingriffen befassten. Unter den Texten finden sich detaillierte Beschreibungen chirurgischer Eingriffe, aber auch Informationen zum Hintergrund von Behandlungsmethoden und deren Begründern. Alle Artikel wurden von mehreren Autoren verfasst und in mehreren Editierungsprozessen revidiert. Auf Wikipedia können prinzipiell die Nutzernamen der Autoren identifiziert werden, soziobiografische Angaben zu ihnen können jedoch nicht gefunden werden, so dass es keinen Hinweis darauf gibt, ob die Autoren Muttersprachler oder Nichtmuttersprachler des Englischen sind. Obwohl die Textauswahl auf die Wikipedia-Homepage beschränkt war, ermöglichte es die große Menge der Beiträge zum Thema *Medical Surgery / Medical Tourism*, eine vergleichsweise hohe Anzahl an Texten für das Korpus zu extrahieren.

Grundsätzlich wurden immer komplette Webseiten beziehungsweise komplette Foren-*Threads* extrahiert. Nichtsdestotrotz gab es eine Reihe von Elementen auf den Webseiten, deren Aufnahme in das Korpus als unnötig oder sogar als hinderlich für die phraseologischen Analysen erachtet wurde und die folglich nicht kopiert wurden. Es folgt nun eine Auflistung aller Elemente, die bei der Extraktion der Texte außer Acht gelassen wurden:

- News-Artikel, wenn sie von externen Quellen stammten.
- Die Kontaktadressen auf den Webseiten. Nur wenn diese in einen laufenden Text integriert waren, wurden sie ebenfalls extrahiert.
- Fragebögen und Anmeldeformulare, wenn diese nicht in vollständigen Sätzen ausformuliert waren.
- Textpassagen, die sich auf verschiedenen Teilen der Webseite wiederholten.
- Hyperlinks, die nicht integraler Bestandteil des laufenden Textes waren.
- Nicht-englische Textpassagen.
- Werbeanzeigen.
- Grafiken, Fotos und Videos.

- Texte, die vor dem Jahr 2004 produziert wurden.³⁵

Für den Bereich Foren ergaben sich zudem weitere Elemente, die ebenfalls nicht in das Korpus übernommen wurden:

- Emoticons.
- Wiederholungen durch direkte Zitation vorheriger Posts.
- Threads, die weniger als fünf Antworten vorwiesen.
- Threads, die vor 2004 gestartet wurden.

Alle Texte wurden aus den jeweiligen Webseiten extrahiert und aufgeteilt in Subkorpora in einem *Plain-Text*-Format abgespeichert. In einem zweiten Schritt wurden die Dateien von nicht benötigten textlichen Elementen (*boilerplate code*) befreit, die durch die Konvertierung von HTML in das *Plain-Text*-Format entstanden sind. Dazu wurde ein Perl-Skript angefertigt, das sowohl alle *HTML-Tags*, als auch die durch die Konvertierung entstandenen Sonderzeichen und Duplikatspassagen aus dem Korpus entfernte. Nach diesem Schritt konnte schließlich die Gesamtgröße des Korpus bestimmt werden (vgl. Tabelle 4).

Web-Genre	Anzahl der Wörter	Prozent
Foren	1.200.422	44%
Blogs	216.449	8%
Kommerzielle Homepages	567.509	21%
Wikipedia-Artikel	740.262	27%
Gesamt	2.724.642	100%

Tabelle 4: Größe der individuellen Komponenten des Web-ELF-Korpus

Die Tabelle illustriert, dass die einzelnen Subkorpora unterschiedlich gewichtet sind. Die größte Komponente stellt die Subkomponente Foren (44%) dar, während die Subkomponente Blogs (8%) den geringsten Teil des Korpus bildet. Die Intention bei der Kompilierung des Korpus war, zwischen Foren / Blogs einerseits und kommerziellen Homepages / Wikipedia-Artikeln andererseits ein ungefähr ausgeglichenes Verhältnis herzustellen. Dies beruht auf der Vorannahme, dass die erste Gruppe eher dem konzeptionell mündlichen Pol von Mündlich-

³⁵ Dies geschah vor dem Hintergrund, dass mit Hilfe des Web-ELF-Korpus synchroner Sprachgebrauch untersucht werden sollte und alle Subkorpora Texte enthalten sollten, die in einem ähnlichen Zeitraum entstanden sind.

keit / Schriftlichkeit zugeordnet werden kann, während die zweite Gruppe eher konzeptionell schriftlichen Sprachgebrauch widerspiegelt.

Die Kompilation und die Annotation des Korpus erwiesen sich aus mehreren Gründen als nicht unproblematisch. Im Folgenden werden die Vorgehensweise bei der Kompilation des Korpus und die damit einhergehenden Probleme, die in den einzelnen Phasen des Kompilationsprozesses auftraten, im Detail beschrieben:

1. Die Textidentifikation

Im ersten Schritt mussten Webseiten ausfindig gemacht werden, die sich für die Aufnahme in das Korpus eigneten. Um sicherzustellen, dass die Texte den erwünschten Anforderungen entsprachen, musste jede Webseite im Detail überprüft werden, denn es zeigte sich immer wieder, dass sich potenzielle Webseiten bei genauerer Betrachtung als ungeeignet erwiesen. Es existierten beispielsweise Blogs, die auf den ersten Blick die Anforderungen erfüllten, bei genauerer Betrachtung jedoch wurde offenbar, dass der thematische Fokus nicht eingehalten wurde. Aus diesem Grund konnte nur ein Teil der ursprünglich identifizierten Blogs verwendet werden. In Foren wurden üblicherweise sowohl Name als auch Herkunft der Autoren angegeben, jedoch musste in einigen Fällen der Wahrheitsgehalt dieser Angaben in Frage gestellt werden. Deshalb wurden die Texte nach weiteren Anhaltspunkten für nichtmuttersprachlichen Sprachgebrauch durchsucht, unter anderem nach der Verwendung markierten grammatikalischen und lexikalischen Sprachgebrauchs.

Insgesamt war es sehr mühsam, genügend geeignete Texte für die beiden Subkorpora Foren und Blogs zu finden, da die Mehrheit aller gefundenen Blogs beziehungsweise Foren-*Threads* nicht alle erforderlichen Kriterien erfüllte. Ein Merkmal des Korpus ist die überproportionale Abbildung asiatischer Webseiten in den Web-Genres Blogs und kommerzielle Homepages. Dies ist dem Fakt geschuldet, dass die Mehrheit aller gefundenen Blogs und kommerziellen Homepages aus Asien stammen. Es muss somit berücksichtigt werden, dass in diesen beiden Web-Genres asiatisch-stämmige Englischnutzer überrepräsentiert sein könnten.³⁶

³⁶ Es muss jedoch ebenfalls betont werden, dass sich in Asien die meisten Nutzer des Englischen befinden. Somit spiegelt dies durchaus die reale globale Sprecherverteilung des Englischen wider.

2. Die Textextraktion und Annotation

Nach der Auswahl der Texte wurden diese im zweiten Schritt extrahiert. Die Extraktion von Webtexten birgt jedoch Risiken. Bei der Konvertierung von HTML-Code in das *Plain-Text*-Format werden nicht alle Zeichen korrekt umgewandelt, so dass je nach Ausgangstext eine mehr oder minder große Menge an *boilerplate code* (Sonderzeichen und Formatierungsannotationen) entsteht. Werden diese Sonderzeichen nicht beseitigt, können sie die Ergebnisse einer phraseologischen Untersuchung maßgeblich verfälschen. Daher musste das Korpus zunächst mit Hilfe mehrerer Perl-Skripte bereinigt werden, um alle Formen des *boilerplate code* vollständig aus dem Korpus zu entfernen.

Nachdem das Korpus erstellt und bereinigt wurde, wurde es mit Hilfe der *Tagging* Software CLAWS (*Constituent-Likelihood Automatic Word-Tagging System*) getaggt. CLAWS ordnet jedem Wort eines Textes eine grammatische Wortklasse zu. Für das *Parts-of-Speech (POS) Tagging* des Web-ELF-Korpus wurde das Tagset C7 verwendet. Dabei mussten abermals Teile des Korpus bereinigt werden, da CLAWS das *Tagging* stoppt, sobald es auf Sonderzeichen trifft. Diese mussten somit in Zeichen umgewandelt werden, die CLAWS identifizieren konnte, so dass das *Tagging* erfolgreich durchgeführt werden konnte. Diese Konvertierung hatte jedoch keinerlei Einfluss auf die Identifizierung von Mehrworteinheiten, da die in der vorliegenden Arbeit eingesetzten Programme, die ein Korpus nach häufig auftretenden Mehrworteinheiten durchsuchen, Sonderzeichen bei der Suche grundsätzlich außer Acht lassen.

5.1.2. Referenzkorpora

In der vorliegenden Arbeit werden wiederholt Ergebnisse mit den Ergebnissen anderer Studien verglichen. Sowohl bei der formalen als auch bei der funktionalen Analyse werden Bezüge zu den Ergebnissen von Biber et al. (1999) hergestellt. Diese haben für ihre formale Studie von *Lexical Bundles* zwei Subkorpora des *Longman Spoken and Written English (LSWE)* Korpus verwendet. Das LSWE-Korpus besteht aus einer Vielzahl von Texten aus unterschiedlichen mündlichen und schriftlichen Registern (*Conversation, Fiction, News, Academic prose, Conversation, Non-conversational speech, General prose*). Für ihre formale Untersuchung haben sie sich dabei auf die beiden Register *Conversation* und *Academic prose* beschränkt. Das Register *Conversation* besteht aus 3.929.500 Wörtern, die ausschließlich britisches Englisch abbilden, während das Register *Academic prose* 5.331.800 Wörter beinhaltet, die sowohl von amerikanischen als auch britischen Schreibern produziert wurden.

Für die funktionale Analyse dienten die Ergebnisse von Biber et al. (2004) als Referenz. Biber et al. (2004) haben für ihre Untersuchung das *TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL Corpus)* verwendet, das zwei Millionen Wörter enthält. In das Korpus wurden ausschließlich Texte aufgenommen, die im universitären Kontext entstanden sind. Die gesprochene Komponente des Korpus enthält Texte aus dem Register *Classroom teaching* (1.248.811 Wörter), die geschriebene Komponente Texte aus dem Register *Textbooks* (760.619 Wörter). Für ihre Studie haben Biber et al. (2004) ebenfalls die beiden Register *Conversation* und *Academic prose* aus dem LSWE-Korpus untersucht.

Für die qualitative Analyse des *4-grams at the end of* wurden die Ergebnisse mit Forchini / Murphy (2008) verglichen, die das *Financial Times Corpus (FT Corpus)* verwendet haben. Dieses Korpus besteht aus Texten, die sie von der Internetpräsenz der *Financial Times* extrahiert haben. Insgesamt enthält das Korpus 35 Millionen Wörter. Es stellt ein spezialisiertes Korpus dar, da es ausschließlich Texte zum Thema Finanzwirtschaft beinhaltet, die von Berufsjournalisten geschrieben wurden. Es scheint insbesondere interessant, zu überprüfen, ob ein Vergleich der vorliegenden Analyse mit der Analyse von Forchini / Murphy (2008) ähnliche Ergebnisse zu Tage bringt, obwohl sich die Korpora thematisch voneinander unterscheiden.

5.2. Linguistische und statistische Analysetools

Das Web-ELF-Korpus diene als textuelle Basis für alle lexikogrammatischen Analysen in der vorliegenden Arbeit. Um jedoch *4-grams* identifizieren und analysieren zu können, wurden linguistische Programme benötigt, deren Funktionen im Folgenden kurz beschrieben werden.

Für die formale Analyse wurde der *Automatic n-gram Analyser (ANGA)* verwendet. Diese spezielle Software wurde von Daniel-Benedikt Heller im Rahmen des LOEWE-Projekts³⁷ „Kulturechniken und ihre Medialisierung“ an der Justus-Liebig-Universität Gießen programmiert. Sie ermöglicht es, eine automatisierte formale Klassifikation von *n-grams* im Sinne von Biber et al. (1999) durchzuführen. Grundsätzlich erlaubt es das Programm, *n-grams* aus einem getaggten Korpus zu extrahieren und in die verschiedenen Kategorien der Taxonomie nach Biber et al. (1999) einzusortieren. Während dieses Prozesses fasst die Software automatisch die Ergebnisse der Analyse zusammen und erstellt eine grafische Abbildung, die einen Vergleich der Ergebnisse mit anderen Korpora ermöglicht. Die Größe der *n-grams* und ihr Mindestvorkommen können dabei frei bestimmt werden.

³⁷ Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz

Der *Automatic n-gram Analyser* (ANGA) basiert auf der Programmiersprache Perl. Die Software ist webbasiert und plattformunabhängig. ANGA ermöglicht es, drei verschiedene Analysen auszuführen. 1. eine Klassifizierung von *n-grams* in die Kategorisierung von Biber et al. (1999: 1001-1023), 2. eine Klassifizierung von *n-grams* in eine komprimierte Kategorisierung von Biber et al. (1999: 996), in der nur die wichtigsten (*major*) 12 Kategorien verwendet werden und 3. eine exhaustive Variante der zweiten Analyse.³⁸

Grundsätzlich ist die Größe der *n-grams*, die klassifiziert werden können, variabel. Um die einzelnen *n-grams* klassifizieren zu können, wurden reguläre Ausdrücke geschrieben, die jedes *n-gram* einer der Kategorien von Biber et al. (1999) zuweisen. Für vier der ursprünglich 26 Kategorien von Biber et al. (1999) konnte kein regulärer Ausdruck programmiert werden, so dass die erste Analyse lediglich einen groben Überblick über die Gesamtverteilung liefern kann. Für die zweite Analyse werden mehrere Kategorien zusammengefasst (analog zu Biber et al. 1999: 996), dementsprechend wurden die regulären Ausdrücke angepasst. Einzig die Kategorie *other expressions* wurde für diese Analyse herausgelassen, da in dieser Kategorie alle verbliebenen *n-grams* einsortiert werden, die sich keiner der anderen Kategorien zuordnen lassen. Da sich in den ersten beiden Analysen Mehrfachkategorisierungen ergaben, wurde die dritte Analyse entwickelt. In dieser sind Mehrfachkategorisierungen nicht möglich; alle *n-grams* können maximal einer Kategorie zugeordnet werden.

Für die Software wurde eine Benutzeroberfläche entwickelt (siehe Abbildung 10). Die Voreinstellungen basieren auf der Analyse von Biber et al. (1999). Sowohl die *n-gram*-Größe als auch der Mindestschwelligwert können verändert werden. Unter *Components* kann schließlich die Art der Analyse ausgewählt werden (eine Analyse der 26 Kategorien, eine Analyse der zusammengefassten Kategorien oder eine exhaustive Analyse). Zudem kann die Hierarchisierung für die Muster verändert werden. Dies ist für die exhaustive Analyse relevant. Wenn *Info* angeklickt wird, werden die Namen der Korpus-Quelldateien, die Größe der untersuchten *n-grams*, die Anzahl der minimalen Vorkommen der *n-grams* pro einer Million Wörter, die Gesamtzahl der Wörter und die relative Häufigkeit von Wörtern in *n-grams* aufgelistet. Ebenso wird angezeigt, wie viele Wörter des Korpus Teil eines *n-grams* sind. Wird die Option *N-Gram-List* angeklickt, wird den Ergebnissen eine Liste hinzugefügt, in der alle *n-grams* des Korpus sortiert nach ihrer Häufigkeit aufgeführt sind. Wird *Ending Analysis* ausgewählt, wird eine zusätzliche Analyse der grammatischen Kategorien des letzten Wortes der *n-*

³⁸ Für eine detailliertere Beschreibung des *Automatic n-gram Analyser* (ANGA) siehe Heller (2013).

grams durchgeführt (analog zu Biber et al. 1999: 997), in der die relative Häufigkeit der vier Kategorien *noun*, *pronoun*, *verb*, *other function word* dargestellt wird.

Sobald die Einstellungen ausgewählt wurden, kann die Analyse durchgeführt und die Ergebnisse können schließlich im HTML-Format ausgegeben werden.

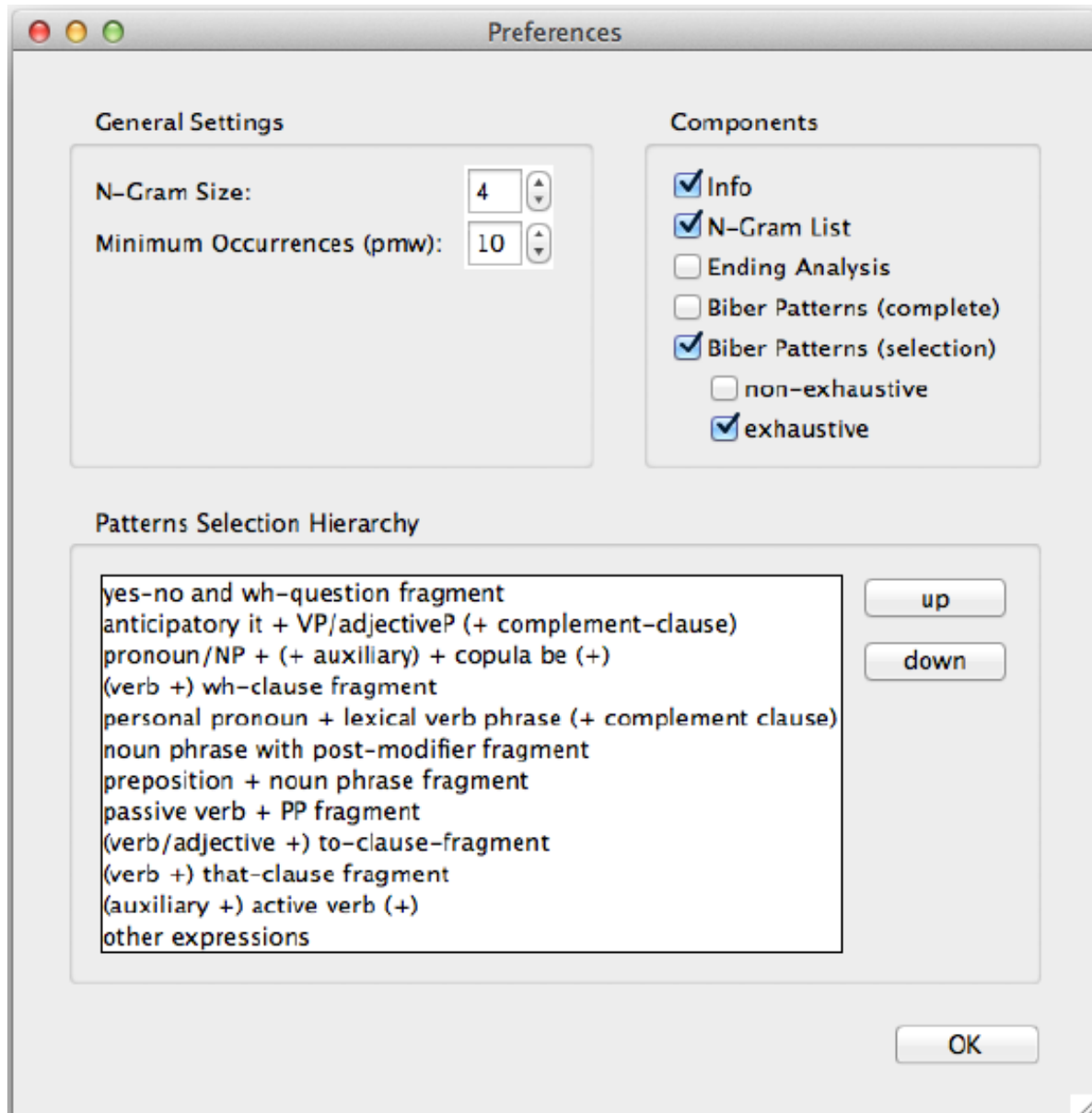


Abbildung 10: Einstellungsmöglichkeiten von ANGA (aus Heller 2013: 49).

Für die funktionale Analyse mussten zunächst die am häufigsten auftretenden *4-grams* aus dem Web-ELF-Korpus identifiziert werden. Dafür wurde ebenfalls der *Automatic n-gram Analyser* verwendet, da das Programm es ermöglicht, ein Korpus systematisch nach häufig

auftretenden Mehrworteinheiten zu durchsuchen und eine Liste anfertigt, in der alle identifizierten *n-grams* geordnet nach ihrer absoluten Häufigkeit im Korpus aufgeführt werden.

Für weitere Analysen wurde WordSmith 5.0 verwendet. WordSmith ist eine linguistische Software, die vielfältige lexikalische Untersuchungen ermöglicht, so zum Beispiel die Erstellung von Wortlisten oder *Keyword*-Listen (vgl. Scott 2008). Für die vorliegende Arbeit habe ich die *Concord*-Funktion genutzt. Diese ermöglicht es, einzelne *4-gram*-Muster in ihrem Kontext zu betrachten und daraus Rückschlüsse für qualitative Analysen zu ziehen, beispielsweise anhand einer Analyse der Kollokate.

Die Ergebnisse der quantitativen Analysen wurden sowohl mit Hilfe der Software Microsoft Excel 2007 als auch mit SPSS auf ihre Signifikanz hin überprüft. Beide Programme ermöglichen es, unterschiedliche Signifikanztests durchzuführen, um die Ergebnisse einer Analyse statistisch zu validieren. Für die vorliegenden Studien wurden mehrere Chi-Quadrat-Tests nach Pearson durchgeführt.

5.3. Methodische Konzeption der quantitativen Studien

Der Fokus der in der vorliegenden Arbeit durchgeführten Analysen liegt auf der Herausstellung lexikogrammatischer Besonderheiten. Für diesen Zweck werden mehrere quantitative und qualitative Studien durchgeführt, die Gemeinsamkeiten und Unterschiede zwischen den einzelnen Web-Genres auf formaler und funktionaler Ebene sichtbar machen sollen. Die methodischen Vorgehensweisen für die einzelnen Studien werden im Folgenden beschrieben.

5.3.1. Quantitative Studie – Formale Klassifizierung von Mehrworteinheiten im Websprachgebrauch

Die erste Studie basiert auf der formalen Klassifizierung von Biber et al. (1999), die nachgewiesen haben, dass es typische Verwendungsmuster (*Structural Patterns*) von Mehrworteinheiten für konzeptionell mündliche Genres gibt, die sich von den Verwendungsstrukturen in konzeptionell schriftlichen Genres unterscheiden. Für die vorliegende Studie wird die Biber-sche Kategorisierung übernommen, um den Sprachgebrauch im Web zu untersuchen. Der Fokus wird dabei ausschließlich auf die Analyse von *4-grams* der vier verschiedenen Web-Genres des Web-ELF-Korpus gelegt. Zudem wird die Untersuchung automatisiert, das heißt mit Hilfe der Software *Automatic n-gram Analyser*, durchgeführt. Für die Analyse werden ausschließlich *4-grams* berücksichtigt, die mindestens zehn Mal auftreten (beziehungsweise

im Fall von Blogs fünf Mal).³⁹ Jedes der *4-grams* wird in eine der von Biber et al. (1999) definierten Kategorien eingeordnet. Die Ergebnisse der Klassifizierung werden schließlich mit den Ergebnissen von Biber et al. (1999) für die Register *Conversation* und *Academic prose* verglichen. Dabei wird die Gesamtzahl aller *4-gram-Types* näher betrachtet und es wird überprüft, inwieweit sich die beiden Register diesbezüglich von den Web-Genres unterscheiden. Im Fokus steht jedoch der direkte Vergleich der Web-Genres im Hinblick auf die Verteilung der *4-gram-Types* in die dreizehn Kategorien. Es zeigt sich dabei, inwiefern manche Web-Genres gemeinsame Tendenzen aufweisen oder ob größere Unterschiede zu anderen Web-Genres sichtbar werden. Ein Vergleich der Ergebnisse für die einzelnen Web-Genres erlaubt es schließlich, generelle Aussagen über die lexikogrammatischen Besonderheiten der einzelnen Web-Genres zu treffen. Des Weiteren ermöglichen die Ergebnisse das Erstellen eines Kontinuums aus konzeptioneller Mündlichkeit / Schriftlichkeit, auf dem die einzelnen Web-Genres verortet werden können.

5.3.2. Quantitative Studie – Funktionale Klassifizierung von Mehrworteinheiten im Websprachgebrauch

Als Ausgangspunkt für die funktionale Analyse dienen die Studien von Biber et al. (1999) und Hyland (2008), die aufzeigen, dass Mehrworteinheiten in verschiedenen Kommunikationssituationen in unterschiedlichem Maße genutzt werden. Sie demonstrieren, dass zwischen "typisch mündlichen" *4-grams* (z.B. *I don 't know what*) und "typisch schriftlichen" *4-grams* (z.B. *on the other hand*) unterschieden werden kann. Biber et al. (2003, 2004) haben in einer weiteren Studie Mehrworteinheiten (*lexical bundles*) systematisch nach ihren vorwiegenden funktionalen Eigenschaften unterschieden. Dazu haben sie die häufigsten Mehrworteinheiten in vier verschiedenen Registern (*Conversation, Classroom teaching, Textbooks, Academic prose*) identifiziert und in drei funktionale Oberkategorien eingruppiert. Das Ergebnis der Studie zeigt, dass in gesprochener Sprache andere funktionale Typen von Mehrworteinheiten bevorzugt werden als in der Schriftsprache. Zudem weisen sie nach, dass zwischen den Registern keine absoluten Unterscheidungen gemacht werden können, sondern dass vielmehr ein

³⁹ Für Blogs wurde die Schwelle niedriger gesetzt, da das Korpus wesentlich weniger Wörter enthielt als die anderen Korpora. Wenn die Schwelle auf zehn belassen worden wäre, wären lediglich 34 *4-grams* identifiziert worden. Kopaczyk (2012: 86) zeigt, dass je nach Studie abhängig von der Korpusgröße andere *Cut-off-points* verwendet werden. Für die Verwendung einer niedrigen Mindestschwelle, siehe Wynne (2005).

Kontinuum von Mündlichkeit / Schriftlichkeit existiert, auf dem sich die verschiedenen Register verorten lassen.

Für die vorliegende Studie werden die 100 häufigsten *4-grams* in den vier Subkorpora mit Hilfe der Software *Automatic n-gram Analyser* identifiziert und in die funktionalen Kategorien von Biber et al. (2004) einsortiert.⁴⁰ Die Ergebnisse dieser Klassifizierung werden dann innerhalb des Korpus miteinander und im Anschluss mit typisch mündlichen und typisch schriftlichen Registern des *T2K-SWAL Corpus* und des *LSWE-Korpus* verglichen. Dies ermöglicht es schließlich, generelle Aussagen über die funktionalen Besonderheiten der einzelnen Web-Genres zu treffen.

Im Anschluss an diese Untersuchung werden die einzelnen *4-gram-Types* im Detail miteinander verglichen. Im Fokus steht die Frage, ob sich *4-gram-Types* identifizieren lassen, die in zwei oder mehreren Web-Genres häufig auftreten. Das Ziel besteht darin aufzuzeigen, dass es je nach Web-Genre mehr oder weniger *4-grams* gibt, die gemeinsam verwendet werden. Die Studie basiert auf der folgenden Grundthese: je mehr gemeinsame *4-gram-Types* verwendet werden, desto verwandter sind die Web-Genres sprachlich. Analog zur formalen Analyse kann für die funktionale Analyse somit ebenfalls ein Kontinuum von konzeptioneller Mündlichkeit / Schriftlichkeit erstellt werden, auf dem sich die Web-Genres verorten lassen.

5.4. Methodische Konzeption der qualitativen Analysen

Neben den quantitativen Analysen werden ebenfalls einige qualitative Studien durchgeführt, in denen prinzipiell nach auffälligen *4-gram*-Mustern gesucht wird, um zu überprüfen, ob auch Unterschiede in der Verwendung einzelne Muster zwischen den einzelnen Web-Genres existieren. Es wird dazu für jedes Web-Genre eine Liste angefertigt, in der die am häufigsten auftretenden *4-gram-Types* aufgeführt werden. Einzelne auffällige Muster werden dann mit Hilfe der *Concord*-Funktion von WordSmith 5.0 in ihrem Kontext qualitativ überprüft. Dabei wird die Analyse auf *4-grams*, die in jedem der vier Web-Genres häufig auftreten, beschränkt. Zunächst wird das Muster *at the end of* näher betrachtet werden. Eine Analyse des Kontexts jedes einzelnen Vorkommens ermöglicht es, aufzuzeigen, dass zwischen den Web-Genres Unterschiede bei der Verwendung des Musters existieren. In einem weiteren Schritt wird das

⁴⁰ Es wurden 100 *4-grams* miteinander verglichen. Die Ergebnisse zeigten, dass dies ausreichend Belege waren, um eindeutige Unterschiede zwischen den Web-Genres aufzuzeigen.

Pattern-Frame (Fletcher 2003-2012) *at the X of* in den verschiedenen Web-Genres überprüft und die Ergebnisse werden im Detail beschrieben.

Neben den beiden Mustern werden zwei weitere *4-grams* untersucht, die in jedem der vier Web-Genres häufig auftreten. Mit Hilfe von WordSmith 5.0 wird jedes Vorkommen der beiden Muster *one of the most* und *to be able to* in seinem Kontext überprüft. Dabei wird eine Liste angefertigt, in der die Kollokate der Muster aus den verschiedenen Web-Genres, geordnet nach funktionalen Kategorien, aufgeführt werden. Es stellt sich die Frage, ob abhängig vom Web-Genre Kollokate bevorzugt werden, die unterschiedlichen funktionalen Kategorien zugehören.

5.5. Zusammenfassung

In dem vorliegenden Kapitel wurden die methodischen Vorgehensweisen, die den quantitativen und qualitativen Studien zu Grunde liegen, näher erläutert. Für die Untersuchungen wird das Web-ELF-Korpus verwendet. Der Kompilationsprozess und die Probleme, die bei der Extraktion und Annotation des Web-ELF-Korpus auftraten, wurden im Detail beschrieben. Zudem wurde ein Überblick über sämtliche linguistische Analysetools gegeben, die für die Analysen verwendet wurden. Im Anschluss wurde eine Zusammenfassung der methodischen Vorgehensweisen für die einzelnen Studien gegeben.

Das folgende Kapitel widmet sich nun der Abbildung der Ergebnisse der quantitativen Studien. Dazu werden die jeweiligen Taxonomien näher beschrieben werden, die als Basis für die Studien dienen und die im Hinblick auf den untersuchten Sprachgebrauch im Medium Internet modifiziert werden müssen. Im Anschluss erfolgt eine ausführliche Darstellung der quantitativen Ergebnisse. Die Ergebnisse der qualitativen Analyse folgen schließlich in Kapitel 7.

6. Quantitative Analyse

In dem vorliegenden Kapitel werden die Ergebnisse der quantitativen Untersuchungen präsentiert und diskutiert. Für die quantitative Analyse werden sowohl formale als auch funktionale Analysen von *4-grams* durchgeführt. Zunächst werden in Kapitel 6.1. die Ergebnisse der formalen Analyse präsentiert, bevor schließlich in Kapitel 6.2. näher auf die Resultate der funktionalen Analyse eingegangen wird. Vor der Darlegung der jeweiligen Ergebnisse wird der konzeptionelle Hintergrund der einzelnen Studien detailliert beschrieben werden. Im Anschluss werden die Resultate mit denen anderer Studien verglichen und vor dem Hintergrund des Modells der konzeptionellen Mündlichkeit / Schriftlichkeit diskutiert.

6.1. Formale *4-gram* Analyse

6.1.1. Konzeptioneller Hintergrund

Die erste Studie fußt auf Biber et al. (1999: 996ff.), die anhand quantitativer Datengrundlage nachgewiesen haben, dass es typische Verwendungsstrukturen (*Structural Patterns*) von Mehrworteinheiten für konzeptionell mündliche und konzeptionell schriftliche Register gibt. Sie haben dazu die *Structural Patterns* in den beiden Registern *Conversation* und *Academic prose* identifiziert und für die jeweiligen Register charakteristische Merkmale herausgestellt. Beispielsweise werden in *Conversation* 44% aller *Structural Patterns* durch das *Pattern personal pronoun + lexical verb phrase (+ complement clause)* realisiert, während dieses *Pattern* in *Academic prose* überhaupt nicht auftritt. Im Gegensatz dazu werden in *Academic prose* 63% aller *Structural patterns* entweder durch das *Pattern noun phrase with post-modifier fragment* oder durch das *Pattern preposition + noun phrase fragment* realisiert, während lediglich 7% aller *4-grams* in *Conversation* durch diese *Patterns* realisiert werden.

Biber et al. (1999) haben für ihre Untersuchung *4-grams* aus dem *Longman Spoken and Written English* (LSWE) Korpus extrahiert, die ein Mindestvorkommen von zehn aufwiesen. Im Anschluss haben sie eine Kategorisierung ausgearbeitet, die es ermöglichte, jedes dieser *4-grams* einer Kategorie zuzuordnen. Insgesamt haben sie neunzehn Kategorien definiert. Allerdings zeigte sich nach Auswertung ihrer Ergebnisse, dass es einige Kategorien gab, die nur eine sehr geringe Zahl von *4-grams* beinhalteten. Aus diesem Grund haben Biber et al. (1999: 996) nur die zwölf häufigsten Kategorien tabellarisch abgebildet (vgl. Tabelle 5).

	CONV	ACAD	Example
patterns more widely used in conversation			
personal pronoun + lexical verb phrase (+ complement clause)	44%	—	I don't know what
pronoun/NP (+ auxiliary) + copula <i>be</i> (+ (auxiliary +) active verb (+)	8%	2%	it was in the
<i>yes-no</i> and <i>wh</i> -question fragment	13%	—	have a look at
(verb +) <i>wh</i> -clause fragment	12%	—	can I have a
	4%	—	know what I mean
patterns more widely used in academic prose			
noun phrase with post-modifier fragment	4%	30%	the nature of the
preposition + noun phrase fragment	3%	33%	as a result of
anticipatory <i>it</i> + VP/adjectiveP (+ complement-clause)	—	9%	it is possible to
passive verb + PP fragment	—	6%	is based on the
(verb +) <i>that</i> -clause fragment	1%	5%	should be noted that
patterns used in both registers			
(verb/adjective +) <i>to</i> -clause fragment	5%	9%	are likely to be
other expressions	6%	6%	
Total	100%	100%	

Tabelle 5: Prozentuale Verteilung der 4-grams in *Structural Patterns*

Die Gegenüberstellung der beiden Register zeichnet ein klares Bild. Es gibt große Unterschiede zwischen *Conversation* und *Academic prose* bezüglich der prozentualen Verteilung der 4-grams. 81% aller 4-grams des Registers *Conversation* wurden einer der ersten fünf Kategorien zugeordnet, während dies für nur 2% aller 4-grams des Registers *Academic prose* der Fall war. Viele 4-grams im Register *Conversation* beinhalteten ein pronominales Subjekt, dem eine Verbalphrase (*verbal phrase*) und der Beginn eines Komplementsatzes folgten (zum Beispiel *I don't know why*). In die nächsten fünf Kategorien wurden hingegen 84% aller 4-grams des Registers *Academic prose* einsortiert, während dies nur für 8% aller 4-grams des Registers *Conversation* der Fall war. Ein großer Teil der 4-grams im Register *Academic prose* beinhaltet Nominalteilsätze (*nominal phrases*) und Präpositionalteilsätze (*prepositional phrases*, wie beispielsweise *the nature of the* und *as a result of*).

In ihrer Untersuchung fanden Biber et al. (1999: 995) außerdem heraus, dass 15% aller 4-grams in *Conversation* komplette strukturelle Einheiten (*Phrases* / Teilsätze oder *Clauses* / Teilsätze) darstellten, während es in *Academic prose* weniger als 5% waren. Für beide Register galt gleichermaßen, dass die meisten der identifizierten 4-grams Elemente aus zwei

Phrases oder *Clauses* beinhalteten.⁴¹ Biber et al. (1999: 994) stellten insgesamt fest, dass *4-grams* sowohl in *Conversation* als auch in *Academic prose* sehr häufig auftraten (in *Conversation* 8500 mal *pmw*, in *Academic prose* 5000 mal *pmw*).⁴² Durchschnittlich kam ein *4-gram* 20 mal *pmw* vor. Gleichzeitig trat nur ein Bruchteil aller *4-grams* mit einer sehr hohen Häufigkeit (über 100 mal *pmw*) auf. Diese konnten vor allem im Register *Conversation* vorgefunden werden, während in *Academic prose* nur sehr wenige *4-grams* sehr häufig auftraten. Dies alles führt Biber et al. (1999: 995) zu dem Schluss, dass ein bedeutender Teil jedes sprachlichen Diskurses aus häufig auftretenden Mehrworteinheiten besteht (30% aller Wörter in *Conversation* und 21% aller Wörter in *Academic prose*). Dies verdeutlicht die Bedeutung struktureller *Patterns* sowohl für den mündlichen als auch für den schriftlichen Sprachgebrauch. Gleichzeitig muss jedoch beachtet werden, dass im Register *Conversation* eine größere Anzahl unterschiedlicher Wortformen von *3-grams* und *4-grams* verwendet werden als im Register *Academic prose* (summiert etwa 4200 *Lexical Bundles*, im Gegensatz zu etwa 3000 *Lexical Bundles* für das Register *Academic prose*).

6.1.2. Definition der Kategorien

Ursprünglich haben Biber et al. (1999) die *4-grams* in neunzehn verschiedene Kategorien eingeteilt. Allerdings haben sie für ihren grafischen Vergleich nur die relevantesten Kategorien ausgewählt, einige davon zusammengefasst und somit lediglich zwölf Kategorien in ihrer Übersicht abgebildet. Grundsätzlich wurde diese reduzierte Klassifizierung für die vorliegende Analyse übernommen, allerdings wurde sie leicht modifiziert, insofern als dass sie um eine dreizehnte Kategorie (*Adverbial clause fragments*) ergänzt wurde. Dies war notwendig, da sich zeigte, dass in einigen der Web-Genres eine hohe Anzahl an *Adverbial clause fragments* existierte und sich diese Kategorie somit als bedeutsam für den Vergleich der Web-Genres erwies.

Im Folgenden werde ich nun die einzelnen Kategorien näher beschreiben und typische Beispiele für jede Kategorie geben. Alle Elemente, die in den Überschriften der einzelnen Kategorien aufgeführt werden, sind obligatorisch. Lediglich die Elemente, die in Klammern aufgeführt werden, sind optional. Neben der Beschreibung der Kategorien werden auch die Ergebnisse für die Bibersche Kategorisierung (vgl. Biber et al. 1999: 996ff.) nochmals aufgegriffen.

⁴¹ Beispielsweise beinhaltet das *4-gram I want to know* Elemente zweier Teilsätze (*Clauses*), während das *4-gram in the case of* Elemente zweier Teilsätze (*Phrases*) enthält.

⁴² *per million words* / pro Million Wörter.

1) personal pronoun + lexical verb phrase (+ complement-clause fragment)

Diese Kategorie deckt einen Großteil aller *4-grams* des Registers *Conversation* ab. *4-grams* in dieser Kategorie sind üblicherweise Teilsätze, die mit einem Subjektpronomen beginnen und von einer Verbalphrase (*verbal phrase*) gefolgt werden. Oftmals wird die Verbalphrase um einen Ergänzungssatz (*complement clause*) erweitert. Viele der *4-grams* in dieser Kategorie folgen dem strukturellen Muster Subjekt in der ersten Person Singular + Zustandsverb (*stative verb*). Die meisten Verben in dieser Kategorie reflektieren die Einstellung (*personal stance*) der Sprecher / Autoren oder drücken ihre persönlichen Gefühle, Gedanken und Wünsche aus. Insbesondere in gesprochener Sprache berichten Sprachnutzer oftmals aus ihrer eigenen Perspektive. Somit ergeben sich typische Konstruktionen mit dem Subjekt *I* und den Verben *know, think, want, say/tell, like, mean* oder modalen und semimodalen Verben. Neben der ersten Person Singular finden sich in dieser Kategorie jedoch auch andere Personalpronomen. Eine sehr häufig auftretende Konstruktion ist beispielsweise *you don't have to*. Typischerweise werden in dieser Kategorie einige *4-grams* sehr häufig verwendet (100 mal *pmw* und häufiger).

Ogleich diese Kategorie einen Großteil aller identifizierten *4-grams* (44%) des Registers *Conversation* enthält, findet sich im Register *Academic prose* kein einziges *4-gram*, das dieser Kategorie zugeordnet werden kann. Dies muss dem Fakt zugeschrieben werden, dass in akademischen Texten eine neutrale Betrachtungsebene beibehalten werden soll. Dies führt dazu, dass dieses *4-gram* Muster in diesem Register nicht verwendet wird.

Typische Beispiele sind *I don't know what, I don't think so, you don't have to, he said to me*.

2) pronoun / noun phrase + be +

Die *4-grams* in dieser Kategorie stellen Teilsätze dar, die zumeist aus einem unpersönlichen Pronomen oder einem Demonstrativpronomen, das als Subjekt fungiert, bestehen und denen das Kopular *be*, das häufig kontraktiert wird (zum Beispiel *it's*), folgt. In manchen Fällen wird das Kopular *be* um die ersten Elemente eines Ergänzungssatzes (*complement clause*) oder einer prädikativen Nominalphrase (*nominal phrase*) ergänzt.

Typische Beispiele sind *it's going to be, that's what I said, there's a lot of, I'm going to be*.

3) verb phrase with active verb

Die *4-grams* in dieser Kategorie bestehen aus Verbalphrasen (*verbal phrases*), die ein aktives Verb enthalten müssen und einen Teil eines ergänzenden Teilsatzes (*phrase*) oder eines Ergänzungssatzes. Eine Vielzahl der *4-grams* in dieser Kategorie ist idiomatisch. Oftmals kann ihre Bedeutung nicht aus den einzelnen Elementen des *4-grams* erschlossen werden.

Typische Beispiele sind *have a look at, get on with it, have a cup of*.

4) yes - no and wh-question fragment

In dieser Kategorie werden zwei Bibersche Unterkategorien (*yes-no fragment* und *wh-question fragment*) zu einer zusammengefasst. Für erstere gilt: die *4-grams* in dieser Kategorie beginnen mit einem Hilfsverb oder einem Modalverb, dem ein Subjektpronomen und ein Verb folgen. Somit bilden diese *4-grams* den beginnenden Teil einer Frage, die nur mit ja oder nein beantwortet werden kann.

Typische Beispiele sind *can I have a, have you got a, do you know what, would you like to*. Für Letztere gilt: Die *4-grams* in dieser Kategorie beginnen mit einem *wh*-Fragewort (zum Beispiel *what, where, how*), dem zunächst ein Hilfsverb oder ein Modalverb folgt, das um ein Subjektpronomen und ein Verb ergänzt wird. Die *4-grams* bilden somit den Beginn einer *wh*-Frage.

Typische Beispiele sind *what are you doing, where are you going, how do you know*.

5) (verb +) wh-clause fragment

In dieser Kategorie befinden sich alle *4-grams*, die den Beginn eines *wh*-Teilsatzes (*wh-clause*) markieren. Optional kann dem *wh*-Teilsatz eine Verbalphrase (*verbal phrase*) vorausgehen. Die meisten *4-grams* in dieser Kategorie sind Teile von größeren *n-grams*, beispielsweise ist das *4-gram don't know what to* Teil des *6-grams I don't know what to get*.

Typische Beispiele sind *don't know what it, see what I mean, when I went to*.

6) adverbial clause fragments

In dieser Kategorie befinden sich zum einen *4-grams*, die mit einem *if*-Teilsatz beginnen, zum anderen *4-grams* mit einem komplexen Subordinator mit dem Muster *as + adverb + as + pronoun*. Viele *if*-Teilsätze haben das Personalpronomen *you* als Subjekt, dem die Hauptverben *want* oder *got/get* folgen. In dieser Kategorie finden sich darüber hinaus auch weitere *4-grams*, die in ihrer Form stark fixiert sind, wie beispielsweise *because I want to*.

Typische Beispiele sind *if you want to, if you can get, as long as you, because I want to*.

7) noun-phrase with post-modifier fragment

Im Register *Academic prose* existiert eine Vielzahl von *4-grams*, die eine Nominalphrase (*noun phrase*) beinhalten, der ein postmodifizierender *of*-Teilsatz (*post-modifier-fragment*) folgt. Diese dienen der physikalischen Beschreibung von Sachverhalten und Gegenständen. Es gibt zudem einige wenige Nominalphrasen (*noun phrases*) mit anderen postmodifizierenden Elementen. Darunter fallen Nominalphrasen mit einem post-nominalen Teilsatz-Fragment und Nominalphrasen mit einem Präpositionalphrasen-Fragment (*prepositional phrase-fragment*). Während viele *4-grams* des Registers *Academic prose* (30%) dieser Kategorie zugehören, lassen sich nur 4% aller *4-grams* des Registers *Conversation* in dieser Kategorie verorten.

Typische Beispiele sind *the end of the, the beginning of the, the way in which, the relationship between the*.

8) preposition + noun phrase fragment

Weiterhin besteht im Register *Academic prose* ein Großteil aller *4-grams* aus einer Präpositionalphrase (*prepositional phrase*) und einer Nominalphrase (*noun phrase*), die von einem *of*-Teilsatz-Fragment (*of-phrase-fragment*) postmodifiziert wird. Die meisten *4-grams* in dieser Kategorie markieren abstrakte, logische Bezüge. Viele von ihnen werden mit den Präpositionen *as* oder *in* gebildet. *4-grams*, die mit *at* beginnen, markieren temporale Bezüge. Es existieren zudem Präpositionalphrasen (*prepositional phrases*) ohne *of*-Teilsatz-Fragment (*of-phrase-fragment*); zumeist beginnen diese mit der Präposition *in*.

Typische Beispiele sind *about the nature of, as a result of, at the end of, as in the case, in the present study, on the other hand*.

9) anticipatory *it* + VP / adjective (+ complement-clause)

Die *4-grams* in dieser Kategorie beginnen mit einem *anticipatory it*, dem eine Adjektivphrase (*adjective phrase*) oder eine Verbalphrase (*verbal phrase*) folgt. Der Großteil der *4-grams* in dieser Kategorie beinhaltet prädikative Adjektive, die einem Komplementsatz vorausgehen (normalerweise einem *to*-Teilsatz). Die Verbalphrasen hingegen sind zumeist passivisiert und gehen einem *that*-Teilsatz voraus.

Typische Beispiele sind *it is possible to, it can be seen, it has been shown*.

10) passive verb + PP fragment

Zu dieser Kategorie zählen *4-grams*, die aus einer passivisierten Verbalphrase (*verbal phrase*) bestehen, der ein Präpositionalphrasen-Fragment (*prepositional phrase-fragment*) folgt. Die Präpositionalphrasen kennzeichnen häufig eine lokative oder logische Beziehung.

Typische Beispiele sind *are shown in table, is shown in figure, is based on the*.

11) (verb +) *that*-clause fragment

In dieser Kategorie gibt es zwei unterschiedliche Ausprägungen von *4-grams*. Die eine Gruppe enthält ein Hauptverb, welches um ein *that*-Teilsatz-Fragment (*that-clause-fragment*) ergänzt wird. Die meisten *4-grams*, die ein Hauptverb enthalten, sind Teil einer größeren Einheit, beispielsweise *it should be noted that*. Die andere Gruppe beinhaltet ausschließlich *that*-Teilsatz-Fragmente. Diese sind Ergänzungssätze in unmarkierten deklarativen Strukturen mit zwei Ausprägungsformen. Es existiert eine Ausprägungsform mit dem *existential there* als Subjekt und einem Präsens Kopular *is* als Verb. Die zweite Ausprägungsform hingegen hat einen extraponierten Teilsatz im *that*-Teilsatz eingebettet, in dem *it* als Subjekt und das Kopular *is* als Verb fungieren.

Typische Beispiele sind *should be noted that, that there is no, that it is not*.

12) (verb / adjective +) *to*-clause fragment

Die *4-grams* in dieser Kategorie können in drei Untergruppen aufgeteilt werden.

1. prädikatives Adjektiv + *to*-Teilsatz
2. (passive) Verbalphrase (*verbal phrase*) + *to*-Teilsatz
3. einfacher *to*-Teilsatz.

Typische Beispiele sind *are likely to be, has been shown to, to be able to, is said to be*.

13) other expressions

In diese Kategorie werden alle *4-grams* einsortiert, die sich keiner der anderen Kategorien zweifelsohne zuweisen lassen. Darunter fallen verschiedenste Konstruktionen wie *than that of the, may or may not* oder *the presence or absence*.

6.1.3. Adaption der Kategorisierung auf Web-Genres

Für die Kategorisierung der vier Web-Genres wird nun die Kategorisierung von Biber et al. übernommen, mit dem Ziel, diese untereinander und mit den Registern *Conversation* und *Academic prose* vergleichen zu können. Die Bibersche Kategorisierung wird jedoch um die

Kategorie 6. *adverbial clause fragments* ergänzt. Die Analyse wird, im Gegensatz zu Biber et al. (1999), automatisiert mit Hilfe der Software *Automatic n-gram Analyser*, die eigens für diesen Zweck entwickelt wurde, durchgeführt. Die Software wurde so konzipiert, dass sie einen beliebigen getaggten Text nach häufig auftretenden *4-gram*-Mustern durchsucht und diese einer der dreizehn vorgegebenen Kategorien zuordnet.

6.1.4. Ergebnisse der formalen Klassifikation – *4-gram-Types*

Der *Automatic n-gram Analyser* gibt für die jeweiligen Subkorpora eine Liste aus, in der alle identifizierten *4-grams*, sortiert nach ihrer absoluten Häufigkeit, aufgelistet werden. Tabelle 6 gibt eine Übersicht über die normalisierte Gesamtzahl aller *Types*, die in den einzelnen Korpora gefunden wurden. In der linken Spalte werden zunächst die einzelnen Kategorien aufgeführt. Die nächsten beiden Spalten zeigen die Verteilung aller *4-grams* in den beiden Registern *Conversation* und *Academic prose*. Die letzten vier Spalten führen alle *4-grams* auf, die in den vier Subkorpora des Web-ELF-Korpus gefunden wurden.

	<i>Conver- sation</i>	<i>Academic prose</i>	Foren	Blogs	kom. Ho- mepages	Wikipedia- Artikel
1. pronoun + lexical verb phrase (+ complement clause)	48	0	200	32	53	0
2. pronoun/noun phrase (+ auxiliary) + copula <i>be</i> (+)	8	1	44	46	53	9
3. (auxiliary +) active verb (+)	14	0	91	69	111	9
4. yes-no and <i>wh</i> -question fragment	12	0	17	9	2	0
5. (verb +) <i>wh</i> -clause fragment	4	0	26	14	18	4
6. adverbial clause fragment	3	1	33	32	21	4
7. noun phrase with post-modifier fragment	4	16	63	231	289	151
8. preposition + noun phrase fragment	3	17	51	189	220	123
9. anticipatory <i>it</i> + noun phrase/adjective phrase (+ complement clause)	0	5	17	14	33	8
10. passive verb + prepositional phrase fragment	0	3	6	0	83	39
11. (verb +) <i>that</i> -clause fragment	1	2	15	0	7	4
12. (verb/adjective +) <i>to</i> -clause fragment	5	5	78	55	92	32
13. other expressions	1	1	297	333	668	140
Gesamt (normalisiert <i>pmw</i>)	104	50	939	1026	1649	525

Tabelle 6: Normalisierte Gesamtzahl (*pmw*) aller *4-gram-Types* in den verschiedenen Korpora⁴³

Die Übersicht offenbart, dass zwischen den einzelnen Registern / Web-Genres enorme Unterschiede bezüglich der normalisierten Gesamtzahl aller kategorisierten *4-gram-Types* bestehen (vgl. Tabelle 6). Zunächst fällt auf, dass der Unterschied zwischen den Biberschen Registern und den Web-Genres groß ist. Die Gesamtzahl aller *4-gram-Types* ist im Register *Conversation* (104 *Types*) höher als im Register *Academic prose* (50 *Types*). Für die Web-Genres hingegen ergeben sich größere Unterschiede. Foren (939 *Types*) und kommerzielle Homepages (1649 *Types*) weisen die größte Anzahl an *4-grams* auf. Blogs weisen mit 1026 *Types* (*pmw*) ebenfalls eine hohe Anzahl auf.

Wikipedia-Artikel (525 *Types*) beinhalten eine vergleichsweise wesentlich geringere Menge an *4-grams*. Dies kann ein Indikator dafür sein, dass der Schreibstil im Web-Genre Wikipedia-Artikel wesentlich homogener ist, als in den Bereichen Foren und kommerzielle

⁴³ Die *Types* wurden auf ganze Zahlen gerundet.

Homepages. Offenbar beschränken sich die Autoren von Wikipedia-Artikeln auf die Verwendung weniger unterschiedlicher *4-gram-Types*. Im Gegensatz dazu wird im Web-Genre kommerzielle Homepages eine viel größere Bandbreite an *4-grams* verwendet. Dies könnte der Tatsache geschuldet sein, dass sich die Autoren in diesem Web-Genre freier ausdrücken können, da sie nicht durch formale Vorgaben (wie insbesondere beim Web-Genre Wikipedia-Artikel) beschränkt werden. Die geringere Anzahl an *4-gram-Types* für die Register *Conversation* und *Academic prose* ist unter anderem damit zu erklären, dass potenzielle *4-grams*, die der Kategorie 13. *other expressions* zugeordnet werden konnten, von Biber et al. (1999) als Teile größerer *n-gram*-Verwendungsstrukturen bewertet wurden und somit nicht für die Gesamtzahl aller *4-gram Types* berücksichtigt wurden.

6.1.5. Ergebnisse der formalen Klassifikation – Distribution der Kategorien

Im nächsten Schritt soll nun überprüft werden, wie sich die einzelnen *4-gram-Types* in den Biberschen Kategorien verteilen. Ein besonderes Augenmerk liegt dabei auf der Frage, inwieweit Distributionsparallelen zwischen den einzelnen Web-Genres und den Biberschen Registern zu erkennen sind. Das Ziel der Kategorisierung ist es, die einzelnen Web-Genres auf dem Kontinuum von Mündlichkeit / Schriftlichkeit im Sinne von Koch / Oesterreicher (1985, 1994) zu verorten. Vor diesem Hintergrund dienen die beiden Biberschen Register *Conversation* und *Academic prose* als oppositionelle Endpunkte auf einem Kontinuum von Mündlichkeit / Schriftlichkeit. Die einzelnen Web-Genres können schließlich mit diesen Endpunkten verglichen und auf dem Kontinuum verortet werden.

Im Folgenden werden nun die einzelnen Subkorpora in Beziehung zu den beiden Registern gesetzt. Dies ermöglicht eine grafische Abbildung der prozentualen Verteilung der *4-gram-Types* in die dreizehn von Biber et al. (1999: 996) definierten Kategorien. Für jede Kategorie werden drei Balken aufgeführt. Der linke Balken zeigt den prozentualen Wert für das Register *Conversation*, der mittlere Balken den Wert für das jeweilige Web-ELF-Subkorpus und der rechte Balken den Wert für das Register *Academic prose* an. Abbildung 11 gibt zunächst das Ergebnis für das Subkorpus Foren wieder.

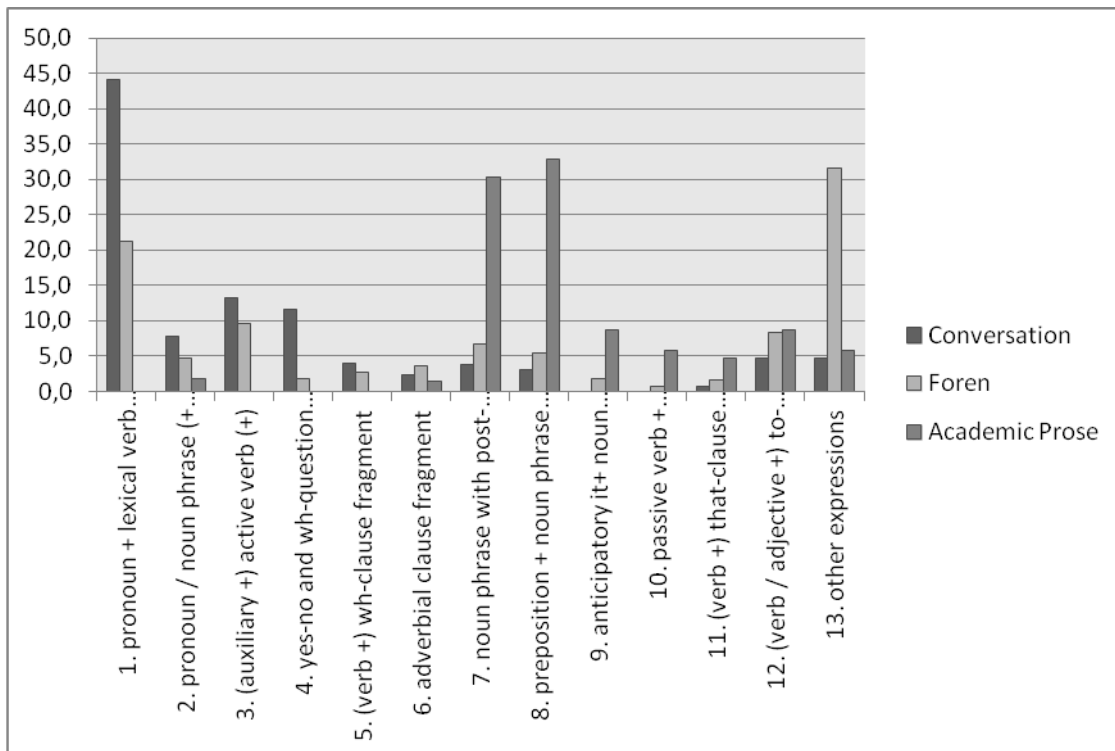


Abbildung 11: Foren im prozentualen Vergleich zu den Registern in Biber et al. (2004)

Die *4-gram-Types* im Subkorpus Foren ähneln in ihrer Distribution eher dem Register *Conversation* als dem Register *Academic prose*, wenngleich die Unterschiede bezüglich der Verteilung der Muster in die einzelnen Kategorien zwischen *Conversation* und *Academic prose* insgesamt hochsignifikant ($p=0.0001$) sind.⁴⁴ Die Mehrheit aller *4-grams* (43,8%) kann den ersten sechs Kategorien zugeordnet werden, die laut Biber et al. (1999) typische Verwendungsstrukturen gesprochener Sprache darstellen. Im Vergleich zu den Registern *Conversation* und *Academic prose* sind die Unterschiede hochsignifikant ($p=0.0001$). Daneben fällt auf, dass ebenfalls ein großer Teil aller *4-grams* (31,7%) in die Kategorie 13. *other expressions* fällt. Dies ist ein Phänomen, das ebenso für die anderen Web-Genres bestätigt werden kann und zum einen damit erklärt werden muss, dass einige *4-grams* mediumspezifisch sind, zum anderen damit, dass die Web-ELF-Texte aus einem sehr spezifischen fachsprachlichen Kontext extrahiert wurden. Die anderen sechs Kategorien (Kategorien 7-12), die laut Biber et al. (1999) eher typische Verwendungsstrukturen schriftlicher Sprache sind, ergeben summiert 24,5%. Im Vergleich zu *Conversation* und *Academic prose* sind die Unterschiede hochsignifikant ($p=0.0001$). Es fällt auf, dass die Kategorie 12. *(verb / adjective +) to-clause fragment*

⁴⁴ Die einzelnen Chi-Quadrat-Tests für die individuellen Kategorien der Web-Genres werden in Appendix B aufgeführt.

einen hohen Anteil (8,3%) aller schriftsprachlichen Kategorien ausmacht. Diesbezüglich sind die Unterschiede zwischen Foren und *Conversation*, *Academic prose* hochsignifikant ($p=0.0001$). Insgesamt enthält das Subkorpus Foren im Vergleich zu den anderen Web-Genres verhältnismäßig viele mündliche Verwendungsstrukturen ($p=0.0001$) und ähnelt in seiner Gesamtverteilung eher dem Register *Conversation*.

Abbildung 12 fasst das Ergebnis für das Subkorpus Blogs zusammen.

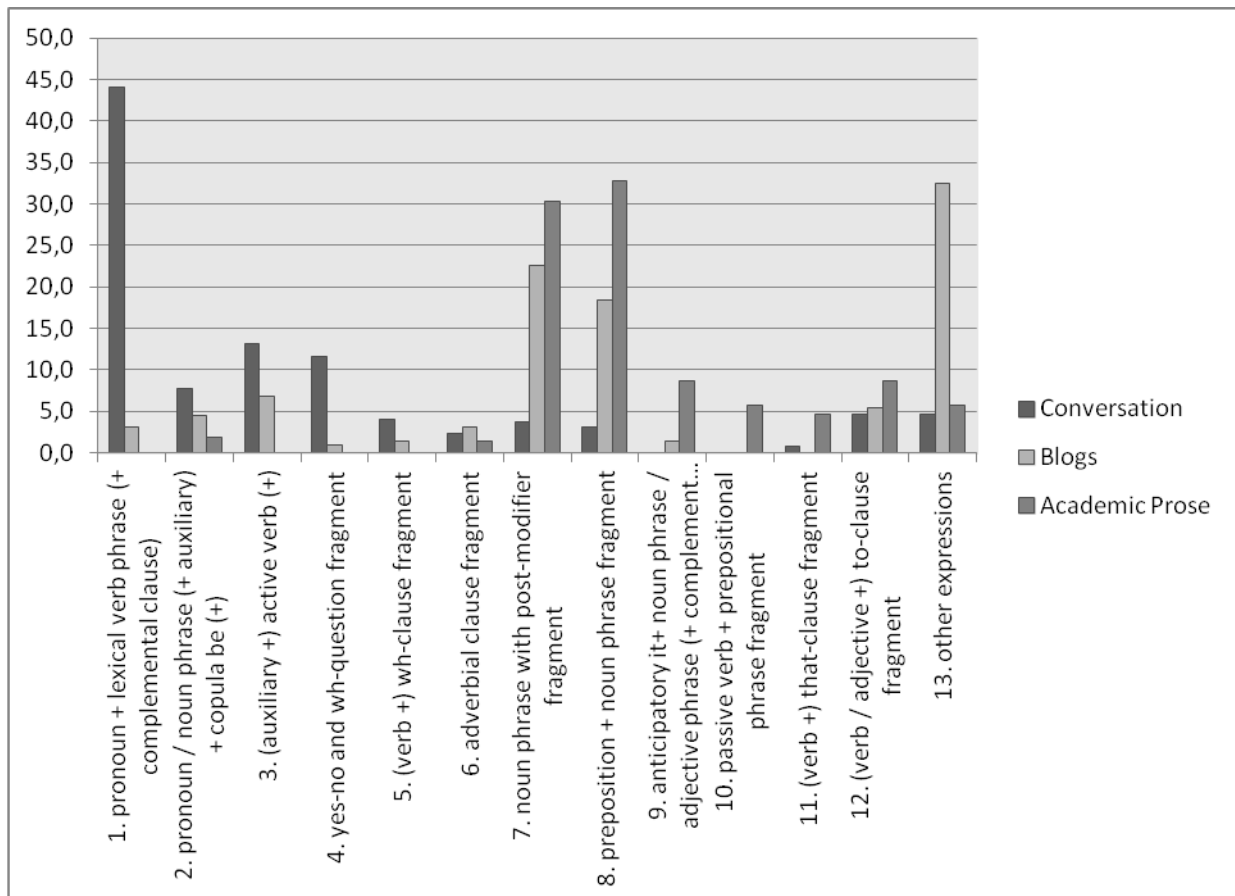


Abbildung 12: Blogs im prozentualen Vergleich zu den Registern in Biber et al. (2004)

Das Subkorpus Blogs unterscheidet sich in seiner Distribution erheblich vom Subkorpus Foren. Es lassen sich eher Parallelen zum Register *Academic prose* herstellen. In die ersten sechs Kategorien können lediglich 20% aller *4-grams* eingeordnet werden. Insbesondere die erste Kategorie beinhaltet im Vergleich zu *Conversation* eine geringe Anzahl an *4-grams* (3,2%, $p=0.0001$). Es befinden sich 47,8% aller *4-grams* in den weiteren sechs Kategorien (Kategorien 7-12). Dabei vereinen die Kategorien 7. *noun-phrase with post-modifier fragment* (22,5%) und 8. *preposition + noun phrase fragment* (18,5%) die Mehrheit von ihnen. Die Kategorie 13. *other expressions* beinhaltet 32,4% aller *4-gram-Types*. Blogs unterscheiden sich

in ihrer Verteilung der mündlichen (1-6) und der schriftsprachlichen (7-12) Kategorien hochsignifikant ($p=0.0001$) sowohl von *Conversation* als auch von *Academic prose*. Dennoch bleibt festzuhalten, dass das Subkorpus Blogs einen wesentlich höheren Anteil schriftsprachlicher 4-gram-Verwendungsstrukturen aufweist und demnach in seiner Verteilung eher dem Register *Academic prose* gleicht.

Abbildung 13 gibt einen Überblick über die Verteilung für das Subkorpus kommerzielle Homepages.

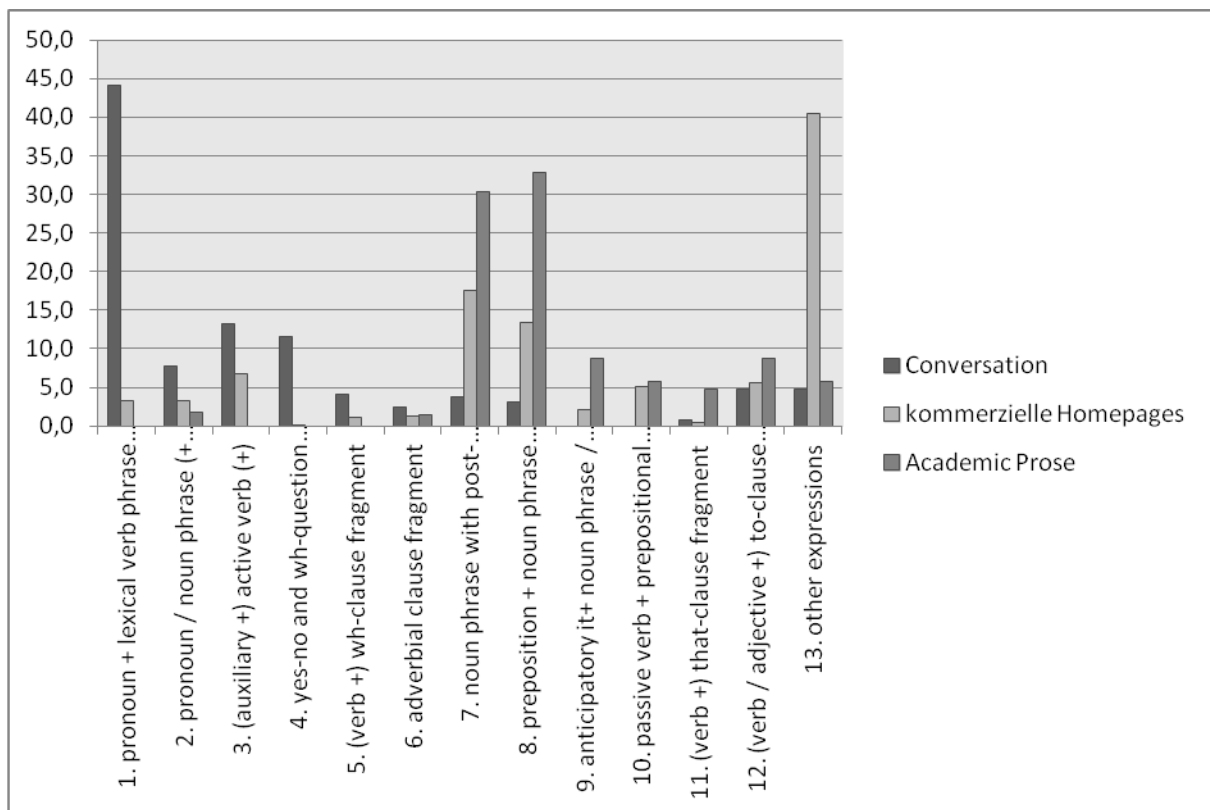


Abbildung 13: Kommerzielle Homepages im prozentualen Vergleich zu den Registern in Biber et al. (2004)

Das Subkorpus ähnelt in seiner Distribution sehr dem Subkorpus Blogs. Im Gegensatz zu Blogs ist die Anzahl der 4-grams in der Kategorie 13. *other expressions* jedoch noch höher und liegt bei 40,5%. Nur ein Bruchteil aller 4-grams (15,6%) lässt sich in die ersten sechs Kategorien einordnen. Diesbezüglich unterscheidet es sich hochsignifikant ($p=0.0001$) von *Conversation* und von *Academic prose*. Die Mehrheit aller 4-grams (43,9%) befindet sich in den nächsten sechs Kategorien (Kategorien 7-12). Analog zum Subkorpus Blogs lässt sich auch in diesem Subkorpus ein Großteil der 4-grams in die beiden Kategorien 7. *noun-phrase with post-modifier fragment* (17,5%) und 8. *preposition + noun phrase fragment* (13,4%) einord-

nen. Auch bezüglich der schriftsprachlichen Kategorien (7-12) sind die Unterschiede zwischen kommerziellen Homepages und *Conversation / Academic prose* hochsignifikant ($p=0.0001$). Insgesamt enthält das Subkorpus kommerzielle Homepages wesentlich mehr schriftsprachliche als mündliche *4-gram*-Verwendungsstrukturen und gleicht in seiner Verteilung eher dem Register *Academic prose*.

Abbildung 14 gibt die Verteilung aller Verwendungsstrukturen für das Subkorpus Wikipedia-Artikel wieder.

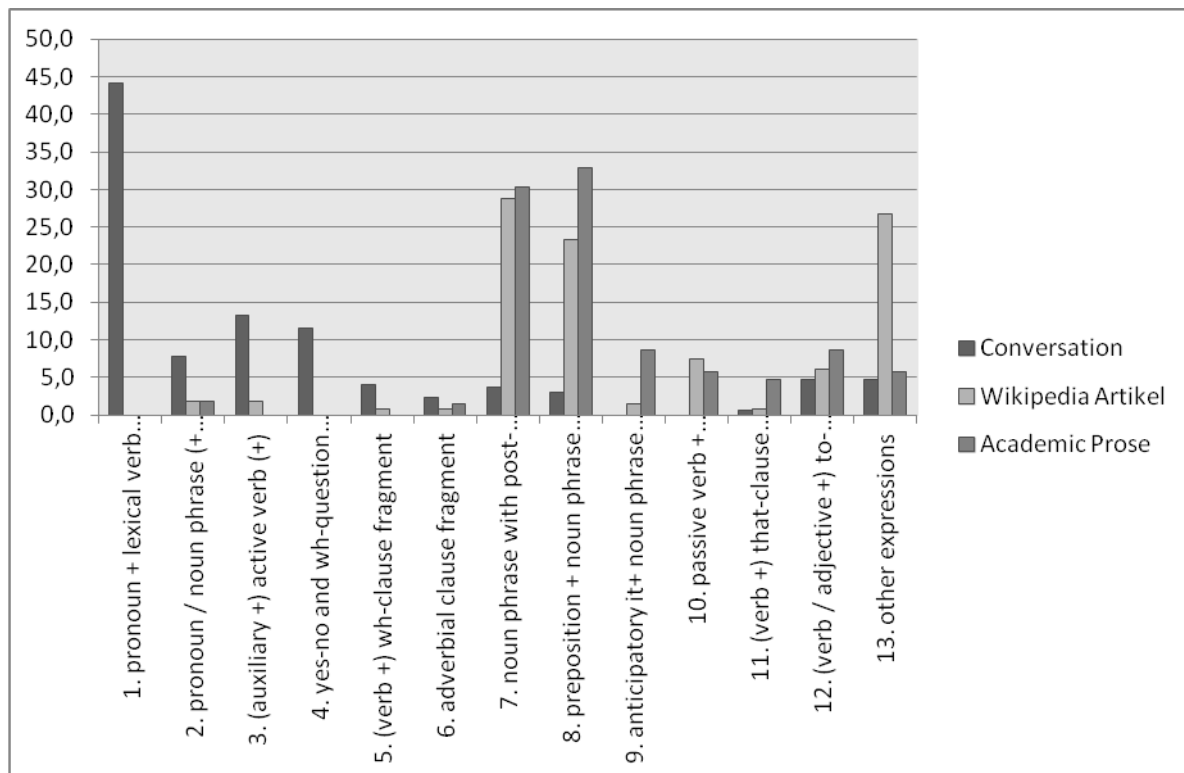


Abbildung 14: Wikipedia-Artikel im prozentualen Vergleich zu den Registern in Biber et al. (2004)

Das Subkorpus ähnelt in seiner Distribution den beiden Subkorpora Blogs und kommerzielle Homepages. Allerdings weist es vergleichsweise weniger mündliche Verwendungsstrukturen und mehr schriftsprachliche Verwendungsstrukturen auf ($p=0.0001$). So können lediglich 5,2% aller *4-grams* den ersten sechs Kategorien zugerechnet werden. In der Kategorie 13. *other expressions* befinden sich 26,7% aller *4-grams*. Den mit Abstand größten Teil (68,2%) machen jedoch die sechs schriftsprachlichen Kategorien (7-12) aus. Auch in diesem Subkorpus zählen die Kategorien 7. *noun-phrase with post-modifier fragment* (28,8%) und 8. *preposition + noun phrase fragment* (23,4%) zu den Kategorien mit den am häufigsten auftretenden *4-grams*. Die Unterschiede zwischen Wikipedia-Artikeln und *Conversation* und

Academic prose sind sowohl im Hinblick auf die mündlichen als auch auf die schriftsprachlichen Kategorien hochsignifikant ($p=0.0001$). Insgesamt enthält das Subkorpus Wikipedia-Artikel einen wesentlich höheren Anteil an schriftsprachlichen *4-gram*-Verwendungsstrukturen und gleicht in seiner Verteilung eher dem Register *Academic prose*.

Ein Vergleich der jeweiligen Subkorpora mit den Biberschen Kategorien vermittelt einen ersten Eindruck darüber, inwieweit die einzelnen Web-Genres sich eher dem konzeptionell mündlichen oder dem konzeptionell schriftsprachlichen Pol zuordnen lassen.

Die formale Analyse ermöglicht einen Vergleich der Web-Genres untereinander, da sie allesamt demselben Korpus entstammen. Abbildung 15 liefert eine Zusammenfassung der Verteilung der *4-gram*-Types in die jeweiligen Kategorien für die vier Subkorpora des Web-ELF-Korpus. Dabei wurde für die Abbildung bewusst die Kategorie 13. *other expressions* weggelassen, um somit nur diejenigen Kategorien zu betrachten, die sich zweifelsfrei einer der 12 mündlichen und schriftsprachlichen Kategorien zuordnen lassen.

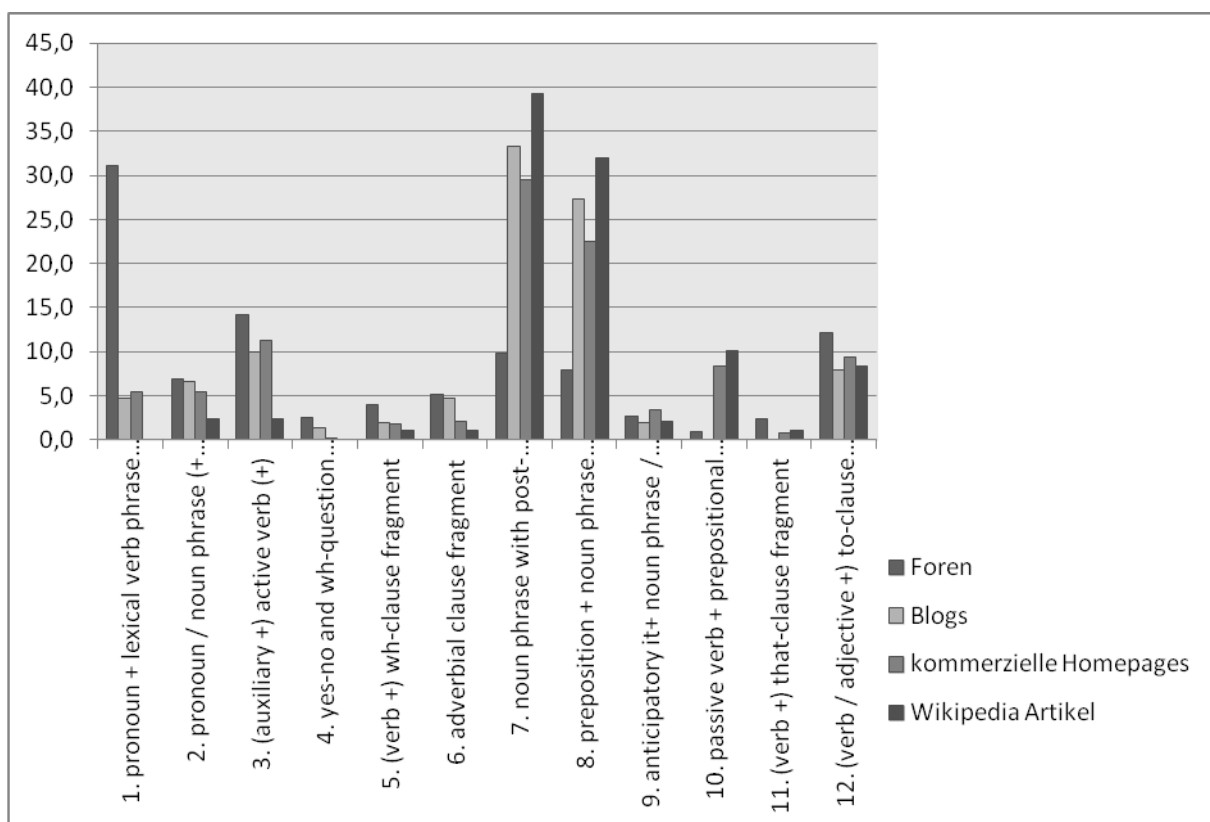


Abbildung 15: Prozentualer Vergleich der Web-Genres untereinander

Die Grafik offenbart, dass Foren mit Abstand die meisten mündlichen *4-gram*-Verwendungsstrukturen aufweisen ($p=0.0001$), wohingegen Wikipedia-Artikel deutlich die meisten schrift-

sprachlichen *4-gram*-Verwendungsstrukturen enthalten ($p=0.0001$). Ein wesentlicher Unterschied ($p=0.0001$) wird vor allem in der ersten Kategorie ersichtlich, in der sich die Mehrheit aller *4-grams* des Bereichs Foren (31,2%) befindet. Die beiden Web-Genres Blogs und kommerzielle Homepages beinhalten im Vergleich zum Web-Genre Foren wesentlich mehr Elemente schriftsprachlicher *4-gram*-Verwendungsstrukturen ($p=0.0001$), jedoch weniger als das Web-Genre Wikipedia-Artikel ($p=0.0001$). Des Weiteren zeigt sich, dass das Web-Genre kommerzielle Homepages zwar insgesamt mehr schriftsprachliche Elemente aufweist als Blogs ($p=0.001$), jedoch in den beiden Hauptkategorien 7 und 8 weniger Elemente (51,8%) als Blogs (60,6%) beinhaltet. Die Unterschiede in den Kategorien 7 ($p=0.19$) und 8 ($p=0.46$) sind jedoch nicht signifikant.

Abbildung 16 fasst die sechs mündlichen Kategorien und die sechs schriftsprachlichen Kategorien weiter in zwei Kategorien zusammen. Daraus ergibt sich ein Kontinuum mit zwei klar identifizierbaren Polen (mündlich – schriftsprachlich). Die Übersicht zeigt auf, dass zwischen den einzelnen Web-Genres graduelle Unterschiede bezüglich ihrer konzeptionellen Mündlichkeit / Schriftlichkeit bestehen.

So ist ersichtlich, dass 64,1% aller *4-gram*-Types im Bereich Foren tendenziell mündliche *4-gram*-Verwendungsstrukturen darstellen. Foren sind somit das einzige Web-Genre des Web-ELF-Korpus, das mehrheitlich mündliche *4-gram*-Verwendungsstrukturen enthält. Die beiden Web-Genres Blogs und kommerzielle Homepages ähneln sich stark in ihrer Gesamtverteilung. Beide enthalten vorwiegend schriftsprachliche *4-gram*-Verwendungsstrukturen (Blogs 29,4%, kommerzielle Homepages 26,3%). Wikipedia-Artikel hingegen können sehr nah am Pol der Schriftlichkeit verortet werden, da ein Großteil aller *4-grams* (92,8%) schriftsprachliche Verwendungsstrukturen aufweist.

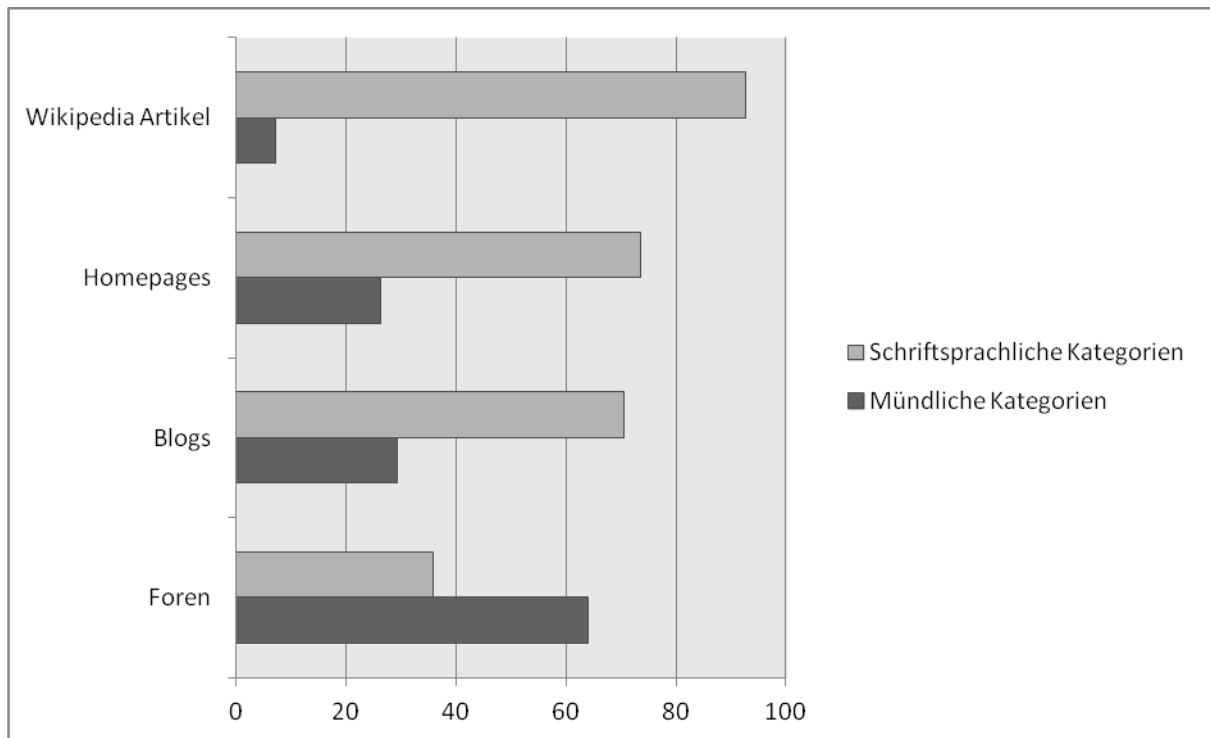


Abbildung 16: Prozentualer Vergleich aller mündlichen und schriftsprachlichen Kategorien der Web-Genres

Neben der Analyse der *4-gram*-Verwendungsstrukturen mündlicher und schriftsprachlicher Register haben Biber et al. (1999: 993f.) die strukturelle Abgeschlossenheit der beiden Register im Detail überprüft. Dabei haben sie gezeigt, dass 15% aller *4-grams* in *Conversation* und weniger als 5% in *Academic prose* komplette *phrases* oder *clauses* bilden. Die meisten *4-grams* stellen somit keine abgeschlossenen Einheiten dar, sondern verbinden zwei strukturelle Einheiten miteinander. Diese Untersuchung wurde ebenfalls für das Web-ELF-Korpus durchgeführt. Dazu wurden die 200 häufigsten *4-gram-Types* jedes Subkorpus im Hinblick auf ihre strukturelle Abgeschlossenheit überprüft. Das Ergebnis zeigt auf, dass auch in den vier Web-Genres nur ein Bruchteil der identifizierten *4-grams* komplette strukturelle Einheiten in Form von *phrases* oder *clauses* darstellt. Für die einzelnen Subkorpora ergibt sich das folgende Bild: Foren (12%), Blogs (6,5%), kommerzielle Homepages (4%) und Wikipedia-Artikel (2%). Auf struktureller Ebene scheint damit ausschließlich das Web-Genre Foren dem Register *Conversation* zu entsprechen, während Blogs, kommerzielle Homepages und Wikipedia-Artikel eher dem Register *Academic prose* gleichen. Dieses Ergebnis ist ein weiterer Indikator dafür, dass sich das Web-Genre Foren im Hinblick auf seine formalen Eigenschaften grundlegend von den drei anderen Web-Genres unterscheidet.

6.1.6. Probleme

Die Analyse hat einige aufschlussreiche Befunde hervorgebracht. Dennoch muss beachtet werden, dass bei einer formalen *4-gram*-Untersuchung Probleme auftreten können, die bei der Betrachtung der Ergebnisse berücksichtigt werden müssen. Ein Problem ist die strukturelle Abgeschlossenheit der untersuchten Mehrworteinheiten. In den meisten Fällen stellen Mehrworteinheiten keine strukturellen Einheiten dar, wie es beispielsweise bei *on the other hand* der Fall ist. Stattdessen überbrücken sie häufig *Phrases* oder *Clauses*. Zum Beispiel besteht eine Vielzahl von *4-grams* im Register *Conversation* üblicherweise aus einem Hauptsatz, der ergänzt wird um den Beginn eines eingebetteten Komplementsatzes. So kann das *4-gram* *But I don't know*⁴⁵ ein Teil des Ausdrucks *But I don't know what you're talking about* sein. Im Register *Academic prose* folgt ein Großteil aller *4-grams* einem anderen Muster. Sie bestehen typischerweise aus einer unvollständigen Nominalphrase (*noun phrase*), die ergänzt wird um den Beginn eines eingebetteten *of*-Teilsatzes (*of-phrase*) (zum Beispiel *the nature of the issues involved*).

Die Beispiele zeigen zudem, dass einige *n-grams* Bestandteil größerer Einheiten sein können. So kann das häufig auftretende *3-gram* (*I don't know*) Bestandteil eines *4-grams* (*I don't know what*) sein. Umgekehrt beinhaltet das *4-gram* *I don't know what* das *3-gram* *I don't know* und das *2-gram* *don't know*. Eine Vielzahl von *3-grams* kann somit um ein Element ergänzt werden, um ein *4-gram* zu bilden. Im selben Maße können *4-grams* Teile von *5-grams*, *6-grams* oder größeren Einheiten sein. Aufgrund dieser Tatsache existieren viele Mehrworteinheiten, die in ihrer Form als *4-gram* keiner der definierten Kategorien zugeordnet werden können, wie zum Beispiel *is one of the*. Diese Tatsache spiegelt sich in der großen Menge an *4-grams* wider, die der Kategorie 13. *other expressions* zugeordnet werden kann. Während bei Biber et al. (1999: 996) nur 6% aller *4-grams* in diese Kategorie einsortiert wurden, ist die Zahl in den einzelnen Web-ELF-Subkorpora wesentlich höher.⁴⁶ Dies hat vielfältige Ursachen. So gibt es in den verschiedenen Web-Genres viele häufig auftretende *4-grams*, die sich in keinem der beiden Register *Conversation* und *Academic prose* finden lassen. Einerseits ist dies vermutlich der Tatsache geschuldet, dass die Texte aus dem Medium Internet ex-

⁴⁵ Kontraktierte Formen (beispielsweise *don't* oder *can't*) werden allgemein als ein Element gezählt, so dass das *n-gram* *I don't know what* als *4-gram* (und nicht etwa als *5-gram*) aufgefasst wird.

⁴⁶ Beispielsweise können in dem Web-Genre Blogs 40,5% aller *4-grams* keiner der ersten zwölf Kategorien zugeordnet werden und müssen somit in die Kategorie 13. *other expressions* einsortiert werden.

trahiert wurden, so dass einige *4-grams* medium-spezifisch sind. Andererseits hat die Beschränkung auf das recht spezifische Thema *Medical Surgery / Medical Tourism* einige häufig auftretende *4-grams* hervorgebracht, die korpuspezifisch sind und sich somit nicht in den beiden Registern finden lassen. Der Hauptgrund für die wesentlich höhere Anzahl an *4-grams* in der Kategorie 13. *other expressions* liegt vermutlich darin, dass Biber et al. (1999) neben *4-grams* auch *5-grams* und *6-grams* untersuchten. *4-grams*, die sich keiner Kategorie zweifelsohne zurechnen ließen, wurden um mehrere Elemente ergänzt, so dass diese sich dann als größere Einheiten (zum Beispiel *5-grams* oder *6-grams*) einer der Kategorien zuordnen ließen. Somit blieben in ihrer Studie kaum *4-grams* übrig, die sich keiner der definierten Kategorien zuordnen ließen. Da ich mich in dieser Studie ausschließlich auf *4-grams* beschränke, bin ich in diesem Punkt von der Methodik von Biber et al. (1999) abgewichen. Aus diesem Grund befindet sich in der vorliegenden Analyse in der Kategorie 13. *other expressions* eine Vielzahl von *4-grams*, die vermutlich Teile größerer *n-grams* sind. Da das Hauptziel der Studie jedoch darin besteht, die einzelnen Web-Genres untereinander zu vergleichen und alle Subkorpora methodisch gleich bearbeitet wurden, ist ein direkter Vergleich der Subkorpora untereinander unproblematisch.

Neben den methodischen Problemen ergaben sich auch einige Probleme, die auf die automatisierte Bearbeitung des Korpus zurückzuführen sind. Hauptsächlich betrifft dies das *Tagging* der *4-grams*. Die Subkorpora wurden allesamt mit CLAWS 4 getaggt. CLAWS eignet sich hervorragend, um schriftsprachliche Korpora zu taggen, da es für sie eine sehr geringe Fehlerquote aufweist. Texte aus dem Web stellen CLAWS jedoch vor wesentlich größere Probleme, so dass die Fehlerquote beim *Tagging* von Webtexten weitaus höher liegt. Im Grunde genommen ordnet CLAWS aus einer geschlossenen Wortdatenbank jedem Wort eines Korpus ein *Tag* zu. Wenn jedoch Abweichungen von der Orthografie auftreten oder Wörter auftauchen, die CLAWS nicht bekannt sind, wird das betreffende Wort häufig als Nomen *getaggt*, was in den meisten Fällen inkorrekt ist. Somit liegt die Fehlerquote beim *Tagging* von Texten aus dem Web bei 20-30%. Da die automatisierte Kategorisierung darauf basiert, dass allen Wörtern die richtigen *Tags* zugewiesen wurden, musste die Liste aller identifizierten *4-grams* überprüft werden und das *Tagging* optimiert werden. Aus diesem Grund mussten die Kategorien für das Skript des *Automatic n-gram Analyser* in mehreren Schritten genauer definiert werden, um sicherzustellen, dass jedes *4-gram* korrekt kategorisiert wurde. Nach mehreren

Korrekturprozessen war der *Automatic n-gram Analyser* schließlich dazu in der Lage, eine vollständig korrekte automatisierte Analyse durchzuführen.⁴⁷

6.1.7. Zusammenfassung

Ein Vergleich der Subkorpora des Web-ELF-Korpus hat aufgezeigt, dass auf formaler Ebene Unterschiede in der Verwendung von *4-gram*-Verwendungsstrukturen existieren. Prinzipiell bestätigte die Studie die Intuition, dass Foren in ihrer Form eher mündlichem Sprachgebrauch entsprechen, während Wikipedia-Artikel eher schriftsprachlichem Sprachgebrauch ähneln. Blogs und kommerzielle Homepages stehen dabei zwischen beiden Extremen, allerdings können sie eher dem Pol der Schriftsprache zugerechnet werden, da sie einen hohen Anteil schriftsprachlicher Verwendungsstrukturen aufweisen. Für das Web-Genre Blogs bedeutet dies, dass die ursprüngliche Annahme, dass auch dieses Subkorpus in seiner Distribution eher dem Web-Genre Foren ähnelt, nicht bestätigt werden konnte. Stattdessen weist es ähnlich viele Elemente schriftsprachlichen Sprachgebrauchs auf wie das Web-Genre kommerzielle Homepages. Inwieweit sich diese Ergebnisse auch auf die funktionale Ebene übertragen lassen, wird im Folgenden mit Hilfe zweier funktionaler Studien überprüft werden.

6.2. Funktionale 4-gram Analyse

6.2.1. Grundkonzeption

Das Ziel der folgenden Studien ist es, Unterschiede in der funktionalen Verwendung von *4-grams* in vier verschiedenen Web-Genres (Foren, Blogs, kommerzielle Homepages, Wikipedia-Artikel) herauszustellen. Zunächst wird eine funktionale Klassifizierung der *4-grams* des Web-ELF-Korpus nach Biber et al. (2004) durchgeführt. Dazu werden im ersten Schritt die häufigsten *4-grams* für die Web-Genres identifiziert. Im zweiten Schritt werden diese dann funktional klassifiziert, um die vier untersuchten Web-Genres im Hinblick auf den Grad ihrer konzeptionellen Mündlichkeit / Schriftlichkeit bewerten zu können. Abschließend werden die Ergebnisse dargelegt und überprüft, inwieweit es Unterschiede in der funktionalen Verwendung von *4-grams* gibt. In einer zweiten Studie wird nach gemeinsamen *4-grams* gesucht, die in zwei oder mehr Web-Genres auftreten. Eine häufige Verwendung derselben

⁴⁷ Für Probleme linguistischer Annotation vgl. auch Leech (2005) und McEnery / Hardie (2011) und für Probleme mit der Bereinigung und Annotation von Webkorpora vgl. Baroni / Kilgarriff (2006).

4-gram-Types kann ein Indikator dafür sein, dass sich bestimmte Web-Genres funktional gleichen. Analog dazu deutet eine lediglich geringe Überschneidung in der Nutzung gemeinsamer *4-gram-Types* darauf hin, dass sich die Web-Genres funktional unterscheiden.

Auf der lexikogrammatischen Ebene sind Genreunterschiede üblicherweise gradueller Natur und können in Form von Häufigkeitsunterschieden im Gebrauch sprachlicher Formen und Strukturen erfasst werden. Um die unterschiedliche Verwendung von Mehrworteinheiten in Web-Genres nachzuweisen wird daher ein großes Textkorpus benötigt, das es erlaubt, anhand dieser empirischen Datengrundlage zuverlässig Frequenzunterschiede zu beschreiben. Für diesen Zweck wird das Web-ELF-Korpus verwendet, da es das erste Korpus dieser Größe ist, das den Web-bezogenen Gebrauch des Englischen in verschiedenen Web-Kommunikationssituationen repräsentiert. Zunächst müssen jedoch mit Hilfe des *Automatic n-gram Analyser* alle *4-grams* aus dem Web-ELF-Korpus identifiziert und extrahiert werden, bevor eine weitergehende Analyse durchgeführt werden kann.

6.2.2. Identifikation von Mehrworteinheiten mit dem *Automatic n-gram Analyser*

Für die vorliegende Studie wird zunächst für jedes der vier Subkorpora des Web-ELF-Korpus eine *4-gram*-Liste erstellt. Tabelle 7 gibt einen Auszug aus einer vom *Automatic n-gram Analyser* generierten Liste des Web-Genres kommerzielle Homepages mit den zehn häufigsten *4-grams* für das Web-Genre kommerzielle Homepages wieder.

Rang	<i>4-gram</i>	Frequenz
1	at the same time	160
2	one of the most	158
3	is one of the	154
4	the best candidates for	118
5	it is important to	118
6	will be able to	109
7	as well as the	100
8	as a result of	95
9	for a few days	93
10	the size of the	91

Tabelle 7: Auszug aus der *4-gram* Liste des Web-Genres kommerzielle Homepages (normalisiert *pmw*)

Um die Web-Genres systematisch miteinander vergleichen zu können, wird für jedes Subkorpus eine Liste erstellt, die die 100 häufigsten *4-grams* enthält.⁴⁸

6.2.3. Die funktionale Klassifikation von *4-grams*

6.2.3.1. Das Erstellen einer Klassifikation

Für die vorliegende Studie wird die von Biber et al. (2003, 2004) entwickelte Taxonomie verwendet, um die eingangs identifizierten und in *4-gram* Listen festgehaltenen Mehrworteinheiten schließlich im Hinblick auf ihre Funktion zu klassifizieren. Biber et al. (2004) haben, angelehnt an frühere frequenzbasierte Studien zu häufig auftretenden Mehrworteinheiten (Altenberg 1998, Biber et al. 1999, 2003), eine Taxonomie erstellt, mit deren Hilfe Mehrworteinheiten funktional klassifiziert werden können. Sie haben für ihre Untersuchung das *T2K-SWAL Corpus* verwendet (*TOEFL 2000 Spoken and Written Academic Language Corpus*), das zwei Millionen Wörter enthält. Dieses besteht aus gesprochenen und geschriebenen Texten, die im universitären Kontext entstanden sind. Sie haben sich jedoch für ihre Studie auf die Untersuchung der beiden Register *Classroom teaching* und *Textbooks* beschränkt und diese anschließend mit zwei weiteren Registern aus dem *Longman Spoken and Written English (LSWE)* Korpus, *Conversation* und *Academic prose*, verglichen.⁴⁹ Die Kategorisierung erlaubte es, die für die Register typischsten *4-grams* zu identifizieren und insgesamt aufzuzeigen, dass sich die einzelnen Register bezüglich der funktionalen Verwendung von *4-grams* voneinander unterscheiden. Aus der Verwendung in den beiden medial mündlichen und den beiden medial schriftlichen Registern konnten Biber et al. (2004) schließlich ableiten, welche *4-grams* typischerweise in gesprochener Sprache und welche in Schriftsprache verwendet werden.

Biber et al. (2003, 2004) unterscheiden in ihrer Taxonomie zwischen drei funktionalen Kategorien, in die die einzelnen *4-grams* einsortiert werden können: *Stance expressions*, *Discourse organising expressions* und *Referential expressions*.

I. *Stance expressions* bereiten einen Rahmen, um die eigentliche Aussage eines Satzes (die Proposition) zu interpretieren. Durch sie moduliert ein Sprecher die Information beziehungsweise die Proposition eines Satzes, um bestimmte subjektive Effekte zu erzielen. Innerhalb

⁴⁸ In Appendix A werden die vier Listen mit den 100 am häufigsten auftretenden *4-grams* abgebildet.

⁴⁹ Die deutschen Entsprechungen der Register lauten universitärer Unterricht, akademische Textbücher, Konversation und akademische Prosa.

der *Stance expressions* wird zwischen *Epistemic stance expressions* und *Attitudinal / Modality stance expressions* unterschieden. Erstere beschreiben den Wissensstand eines Sprechers gegenüber der folgenden Proposition. Beispiele dafür sind *I don't know if your brother received the message in time* und *I don't think that Crystal Palace will win the championship this year*. Die meisten *Epistemic stance expressions* sind personenbezogen und drücken Unsicherheit aus (*I don't know what* und *I don't think that*). *Attitudinal / Modal stance expressions* drücken die Einstellung des Sprechers gegenüber der sich anschließenden Proposition aus. Beispiele für *Epistemic stance expressions* sind *I want you to come to my daughter's birthday* und *I'm not going to let you win this time*.

II. Discourse organizers haben zwei Hauptfunktionen: zum einen dienen sie der Einführung eines Themas, zum anderen der Verdeutlichung und der Elaboration eines Themas. Typische Beispiele für diese Kategorie sind *I want to talk about Obama's new foreign policy* und *Today we are going to talk about the history of Iceland*.

III. Referential expressions identifizieren und spezifizieren physische und abstrakte Objekte und geben Orts- und Zeitreferenzen. Beispiele dafür sind *For those of you who come late, there will be no dinner*; *Let's talk a little bit about the company's internal organization* und *Many people in the United States have no social security*.

Um die jeweiligen *4-grams* nach dieser Taxonomie klassifizieren zu können, müssen sie zusammen mit ihrem Kontext betrachtet werden. In Tabelle 8 sieht man typische Beispiele für *4-grams* aus den vier Web-Genres, die sich anhand ihres Kontextes zweifelsfrei bestimmten funktionalen Kategorien zuordnen lassen.

I. Stance Expressions	Beispiele
A) Epistemic Stance I don't know if	<u>I don't know</u> if this will help, but I think it may. (Forum)
I don't think she	<u>I don't think she</u> should be working with such troubled people. (Forum)
B) Attitudinal / Modality Stance I would like to	<u>I would like to</u> do something more than just make money for myself. (Forum)
if you want to	<u>If you want to</u> have a healthy life during your retirement years... (Blog)
II. Discourse Organizers	Beispiele
keep in mind that as a result of	Try to <u>keep in mind</u> that this is normal and will subside as... (k. Homepage) <u>As a result of</u> the bleeding, the body releases large amounts of... (Wikip.)
III. Referential Expressions	Beispiele
the rest of the one of the most	I'll definitely be watching <u>the rest of the</u> season. (Forum) Price is <u>one of the most</u> unimportant factors in the end. (k. Homepage)

Tabelle 8: Typische 4-grams aus verschiedenen Web-Genres des Web-ELF-Korpus für die einzelnen funktionalen Kategorien (vgl. Saage et al. 2012: 146)

In ihrer Analyse haben Biber et al. (2004) alle identifizierten 4-grams der vier Register in die drei Kategorien einsortiert. Eine grafische Zusammenfassung dieser quantitativen Distribution der 4-grams für die einzelnen Register liefert Abbildung 17. Insgesamt fällt auf, dass in den medial mündlichen Registern *Conversation* und *Classroom teaching* viele *Stance expressions* verwendet werden, während in den medial schriftlichen Registern *Textbooks* und *Academic prose* vorrangig *Referential expressions* verwendet werden. *Discourse organizers* wiederum werden wesentlich häufiger in *Conversation* und *Classroom teaching* genutzt als in *Textbooks* und *Academic prose*. Als Fazit aus der Studie von Biber et al. (2004) lässt sich festhalten, dass *Referential expressions* eher typisch für Schriftsprache sind, während *Stance expressions* und *Discourse organizers* eher typisch für mündlichen Sprachgebrauch sind.

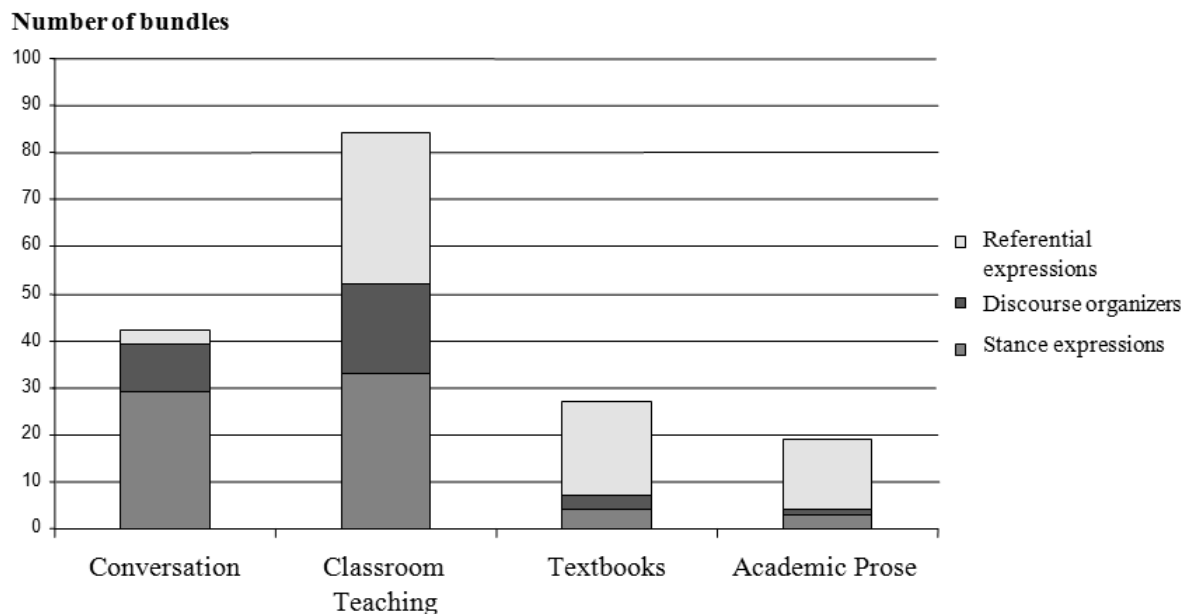


Abbildung 17: Verteilung von *Lexical Bundles* in funktionalen Kategorien (Biber et al. 2004: 397)

Vor diesem Hintergrund wird nun im Folgenden überprüft werden, inwieweit sich die vier Web-Genres in ihrer *4-gram*-Distribution gleichen. Dazu werden die 100 häufigsten *4-grams* jedes Subkorpus überprüft und in eine der drei Biberschen Kategorien (Biber et al. 2003, 2004) einsortiert.

6.2.3.2. Klassifikation der *4-grams* der vier Web-Genres

Grundsätzlich wird jedes *4-gram*, das in den vom *Automatic n-gram Analyser* generierten Listen aufgeführt wird, einer der drei funktionalen Kategorien zugeordnet. So werden beispielsweise sechs der häufigsten *4-grams* des Web-Genres Foren (*I would like to, if you want to, I would love to, to be able to, I don't want to, I don't know if*) in die Kategorie *Stance expressions* einsortiert, während *when it comes to* und *at the same time* der Kategorie *Discourse organizing expressions* zugeordnet werden (vgl. Tabelle 9).

Es muss jedoch betont werden, dass eine Vielzahl häufig auftretender *4-grams* aufgelistet wird, die sich keiner der drei Kategorien einwandfrei zuordnen lässt und die herausgefiltert werden muss. Dazu zählen unter anderem die sehr häufig vorkommenden Konstruktionen *had a nose job* (64 mal *pmw*) und *she looks like a* (57 mal *pmw*). Nur diejenigen *4-grams*, die sich einer der Kategorien zuordnen lassen, werden für die Klassifikation berücksichtigt. Aus

diesem Grund wurde solange nach kategorisierbaren *4-grams* in den einzelnen Listen gesucht, bis die Schwelle von 100 klassifizierten *4-grams* für jedes Subkorpus erreicht wurde.

Rang	<i>4-gram</i>	Frequenz
1	I would like to	120
2	if you want to	97
3	I would love to	79
4	to be able to	78
5	I don't want to	74
6	when it comes to	71
7	at the same time	64
8	had a nose job	64
9	I don't know if	62
10	she looks like a	57

Tabelle 9: Auszug aus der *4-gram* Liste des Web-Genres Foren (normalisiert *pmw*)

Insgesamt konnten etwa 64% aller *4-grams* des Web-ELF-Korpus keiner Kategorie zugeordnet werden, das heißt sie wurden herausgefiltert und fanden keine Berücksichtigung für die Kategorisierung der 100 häufigsten Muster jedes Web-Genres. Dabei zeigte sich jedoch, dass zwischen den Web-Genres eklatante Unterschiede bestanden. Während in Foren 37% aller *4-grams* keiner Kategorie zugeordnet werden konnte, lag der Anteil bei den anderen Web-Genres höher (Blogs 63%, kommerzielle Homepages 80%, Wikipedia-Artikel 51%). Insbesondere im Web-Genre kommerzielle Homepages wird dabei eine Vielzahl häufig auftretender *4-grams* verwendet, die sich keiner der drei Kategorien zuordnen lassen.⁵⁰

Die Analyse von Webtexten stellt aus unterschiedlichen Gründen eine Herausforderung dar. So sind abhängig vom Web-Genre mehr oder weniger wiederholende Textpassagen zu finden. Auf kommerziellen Homepages werden beispielsweise sehr oft bestimmte Formulierungen wiederholt, um Kernaussagen zu betonen und diese beim Rezipienten zu verfestigen. Foren wiederum zeichnen sich dadurch aus, dass bestimmte Textpassagen wieder aufgegriffen werden, um einen Bezug zu einem vorherigen oder laufenden Diskurs herstellen zu können. Diese Tatsache spiegelt sich ebenfalls in den Texten des Web-ELF-Korpus wider. Offenkundige idiosynkratische *4-grams* wurden für die Klassifikation ebenso wenig berücksichtigt wie *4-grams*, die sich keiner der drei Kategorien zweifelsfrei zuordnen ließen.

⁵⁰ Die *4-grams*, die sich nicht zuordnen ließen, konnten den Kategorien in keinsten Weise zugeordnet werden. Beispiele dafür sind *the United States and, has been edited last, had a tummy tuck* oder *of the eye and*.

6.2.3.3. Ergebnisse der Klassifikation

Die Ergebnisse der funktionalen Klassifikation werden in Tabelle 10 zusammenfassend dargestellt. Zum besseren Vergleich werden in Tabelle 11 die Resultate aus Biber et al. (2004) in der gleichen Darstellungsform wiedergegeben (vgl. auch Saage et al. 2012).

	Foren	Blogs	Kommerzielle Homepages	Wikipedia-Artikel
<i>Stance expressions</i>	54%	23%	29%	13%
<i>Discourse organizers</i>	16%	12%	13%	12%
<i>Referential expressions</i>	30%	66%	58%	75%
Total	100%	100%	100%	100%

Tabelle 10: Prozentuale Verteilung der *4-grams* entlang der funktionalen Kategorien im Web-ELF-Korpus

	<i>Conversation</i>	<i>Classroom teaching</i>	<i>Textbooks</i>	<i>Academic prose</i>
<i>Stance expressions</i>	69%	39%	15%	16%
<i>Discourse organizers</i>	24%	23%	11%	5%
<i>Referential expressions</i>	7%	38%	74%	79%
Total	100%	100%	100%	100%

Tabelle 11: Prozentuale Verteilung der *4-grams* entlang der funktionalen Kategorien in Biber et al. (2004)

Ein Vergleich der beiden Tabellen zeigt einige Parallelen auf. Unter anderem gehört die Mehrheit der *4-grams* im Web-Genre Foren (54%) der Kategorie *Stance expressions* an. Dies stellt eine Parallele zu den Registern *Conversation* und *Classroom teaching* von Biber et al. (2004) dar. Im Gegensatz zu *Conversation* (7%) werden in Foren (30%) jedoch vergleichsweise mehr *Referential expressions* verwendet. In den Web-Genres Blogs (66%), kommerzielle Homepages (58%) und Wikipedia-Artikeln (75%) werden vorrangig *Referential expressions* verwendet. Diesbezüglich sind die Unterschiede zwischen diesen Web-Genres und Foren hochsignifikant. Ein Vergleich mit Tabelle 11 verdeutlicht, dass auch in den beiden Registern *Textbooks* und *Academic prose* mehrheitlich *Referential expressions* verwendet werden. Das Web-Genre Wikipedia-Artikel gleicht dem Register *Textbooks* in seiner funktionalen Distribution beinahe völlig. Zur besseren Visualisierung der Ergebnisse dient Abbildung 18. Die Web-Genres werden dabei von den mündlichen und schriftlichen Registern *Conversation* und *Academic prose* eingerahmt.

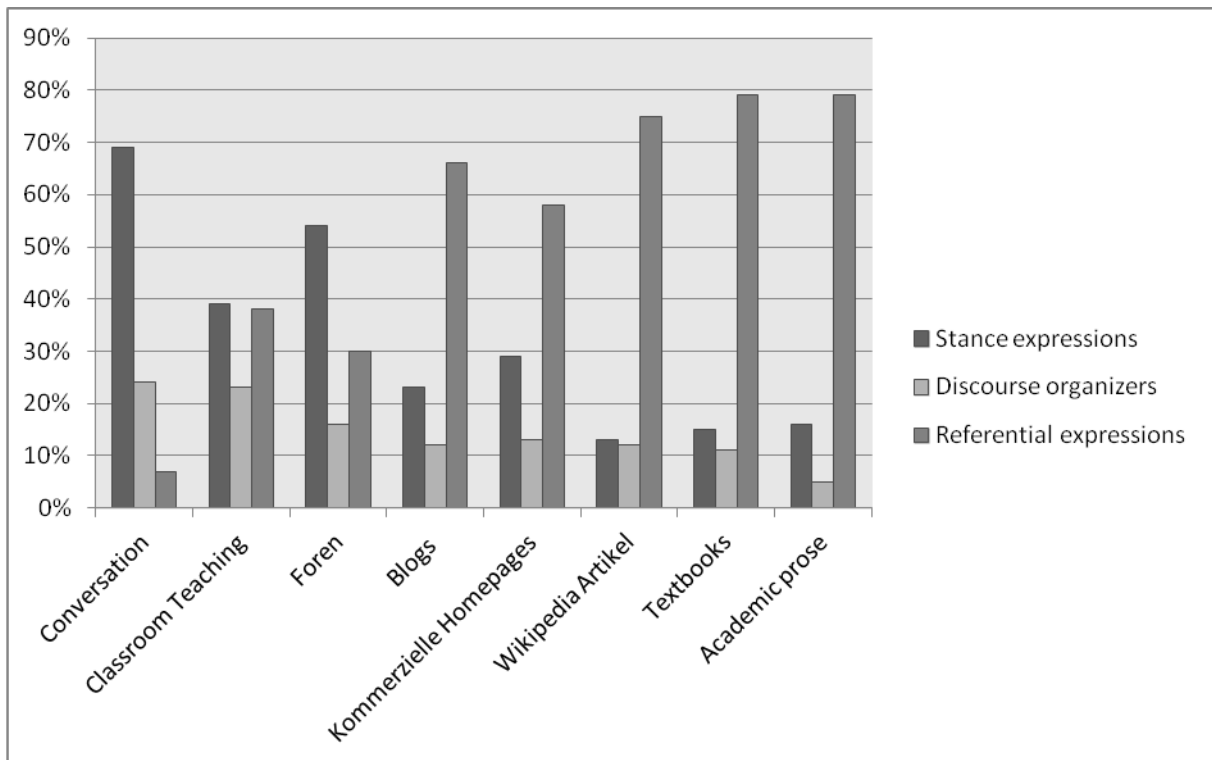


Abbildung 18: Zusammenfassung der funktionalen Klassifikation

Die Abbildung verdeutlicht, dass Web-Genres bezüglich der Verteilung der *4-grams* in funktionale Kategorien durchaus mündlichen und schriftlichen Registern ähneln. Mehr noch, es lassen sich einige Beobachtungen machen, die umso interessanter erscheinen, wenn die Ergebnisse der formalen Klassifikation mitberücksichtigt werden.

Ebenso wie für die formale Klassifikation lässt sich auch für die funktionale Klassifikation ein Kontinuum von Mündlichkeit / Schriftlichkeit formulieren. Der mündliche Pol zeichnet sich dabei durch eine hohe Anzahl an *Stance expressions* und eine sehr geringe Anzahl an *Referential expressions* aus. Der schriftliche Pol hingegen definiert sich durch eine hohe Anzahl an *Referential expressions* und eine geringe Anzahl an *Stance expressions*. Vor diesem Hintergrund zeigt sich, dass Forumstexte eher Parallelen zur mündlichen Sprachverwendung aufweisen, während Wikipedia Texte eher am schriftsprachlichen Pol des Mündlichkeits-/Schriftlichkeitskontinuums zu verorten sind. Kommerzielle Homepages und Blogs weisen sowohl Elemente mündlichen als auch schriftsprachlichen Sprachgebrauchs auf. Insgesamt enthalten sie jedoch mehr schriftsprachliche Elemente. Die Mehrheit der *4-grams* im Web-Genre kommerzielle Homepages (58%) gehört zwar der Kategorie *Referential expressions* an, es wird jedoch im Vergleich zu schriftsprachlichen Registern ebenfalls ein hoher Anteil an

Stance expressions (29%) verwendet. Blogs weisen mit 23% ein im Vergleich zu den schriftsprachlichen Registern höheres Maß an *Stance expressions* aus.

Die Ergebnisse der Klassifikation liefern einige klare Befunde, die im Folgenden weiter überprüft werden müssen. Aus diesem Grund wird für die folgenden Analysen ein besonderes Augenmerk auf einzelne *4-grams* in den Web-Genres gelegt. Es wird dabei überprüft, inwieweit es *4-grams* gibt, die in ähnlicher Häufigkeit ebenfalls in anderen Web-Genres verwendet werden. Auf diese Weise können Verbindungen zwischen den einzelnen Web-Genres aufgezeigt werden. Die Verwendung gemeinsamer *4-grams* zeigt, dass sich die Web-Genres funktional gleichen.

6.2.3.4. Eine funktionale Klassifikation gemeinsamer *4-grams* der vier Web-Genres

In ihrer Studie haben Biber et al. (2004) nach *4-grams* gesucht, die in mehreren Registern verwendet werden. Dabei stellten sie deutlich heraus, dass es eine Vielzahl von *4-grams* gibt, die in mehreren Registern verwendet werden. Tabelle 12 gibt die Gesamtzahl aller *4-gram-Types* an, die in zwei oder mehr Registern auftreten.

	<i>Conversation</i>	<i>Classroom teaching</i>	<i>Textbooks</i>	<i>Academic prose</i>
<i>Stance expressions</i>	29	33	4	3
<i>Discourse organizers</i>	10	19	3	1
<i>Referential expressions</i>	3	32	20	15

Tabelle 12: Distribution gemeinsamer *4-grams* in verschiedenen Funktionen (Biber et al. 2004: 396)

Es zeigt sich, dass das Register *Classroom teaching* 33 *Stance expressions* enthält, die ebenfalls in einem (oder mehreren) der drei Register *Conversation*, *Textbooks* oder *Academic prose* auftreten. Die Mehrheit der gemeinsamen *Stance expressions* teilt sich das Register *Classroom teaching* mit dem Register *Conversation*. Im Gegensatz dazu gibt es lediglich drei *Stance expressions*, die sowohl im Register *Academic prose* als auch in einem der anderen Register verwendet werden. Analog zu der Kategorie *Stance expressions* finden sich auch für die Kategorie *Discourse organizers* die meisten gemeinsamen *4-grams* in den Registern *Conversation* (10) und *Classroom teaching* (19). Gleichzeitig werden in den Registern *Textbooks* (3) und *Academic prose* (1) ebenfalls nur wenige gemeinsame *4-grams* verwendet. Die Kategorie *Referential expressions* hingegen divergiert in ihrer Distribution von den anderen Kate-

gorien, da sich nur wenige gemeinsame *4-grams* identifizieren lassen, die sowohl in *Conversation* als auch in einem der drei anderen Register häufig auftreten. Stattdessen existiert eine Vielzahl gemeinsamer *4-grams* in den Registern *Classroom teaching*, *Textbooks* und *Academic prose*.

Biber et al. (2004) grenzen die vier Register auf funktionaler Ebene voneinander ab. *Conversation* gilt als typisch für mündlichen Sprachgebrauch, die beiden Register *Textbooks* und *Academic prose* als typisch für schriftlichen Sprachgebrauch. Das Register *Classroom teaching* befindet sich zwischen diesen beiden Polen. Dies liegt darin begründet, dass dieses Register einerseits mündlich vorgetragen wird, andererseits eher monologisch und wissenschaftlich formuliert präsentiert wird.

Analog zu Biber et al. (2004) wurde auch für die vorliegende Studie eine Klassifikation gemeinsamer *4-grams* der Web-Genres durchgeführt. Grundsätzlich basiert die Kategorisierung auf Biber et al. (2004), die in ihrer Taxonomie zwischen den drei Kategorien *Stance expressions*, *Discourse organizers* und *Referential expressions* unterscheiden. Für den Vergleich gemeinsamer *4-grams* haben sie jedoch für jede der Kategorien weitere Unterkategorien definiert, um den einzelnen *4-grams* noch genauere Funktionen zuzuschreiben. Diese Unterkategorien wurden auch für den Web-Genre-Vergleich in der vorliegenden Untersuchung übernommen, die Taxonomie wurde jedoch modifiziert und erweitert durch die feingliedrigere Klassifikation von Simpson-Vlach / Ellis (2010). Im Folgenden werden die einzelnen Unterkategorien aufgeführt. In allen Fällen, in denen Kategorien ergänzt oder modifiziert wurden, wird dies genauer erläutert.

I. Die Kategorie *Referential expressions* wird in die Unterkategorien A. *Specification of attributes*, B. *Identification and focus*, C. *Contrast and comparison*, D. *Deictics and locatives* und E. *Vagueness markers* unterteilt.⁵¹

II. Die Oberkategorie *Stance expressions* wird in die Unterkategorien A. *Hedges*, B. *Epistemic stance*, C. *Obligation and directive*, D. *Ability and possibility*, E. *Evaluation* und F. *Intention / Volition* unterteilt.⁵²

⁵¹ Die Kategorisierung von Biber et al. (2004) wurde von Simpson-Vlach / Ellis (2010) um die Unterkategorie C. *Contrast and comparison* ergänzt.

⁵² Die Kategorien A. *Hedges* und E. *Evaluation* wurden von Biber et al. (2004) nicht explizit aufgeführt. Zudem wurden die beiden Kategorien *Desire* und *Intention / Prediction* in der Kategorie F. *Intention / Volition* zusammengefasst.

III. Die Oberkategorie *Discourse organizers* wird in die Unterkategorien A. *Metadiscourse and textual reference*, B. *Topic introduction and focus*, C. *Topic elaboration* und D. *Discourse markers* unterteilt.⁵³

Es muss ausdrücklich betont werden, dass einige der *4-grams* in den funktionalen Kategorien durchaus mehrere Funktionen erfüllen können und somit theoretisch in mehr als eine Kategorie einsortiert werden könnten. In jedem dieser Fälle wurde ihr Kontext überprüft, um sie in die funktionale Kategorie einzusortieren, in der sie am häufigsten verwendet wurden. Tabelle 13 liefert eine Übersicht über alle *4-grams*, die in zwei oder mehr Web-Genres verwendet werden.

	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
<i>Stance expressions</i>	7	9	9	5
<i>Discourse organizers</i>	6	5	6	4
<i>Referential expressions</i>	9	22	29	29

Tabelle 13: Distribution gemeinsamer *4-grams* der Web-Genres

Tabelle 13 veranschaulicht, dass sich Web-Genres in ihrer Verteilung gemeinsamer *4-grams* entscheidend von den Registern in Biber et al. (vgl. Tabelle 12) unterscheiden. Insgesamt werden weniger gemeinsame *4-gram-Types* in den Web-Genres verwendet. Lediglich in der Kategorie *Referential expressions* ist eine Vielzahl gemeinsamer *Types* zu finden, die sowohl in den Web-Genres Blogs, kommerzielle Homepages und Wikipedia-Artikel verwendet werden. In den beiden anderen Kategorien werden hingegen nur sehr wenige gemeinsame *Types* verwendet. In diesem Punkt unterscheidet sich das Web-Genre Foren von den mündlichen Registern *Conversation* und *Classroom teaching*. Die geringe Menge gemeinsamer *4-grams* ist der Tatsache geschuldet, dass in den anderen Web-Genres nur sehr wenige typisch mündliche *4-grams* verwendet werden und es somit nur sehr geringe Überschneidungen mit dem Web-Genre Foren gibt. Dieses Resultat verdeutlicht somit noch einmal, dass Foren deutlich von den anderen Web-Genres abgegrenzt werden müssen.

Im Folgenden möchte ich nun einen genaueren Blick auf die einzelnen *4-gram-Types* richten. Dies ermöglicht es, im Detail zu überprüfen, welche Web-Genres eher gemeinsame *4-gram-Types* verwenden und sich somit in ihrer funktionalen Verwendung ähneln. Dies

⁵³ Die beiden Unterkategorien A. *Metadiscourse and textual reference* und D. *Discourse markers* wurden neu hinzugefügt. Zusätzlich wurde die Kategorie B. *Topic introduction and focus* um den Aspekt *Cause and effect* ergänzt.

basiert auf der Annahme, dass Web-Genres funktional enger miteinander verbunden sind, je mehr gemeinsame *4-gram-Types* sie verwenden. Um die Nähe / Distanz der Web-Genres zueinander zu bestimmen, wird verglichen, wie häufig ein gemeinsames *4-gram* zwischen zwei Web-Genres verwendet wird und diese Zahl in Verhältnis gesetzt mit der Zahl an *4-grams*, die das Web-Genre mit allen anderen Web-Genres teilt. Bei der Berechnung der Gesamtzahl wird berücksichtigt, dass manche *4-grams* in mehr als zwei Web-Genres auftreten.

Tabelle 14 gibt zunächst die Ergebnisse für die Kategorie *Stance expressions* wieder. Für jedes *4-gram* ist zudem angegeben, wie häufig es in den einzelnen Web-Genres auftritt (*pmw*).

Web-Genre	Foren	Blogs	komm. Homepages	Wikipedia-Artikel
I. Stance Expressions				
A. Hedges				
you may want to		33	23	
B. Epistemic stance				
C. Obligation and directive				
you will have to	30	28		
you are going to	26	32		
D. Ability and possibility				
to be able to	78	23	39	19
will be able to	41	56	109	
should be able to	24		62	
can be used to		37	59	75
it is possible to			89	22
can also be used			26	43
E. Evaluation				
it is important to		70	118	30
F. Intention / volition				
I am going to	50	37		
I just want to	41	23	23	

Tabelle 14: Gemeinsame *4-grams* für die Kategorie *Stance expressions*

Die Tabelle lässt bereits eine Tendenz bezüglich der Gruppierung der Web-Genres erkennen. Foren und Blogs haben 6 gemeinsame *4-grams* (zum Vergleich Foren – kommerzielle Homepages 3; Foren – Wikipedia-Artikel 1). Dies bedeutet, dass sich Foren 60% aller *4-grams* mit Blogs teilen. Diese Tatsache lässt bereits vermuten, dass Foren und Blogs sich konzeptionell ähnlicher sind als beispielsweise Foren und Wikipedia-Artikel. Gleichermaßen nutzen Wiki-

pedia-Artikel und kommerzielle Homepages 5 gemeinsame *4-grams*. (zum Vergleich Wikipedia-Artikel – Foren 1; Wikipedia-Artikel – Blogs 3). Blogs und Kommerzielle Homepages teilen sich 5 gemeinsame *4-grams*.

Die Überprüfung der Unterkategorien zeigt, dass sich insbesondere in der Kategorie D. *Ability and possibility* viele gemeinsame *4-grams* nachweisen lassen. In der Kategorie B. *Epistemic stance*, die stereotypisch für mündlichen Sprachgebrauch ist, gibt es hingegen kein einziges gemeinsames *4-gram*-Muster. Obwohl eine Vielzahl der *4-grams* im Web-Genre Foren *Epistemic Stance expressions* darstellen, werden diese in den anderen Web-Genres gar nicht oder nur sehr selten verwendet.

Als nächstes gilt es, zu überprüfen, inwieweit die Tendenzen, die sich für die erste Kategorie abzeichnen, für die zweite Kategorie bestätigt werden können. In Tabelle 15 werden die Ergebnisse für die Kategorie *Discourse organizers* aufgeführt.

Web-Genre	Foren	Blogs	komm. Homepages	Wikipedia-Artikel
II. Discourse organising expressions				
A. Metadiscourse and textual reference				
B. Topic introduction and focus				
if you have any	49		59	
in the first place	38	23		
if you have a	37		37	
C. Topic elaboration				
when it comes to	71	75	35	
for the first time	33	28		25
as a result of		37	95	101
in most cases the			25	18
D. Discourse markers				
at the same time	64	56	160	41

Tabelle 15: Gemeinsame *4-grams* für die Kategorie *Discourse organizers*

Ein Vergleich mit Tabelle 14 veranschaulicht, dass es durchaus Distributionsparallelen gibt. Foren und Blogs haben 4 gemeinsame *4-grams* (44%) (zum Vergleich Foren – kommerzielle Homepages 4; Foren – Wikipedia-Artikel 1). Wikipedia-Artikel und kommerzielle Homepages teilen sich 3 gemeinsame *4-grams* (42%) (zum Vergleich Wikipedia-Artikel – Foren 1; Wikipedia-Artikel – Blogs 3). Blogs und kommerzielle Homepages teilen sich 3 gemeinsame *4-grams* (30%). Dieses Resultat lässt vermuten, dass sich Blogs, kommerzielle Homepages

und Wikipedia-Artikel relativ ähneln, während Web- Foren und Wikipedia-Artikel (11%) sich funktional kaum gleichen.

Insbesondere in den Unterkategorien B. *Topic introduction and focus* und C. *Topic elaboration* befinden sich fast alle der identifizierten gemeinsamen *4-grams*. In der Kategorie A. *Metadiscourse and textual reference* hingegen lassen sich keine gemeinsamen *4-grams* ausmachen. Die Unterkategorie D. *Discourse markers* stellt die einzige Kategorie dar, in der sich ein *4-gram* befindet, das in allen Web-Genres gleichermaßen auftritt. Dabei fällt auf, dass das *4-gram at the same time* im Web-Genre kommerzielle Homepages überproportional oft vertreten ist.

Zuletzt wird das Resultat für die Kategorie *Referential expressions* präsentiert (vgl. Tabelle 16). Es fällt zunächst auf, dass diese Kategorie, im Vergleich zu den beiden anderen, eine wesentlich höhere Anzahl gemeinsamer *4-grams* enthält. Des Weiteren zeichnet sich im Hinblick auf die funktionale Ähnlichkeit ein etwas differenzierteres Bild ab als bei den beiden vorherigen Kategorien. Die Mehrheit aller *4-grams* teilen sich Foren mit kommerziellen Homepages (8), allerdings dicht gefolgt von den Web-Genres Blogs (7) und Wikipedia-Artikel (6). Bei genauerer Betrachtung wird dabei ersichtlich, dass vier *4-grams* im Web-Genre Foren existieren, die in allen Web-Genres verwendet werden. Auch für diese funktionale Kategorie zeigt sich, dass Foren die wenigsten *4-grams* mit dem Web-Genre Wikipedia-Artikel teilen. Wenn jedoch die Gesamtzahl aller identifizierten *4-grams* für die Kategorie *Referential expressions* berücksichtigt wird, wird deutlich, dass die Zahl aller *4-grams*, die sich Foren mit den anderen Web-Genres teilen, insgesamt verhältnismäßig gering ist. Im Gegensatz zum Web-Genre Foren ergibt sich für das Web-Genre Wikipedia-Artikel ein wesentlich klareres Bild. Es gibt 24 *4-grams*, die diese mit dem Web-Genre kommerzielle Homepages teilen (55%) (zum Vergleich Wikipedia-Artikel – Foren 6; Wikipedia – Artikel-Blogs 14). Blogs und kommerzielle Homepages teilen sich 16 *4-grams*. Es fällt auf, dass die große Mehrheit aller gemeinsamen *4-grams* der Unterkategorie A. *Specifications of attributes* angehört.

Web-Genre	Foren	Blogs	komm. Homepages	Wikipedia- Artikel
III. Referential expressions				
A. Specifications of attributes				
a lot of people	53	18		
the end of the	50		37	47
the rest of the	43		43	38
one of the most	38	132	158	76
one of the best	33	23	23	
in the field of		80	55	43
the surface of the		61	26	36
of the most common		56		23
the shape of the		42	35	26
a wide range of		33	23	
in the form of		33	35	34
a small amount of		28	26	22
the quality of the		28	32	
a member of the		23	43	31
the extent of the		23	66	
a wide variety of		18	28	20
are a number of		18		18
the size of the			91	48
the area of the			61	26
the back of the			53	36
with the help of			43	19
with the use of			34	31
in the treatment of			28	76
on the type of			28	20
over a period of			25	25
the head of the			23	36
the nature of the			23	20
B. Identification and focus				
is one of the	54	151	154	90
is the most common		33		50
in some cases the		28		33
C. Contrast and comparison				
on the other hand	40	33	32	20
as well as the	29	80	100	66
as well as a			26	25
D. Deictics and locatives				
in the United States	22	358	59	
at the age of		37		25
E. Vagueness markers				

Tabelle 16: Gemeinsame 4-grams für die Kategorie *Referential expressions*

Der Vergleich der einzelnen Web-Genres hat bereits einen ersten Eindruck vermittelt, welche Web-Genres sich funktional gleichen und welche von ihnen funktional verschieden sind. Um diesen Eindruck zu bestätigen, wurde ein Chi-Quadrat Test nach Pearson durchgeführt, um zu überprüfen, ob die Gemeinsamkeiten und Unterschiede zwischen den einzelnen Web-Genres statistisch signifikant sind. In Tabelle 17 werden dazu die Korrelationen der einzelnen Web-Genres zueinander und die Signifikanz dieser Korrelationen abgebildet.

Es zeigt sich, dass es kaum statistisch signifikante Ergebnisse gibt. Dies kann darin begründet sein, dass die Gesamtzahl der Vorkommen (55) zu gering sein könnte, um klare Unterschiede zwischen den Web-Genres aufzuzeigen. Trotz dieser Einschränkung ermöglicht die Übersicht dennoch graduelle Unterschiede zwischen den Web-Genres zu veranschaulichen (vgl. Tabelle 17).

Korrelationen					
		Foren	Blogs	komm. Homepages	Wikipedia-Artikel
Foren	Korrelation nach Pearson	1	,203	-,148	-,421 ^{**54}
	Signifikanz (2-seitig)		,137	,279	,001
	N	55	55	55	55
Blogs	Korrelation nach Pearson	,203	1	-,363 ^{**}	-,206
	Signifikanz (2-seitig)	,137		,006	,131
	N	55	55	55	55
Kommerzielle Homepages	Korrelation nach Pearson	-,148	-,363 ^{**}	1	,115
	Signifikanz (2-seitig)	,279	,006		,404
	N	55	55	55	55
Wikipedia-Artikel	Korrelation nach Pearson	-,421 ^{**}	-,206	,115	1
	Signifikanz (2-seitig)	,001	,131	,404	
	N	55	55	55	55

Tabelle 17: Statistische Auswertung der funktionalen Ähnlichkeit der Web-Genres

Es fällt auf, dass lediglich zwischen Foren und Wikipedia-Artikeln ein hochsignifikanter Unterschied besteht (Korrelation -,421/p=0.001). Alle anderen Ergebnisse zeigen tendenziell negative oder positive Korrelationen, diese sind jedoch nicht signifikant.

⁵⁴ **. Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Sowohl die formale als auch die funktionale Analyse haben bestätigt, dass sich Foren am mündlichen Ende des Mündlichkeits- / Schriftlichkeitskontinuums befinden. Die vorliegende Analyse zeigt, dass sich Foren und Blogs ähneln. Es zeigt sich ebenso, dass Wikipedia-Artikel den schriftlichen Gegenpol zu Foren bilden, da sie kaum gemeinsame *4-grams* mit Foren teilen. Stattdessen ähneln Wikipedia-Artikel kommerziellen Homepages. Sowohl Blogs als auch kommerzielle Homepages befinden sich zwischen den beiden Extremen. Allerdings können sie tendenziell eher dem schriftlichen Pol zugeordnet werden, da die große Mehrheit aller von ihnen geteilten *4-gram-Types* typisch schriftlichen Kategorien zugewiesen werden kann (*Referential expressions*). Dennoch befinden sich kommerzielle Homepages auf dem Kontinuum im Vergleich zu Blogs etwas näher am schriftlichen Pol, da sie sich mehr gemeinsame *4-grams* mit dem Web-Genre Wikipedia-Artikel teilen.

Die Studien haben bestätigt, dass, neben den graduellen Unterschieden auf formaler Ebene, auch graduelle Unterschiede auf funktionaler Ebene zwischen den verschiedenen Web-Genres existieren. Anhand der funktionalen Analyse nach Biber et al. (2004) und dem Abgleich von gemeinsam verwendeten *4-grams* wurde dies systematisch dargelegt. Als nächstes wird nun im Detail dargelegt, inwieweit sich Unterschiede nachweisen lassen, die sowohl Form als auch Funktion betreffen.

6.2.3.5. Verbindungen von Form und Funktion

Im Folgenden wird die formale Struktur der *4-grams* innerhalb der Kategorien überprüft und mit Biber et al. (2004) verglichen. Biber et al. (2004: 397f.) fanden in ihrer Arbeit heraus, dass sich eine sehr starke Verbindung zwischen funktionalen Kategorien und den strukturellen Eigenschaften innerhalb dieser Kategorien nachweisen lässt (vgl. Abbildung 19). Demnach besteht die Mehrheit aller *Stance expressions* aus Nebensatzfragmenten (*Dependent clause bundles*), während die Mehrzahl der *Referential expressions* aus Nominalphrasen (*noun phrases*) und Präpositionalphrasen-Fragmenten (*NP / PP-based bundles*) besteht. Die gesondert aufgeführte Unterkategorie *Stance-prediction* beinhaltet ausschließlich Verbalphrasen-Fragmente (*VP-based bundles*), sehr oft in Kombination mit der Semimodal-Konstruktion *be going to*. In der Kategorie *Discourse organizers* hingegen werden alle Strukturtypen verwendet.

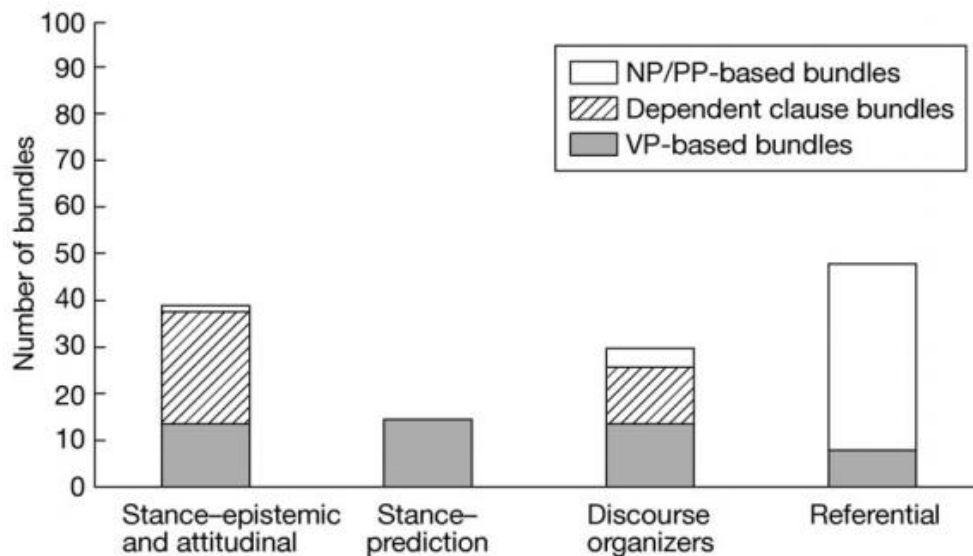


Abbildung 19: Interaktion struktureller und funktionaler Kategorien (Biber et al. 2004: 398)

Biber et al. (2004: 398) verweisen darauf, dass die strukturellen Muster in enger Verbindung zu den untersuchten Registern stehen. Sie haben in ihrer Studie untersucht, welche strukturellen Muster in jedem ihrer vier Register bevorzugt werden. Sie kamen zum Ergebnis, dass im Register *Conversation*, das als typisch mündlicher Sprachgebrauch gilt, mehrheitlich *VP-based bundles* und auch *Dependent clause bundles* verwendet werden, während im Register *Academic prose*, das als typisch schriftsprachlicher Sprachgebrauch gilt, mehrheitlich *NP / PP-based bundles* verwendet werden. Das Register *Classroom teaching* ähnelt in seiner Struktur dem Register *Conversation*, allerdings werden zu referentiellen Zwecken auch viele *NP / PP-based bundles* verwendet. Diese Beobachtungen veranlassen Biber et al. (2004: 398) zur Schlussfolgerung, dass bei Mehrworteinheiten eine direkte Verbindung zwischen Form und Funktion besteht. Als Beispiel nennen sie, dass *Complement clause-based bundles* verwendet werden, um Einstellungen (*stance*) auszudrücken, da diese in der englischen Sprache eines der produktivsten grammatischen Mittel darstellen, um dies zu tun (vgl. Biber et al. 1999: 965ff). Folglich ergibt es Sinn, dass eine Vielzahl der am häufigsten genutzten Mehrworteinheiten in ihrer Form als *Stance expression* fixiert wird. In ähnlicher Weise stellen Nominalphrasen (*noun phrases*) / Präpositionalphrasen (*prepositional phrases*) die produktivsten grammatischen Mittel dar, um referentielle Äußerungen zu tätigen. Daher bilden sich ebenfalls Mehrworteinheiten aus, die in ihrer Form fest fixiert sind. Biber et al. (2004: 398) schlussfolgern aus diesen Ergebnissen, dass eine komplexe Interaktion zwischen

struktureller Form, Diskursfunktion und den typischen Verwendungszwecken und situativen Charakteristika von Registern stattfindet.

Analog zu Biber et al. (2004) wird nun überprüft, inwieweit sich die einzelnen Web-Genres bezüglich ihrer strukturellen Eigenschaften unterscheiden. Zu diesem Zweck werden die 100 häufigsten *4-grams* aller Web-Genres diesbezüglich überprüft. Die Ergebnisse werden in Abbildung 20 wiedergegeben.

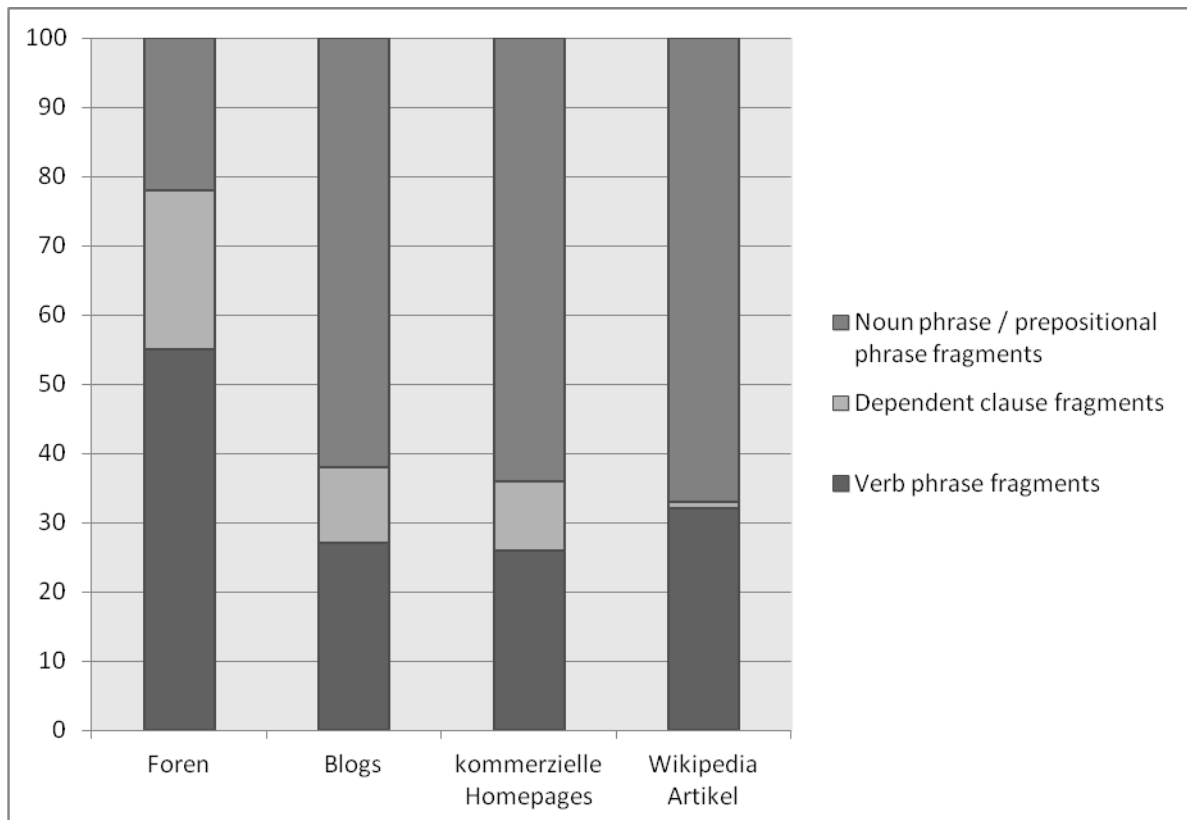


Abbildung 20: Ergebnisse der Analyse von Form und Funktion

Das Ergebnis der Analyse von Form und Funktion bestätigt die Vermutung, dass sich die einzelnen Web-Genres sowohl auf formaler als auch auf funktionaler Ebene voneinander unterscheiden. Es zeigt sich somit zum wiederholten Mal, dass Foren eine Sonderstellung einnehmen. 55% aller *4-grams* im Web-Genre Foren bestehen aus *Verb phrase fragments*, 23% aus *Dependent clause fragments* und 22% aus *Noun phrase / Prepositional phrase fragments*. Demgegenüber stellt sich die Distribution in den drei weiteren Web-Genres unterschiedlich dar. Blogs und kommerzielle Homepages gleichen sich in ihrer Verteilung fast vollständig. Ebenso zeichnen sich zwischen Wikipedia-Artikeln und Blogs / kommerziellen Homepages Parallelen ab. Ein Großteil der *4-grams* in diesen drei Web-Genres besteht aus *Noun phrase /*

Prepositional phrase fragments (Blogs 62%, kommerzielle Homepages 64%, Wikipedia-Artikel 67%). *Verb phrase fragments* machen hingegen nur einen geringeren Anteil aus (Blogs 27%, kommerzielle Homepages 26%, Wikipedia-Artikel 32%). Die Unterschiede zwischen diesen drei Web-Genres sind nicht signifikant. Ebenso weist sich der Anteil der *Dependent clause fragments* im Vergleich zu Foren geringer aus (Blogs 11%, kommerzielle Homepages 10%, Wikipedia-Artikel 1%). Es fällt dabei auf, dass der Anteil an *Dependent clause fragments* für das Web-Genre Wikipedia-Artikel mit Abstand am geringsten ist. Der Unterschied zwischen Blogs / kommerziellen Homepages und Wikipedia-Artikeln ist diesbezüglich signifikant ($p=0.01$).

Als Fazit bleibt festzuhalten, dass die Analyse von Form und Funktion ebenfalls Unterschiede zwischen den einzelnen Web-Genres offenkundig gemacht hat. Es lässt sich wiederum ein Kontinuum ableiten, an dessen oppositionellen Enden sich Foren und Wikipedia-Artikel gegenüberstehen. Ein Vergleich mit den Registern von Biber et al. (2004) zeigt, dass Foren das einzige Web-Genre sind, das in seiner Struktur mündlicher Sprache ähnelt, während die drei anderen Web-Genres in ihrer Struktur typischer Schriftsprache ähneln. Die verwendeten Strukturen im Web-Genre Foren gleichen denen des mündlichen Registers *Conversation*. Blogs und kommerzielle Homepages gleichen sich bezüglich der Verwendung von *4-grams* fast vollständig und befinden sich tendenziell am schriftsprachlichen Pol. Wikipedia-Artikel hingegen weisen eine *4-gram*-Struktur auf, die der Struktur schriftsprachlicher Register sehr stark ähnelt.

6.2.4. Zusammenfassung

Die vorliegenden Studien haben deutlich aufgezeigt, dass sich auf quantitativer Ebene große Unterschiede zwischen den einzelnen Web-Genres nachweisen lassen. Dabei offenbart sich, dass die untersuchten Web-Genres sowohl auf formaler als auch auf funktionaler Ebene voneinander divergieren. Das Ziel jeder der Analysen war es, graduelle Unterschiede zwischen den einzelnen Web-Genres aufzuzeigen. Durch einen Vergleich mit typischen mündlichen und typisch schriftlichen Registern sollte aufgezeigt werden, dass Web-Genres sich mit Bezug auf ihre formalen und funktionalen Eigenschaften auf einem Kontinuum von Mündlichkeit / Schriftlichkeit voneinander abgrenzen lassen. Die Ergebnisse der einzelnen Studien zeichneten diesbezüglich ein klares Bild. Foren sind das einzige Web-Genre, das viele Elemente konzeptioneller Mündlichkeit aufweist und das sehr nah am Pol der konzeptionellen Mündlichkeit verortet werden kann. Wikipedia-Artikel hingegen gleichen in ihrer Form und Funktion sehr

stark schriftsprachlichen Registern und stehen sehr nah am Pol der konzeptionellen Schriftlichkeit. Blogs und kommerzielle Homepages gleichen sich stark in Form und Funktion und weisen viele Elemente konzeptioneller Schriftlichkeit auf. Sie können näher am Pol der konzeptionellen Schriftlichkeit verortet werden. Im nachfolgenden Kapitel soll nun überprüft werden, inwiefern auch im Hinblick auf einzelne *4-gram*-Muster Unterschiede zwischen den vier Web-Genres offengelegt werden können.

7. Qualitative Analyse

7.1. Häufigkeitsliste 4-grams

Im vorliegenden Kapitel soll nun der Blick auf einzelne *4-gram-Types* gerichtet werden. Im Erkenntnisinteresse steht dabei die Frage, inwieweit gemeinsame *4-gram-Muster* existieren, die unterschiedlich konnotiert sind. Dazu werden einzelne *4-gram-Types* ausgewählt und ihre Verwendung im Kontext der jeweiligen Web-Genres überprüft. Als Ausgangspunkt für die Analyse wird für jedes Web-Genre eine Tabelle erstellt, in der die häufigsten *4-grams* aufgelistet werden.

Im Folgenden wird ein Auszug jedes Web-Genres abgebildet, in dem die 20 häufigsten *4-gram-Types* aufgelistet werden. Diese geben bereits einen ersten Eindruck, welche Muster in den einzelnen Web-Genres präferiert werden. Eine Liste der 100 häufigsten *4-grams* jedes Web-Genres findet sich in Appendix A. In Tabelle 18 werden die zwanzig häufigsten *4-grams* für den Bereich Foren aufgelistet.

Rang	4-gram	Frequenz	Frequenz <i>pmw</i>
1.	I would like to	142	120
2.	if you want to	115	97
3.	I would love to	94	79
4.	to be able to	93	78
5.	I don't want to	88	74
6.	when it comes to	84	71
7.	at the same time	76	64
8.	had a nose job	76	64
9.	I don't know if	74	62
10.	she looks like a	68	57
11.	at the end of	64	54
12.	is one of the	64	54
13.	a lot of people	63	53
14.	to get rid of	62	52
15.	I am going to	62	62
16.	I had my surgery	60	50
17.	the end of the	59	50
18.	if you have any	59	50
19.	I don't know what	55	46
20.	I can tell you	54	45

Tabelle 18: Die zwanzig häufigsten *4-gram-Types* des Web-Genres Foren

Es ist auffällig, dass ein Großteil aller *4-gram-Types* formal dem Muster personal pronoun + lexical verb phrase (+ complement-clause fragment) entspricht. Dementsprechend lässt sich auf der funktionalen Ebene die Mehrheit aller *4-gram-Types* der Kategorie *Stance Expressions* zuordnen. Somit sind viele der *4-gram-Types* Manifestationen konzeptioneller Mündlichkeit. In fast allen dieser *4-grams* wird die erste Person Singular als Personalpronomen verwendet (unter anderem 1. *I would like to*, 3. *I would love to*, 5. *I don't want to*, 9. *I don't know if*). Es zeigt sich jedoch auch, dass einige häufig auftretende *4-gram-Types* auf die Textsorte *Medical Surgery / Medical Tourism* zurückzuführen sind, so zum Beispiel 8. *had a nose job* und 16. *I had my surgery*.

Insgesamt existieren 1127 *4-gram-Types* für das Web-Genre Foren mit einem Mindestvorkommen von 10. Von ihnen tritt nur ein Bruchteil sehr häufig auf. So gibt es nur 28 *4-grams*, die häufiger als 40 mal *pmw* vorkommen (vgl. Biber et al. 1999). In den anderen Web-Genres befinden sich weniger *4-gram-Types*. Gleichzeitig beinhalten diese jedoch eine größere Anzahl von *4-gram-Types*, die mindestens 40 mal *pmw* auftreten: Blogs 44 (bei 222 insgesamt), kommerzielle Homepages 70 (bei 936 insgesamt) und Wikipedia-Artikel 41 (bei 389 insgesamt).

In Tabelle 19 werden die zwanzig häufigsten *4-gram-Types* für das Web-Genre Blogs aufgeführt. Es ist offenkundig, dass sich diese Liste stark von der Liste des Web-Genres Foren unterscheidet. Die Mehrheit von ihnen gehört entweder der Kategorie noun-phrase with post-modifier fragment oder der Kategorie preposition + noun-phrase fragment an. Diese sind typisch schriftsprachliche *4-gram-Verwendungen*. Es finden sich hingegen nur sehr wenige typisch mündliche *4-gram-Verwendungen*, die sehr häufig auftreten (beispielsweise 2. *is one of the*). Ebenso wie im Web-Genre Foren gibt es auch *4-grams*, die der Spezifität des Themas geschuldet sind (10. *the medical tourism industry*, 11. *medical tourism in India*). Insgesamt ist die Zahl der fachspezifischen *4-gram-Types* unter den 100 häufigsten *4-grams* höher als beim Web-Genre Foren. Dies gilt ebenso für die Web-Genres kommerzielle Homepages und Wikipedia-Artikel. Funktional lässt sich ein Großteil aller häufig auftretenden *4-gram-Types* der Kategorie *Referential expressions* zuordnen. Auffällig ist zudem die sehr häufige Verwendung des referentiellen *4-grams in the United States*. Generell werden sehr viele räumliche Bezüge hergestellt (1. *in the United States*, 5. *at the university of*, 11. *medical tourism in India*, 12. *the United States and*). Dies ist der Tatsache geschuldet, dass ein Großteil der Blogs thematisch darauf fokussiert ist, Informationen über medizinische Eingriffe an verschiedenen Orten der Welt bereitzustellen.

Rang	4-gram	Frequenz	Frequenz <i>pmw</i>
1.	in the United States	76	358
2.	is one of the	32	151
3.	one of the most	28	132
4.	at the university of	21	99
5.	in the field of	17	80
6.	as well as the	17	80
7.	when it comes to	16	75
8.	a fraction of the	15	70
9.	it is important to	15	70
10.	the medical tourism industry	14	66
11.	medical tourism in India	13	61
12.	the United States and	13	61
13.	on top of the	13	61
14.	the surface of the	13	61
15.	fraction of the cost	13	61
16.	of the most common	12	56
17.	will be able to	12	56
18.	of the inner ear	12	56
19.	at the same time	12	56
20.	to make sure that	12	56

Tabelle 19: Die zwanzig häufigsten *4-gram-Types* des Web-Genre Blogs

Analog zu den vorherigen Tabellen spiegelt Tabelle 20 die Ergebnisse für den Bereich kommerzielle Homepages wider. Auch für dieses Web-Genre gilt, dass die Mehrheit aller *4-gram-Types* entweder der Kategorie *noun-phrase with post-modifier fragment* oder der Kategorie *preposition + noun-phrase fragment* angehört. Als Beispiel kann das am häufigsten auftretende *4-gram at the same time* genannt werden. Auffällig ist, dass einige der am häufigsten auftretenden *4-grams* einen temporalen Bezug aufweisen (9. *for a few days*, 14. *on the day of*, 17. *a day or two*). Die Annahme, dass in Blogs vor allem *4-grams* verwendet werden, die Personalpronomen beinhalten, konnte jedoch nicht bestätigt werden. Es zeigt sich somit, dass die Autoren durch die Fokussierung auf das Thema *Medical Surgery / Medical Tourism* in den Hintergrund treten. Ebenso wie für Blogs lässt sich die Mehrheit aller häufig auftretenden *4-gram-Types* der funktionalen Kategorie *Referential expressions* zuordnen.

Rang	4-gram	Frequenz	Frequenz <i>pmw</i>
1.	at the same time	89	160
2.	one of the most	88	159
3.	is one of the	86	154
4.	the best candidates for	66	118
5.	it is important to	66	118
6.	will be able to	61	109
7.	as well as the	56	100
8.	as a result of	53	95
9.	for a few days	52	93
10.	the size of the	51	91
11.	it is possible to	50	89
12.	surgery in Cape Town	41	73
13.	you will be able	40	71
14.	on the day of	39	70
15.	is a surgical procedure	39	70
16.	plastic and reconstructive surgery	38	68
17.	a day or two	37	66
18.	the extent of the	37	66
19.	on an outpatient basis	37	66
20.	on the extent of	37	66

Tabelle 20: Die zwanzig häufigsten 4-gram-Types des Web-Genres kommerzielle Homepages

Zuletzt führt Tabelle 21 die am häufigsten auftretenden 4-gram-Types für das Web-Genre Wikipedia-Artikel auf. Es zeichnen sich auch für dieses Web-Genre Parallelen zu Blogs und kommerziellen Homepages ab, da ebenfalls die Mehrheit aller 4-gram-Types entweder der Kategorie *noun-phrase with post-modifier fragment* oder der Kategorie *preposition + noun-phrase fragment* angehört. Auffällig ist die sehr hohe Anzahl an Vorkommen für das häufigste 4-gram *in the United States*. Unter den 20 häufigsten Mustern lassen sich weitere 4-grams mit räumlichem Bezug identifizieren (6. *at the university of*, 16. *in the United Kingdom*). Im Gegensatz zum Web-Genre kommerzielle Homepages gibt es keine 4-grams mit temporalem Bezug. Auffällig ist zudem die häufige Verwendung von 4-grams, die sich auf den Operationsprozess selbst und den Behandlungsprozess bei Operationen beziehen (2. *is a surgical procedure*, 8. *for the treatment of*, 9. *in the treatment of*, 12. *operations and other procedures*, 18. *and other procedures on*, 19. *other procedures on the*). Funktional lässt sich, analog zu Blogs und kommerziellen Homepages, ein Großteil aller häufig auftretenden 4-grams der Kategorie *Referential expressions* zuordnen.

Rang	4-gram	Frequenz	Frequenz <i>pmw</i>
1.	in the United States	230	319
2.	is a surgical procedure	87	120
3.	the Royal College of	86	119
4.	as a result of	73	101
5.	is one of the	65	90
6.	at the university of	60	83
7.	Royal College of Surgeons	59	82
8.	for the treatment of	59	82
9.	in the treatment of	55	76
10.	one of the most	55	76
11.	in the case of	55	76
12.	operations and other procedures	55	76
13.	can be used to	54	75
14.	of the Royal College	52	72
15.	as well as the	48	66
16.	at the time of	44	61
17.	in the United Kingdom	43	59
18.	has been shown to	42	58
19.	and other procedures on	42	58
20.	other procedures on the	42	58

Tabelle 21: Die zwanzig häufigsten *4-gram*-Types des Web-Genres Wikipedia-Artikel

Die Tabellen 18-21 konnten bereits einen ersten Eindruck vermitteln, welche Ausprägungen von *4-gram*-Types am häufigsten in den einzelnen Web-Genres auftreten. Die Häufigkeitslisten dienen dabei als erste Anhaltspunkte für auffällige und untypische Verwendungsweisen. Für die folgenden Analysen werden *4-grams* identifiziert, die in jedem der vier Web-Genres häufig auftreten. Die so gefundenen *4-grams* werden dann in ihrem weiteren Kontext betrachtet. Eines der Muster, das diese Voraussetzung erfüllt, ist *at the end of*.

7.2. Analyse von *at the end of*

Die Verwendung des *4-grams at the end of* wurde bereits von Forchini / Murphy (2008) genauer betrachtet. Sie haben dieses *4-gram* näher untersucht, da es sowohl im *British National Corpus* als auch in ihrem selbst erstellten thematisch spezialisierten *Financial Times Corpus*, das aus etwa 35 Millionen Wörtern besteht, sehr häufig auftritt (*BNC* 93,41 mal *pmw*; *FTC* 100,69 *pmw*). Syntaktisch kann die Form *at the end of* auf unterschiedliche Weise betrachtet werden. Einerseits stellt sie als "recurrent string of orthographic word-forms" (Stubbs 2007: 90) ein *4-gram* dar. Andererseits kann sie als Realisierung des *Phrase-Frames at the X of* ver-

standen werden (vgl. Fletcher 2003-2012), in welchem der variable Slot (X) durch *end* ausgefüllt wird. Forchini / Murphy (2008: 255) haben eine semantische Analyse von *at the end of* im *Financial Times Corpus* durchgeführt und festgestellt, dass die Form in 95% aller Fälle in einem temporalen Sinn verwendet wird, während sie nur in 4% aller Fälle in einem räumlichen Sinn gebraucht wird. In 1% aller Fälle ist die Verwendung nicht eindeutig, da sie beispielsweise in einem metaphorischen Sinn gebraucht werden kann. Die Beispiele (1) und (2) zeigen die Verwendung von *at the end of* in einem temporalen beziehungsweise räumlichen Sinn. Beispiel (3) zeigt, dass in einigen Fällen die Grenze zwischen temporalem und räumlichem Sinn verschwinden kann und es in beiderlei Sinn interpretiert werden kann. Alle Beispiele stammen aus dem *Financial Times Corpus (FTC)* (vgl. Forchini / Murphy 2008: 255).

(1) **Temporal:** I hope to announce the figure at the end of April but thank you all for being so generous once again. (FTC)

(2) **Räumlich:** ...they found the right one: a small cabin on sloping land overlooking a lake at the end of a quiet street. (FTC)

(3) **Temporal / Räumlich:** ...and I'll never have to buy shirts ever again, he says, his voice drifting up at the end of each sentence in that Canadian way. (FTC)

In allen vier Subkorpora des Web-ELF-Korpus gehört *at the end of* ebenfalls zur Liste häufig auftretender *4-grams*, allerdings in unterschiedlichem Maße (Foren 54 mal *pmw*, Blogs 14 mal *pmw*, kommerzielle Homepages 32 mal *pmw*, Wikipedia-Artikel 40 mal *pmw*). In Foren (Nr. 11) und Wikipedia-Artikeln (Nr. 38) gehört *at the end of* zu den am häufigsten auftretenden *4-grams*. In kommerziellen Homepages (Nr. 132) werden sie etwas seltener verwendet, in Blogs (Nr. 547) hingegen sogar nur sehr selten. Im Vergleich zu dem *British National Corpus (BNC)* und dem *Financial Times Corpus*, in denen *at the end of* das zweithäufigste (93,41 mal *pmw*) beziehungsweise das neunthäufigste (100,69 mal *pmw*) *4-gram* darstellt, wird das Muster in allen Web-Genres seltener verwendet.

Die Analyse der semantischen Verwendungen von *at the end of* in den beiden Korpora *BNC* und *FTC* hat gezeigt, dass das *4-gram* auf unterschiedliche Weise verwendet werden kann. Es bietet sich nun an, den Verwendungskontext von *at the end of* in den einzelnen Web-Genres zu betrachten. Dabei wird auch überprüft, inwiefern Parallelen zu den beiden Referenzkorpora (*BNC / FTC*) sichtbar werden. Das Beispiel (4) zeigt die Verwendung von *at the*

end of in einem temporalen Sinn (aus einem Auszug des Subkorpus kommerzielle Homepages). Das Beispiel (5) hingegen zeigt die Verwendung des *4-grams* in einem räumlichen Sinn.

(4) At the end of 3 months, the new hair begins to increase, starting from the roots that were transplanted. (kommerzielle Homepages)

(5) You will see the terms and conditions at the end of this page. (kommerzielle Homepages)

Für die Analyse wird jede Form von *at the end of* in ihrem Kontext überprüft und diese einer der drei Kategorien 1. Temporal, 2. Räumlich, 3. Temporal / Räumlich zugeordnet. Tabelle 22 gibt einen Überblick über die Ergebnisse für alle vier Web-Genres.

<i>at the end of</i>	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
1. Temporal	79,7%	66,7%	88,9%	75,9%
2. Räumlich	3,1%	33,3%	5,6%	24,1%
3. Temporal/Räumlich	17,2%	0%	5,6%	0%
Gesamt (Prozent)	100%	100%	100%	100%
Gesamtzahl Tokens	64	3	18	29

Tabelle 22: Ergebnis der semantischen Analyse von *at the end of*

Zunächst fällt auf, dass sich alle Subkorpora im Hinblick auf die Verteilung dieses Musters von den Ergebnissen des *Financial Times Corpus* unterscheiden. Das *Financial Times Corpus* hat die folgende prozentuale Verteilung: Temporal 95%, Räumlich 4%, Temporal / Räumlich 1%. Ein Grund dafür ist sicherlich die zum Teil extrem unterschiedliche Gesamtzahl an *Tokens*. So finden sich beispielsweise im Web-Genre Blogs lediglich drei Vorkommen von *at the end of*. Selbstverständlich muss dies bei der Interpretation der Ergebnisse berücksichtigt werden. Nichtsdestotrotz liefert die Tabelle einige interessante Befunde.

Einzig das Web-Genre kommerzielle Homepages weist einen ähnlichen Anteil an temporalen Verwendungen von *at the end of* auf (88,9% im Vergleich zu 95% im *Financial Times Corpus*). Die anderen drei Web-Genres haben eine geringere Anzahl an temporalen Verwendungen von *at the end of*. Es ist dabei auffällig, dass in Wikipedia-Artikeln lediglich 24,1% aller *4-grams* in einem räumlichen Sinne verwendet werden. In jedem dieser Fälle handelt es

sich um eine detaillierte Beschreibung chirurgischer Eingriffe. Als typische Beispiele dafür fungieren (6) und (7).

(6) These simple retractors may be handheld, clamped in situ, or suspended at the end of a robotic arm. (Wikipedia-Artikel)

(7) The method works by inserting a titanium bolt into the bone at the end of the stump. (Wikipedia-Artikel)

Für das Web-Genre Foren zeichnet sich eine andere Besonderheit ab. 17,2% aller Vorkommen werden in einem temporalen / räumlichen Sinn verwendet. Es zeigt sich dabei, dass es eine Vielzahl von metaphorischen Verwendungen gibt. Insbesondere die Metapher *at the end of the tunnel* wird auffallend häufig verwendet (vgl. auch Beispiele 8 und 9).

(8) One of the major issues here is that there isn't really a light at the end of the tunnel. (Foren)

(9) The Inspiration [sic] for me is the fact that I see a light at the end of the tunnel. (Foren)

Die Ergebnisse der Tabelle 23 verdeutlichen, dass es Unterschiede zwischen den einzelnen Web-Genres bezüglich der semantischen Verwendung von *at the end of* gibt. Es erscheint mir nun im nächsten Schritt sinnvoll, zu überprüfen, inwieweit in den jeweiligen Web-Genres andere Kollokate präferiert werden. Im Folgenden werden alle Kollokate von *at the end of* gesondert für die einzelnen Web-Genres aufgelistet. Dabei habe ich mich auf Kollokate beschränkt, die rechts vom *Node-Word* stehen.⁵⁵ Es werden für die Analyse ausschließlich Nomen aufgeführt. Ein Vergleich der Kollokate von *at the end of* erlaubt es zu überprüfen, inwieweit sich die Wortfelder (vgl. Trier 1931) der Web-Genres ähneln. Zunächst wird das Ergebnis für das Web-Genre Foren in Tabelle 23 präsentiert.

⁵⁵ Es wurden alle Nomen berücksichtigt, die bis zu fünf Wörter entfernt waren von *at the end of*.

Kollokate (rechts)	Beispiele
Zeitperioden	day, 7 weeks, month, year, semester, season
Monatsnamen	March, October, November
Nomen mit Temporalbezug	show, university, course, report, church service
Nomen mit räumlichem Bezug	ear, thread
Metaphorische Ausdrücke	tunnel, road

Tabelle 23: Kollokate von *at the end of* im Web-Genre Foren

Es zeigt sich, dass die Mehrzahl aller Kollokate von *at the end of* einem gewissen Muster folgt. Es werden einerseits konkrete Zeitperioden oder Monatsnamen genannt (unter anderem *day, month, year*), andererseits einige Nomen, die einen klaren temporalen Bezug aufweisen (unter anderem *show, university, course*). Einen typischen Auszug aus einer temporalen Verwendung von *at the end of* liefert Beispiel 10.

- (10) At the end of the show I was left wondering if that show followed Mondays just to show the doctors flip flopping on their protocol. (Foren)

Gleichzeitig lassen sich nur zwei Nomen mit räumlichem Bezug identifizieren (*ear, thread*) (vgl. Beispiel 11). Im Gegensatz zu den anderen Web-Genres lassen sich auch Kollokate finden, die in einem metaphorischen Sinn verwendet werden. Die elf *Tokens* verteilen sich allerdings ausschließlich auf die beiden Nomen *tunnel* und *road* (vgl. Beispiel 12).

- (11) The above was in answer to the original thread question - why is plastic surgery frowned upon? (not a response to the personal stories told at the end of the thread). (Foren)

- (12) I am so sad - I think I am at the end of the road. I needed the restriction of the band and now I dont [sic] have it - I know i cant [sic] do it. (Foren)

Für Blogs bietet sich eine derartige Analyse nur bedingt an, da *at the end of* insgesamt nur dreimal auftritt, so dass es sich als schwierig erweist, aus dieser geringen Anzahl an *Tokens* aussagekräftige Ergebnisse zu generieren. Tabelle 24 gibt das Resultat für das Web-Genre Blogs wieder. Zwei Kollokate haben einen temporalen (*life, 4 days*) Bezug, während ein Kollokat einen räumlichen Bezug (*catheter*) hat.

Kollokate (rechts)	Beispiele
Zeitperioden	life, 4 days
Nomen mit räumlichem Bezug	catheter

Tabelle 24: Kollokate von *at the end of* im Web-Genre Blogs

Für das Web-Genre kommerzielle Homepages ergibt sich ein anders geartetes Bild (vgl. Tabelle 25). Es werden fast ausschließlich Kollokate in einem temporalen Sinn verwendet. Einerseits werden dabei konkrete Zeitperioden genannt (*year, period, sessions*), andererseits werden Nomen verwendet, die ebenfalls einen temporalen Bezug haben (*procedure, tour, surgery*). Ein räumlicher Bezug kann in zwei Fällen (*page, telephone*) hergestellt werden.

Kollokate (rechts)	Beispiele
Zeitperioden	year, period, sessions, 19 th century, 3 months
Nomen mit Temporalbezug	procedure, tour, surgery
Nomen mit räumlichem Bezug	page, telephone

Tabelle 25: Kollokate von *at the end of* im Web-Genre kommerzielle Homepages

Abschließend gibt Tabelle 26 das Ergebnis für das Web-Genre Wikipedia-Artikel wieder. Analog zu den Web-Genres Foren und kommerzielle Homepages lassen sich viele Kollokate identifizieren, die einen temporalen Bezug aufweisen. Darunter fallen Kollokate, die auf konkrete Zeitperioden verweisen (unter anderem *day, year, phase*) und Kollokate mit Temporalbezug (unter anderem *procedure, treatment, surgery*). Im Gegensatz zu allen anderen Web-Genres lässt sich jedoch ebenso eine Vielzahl von Kollokaten finden, die einen räumlichen Bezug haben. Diese stellen allesamt Fachtermini mit spezialisierten fachsprachlichen Bedeutungen dar, die auf chirurgische Eingriffe verweisen (unter anderem *laparoscope, robotic arm, bone*).

Kollokate (rechts)	Beispiele
Zeitperioden	day, year, phase, cycle, minutes, the 1970s
Nomen mit Temporalbezug	procedure, treatment, surgery, encounter, terms, transfusion, expiration, war
Nomen mit räumlichem Bezug	laparoscope, robotic arm, bone, rubber tube, stump, catheter, SEMS

Tabelle 26: Kollokate von *at the end of* im Web-Genre Wikipedia-Artikel

Die Ergebnisse offenbaren, dass je nach Web-Genre ein anderer semantischer Bezug hergestellt wird. Das Subkorpus kommerzielle Homepages ist das einzige, das in seiner Distribution dem *Financial Times Corpus* gleicht. In beiden Fällen wird *at the end of* zu einem großen Teil in einem temporalen Sinn verwendet. Foren zeichnen sich hingegen dadurch aus, dass *at the end of* vergleichsweise häufig in einem metaphorischen Sinn verwendet wird. Wikipedia-Artikel stellen das einzige Web-Genre dar, in dem *at the end of* vergleichsweise häufig in einem räumlichen Sinn verwendet wird.

Im Hinblick auf die verwendeten Kollokate zeichnen sich ebenfalls klare Tendenzen ab. Es ist generell auffällig, dass nur wenige Kollokate existieren, die in mehreren Web-Genres vorkommen. Dies sind allesamt Kollokate mit temporalem Bezug, vor allem Zeitperioden (*day, year*), aber auch Nomen, die sich auf chirurgische Eingriffe beziehen (*procedure, surgery*). Letztere werden sowohl in kommerziellen Homepages als auch in Wikipedia-Artikeln verwendet.

7.3. Analyse des *Phrase-Frames at the X of*

Die Analyse des *4-grams at the end of* hat gezeigt, dass sich die Web-Genres in der Verwendung dieses Musters voneinander unterscheiden. In einer weiteren Analyse wird nun überprüft, inwieweit auch Unterschiede bezüglich des generellen *Phrase-Frames at the X of* existieren. Biber et al. (1999: 1018) haben bereits festgestellt, dass *Lexical Bundles*, die mit der Präposition *at* beginnen, in den meisten Fällen dazu verwendet werden, um temporale Beziehungen zu markieren. Aus diesem Grund muss davon ausgegangen werden, dass auch in der Mehrheit aller Fälle das *Phrase-Frame at the X of* in einem temporalen Sinn verwendet wird. Es gilt nun zu überprüfen, ob sich diese Beobachtung auch auf Web-Genres übertragen lässt. Als Referenz dient wiederum das *Financial Times Corpus* von Forchini / Murphy (2008: 356). Sie haben für ihre Studien alle Vorkommen des *Phrase-Frames at the X of* überprüft, in denen Nomen den freien Slot ausgefüllt haben. Dabei fanden sie heraus, dass 50% aller Vorkommen in einem temporalen und 38% aller Vorkommen in einen räumlichen Sinn verwendet wurden. In 12% aller Fälle konnten die *4-grams* keiner der beiden Alternativen zugeordnet werden. Stattdessen wurde *at the end of* in diesen Fällen verwendet, um auf "processes or measurements" (Prozesse oder Mengenangaben) zu verweisen (Forchini / Murphy 2008: 356). Das Beispiel 13 zeigt die Verwendung des *Phrase-Frames at the X of* in diesem Sinn. In diese letzte Kategorie müssen jedoch auch alle anderen Verwendungsweisen einsortiert werden, so auch die metaphorischen.

(13) I watched like a helpless child as some of my fellow students juggled the entries on their Excel spreadsheets at the speed of lightning. (*Financial Times Corpus*)

Im Folgenden soll nun das *Phrase-Frame at the X of* für alle vier Subkorpora im jeweiligen Kontext betrachtet werden. Tabelle 27 illustriert die prozentuale Verteilung aller gefundenen *Types*, die sich in die drei Kategorien einordnen lassen.

<i>at the X of</i>	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
1. Temporal	49,0	33,3	36,3	35,9
2. Räumlich	27,5	58,3	54,8	53,8
3. Andere	23,5	8,3	8,9	10,3
Gesamt	100	100	100	100

Tabelle 27: Prozentuale Verteilung des *Phrase-Frames at the X of* in die semantischen Kategorien

Tabelle 27 zeigt, dass die Distribution der *Phrase-Frames* für die Web-Genres Blogs, kommerzielle Homepages und Wikipedia-Artikel nahezu identisch ist. Insbesondere kommerzielle Homepages und Wikipedia-Artikel gleichen sich in ihrer Verteilung fast vollständig. Lediglich das Web-Genre Foren hat eine andere prozentuale Verteilung. In Foren wird *at the X of* in etwa der Hälfte aller Fälle (49%) in einem temporalen Sinn verwendet, während es nur in 27,5% aller Fälle in einem räumlichen Sinn verwendet wird. Auffällig ist des Weiteren die vergleichsweise häufige Verwendung in einem "anderen" Sinn, also metaphorisch oder um auf Prozesse oder Mengenangaben zu verweisen. Die Ergebnisse der Analyse unterstreichen somit, dass Foren sich von den drei anderen Web-Genres unterscheiden und betonen gleichzeitig die relative Homogenität der anderen drei Web-Genres im Hinblick auf die Verteilung der semantischen Kategorien für *at the X of*.

7.4. Analyse von *one of the most*

Die Suche nach gemeinsamen *4-grams*, die in mehreren Web-Genres verwendet werden, zeigt, dass *one of the most* eines der wenigen Muster darstellt, das in allen vier Web-Genres häufig verwendet wird. In drei von vier Web-Genres gehört es zu den zehn am häufigsten auftretenden *4-grams* (Blogs Nr. 3, kommerzielle Homepages Nr.2, Wikipedia-Artikel Nr. 10, Foren Nr. 31). Im Folgenden wird nun untersucht, inwieweit es auch für dieses *4-gram* Gebrauchsunterschiede gibt. Der Fokus liegt dabei auf der Analyse der Kollokate, die zusammen

mit dem Muster auftreten. Dabei wurden ausschließlich Kollokate betrachtet, die sich rechts von dem Muster befinden. Für die Auswertung wurden alle Muster in drei Kategorien unterteilt: Abstraktionen und Dinge, Orte und Personen. Dabei folgen alle Muster der Struktur *one of the most* + Adjektiv + Nomen. In Tabelle 28 werden alle Kollokate der drei Kategorien aufgelistet.

Kollokate (rechts)	Beispiele
Abstraktionen und Gegenstände	amazing and subtle transformations; basic requirements; common causes; dangerous substances; difficult languages; exposed areas; fascinating things; important/retarded things; important nutrients/concern; obvious examples; painful surgeries; pertinent issues; popular procedures.
Orte	interesting small cities; spectacular cities, unstable former Soviet republics; respected faculties.
Personen	acclaimed beauties; beautiful people/faces ⁵⁶ /women; compliant patients; concerned, loving, caring, dedicated, helpful doc; controversial psychiatrists; hopelessly ugly faces; impatient people; inspiring and supportive posters; productive human beings; prominent surgeons, respected lecturing pediatricians; unfeminine women.

Tabelle 28: Kollokate von *one of the most* im Web-Genre Foren

Es fällt auf, dass es zwischen den Kategorien einen Unterschied in der Konnotation der Muster gibt. In der Mehrheit aller Fälle, in denen *one of the most* in einem abstrakten Sinne verwendet wird oder genutzt wird, um auf Gegenstände zu verweisen, zeigt sich eine neutrale Konnotation (vgl. Beispiele 14 und 15).

(14) Also, one of the most important things, they provide stafford loans. (Foren)

(15) Basically, what they are saying is that the school is IMED listed - one of the most basic requirements for an institution teaching medicine. (Foren)

⁵⁶ Das Nomen *face* wurde zu der Kategorie Personen hinzugerechnet, da es einen Teil einer Person darstellt.

In einigen wenigen Fällen haben die Muster jedoch auch eine negative oder eine positive Konnotation (vgl. Beispiele 16 und 17).

(16) I have had a tummy tuck and I know that's one of the most painful surgeries you can have and I did well with it but my nipples are very sensitive. (Foren)

(17) This is one of the most amazing and subtle transformations ever - and all done in the 1940's! (Foren)

Im Vergleich dazu wird deutlich, dass die Muster in der Kategorie Personen mehrheitlich positiv konnotiert sind. Es gibt nur wenige negativ konnotierte Fälle. Häufig werden Adjektive verwendet, die eine große positive oder negative Emotionalität seitens der Schreiber ausdrücken (beispielsweise *amazing*, *retarded*, *spectacular*). Dies ist ein typisches Merkmal gesprochener Sprache. In der Kategorie Orte werden ebenfalls vorrangig positiv konnotierte Muster verwendet. Dass in diesen beiden Kategorien vorrangig positiv konnotierte Muster verwendet werden, kann darauf zurückzuführen sein, dass die Autoren andere Forennutzer ermutigen wollen, sich medizinischen Eingriffen zu unterziehen.

Es bietet sich nun an, diese Befunde mit denen der anderen Web-Genres zu vergleichen. In Tabelle 29 werden alle Kollokate des Web-Genres Blogs aufgeführt.

Kollokate (rechts)	Beispiele
Abstraktionen und Gegenstände	beautiful creations; common complications/methods; difficult problems; efficient health care system; frequently performed plastic surgery procedures; important parts of your consultation/organ/phrases/things; popular destinations/plastic surgery procedures; prevalent types of eye surgery; sensitive areas of the body; serious consequences; significant factors; visited destination/place; widely accepted procedures.
Orte	favorite travel destinations.

Tabelle 29: Kollokate von *one of the most* im Web-Genre Blogs

Im Gegensatz zum Web-Genre Foren beziehen sich beinahe alle Kollokate auf Abstraktionen und Gegenstände. Auf Personen wird nicht verwiesen, jedoch in einem Fall auf Orte. Die meisten Kollokate haben eine neutrale Konnotation, unter anderem das Beispiel 18.

- (18) The scratch test, one of the most common methods, involves placing a small amount of suspected allergy-causing substances on the skin, usually the forearm, upper arm, or the back. (Blogs)

Im Vergleich zu Foren fällt auf, dass in Blogs eine oft nüchternere Beschreibung stattfindet. Es werden weder stark positive noch stark negative Adjektive verwendet, während in Foren sehr emotionale Adjektive (zum Beispiel *amazing* oder *spectacular*) verwendet werden. Auffällig ist, dass viele Muster sich auf den chirurgischen Eingriff selbst beziehen (*complications, methods, problems, procedures*). Die Häufung dieser Kollokate lässt vermuten, dass die Autoren in erster Linie die Rezipienten über medizinische Eingriffe und deren Risiken aufklären wollen.

Für das Web-Genre kommerzielle Homepages ergibt sich das folgende Bild (vgl. Tabelle 30). In allen drei Kategorien lässt sich eine Vielzahl von Beispielen finden, die vorrangig positiv oder neutral konnotiert sind. In den Kategorien Orte und Personen befinden sich fast ausschließlich positiv konnotierte Muster (vgl. Beispiele 19 und 20). Es gibt für diese beiden Kategorien kein einziges Beispiel, in dem eine negative Konnotation nachgewiesen werden kann. In der Kategorie Abstraktionen und Gegenstände werden sowohl neutral als auch positiv konnotierte Muster verwendet.

- (19) Set in one of the most pristine areas of South Africa, Pezula is also a family friendly hotel that has fantastic children's facilities and access to a private beach, ...
(kommerzielle Homepages)

- (20) Dr. Manuel Sanmiguel is one of the most prestigious and well-known plastic surgeons in Northern Mexico. (kommerzielle Homepages)

Analog zum Web-Genre Blogs wird eine Vielzahl von Mustern verwendet, die sich auf den chirurgischen Eingriff selbst und die Methoden beziehen (*procedure, methods, surgery*).

Kollokate (rechts)	Beispiele
Abstraktionen und Gegenstände	<p>common procedures/contraindication/hand conditions; causes of cavities/surgeries/plastic surgery procedure/dental complaints/surgeries; commonly used forms of anesthesia; complicated surgeries/operations; difficult areas to treat; effective teeth whitening methods; efficient treatment methods; frequent surgeries; frequently performed cosmetic surgery/operations; important textbooks/aspects/question/notions/skills/factors/advantages/elements; intricate and rewarding dramas; joyful experiences; luxurious trains; mobile joints; perfect towers; popular surgical procedures/methods/forms/procedures; relaxing activates; serious complications; significant improvements; sophisticated programs; sought plastic surgeries/cosmetic surgeries; successful procedures/operations; successful applications; unimportant factors; well-known of cosmetic surgeries.</p>
Orte	<p>beautiful and bio-diverse areas; beautiful and welcoming establishments; beautiful and tranquil cities; beautifully verdant regions; beautiful areas of the Cape; cutting edge fertility clinics; important Hanseatic cities; modern clinics/establishments/private healthcare centre/Polish private centre; picturesque mountain lakes; popular destinations; preferred destinations; pristine areas in South Africa; sought-after areas; unique and splendid destinations.</p>
Personen	<p>distinguished plastic surgeons; experienced orthopaedic surgeons/pelvic surgeons; famous LASIK surgeons; important plastic surgeons; personable doctors; prestigious and well-known plastic surgeons; prolific inventors; qualified plastic surgeon.</p>

Tabelle 30: Kollokate von *one of the most* im Web-Genre kommerzielle Homepages

Die Muster heben sich insgesamt auch durch eine vergleichsweise geringere stilistische Varianz hervor. Einige wenige Adjektive (*beautiful, common, important, popular*) werden sehr häufig verwendet. Auffällig ist insbesondere, dass in den beiden Kategorie Orte und Personen in einigen Fällen zwei Adjektive verwendet werden (*beautiful and welcoming, beautiful and tranquil, prestigious and well-known, unique and splendid*) (vgl. auch Beispiel 21).

(21) Allow us to arrange your stay at one of the most beautiful and welcoming establishments in the region. (kommerzielle Homepages)

Durch das Attribuieren zweier Adjektive zu einem Nomen soll dessen Aussagekraft verstärkt werden. Das Ziel kommerzieller Homepages ist es schließlich, potenzielle Kunden von einer Operation an dem beworbenen Ort zu überzeugen. Im Fall von Personen soll damit beispielsweise die Kompetenz des Chirurgen betont werden. Da alle Texte des Web-Genres kommerzielle Homepages mit dem Ziel formuliert wurden, potenzielle Kunden von einem chirurgischen Eingriff zu überzeugen, wurden vorrangig positiv konnotierte Adjektive verwendet, um auf die Eingriffe selbst, die Kliniken und die ausführenden Chirurgen zu verweisen. Eine negative Konnotation wäre vor diesem Hintergrund nicht zielführend.

Abschließend werden die Ergebnisse für das Web-Genre Wikipedia-Artikel in Tabelle 31 veranschaulicht. Es fällt auf, dass in diesem Web-Genre ausschließlich Muster identifiziert werden konnten, die den Kategorien Abstraktionen und Gegenstände und Personen zugeordnet wurden. Auf Orte wird nicht verwiesen. Zudem gehört die große Mehrheit aller Muster der ersten Kategorie an. Fast alle Muster in der Kategorie Personen sind positiv konnotiert. In der Kategorie Abstraktionen und Gegenstände hingegen ist die überwiegende Mehrheit aller Muster neutral konnotiert. Es gibt einige wenige Muster, die positiv konnotiert sind und eine noch geringere Anzahl an Mustern die negativ konnotiert sind. Auffällig ist die vergleichsweise häufige Nutzung des Adjektivs *common*. Fast alle Muster der Kategorie Abstraktionen und Gegenstände beziehen sich auf die chirurgischen Eingriffe und die angewandten Methoden (*procedures, methods, approaches, operations*).

Grundsätzlich sind die Muster mit positiver oder negativer Konnotation weniger emotionalisiert als die Muster im Web-Genre Foren. Die einzige Ausnahme stellt das Beispiel 22 dar.

(22) After the introduction of the antipsychotic chlorpromazine (Thorazine), lobotomies fell out of common use and the procedure has since been characterized "as one of

the most barbaric mistakes ever perpetrated by mainstream medicine". (Wikipedia-Artikel)

Der Kontext zeigt jedoch, dass diese Ausnahme der Tatsache geschuldet ist, dass direkt zitiert wird. Ansonsten spiegeln die Ergebnisse den nüchternen Stil der Online-Enzyklopädie Wikipedia wider. Das Ziel der Artikel ist es, objektiv und wissenschaftlich zu beschreiben. Aus diesem Grund werden vorrangig neutral konnotierte Adjektive verwendet.

Kollokate (rechts)	Beispiele
Abstraktionen und Gegenstände	accepted means of function; barbaric mistakes; basic type of surgery; challenging complications; challenging and complex areas; common causes/transplant procedures/eye diseases/surgical/procedures/ complications/operations/topics/procedures/birth defects/graft sites/treatment recommendations/elective surgical procedures; commonly used intravenous drugs; commonly performed interventions; demanding tasks; difficult procedures; effective operations/procedures/Mini-Systems; enduring human ailments; extensive research studies; famous fictional cyborgs/instances of self-trepanation; frequently performed surgical operations; important factors; long-lived drugs; painful incisions; prevalent types; prominent medical uses; promising new approaches; reliable forms; research-oriented specialties; significant barriers; studied procedures; technically challenging operations; used techniques; widely known systems.
Personen	acclaimed surgeons; accomplished field delegates; common medical practitioners; gifted tactical commanders; profound philosophers; prominent students; unique characters; widely acclaimed surgeons.

Tabelle 31: Kollokate von *one of the most* im Web-Genre Wikipedia-Artikel

In Tabelle 32 wird eine abschließende prozentuale Übersicht über die Verteilung der Muster in den verschiedenen Kategorien dargestellt. Diese erlaubt es, die Unterschiede zwischen den einzelnen Web-Genres besser zu visualisieren. Die Übersicht belegt, dass für kommerzielle Homepages eine Vielzahl positiv konnotierte Muster verwendet wird. In Foren werden ebenfalls viele positiv konnotierte Muster verwendet; gleichzeitig ist dieses Web-Genre jedoch das Einzige, in dem eine vergleichsweise hohe Anzahl an negativ konnotierten Mustern gefunden werden kann. Blogs und Wikipedia-Artikel ähneln sich in ihrer Verteilung. In beiden Web-Genres werden mehrheitlich neutral konnotierte Muster verwendet. Daraus könnte man schließen, dass die Autoren in diesen Web-Genres die Absicht verfolgen, ihre Standpunkte objektiv darzulegen.

Abstraktionen und Gegenstände			
	positiv	neutral	negativ
Foren	30%	55%	15%
Blogs	29%	62%	10%
kommerzielle Homepages	41%	52%	7%
Wikipedia-Artikel	20%	76%	4%
Orte			
	positiv	neutral	negativ
Foren	80%	0%	20%
Blogs	50%	50%	0%
kommerzielle Homepages	95%	5%	0%
Wikipedia-Artikel	0%	0%	0%
Personen			
	positiv	neutral	negativ
Foren	79%	0%	21%
Blogs	0%	0%	0%
kommerzielle Homepages	92%	8%	0%
Wikipedia-Artikel	87,5%	12,5%	0%
Gesamt			
	positiv	neutral	negativ
Foren	57%	25%	18%
Blogs	33%	59%	7%
kommerzielle Homepages	60%	36%	5%
Wikipedia-Artikel	30%	67%	4%

Tabelle 32: Übersicht der prozentualen Verteilung der konnotierten *4-grams*

7.5. Analyse von *to be able to*

Abschließend wird ein weiteres *4-gram* identifiziert, das relativ häufig in allen vier Web-Genres auftritt. Das Muster *to be able to* erfüllt diese Kriterien.⁵⁷ Analog zur vorherigen Studie wird somit ebenfalls eine Analyse der Kollokate durchgeführt, die rechts von dem Muster auftreten. Alle Muster folgen der Struktur *to be able to* + Verb. Die Ergebnisse der Untersuchung ermöglichen es, die nachfolgenden Verben in verschiedene Kategorien einzusortieren. In allen Fällen wird jedes einzelne Muster in dessen Kontext betrachtet, um sicherzustellen, dass es auch in die passende Kategorie einsortiert wird. Insgesamt gibt es elf Kategorien. Die Mehrzahl dieser Kategorien bedarf keiner näheren Erläuterung, so dass im Folgenden nur einige von ihnen näher beschrieben werden. In der Kategorie *Beratung / Vorabinformation* befinden sich alle Kollokate, die vorrangig dazu dienen, konkrete Vorabinformationen über einzelne medizinische Eingriffe zu präsentieren, um die Leser über Chancen und Risiken zu informieren. Die Kategorie *Reise / Studienreise* enthält Kollokate, in denen konkrete Reisen oder längere Studienaufenthalte an anderen Orten beschrieben werden. Alle Kollokate in der Kategorie *Operation* beziehen sich konkret auf einen chirurgischen Eingriff. In der Kategorie *Selbstreferenz / Referenz zu Mitmenschen* befinden sich alle kollokierenden Muster, in denen die Autoren referentielle Bezüge zu ihrem eigenen Leben herstellen oder einen Bezug zu ihren Mitmenschen herstellen. Die Kategorie *Sonstiges* enthält alle Muster, die sich keiner der Kategorien zweifelsohne zuordnen ließen. Zunächst soll der Blick auf das Web-Genre Foren gerichtet werden (vgl. Tabelle 33).

⁵⁷ Das *4-gram* tritt je nach Web-Genre unterschiedlich oft auf, gehört jedoch in allen Web-Genres zu den 200 häufigsten Mustern (Foren Nr. 4, Blogs Nr. 191, kommerzielle Homepages Nr. 181, Wikipedia-Artikel Nr. 82).

Kollokate (rechts)	Beispiele
Beratung / Vorabinformation	continue into the 2 nd year; exercise; fix the source; get a hold of; have the gastric bypass surgery; judge an area; loose the weight; remove at least 500g; use these things.
Bewegung	go to the beach; go swimming; jog; lift him; move; run around; walk.
Essen	be eaten all the time; eat more than 2500 calories; eat better as I go along.
Finanzen	afford the tuition fees; buy pretty clothes; defer loans; financially take care of; pay to have the surgery; pay for my surgery; save the full amount for surgery.
Kommunikation	answer that question; converse with the patients; express your anger; say that; speak German; speak another language; tell him anything.
Reise / Studienreise	allow students to do more; accomidate obese patients; come home; do medschool in france; practice in California; spend two summers during college; study in Poland.
Resultat	achieve proper restrictions; accomplish training; have done that; pass the test.
Selbstreferenz / Referenz zu Mitmenschen	absorb what is offered; admit when another girl is hot; alter their appearances; care for someone; control the behaviors; do this surgery; do it post surgery; enjoy Germany; finally gain control; get his hands on; get over it; have a family portrait; have a tool to use; have some of the needed; have this surgery; help other people; keep my options open; live to see my kids grow up; look good in photos; make a difference; make them smaller; play with my kids; save lives; stop that cycle; support this weight; surgically improve; take advantage of; take the face back; touch the future; wear a bikini/cute clothes/certain things; wear up styles.
Wahrnehmung	look down and see.
Sonstige	evaluated; restrict them; send PMs.

Tabelle 33: Kollokate von *to be able to* im Web-Genre Foren

Im Gegensatz zu den anderen Web-Genres weisen Foren eine Vielzahl unterschiedlicher Kategorien auf. Es fehlen lediglich Kollokate, die sich auf einen konkreten chirurgischen Eingriff beziehen und damit der Kategorie *Operation* angehören. Neun Kollokate gehören der Kategorie *Beratung / Vorabinformation* an. Die Autoren berichten in diesen Fällen oftmals von ihren eigenen Erfahrungen und geben Tipps, wie man sich auf chirurgische Eingriffe vorbereiten soll (vgl. Beispiel 24).

(24) They always seem to be able to use these things on the butt, but for some reason they always feel like it cannot work well on the hips. (Foren)

Auffällig ist ebenfalls, dass einige Kollokate sich auf Reisen und Studienreisen beziehen (vgl. Beispiel 25).

(25) [...] for sure, you will need an excellent level of french to be able to do medschool france, because there is enough course material as it is, let alone needing to struggle with the language as well. (Foren)

Die Kategorie *Selbstreferenz / Referenz zu Mitmenschen* enthält mit Abstand die meisten Kollokate. Oftmals werden dort selbstreferentielle Aussagen über das eigene Befinden gemacht oder die Meinung zu einem Thema wiedergegeben (vgl. Beispiel 26). In manchen Fällen wird aber auch Bezug auf andere Menschen genommen und dabei die Meinung des Autors ausgedrückt (vgl. Beispiel 27).

(26) I'd love to be able to take advantage of the WLS tool, but at the moment I don't qualify. (Foren)

(27) She has lost some weight, I just thought it was way too drastic on the bottom for her to be able to have done that through diet and excercise [sic]. (Foren)

Insgesamt fällt auf, dass die Autoren bei einer Vielzahl der Konstruktionen einen direkten Bezug zu ihrem eigenen Leben herstellen. Der Fokus liegt auf einer Darstellung der eigenen Meinung und der eigenen Erfahrung. Dabei umreißen sie einige für sie relevante Themen, wie zum Beispiel die finanziellen Konsequenzen eines chirurgischen Eingriffes, die Möglichkeit zum Reisen oder die möglichen Konsequenzen einer Operation.

Aufgrund der wenigen Vorkommen des *4-grams to be able to* müssen die Ergebnisse für das Web-Genre Blogs relativiert werden (vgl. Tabelle 34). Trotz dieser Einschränkung lässt sich beobachten, dass in keinem der Vorkommen ein selbstreferentieller Bezug hergestellt wird. Stattdessen werden im Kern objektive Informationen zu Eingriffen gegeben, die nicht persönlich bewertet werden.

Kollokate (rechts)	Beispiele
Beratung / Vorbereitung	decrease their dependancy.
Operation	get the earring back in.
Wahrnehmung	see treatment options; see a ball coming out of the light.
Sonstige	work autonomously.

Tabelle 34: Kollokate von *to be able to* im Web-Genre Blogs

Für das Web-Genre kommerzielle Homepages ergibt sich ein etwas anderes Bild (vgl. Tabelle 35). Obwohl in diesem Web-Genre ebenfalls scheinbar objektive Informationen gegeben werden, gibt es dennoch einige Unterschiede zu Blogs. Es werden mehr Vorabinformationen gegeben und einzelne Eingriffe im Detail erklärt. Ein Herausstellungsmerkmal ist darüber hinaus die Betonung des positiven Resultats einer Operation. Es finden sich demzufolge einige kollokierende Muster in der Kategorie *Resultat*.

Kollokate (rechts)	Beispiele
Beratung / Vorbereitung	evaluate the latest technical developments; manage stress; effectively program everything.
Bewegung	swim with clear vision.
Finanzen	afford it.
Operation	avail themselves of this procedure; examine skin quality.
Reise / Studienreise	follow natural curves; harmonize the area around the mouth. return to your country; return home; seek treatment abroad.
Resultat	demonstrate your new nose; do low-impact activities; live long enough; work the day after having the implant installed.
Selbstreferenz / Referenz zu Mitmenschen	accompany Norma.
Sonstige	keep in contact with you; shop for clothes.

Tabelle 35: Kollokate von *to be able to* im Web-Genre kommerzielle Homepages

Zuletzt wird das Ergebnis für das Web-Genre Wikipedia-Artikel wiedergegeben (vgl. Tabelle 36). Auffällig ist, dass sich mehr als die Hälfte aller kollokierenden Muster der Kategorie *Operation* zuordnen lassen. Die Verteilung für das Web-Genre Wikipedia-Artikel ähnelt der Verteilung für kommerzielle Homepages. Allerdings wird deutlich, dass sich in der Kategorie *Resultat* nur ein Kollokat finden lässt, während sich die Mehrheit der Kollokate der Kategorie *Operation* zuordnen lässt.

Kollokate (rechts)	Beispiele
Beratung / Vorbereitung	perform the same thing.
Finanzen	cover all costs.
Operation	cause these follicle stem cells; later realign the virtual dataset; place dental implants; detect ventricular tachycardia; feel different tissue densities; place dental implants; prevent the spasm of blood vessels; resist a force.
Reise / Studienreise	return to practicing medicine.
Resultat	carry out activities.
Sonstige	read; practice.

Tabelle 36: Kollokate von *to be able to* im Web-Genre Wikipedia-Artikel

Im Gegensatz zur Analyse des *4-grams one of the most* werden nun auch die Kollokate, die sich links vom Muster befinden, betrachtet. Dabei werden die Ergebnisse für die vier Web-Genres zusammengefasst und Dopplungen bestimmter Kollokate durch Klammern angezeigt: *need* (4) bedeutet beispielsweise, dass das Verb *need* viermal auftrat. In Tabelle 37 wird eine Übersicht über die Ergebnisse gegeben. Es muss betont werden, dass in der Tabelle nur diejenigen Kollokate aufgeführt werden, die die Konstruktion *to be able to* prämodifizieren. In allen Fällen, in denen *to be able to* verwendet wird, um eine *Phrase* oder eine *Clause* einzuleiten, werden die Kollokate nicht beachtet. In dieser Hinsicht unterscheiden sich die Web-Genres zum Teil beträchtlich voneinander (Foren 36 mal, Blogs 3 mal, kommerzielle Homepages 4 mal, Wikipedia-Artikel 0 mal). Es fällt auf, dass insbesondere im Web-Genre Wikipedia-Artikel *to be able to* in keinem einzigen Fall verwendet wird, um eine *Phrase* oder eine *Clause* einzuleiten.

Für das Web-Genre Foren zeigt sich, dass häufig aktive Verben verwendet werden, die zudem die Einstellung der Autoren ausdrücken (*hope, like, want, would love*). Oftmals nutzen

die Autoren dabei auch *Emotive verbs*, um starke Emotionen auszudrücken. In der Mehrheit aller Fälle werden diese von einem Pronomen in der ersten Person Singular prämodifiziert. Generell folgen fast alle Kollokate dem Muster Verb + *to be able to*. Im Web-Genre Blogs existieren nur zwei Muster, die die Voraussetzungen erfüllen. Für das Web-Genre kommerzielle Homepages ergibt sich ein differenzierteres Bild. Dort werden mehrheitlich aktive Verben genutzt, die jedoch bis auf zwei Ausnahmen (*I am happy, I was happy*) nicht die persönliche Einstellung der Autoren ausdrücken. Generell geht in einigen Fällen dem *4-gram* kein Verb voraus. Für das Web-Genre Wikipedia-Artikel zeigt sich hingegen ein klares Bild. Im Gegensatz zu allen anderen Web-Genres sind fast alle Kollokate passivisiert.

Kollokate (links)	Beispiele
Foren	and; but (2); fortune enough; getting him prepared; going (6); have (5); hope (2); like; need (4); ought; seem (2); want/-s (4); wishes; would love (7); am hoping; can loose weight; can not wait; cant wait; was supposed; in order (3); is a pleasure/our pleasure; lose; tend; will have; you need/ought.
Blogs	allow dental tourists; who needed.
kommerzielle Homepages	are considered perfect; but; expect; for us/you; I am/was happy; in order; is important (2); is nice; not only; or; you want.
Wikipedia-Artikel	and; claim/-s/-ed (2); expected; in order (2); programmed; required (2); thought (2).

Tabelle 37: Kollokate aller Web-Genres links vom *Node-Word*

Die Analyse der Kollokate des *4-grams to be able to* hat ergeben, dass die Muster in unterschiedlichen Kontexten verwendet werden. Je nach Web-Genre werden andere Kollokationen präferiert.

7.6. Zusammenfassung

In dem vorliegenden Kapitel wurden vier qualitative Studien dargestellt, die sich mit der Verwendung von Mehrworteinheiten in den vier untersuchten Web-Genres befassen. Es konnte hierbei in jeder der vier Studien aufgezeigt werden, dass sich die Web-Genres deutlich voneinander unterscheiden. Foren und Wikipedia-Artikel erweisen sich in jedem Fall als besonders unterschiedlich. Zwischen den Web-Genres Blogs und kommerzielle Homepages gibt es

weniger starke Unterscheidungen, jedoch können auch diese Unterschiede deutlich herausgestellt werden.

Die Analyse der Kollokate der betrachteten *4-grams* hat gezeigt, dass sich der Schreibstil der Autoren je nach Web-Genre mehr oder weniger voneinander unterscheidet. Die Kollokate zeigen, dass Autoren von Wikipedia-Artikeln eine objektive und nüchterne Schreibweise bevorzugen. In Blogs versuchen die Autoren, eine ebenso nüchterne objektive Beschreibung medizinischer Eingriffe zu geben. Die Autoren des Web-Genres kommerzielle Homepages wählen einen Schreibstil, der Objektivität widerspiegeln soll. Im Fokus steht jedoch das Ziel, potenzielle Kunden von einem chirurgischen Eingriff zu überzeugen. Aus diesem Grund werden die positiven Resultate einer Operation betont und dies unter Verwendung vieler positiv-konnotierter Wörter erreicht. Foren hingegen unterscheiden sich sehr deutlich von den drei anderen Web-Genres. In ihnen wird sehr oft die persönliche Meinung der Autoren ausgedrückt und eine persönliche, subjektive Sicht wiedergegeben.

8. Diskussion der Ergebnisse und Resümee

8.1. Zusammenfassung der beobachteten Ergebnisse

Aus den vorliegenden Studien lassen sich einige Befunde ableiten. Es zeigte sich, dass sich die vier untersuchten Web-Genres in der Verwendung von *4-grams* auf formaler und funktionaler, aber auch in der Verwendung einzelner Muster stark voneinander unterscheiden. Die beobachteten Unterschiede waren, ebenso wie es bei traditionell mündlichen und schriftlichen Registern der Fall ist, graduell. In Tabelle 38 werden die Hauptkenntnisse der Studien knapp zusammengefasst dargestellt.

	Studien	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
	Formale Analyse				
1.	Formale <i>4-gram</i> -Klassifikation	→mehrheitlich mündliche Kategorien verwendet →bilden Gegenpol zu Wikipedia-Artikeln	→tendenziell Schriftsprachliche Kategorien verwendet →gleichen kommerziellen Homepages fast vollständig	→tendenziell Schriftsprachliche Kategorien verwendet →gleichen Blogs fast vollständig	→mehrheitlich schriftsprachliche Kategorien verwendet →bilden Gegenpol zu Foren
	Funktionale Analysen				
2.	Funktionale <i>4-gram</i> -Klassifikation	→ähneln in ihrer Verteilung dem mündlichen Register <i>Conversation</i>	→tendenziell schriftsprachlich →ähneln kommerziellen Homepages	→tendenziell schriftsprachlich →ähneln Blogs	→gleichen in ihrer Verteilung fast vollständig dem schriftsprachlichen Register <i>Textbooks</i>
3.	Verwendung gemeinsamer Muster	→teilen kaum <i>4-grams</i> mit Wikipedia-Artikeln →ähneln Blogs →nutzen sowohl mündliche als auch schriftsprachliche Kategorien	→ähneln Foren →unterscheiden sich von kommerziellen Homepages →nutzen mehrheitlich schriftsprachliche Kategorien	→ähneln Wikipedia-Artikeln →unterscheiden sich von Blogs →nutzen mehrheitlich schriftsprachliche Kategorien	→teilen kaum <i>4-grams</i> mit Foren →ähneln kommerziellen Homepages →nutzen mehrheitlich schriftsprachliche Kategorien
4.	Verbindung von Form und Funktion	→nutzen mehrheitlich mündliche Strukturen und gleichen mündlichen Registern	→nutzen tendenziell schriftsprachliche Kategorien →gleichen kommerziellen Homepages fast vollständig in ihrer Verteilung	→nutzen tendenziell schriftsprachliche Kategorien →gleichen Blogs fast vollständig in ihrer Verteilung	→nutzen mehrheitlich schriftsprachliche Strukturen

Qualitative Analysen					
5.	Analyse von <i>at the end of</i>	→ wird vorrangig in einem temporalen Sinn gebraucht, auch metaphorische Nutzung ist üblich	→ wird vorrangig in einem räumlichen Sinn gebraucht → ähneln kommerziellen Homepages und Wikipedia-Artikeln stark	→ wird vorrangig in einem räumlichen Sinn gebraucht → gleichen Wikipedia-Artikeln fast vollständig → ähneln Blogs stark	→ wird vorrangig in einem räumlichen Sinn gebraucht → gleichen kommerziellen Homepages fast vollständig → ähneln Blogs stark
6.	Analyse des <i>Phrase-Frames at the X of</i>	→ wird vorrangig in einem temporalen Sinn gebraucht	→ wird vorrangig in einem temporalen Sinn gebraucht → insgesamt nur geringe Anzahl an Vorkommen	→ wird vorrangig in einem temporalen Sinn gebraucht	→ wird vorrangig in einem räumlichen Sinn gebraucht
7.	Analyse von <i>one of the most</i>	→ ist vorrangig positiv konnotiert → aber auch Fälle mit negativer Konnotation (18%)	→ ist vorrangig neutral konnotiert	→ ist vorrangig positiv konnotiert	→ ist vorrangig neutral konnotiert
8.	Analyse von <i>to be able to</i>	→ Autoren drücken oftmals persönliche Meinung aus und geben Informationen → hohe Varianz an Kollokaten → Kollokate größtenteils in Verbindung mit aktiven Verben	→ Autoren nehmen keine persönlichen Bewertungen vor, sondern geben vorrangig Informationen	→ Autoren geben vorrangig Informationen und betonen die positiven Resultate eines chirurgischen Eingriffs	→ Autoren beschreiben die Operationsprozesse → Kollokate sind vorrangig passivisiert

Tabelle 38: Zusammenfassung der Hauptkenntnisse der einzelnen Studien

In allen vier quantitativen Studien wurde gezeigt, dass Foren und Wikipedia-Artikel Gegenpole bilden. Foren ähneln in jeder der Studien mündlicher Konversation sowohl auf formaler als auch auf funktionaler Ebene. Allerdings gleichen sie dem mündlichen Register *Conversation* sowohl in Form als auch Funktion nicht völlig, da in Foren ebenfalls einige schriftsprachliche Verwendungen von *4-grams* gefunden werden können. Studienübergreifend zeigte sich, dass Foren das einzige Web-Genre darstellen, das in Form und Funktion Distributionsparallelen zu mündlichem Sprachgebrauch aufweist. Die drei anderen Web-Genres gleichen diesbezüglich eher schriftsprachlichem Sprachgebrauch. Die Unterschiede zwischen den drei Web-Genres sind jedoch gradueller Natur.

Anhand der qualitativen Analysen lassen sich deutliche Unterschiede zwischen den einzelnen Web-Genres belegen. Es stellt sich nun die entscheidende Frage nach den Ursachen, die für Unterschiede zwischen den Web-Genres verantwortlich gemacht werden können.

8.2. Erklärungsansätze für die sprachliche Heterogenität der Web-Genres im Hinblick auf die Verwendung von 4-grams

Die Resultate der Studien lassen sich nicht monofaktoriell erklären, sondern erfordern die Betrachtung verschiedener Erklärungsansätze. Zunächst werden jedoch mögliche Ursachen ausgeschlossen, die meines Erachtens keinen Einfluss auf die sprachliche Heterogenität der Web-Genres haben. Unter anderem können die Ergebnisse nicht auf thematische Unterschiede der Subkorpora zurückgeführt werden, da der inhaltliche Fokus aller Web-Genres gleichgeartet war.

Im Gegensatz zu ihren Kommunikationszielen unterliegen die einzelnen Web-Genres unterschiedlichen Kommunikationsbedingungen, worin der Hauptgrund für die Unterschiede zwischen den Web-Genres zu sehen ist. Einzelne Parameter, wie beispielsweise der Grad der Vertrautheit der Kommunikationspartner oder der Grad der emotionalen Beteiligung haben grundsätzlich einen großen Einfluss auf den Schreibprozess. Dabei gilt es im Detail zu überprüfen, welche der Kommunikationsbedingungen sich hauptsächlich für die Heterogenität der Web-Genres verantwortlich zeichnen. Dazu werden diese nochmals zusammengefasst dargestellt (vgl. Tabelle 39).

Grundsätzlich haben die aufgeführten Kommunikationsbedingungen vor dem Hintergrund des Mediums Internet einen unterschiedlich großen Einfluss auf den Sprachgebrauch. Der Faktor a) Vertrautheit der Partner hat für das Medium Internet keine übermäßige Relevanz, da die Kommunikationsteilnehmer in allen untersuchten Web-Genres anonym sind. Der Faktor d) Physische Nähe der Kommunikationspartner ist im Kontext des Mediums Internet völlig unerheblich, so dass er keinen Erklärungsansatz für die Unterschiede zwischen den Web-Genres liefern kann. Ebenso ist der Grad der Themenfixierung für die Web-Genres des Web-ELF-Korpus unerheblich, da dieser in allen Subkorpora prinzipiell hoch ist.

Im Gegensatz zu den bereits genannten Faktoren, existieren jedoch Faktoren, die meines Erachtens einen großen Einfluss auf die Heterogenität des Sprachgebrauchs der Web-Genres haben und die miteinander in Verbindung stehen. In erster Linie zeichnen sich die Faktoren b) Grad der emotionalen Beteiligung, c) Grad der Situations- und Handlungseinbindung, e) Grad der Kooperation, f) Grad der Dialogizität und g) Grad der Spontaneität dafür verantwortlich.

Parameter	Foren	Blogs	komm. Homepages	Wikipedia-Artikel
a) Grad der Vertrautheit der Partner	gering	gering	gering	sehr gering
b) Grad der emotionalen Beteiligung	hoch	mittel	gering	sehr gering
c) Grad der Situations- und Handlungseinbindung	sehr hoch	mittel	gering	sehr gering
d) Physische Nähe der Kommunikationspartner	nicht existent	nicht existent	nicht existent	nicht existent
e) Grad der kommunikativen Kooperation	hoch	gering	gering	gering
f) Grad der Dialogizität	sehr hoch	gering	sehr gering	nicht existent
g) Grad der Spontaneität	sehr hoch	gering	sehr gering	sehr gering
h) Grad der Themenfixierung	mittel	hoch	sehr hoch	sehr hoch

Tabelle 39: Vergleich der Kommunikationsbedingungen der Web-Genres des Web-ELF-Korpus

Zunächst führt eine hohe emotionale Beteiligung dazu, dass Gefühle und Einstellungen ausgedrückt werden. Dies geschieht vorrangig im Web-Genre Foren. Dieser Faktor ist eng verknüpft mit der Frage, ob die Kommunikation grundsätzlich monologisch oder dialogisch ist. Foren stellen das einzige Web-Genre dar, in dem ein dialogischer Austausch stattfindet. Die Webnutzer tauschen in ihnen persönliche Erfahrungen aus und beraten andere Nutzer. In den anderen Web-Genres findet hingegen eine monologische Kommunikation statt. Dies erklärt auch, warum der Grad der kommunikativen Kooperation in diesen Web-Genres gering ist, da es eine Trennung von Produktion und Rezeption in der Kommunikation gibt. Ein wesentlicher Unterschied besteht in dem Grad der Spontaneität. In dialogischer Kommunikation ist dieser wesentlich höher als in monologischer Kommunikation. Insbesondere die beiden untersuchten Web-Genres kommerzielle Homepages und Wikipedia-Artikel sind davon geprägt, dass Texte üblicherweise mehreren Revisionsprozessen unterliegen, da orthografische Fehler und inhaltliche Fehler von den Rezipienten mit mangelnder Professionalität gleichgesetzt werden und die Autoren daher bemüht sind, fehlerfreie Texte zu produzieren. Für das Web-Genre Blogs muss angenommen werden, dass die verfassten Texte ebenfalls einem größeren Planungsprozess unterstehen als zum Beispiel Beiträge in Foren, da die Autoren darum bemüht sind, eine möglichst große Leserschaft zu erreichen und anderen Webnutzern nützliche Informationen zu vermitteln.

Alle fünf genannten Faktoren können als Maßstab fungieren, um gesprochene und geschriebene Sprache voneinander abzugrenzen. Die genannten Faktoren stehen in enger Ver-

bindung zueinander. In Abbildung 21 wird ein Überblick über alle relevanten Parameter gegeben, in die die vier untersuchten Web-Genres verortet wurden. Die Abbildung verdeutlicht die graduellen Unterschiede zwischen den vier Web-Genres im Hinblick auf deren Kommunikationsbedingungen.

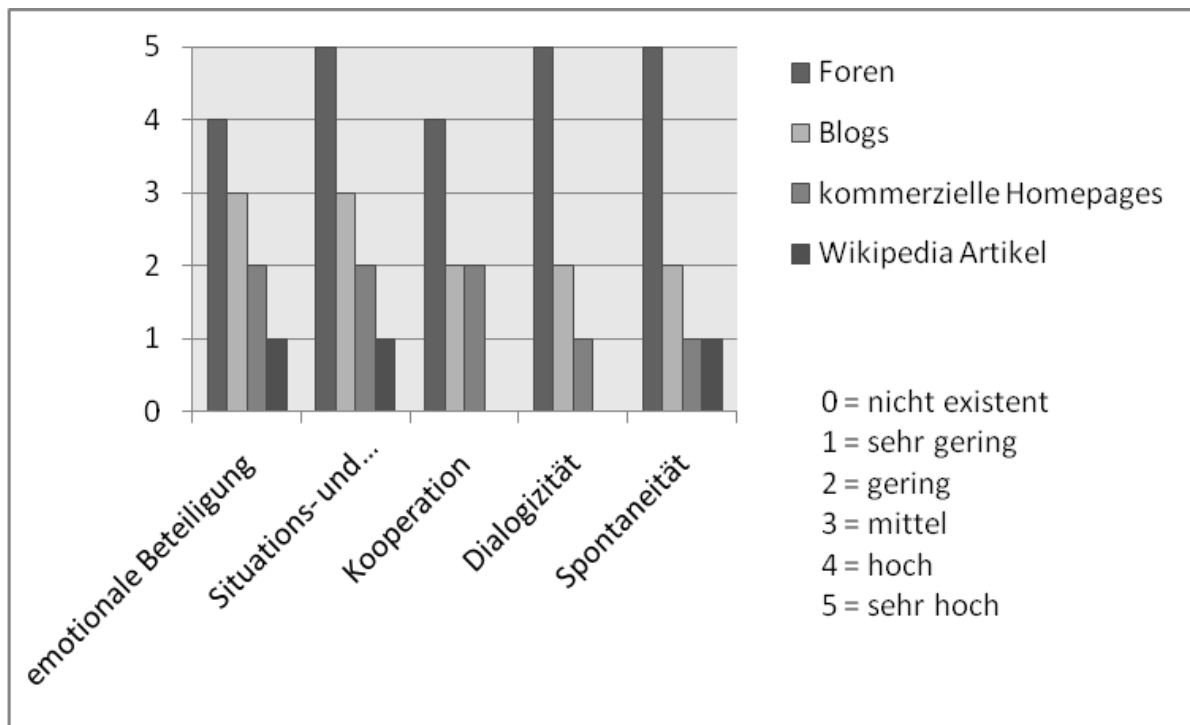


Abbildung 21: Übersicht relevanter Kommunikationsbedingungen für die Web-Genres

Es lässt sich festhalten, dass sowohl Form als auch Funktion in hohem Maße von den fünf Parametern bestimmt werden. Dabei erstaunt, dass die Graduierung der Web-Genres bezüglich der Kommunikationsbedingungen eine sehr ähnliche Verteilung aufweist, wie die Graduierung der Web-Genres anhand der Ergebnisse der quantitativen Studien. Daraus lässt sich schlussfolgern, dass Autoren von den Kommunikationsbedingungen in ihrem Schreibprozess beeinflusst werden. Die Ergebnisse deuten darauf hin, dass sich auch für andere Web-Genres mit Hilfe der Analyse von Form und Funktion graduelle Unterschiede herausstellen lassen.

8.3. Überprüfung der Hypothesen

Im Folgenden wird der Blick auf die in der Arbeit aufgeworfenen Forschungshypothesen gerichtet. Die Ergebnisse der Analysen erlauben es, die in der Arbeit aufgeworfenen Hypothesen zu betrachten und diese auf ihre Richtigkeit hin zu überprüfen.

1. Web-Genres unterscheiden sich in Bezug auf ihre formalen und funktionalen Verwendungen von Mehrworteinheiten entscheidend voneinander. Während Foren und Blogs mehr Elemente konzeptioneller Mündlichkeit enthalten, enthalten kommerzielle Homepages und Wikipedia-Artikel mehr Elemente konzeptioneller Schriftlichkeit.

Die Hypothese konnte grundsätzlich bestätigt werden. Sowohl die quantitativen als auch die qualitativen Studien haben belegt, dass sich die untersuchten Web-Genres entscheidend voneinander unterscheiden. Gleichzeitig konnte nicht bestätigt werden, dass Foren und Blogs, auf der einen und kommerzielle Homepages und Wikipedia-Artikel auf der anderen Seite oppositionelle Paare darstellen. Lediglich Foren wiesen mehrheitlich *4-grams* als Manifestationen konzeptioneller Mündlichkeit auf, während die anderen drei Web-Genres mehrheitlich *4-grams* als Manifestationen konzeptioneller Schriftlichkeit enthielten. Die Ergebnisse der Analysen zeigen, dass jedes Web-Genre gesondert betrachtet werden muss.

Verallgemeinernde Aussagen lassen sich in einem heterogenen Medium wie dem World Wide Web nicht treffen.

2. Der Sprachgebrauch im Web unterscheidet sich entscheidend von traditionellem konzeptionell mündlichem und konzeptionell schriftlichem Sprachgebrauch. Obwohl einige Web-Genres auf der formalen und funktionalen Ebene typisch mündlichen und schriftlichen Registern ähneln, können sie dennoch nicht mit ihnen gleichgesetzt werden, da die vier Web-Genres Hybride aus mündlichem und schriftlichem Sprachgebrauch darstellen.

Die Hypothese konnte entkräftet werden, da sie für drei Web-Genres zutraf, jedoch für ein Web-Genre widerlegt werden konnte. Das Web-Genre Foren teilt auf der lexikogrammatistischen Ebene grundsätzlich viele Elemente konzeptioneller Mündlichkeit mit dem Register *Conversation*, dennoch werden auch klare Unterschiede deutlich, da Foren vergleichsweise mehr *4-grams* als Manifestationen konzeptioneller Schriftlichkeit aufweisen. Die Web-Genres Blogs und kommerzielle Homepages stellen Hybride dar, da sie mehrheitlich Elemente konzeptioneller Schriftlichkeit aufweisen, aber dennoch auch Elemente konzeptioneller Münd-

lichkeit enthalten. Das Web-Genre Wikipedia-Artikel divergiert in dieser Hinsicht jedoch von den drei anderen Web-Genres. Es ähnelt in Form und Funktion fast vollständig dem Register *Academic prose*. Aus diesem Grund kann es auf konzeptioneller Ebene durchaus mit diesem Register gleichgesetzt werden. Insgesamt zeigt das Beispiel, dass jedes Web-Genre im Detail analysiert werden muss, um seine Position auf dem Kontinuum von Mündlichkeit / Schriftlichkeit zu bestimmen. Gleichzeitig belegt dieses Beispiel, dass Web-Genres eine höhere konzeptionelle Schriftlichkeit besitzen können als typisch schriftsprachliche Register (beispielsweise *Textbooks*). Somit stellen nicht alle Web-Genres Hybridwesen aus mündlichem und schriftlichem Sprachgebrauch dar, so wie es das Konzept eines *Netspeak* im Sinne von Crystal (2006) vermuten lässt.

3. Obwohl sich die Web-Genres strukturell stark voneinander unterscheiden, erlaubt es der thematische Fokus auf *Medical Surgery* / *Medical Tourism*, die einzelnen Web-Genres miteinander zu vergleichen.

Die Hypothese konnte bestätigt werden. Die Analyse der Subkorpora hat gezeigt, dass der thematische Fokus es erlaubt, Korpora miteinander zu vergleichen. Unabhängig vom Web-Genre haben die Autoren den gleichen thematischen Fokus. Dies ist im Kontext des Web nicht selbstverständlich. Es ist insbesondere in dialogischen Web-Genres, wie beispielsweise in Foren üblich, dass der Fokus auf das Thema nicht beibehalten wird. Dies muss somit dem sehr speziellen Thema geschuldet sein, das sich durch seine Ernsthaftigkeit auszeichnet. Daraus ergibt sich die Schlussfolgerung, dass der thematische Schwerpunkt bei der Korpuserstellung von entscheidender Bedeutung ist. Inwieweit Subkorpora vergleichbar wären, wenn der thematische Schwerpunkt weiter gefasst und der Diskurs weniger fokussiert wäre, müsste anhand weiterer Studien überprüft werden.

8.4. Resümee und Ausblick

Die vorliegende Arbeit zielte darauf ab, lexikogrammatische Besonderheiten des Englischen im World Wide Web herauszustellen. Dabei habe ich mich bewusst auf die Analyse von *4-grams* in vier Web-Genres (Foren, Blogs, kommerzielle Homepages und Wikipedia-Artikel) beschränkt.

In den Kapiteln 2, 3 und 4 wurden die theoretischen Hintergründe der Arbeit beschrieben. In Kapitel 2 wurde der *Lingua Franca*-Begriff, der dieser Arbeit zugrunde liegt, geklärt. Die Kommunikationssituation im Web kann nicht mit einem globalen *Lingua Franca*-Konzept im

Sinne von Seidlhofer (2001) und Jenkins (2000) gleichgesetzt werden. Es wurde deutlich gezeigt, dass der Sprachgebrauch im Web als ein Konglomerat von *ESP*-Registern im Medium Internet verstanden werden muss. In Kapitel 3 wurde der Sprachgebrauch im World Wide Web aus verschiedenen Perspektiven beleuchtet. Es wurde in diesem Kapitel deutlich herausgestellt, dass das Internet ein sehr heterogenes Medium mit höchst unterschiedlichen Web-Genres ist. Die Existenz eines *Netspeak* konnte nicht belegt werden. Stattdessen wurde der Blick auf die einzelnen Kommunikationssituationen im Web gerichtet und jedes Web-Genre gesondert betrachtet. Grundsätzlich erlaubte das Nähe-Distanz-Modell (Koch / Oesterreicher 1985, 1994) Parameter aufzuzeigen anhand derer Web-Genres unterschieden werden können. Schließlich konnten die Web-Genres im Nähe-Distanz-Modell verortet werden.

In der vorliegenden Arbeit wurde der Fokus auf die Analyse von Mehrworteinheiten in Form von *4-grams* gelegt. In Kapitel 4 wurde die Entwicklung der Forschung zu Mehrworteinheiten nachgezeichnet und die Bedeutung von Mehrworteinheiten für den Sprachgebrauch herausgestellt. Es zeigte sich dabei, dass zahlreiche Studien belegen, dass in mündlichen Registern andere Mehrworteinheiten präferiert werden als in schriftlichen Registern. Gleichzeitig wurde verdeutlicht, dass bisher keine Studien existieren, die Web-Genres bezüglich der Verwendung von *4-grams* untersuchen und dass die vorliegende Arbeit einen Teil dazu beiträgt, diesen Mangel an quantitativen und qualitativen Studien zu beheben. Abschließend wurden drei Hypothesen, die der Arbeit zugrunde liegen, formuliert.

Im Kapitel 5 wurden die methodischen Konzeptionen der einzelnen quantitativen und qualitativen Studien näher erläutert. Da die Studien allesamt auf dem Web-ELF-Korpus beruhen, wurde zudem dessen Kompilationsprozess und dessen Zusammensetzung im Detail beschrieben.

In Kapitel 6 wurden schließlich die Ergebnisse der quantitativen Studien dargelegt. Zunächst wurde mit Hilfe der formalen Analyse aufgezeigt, dass sich die Web-Genres in ihrer Form voneinander unterscheiden. Foren und Wikipedia-Artikel bildeten Gegenpole, da Foren in ihrer Form konzeptionell mündlichem Sprachgebrauch glichen, während Wikipedia-Artikel eher konzeptionell schriftlichem Sprachgebrauch entsprachen. Blogs und kommerzielle Homepages befanden sich zwischen diesen beiden Extremen, ließen sich jedoch näher am Pol der konzeptionellen Schriftlichkeit verorten. Neben der formalen Analyse wurden mehrere funktionale Analysen durchgeführt, in denen die Beobachtungen der formalen Analyse größtenteils bestätigt werden konnten. Die Ergebnisse bestätigten, dass Foren und Wikipedia-Artikel Gegenpaare bildeten und dass sich Blogs und kommerzielle Homepages zwischen diesen

beiden Extremen befanden. Insgesamt zeichneten alle quantitativen Studien ein aussagekräftiges Bild.

Um die Ergebnisse der quantitativen Studien zu untermauern, wurden in Kapitel 7 mehrere qualitative Analysen durchgeführt. Dabei wurden zum einen die Konnotationen der Muster in den vier Web-Genres miteinander verglichen, zum anderen wurden die Kollokationen bestimmter Muster miteinander verglichen. Die Ergebnisse bestätigten die Beobachtungen, die bereits für die quantitativen Analysen gemacht werden konnten. Foren und Wikipedia-Artikel bilden Gegenpole, Blogs und kommerzielle Homepages standen zwischen den Polen, wiesen jedoch mehr Elemente konzeptioneller Schriftlichkeit auf. Es konnten ebenso lexikogrammatistische Besonderheiten für die einzelnen Web-Genres herausgestellt werden, so zum Beispiel, dass in Foren und Wikipedia-Artikeln unterschiedliche Kollokationen bevorzugt werden.

In Kapitel 8 wurde zunächst ein Überblick über die Ergebnisse aller durchgeführten Studien gegeben. Es wurden dabei Erklärungsansätze für die sprachliche Heterogenität der Web-Genres formuliert. Es stellte sich heraus, dass einige Kommunikationsbedingungen existieren, die einen maßgeblichen Einfluss auf den Sprachgebrauch in den Web-Genres nehmen. Im Anschluss wurden die Forschungshypothesen, die der Arbeit zugrunde liegen, überprüft. Es konnte belegt werden, dass sich die untersuchten Web-Genres sowohl auf formaler als auch auf funktionaler Ebene voneinander unterscheiden. Gleichzeitig konnte gezeigt werden, dass das Web-Genre Wikipedia-Artikel in Form und Funktion fast vollständig dem schriftlichen Register *Academic prose* entsprach. Dies belegte, dass Web-Genres sich nicht per se von mündlichen und schriftlichen Registern unterscheiden müssen. Zudem wurde aufgezeigt, dass anhand eines eingeschränkten thematischen Fokus unterschiedliche Web-Genres miteinander verglichen werden können.

Die vorliegende Arbeit hat deutlich gemacht, dass der Sprachgebrauch im Web nicht verallgemeinert werden kann. Es gibt vor dem Hintergrund höchst unterschiedlicher Web-Genres keine Vereinheitlichungstendenzen. Meines Erachtens lässt es die Heterogenität der Web-Genres nicht zu, dass sich ein gemeinsamer Kern aus sprachlichen Charakteristika herausbilden kann, der als kleinster gemeinsamer Nenner aller Web-Genres fungiert. Der Sprachgebrauch im Web muss stattdessen als ein *Community of practice*-Sprachgebrauch aufgefasst werden, der völlig frei ist von kulturellen Werten und vom sprachlichen Hintergrund der Webnutzer. Der Sprachgebrauch im Web kann somit nicht als Lingua Franca-Sprachgebrauch im Sinne von Seidlhofer (2001) und Jenkins (2007) verstanden werden. Er wird stattdessen maßgeblich beeinflusst vom thematischen Fokus und den individuellen Webnutzern. In Foren

hat die Zusammensetzung der Webnutzer einen entscheidenden Einfluss auf den Sprachgebrauch. Beispielsweise hat die Ernsthaftigkeit des Themas *Medical Surgery / Medical Tourism* dazu geführt, dass der thematische Fokus in den untersuchten Foren beibehalten wurde. Alle Texte des Web-ELF-Korpus mussten grundsätzlich so verfasst sein, dass sie von einer globalen Leserschaft rezipiert werden können. Das Ziel aller untersuchten Web-Genres war es, Rezipienten weltweit zu informieren oder sie von einem chirurgischen Eingriff zu überzeugen. Dies hat sicherlich einen Einfluss auf den Sprachgebrauch der Autoren gehabt.

Inwiefern sich im Web *ESP*-Register im Sinne von Brutt-Griffler (1998) etablieren, müsste anhand weiterer Studien überprüft werden. Es ist durchaus denkbar, dass sich in bestimmten Web-Genres stabile sprachliche Merkmale herausbilden können. Dies hängt jedoch mit den individuellen Web-Genres zusammen. Stabile sprachliche Merkmale können sich meines Erachtens nur in bestimmten Web-Genres ausbilden. Unter anderem zeigte sich, dass in Wikipedia-Artikeln eine wesentlich geringere Anzahl unterschiedlicher *4-gram-Types* verwendet wurden. Dies kann ein Indikator dafür sein, dass der Sprachgebrauch dort wesentlich homogener ist als beispielsweise in Foren, in denen eine wesentlich höhere Anzahl unterschiedlicher *4-gram-Types* verwendet wurde. Es stellt sich somit die Frage, ob die Homogenisierungstendenz ein potenzielles Anzeichen dafür sein kann, dass sich in Wikipedia-Artikeln ein Sprachgebrauch mit stabilen sprachlichen Merkmalen herausbildet. Um diese Vermutung zu stützen, müssten jedoch weitere Studien durchgeführt werden.

Eine Reduktion des Sprachgebrauchs konnte anhand der vorliegenden Ergebnisse nicht beobachtet werden. Es können lediglich Parallelen gezogen werden zu typisch mündlichem / schriftsprachlichem Sprachgebrauch. Foren ähneln in Form und Funktion eher mündlichem, Blogs, kommerzielle Homepages und Wikipedia-Artikel eher schriftlichem Sprachgebrauch.

Anhand dieser Arbeit wurde verdeutlicht, dass es keinen homogenen Web-Genre-übergreifenden Sprachgebrauch in Form eines *Netspeak* gibt. Das Web besteht aus zahlreichen Web-Genres, die sich aufgrund ihrer Kommunikationsbedingungen und ihrer Kommunikationsziele stark voneinander unterscheiden. Dabei gibt es bestimmte Faktoren, die einen maßgeblichen Einfluss auf den Sprachgebrauch nehmen. Der Sprachgebrauch im Web muss ebenso wie der nicht-virtuelle Sprachgebrauch bewertet werden. Es ist nicht sinnvoll, für die unterschiedlichen Ausprägungen von medial mündlichen und medial schriftlichen Sprachgebrauch einen gemeinsamen sprachlichen Kern zu formulieren. Web-Genres lassen sich ebenso wie mündliche und schriftsprachliche Register voneinander unterscheiden. Sie können anhand von for-

malen und funktionalen Eigenschaften auf einem Kontinuum von Mündlichkeit / Schriftlichkeit verortet werden.

In der vorliegenden Studie wurde eine beschränkte Auswahl an Web-Genres in einem sehr speziellen Themengebiet analysiert. Trotz dieser Einschränkung konnten aussagekräftige Ergebnisse erarbeitet werden. Für zukünftige Studien würde es sich anbieten, einen anderen thematischen Fokus zu wählen, um die Ergebnisse mit denen der vorliegenden Analysen zu vergleichen. Dabei kann überprüft werden, ob sich ähnliche sprachliche Tendenzen bestätigen lassen und ob das Thema einen größeren Einfluss auf das Sprachverhalten nehmen kann.

Ebenso erscheint es lohnenswert, eine diachrone Studie durchzuführen, um zu beobachten, inwieweit die gefundenen Ergebnisse sich bestätigen lassen und um zu überprüfen, ob ein sprachlicher Wandel in den einzelnen Web-Genres stattgefunden hat. Es wäre ebenfalls lohnenswert, andere Web-Genres anhand ihrer formalen und funktionalen Eigenschaften zu untersuchen und die Ergebnisse mit den vorliegenden Resultaten der Analysen der Web-Genres zu vergleichen.

Der Einfluss des Webs auf den Sprachgebrauch wird weiterhin ein bestimmendes Thema sein, da es den Alltag der meisten Menschen in hohem Maße beeinflusst. In Debatten über den Einfluss des Webs werden Sprachpuristen weiterhin generalistische Aussagen darüber machen, dass der Sprachgebrauch im Web den Niedergang der Hochsprache befördert. Das Ziel dieser Arbeit war es, nachzuweisen, dass generalistische Aussagen über *den* Sprachgebrauch im Web nicht zulässig sind. Ich hoffe, dass die Ergebnisse meiner Arbeit ihren Teil dazu beitragen, dass der Sprachgebrauch im Web differenzierter betrachtet wird und dass der Erforschung einzelner Kommunikationsformen des Webs in Zukunft mehr Aufmerksamkeit geschenkt wird.

Literaturverzeichnis

- Àgel, V. / Hennig, M. (2006): "Theorie des Nähe- und Distanzsprechens". In V. Àgel / M. Hennig (Hrsg.) *Grammatik aus Nähe und Distanz*. Tübingen: Niemeyer, 3-31.
- Akkinaso, F. (1982): "On the differences between spoken and written language". *Language and Speech* 25(2), 97-125.
- Alexa (2012): "Global top sites". http://www.alex.com/site/ds/top_sites (30.12.2012).
- Altenberg, B. (1990): "Speech as linear composition". In G. Caie / K. Haastrup / A. L. Jakobsen / J. E. Nielsen / J. Sevaldsen / H. Specht / A. Zettersen (Hrsg.) *Proceedings from the Fourth Nordic Conference for English Studies, Vol. 1*. Copenhagen: Dept. of English, University of Copenhagen, 133-143.
- Altenberg, B. (1998): "On the phraseology of spoken English: the evidence of recurrent word-combinations". In A. Cowie (Hrsg.) *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press, 101-122.
- Androutsopoulos, J. / Schmidt, G. (2001): "SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe". *Zeitschrift für Angewandte Linguistik* 36, 49-79.
- Ammon, U. (1998): *Ist deutsch noch internationale Wissenschaftssprache? Englisch auch für die Lehre an den deutschsprachigen Hochschulen*. Berlin, New York: Walter de Gruyter.
- Bamgbose, A. (1998): "Torn between the norms: innovations in world Englishes". *World Englishes* 17(1), 1-14.
- Baron, N. S. (2000): *Alphabet to Email. How Written English Evolved and Where It's Heading*. London: Routledge.
- Baron, N. S. (2005): "The future of written culture. Envisioning language in the new millennium". *Ibérica, Journal of the European Association of Languages for Specific Purposes, Special Issue 9*, 7-31.
- Baron, N. S. (2008): *Always On: Language in an Online and Mobile World*. Oxford: OUP.
- Baron, N. S. (2010): "Discourse structures in Instant Messaging: the case of utterance breaks". *Language@Internet*, 7, Article 3.
- Baroni, M. / Kilgarriff, A. (2006): "Large linguistically-processed Web corpora for multiple languages.". Online verfügbar: <http://acl.ldc.upenn.edu/E/E06/E06-2001.pdf> (30.12.2012).
- Barotchi, M. (1994): "Lingua franca". *The Encyclopedia of Language and Linguistics*, 4. Oxford: 2211.
- Baumgarten, N. (2011): "English as a lingua franca – (where) does it exist?". Gastvorlesung: English Department der Justus-Liebig-Universität Gießen, 3. Februar 2011.
- Bell, M. (2007): "The transformation of the encyclopedia: a textual analysis and comparison of the Encyclopaedia Britannica and Wikipedia". *Unpublished master's thesis, Ball State University, Muncie, IN*. Online verfügbar: http://storygeek.com/?page_id=85 (30.12.2012).

- Beneke, J. (1991): "Englisch als lingua franca oder als Medium interkultureller Kommunikation". In R. Grebing (Hrsg.) *Grenzenloses Sprachenlernen. Festschrift für Reinhold Freudenstein*. Berlin: Cornelsen, 54-66.
- Bernaisch, T. / Koch, C. / Schilk, M. / Mukherjee, J. (2011): *Manual to the South Asian Varieties of English (SAVE) Corpus*. Gießen: Justus-Liebig-Universität, Department of English.
- Berners-Lee, T. / Hendler, J. / Lassila, O. (2001): "The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *Scientific American* 284(5), 34–43.
- Biber, D. (1988): *Variation Across Speech and Writing*. Cambridge: CUP.
- Biber, D. (2006): *University Language: A Corpus-Based Study of Spoken and Written University Registers*. Amsterdam: John Benjamins.
- Biber, D. (2009): "A corpus-driven approach to formulaic language in English. Multi-word patterns in speech and writing". *International Journal of Corpus Linguistics* 14(3), 275–311.
- Biber, D. / Conrad S. (2009): *Register, Genre, and Style*. Cambridge: CUP.
- Biber, D. / Conrad, S. / Cortes, V. (2003): "Lexical bundles in speech and writing: an initial taxonomy". In A. Wilson / P. Rayson / T. McEnery (Hrsg.) *Corpus Linguistics by the Lune*. Frankfurt am Main: Peter Lang, 71-93.
- Biber, D. / Conrad, S. / Cortes, V. (2004): "If you look at...: lexical bundles in university teaching and textbooks". *Applied Linguistics* 25, 371-405.
- Biber, D. / Johansson, S. / Leech, G. / Conrad, S. / Finegan, E. (1999): *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, D. / Kurjian, J. (2007): "Towards a taxonomy of web registers and text types: a multidimensional analysis". In M. Hundt / N. Nesselhauf / C. Biewer (Hrsg.) *Corpus Linguistics and the Web*. Amsterdam, New York: Rodopi, 109-131.
- Biewer, C. (2011): "Modal auxiliaries in second language varieties of English". In M. Hundt / J. Mukherjee (Hrsg.) *Exploring Second-language Varieties of English and Learner Englishes*. Amsterdam: John Benjamins, 7-33.
- Bittner, J. (2003): *Digitalität, Sprache, Kommunikation. Eine Untersuchung zur Medialität von digitalen Kommunikationsformen und Textsorten und deren varietätenlinguistischer Modellierung*. Berlin: Erich Schmidt.
- Bloomfield, L. (1933): *Language*. London: Allen & Unwin.
- Bolinger, D. (1976): *Aspects of language* (2. Edition). New York: Harcourt Brace Jovanovich.
- Breiteneder, A. (2005): "The naturalness of English as a European lingua franca: the case of the 'third person –s'". *Vienna English Working Papers* 14(2), 3-26.
- Breiteneder, A. (2009): *English as a Lingua Franca in Europe. A Natural Development*. Saarbrücken: VDM-Verlag Müller.
- Bruthiaux, P. (2003): "Squaring the circles: issues in modeling English worldwide". *International Journal of Applied Linguistics* 13(2), 159-178.

- Brutt-Griffler, J. (1998): "Conceptual questions in English as a world language: taking up an issue". *World Englishes* 17(3), 381-392.
- Brutt-Griffler, J. (2002): *World English*. Clevedon: Multilingual Matters.
- Brutt-Griffler, J. (2008): "'Who do you think you are, where do you think you are?': language policy and the political economy of English in South Africa". In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom* (2. Edition). Tübingen: Narr, 27-40.
- Canagarajah, A. S. (2006): "Negotiating the local in English as a lingua franca". *Annual Review of Applied Linguistics* 26, 197-218.
- Chesher, C. (2005): "Blogs and the crisis of authorship". *Präsentation auf der Blogtalk Downunder Konferenz in Sydney, Australien*, 19.-22. Mai 2005. Online verfügbar: http://incsub.org/blogtalk/?page_id=40 (30.12.2012).
- Claridge, C. (2007): "Constructing a corpus from the web: message boards". In M. Hundt / N. Nesselhauf / C. Biewer (Hrsg.) *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 87-108.
- Clarke, A. / Gatineau, M. / Grimaud, O. / Royer-Devaux, S. / Wyn-Roberts, N. / Le Bis, I. / Lewison, G. (2007): "A bibliometric overview of public health research in Europe". *European Journal of Public Health Suppl.* 1, 43-49.
- Code, C. (1994): "Speech automatism production in aphasia". *Journal of Neurolinguistics* 8(2), 135-148.
- Cowie, A. P. (1988): "Stable and creative aspects of vocabulary use". In R. Carter / M. McCarthy (Hrsg.) *Vocabulary and language teaching*. London, New York: Longman, 126-139.
- Conklin, K. / Schmitt, N. (2008): "Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers?". *Applied Linguistics* 29, 72-89.
- Conrad, S. / Biber D. (2001): *Variation in English: Multi-Dimensional Studies*. London: Longman.
- Croft, W. (2001): *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: OUP.
- Crossley, S. A. / Louwse M. (2007): "Multi-dimensional register classification using bigrams". *International Journal of Corpus Linguistics* 12(4), 453-478.
- Crystal, D. (1995): *The Cambridge encyclopedia of the English language*. Cambridge: CUP.
- Crystal, D. (1997): *English as a Global Language*. Cambridge: CUP.
- Crystal, D. (2001): *Language and the Internet*. Cambridge: CUP.
- Crystal, D. (2003a): *English as a Global Language* (2. Edition). Cambridge: CUP.
- Crystal, D. (2003b): *The Cambridge Encyclopedia of the English language* (2. Edition). Cambridge: CUP.
- Crystal, D. (2006): *Language and the Internet* (2. Edition). Cambridge: CUP.
- Crystal, D. (2011): *Internet Linguistics*. Routledge: New York.

- Danet, B. / Herring, S. C. (2007): *The Multilingual Internet: Language, Culture and Communication Online*. New York, Oxford: OUP.
- Davis, B. H. & Brewer, J. P. (1997): *Electronic Discourse: Linguistic Individuals in Virtual Space*. New York: State University of New York Press.
- De Cock, S. (2004): "Preferred sequences of words in NS and NNS speech". *Belgian Journal of English Language and Literatures*, 225-246.
- Divjak, D. / Gries S. T. (2008): "Clusters in the mind? Converging evidence from near synonymy in Russian". *The Mental Lexicon* 3(2), 188-213.
- Drieman, G. (1962): "Differences between *spoken* and *written* language: quantitative approach". *Acta Psychologica* 20, 36-57.
- Dröschel, Y. (2011): *Lingua Franca English. The Role of Simplification and Transfer*. Bern: Peter Lang.
- Durrant, P. / Schmitt, N. (2009): "To what extent do native and non-native writers make use of collocations?". *International Review of Applied Linguistics* 47, 157-177.
- Dürscheid, C. (2003a): "Netzsprache – ein neuer Mythos". Online verfügbar: www.ds.uzh.ch/lehrstuhlduerscheid/docs/netzsprache.pdf (30.12.2012).
- Dürscheid, C. (2003b): "Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit. Theoretische und empirische Probleme". *Zeitschrift für Angewandte Linguistik* 38, 37-56.
- Eckert, P. / McConnell-Ginet, S. (1992): "Think Practically and Look Locally: Language and Gender as Community-Based Practice". *Annual Review of Anthropology* 21, 461-490.
- Elia, A. (2007): "An analysis of Wikipedia digital writing". Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, 16-21. Online verfügbar: http://www.sics.se/jussi/newtext/working_notes/04_elia_new.pdf (30.12.2012).
- Ellis, N. C. / Simpson-Vlach, R. / Maynard, C. (2008): "Formulaic language in native and second-language speakers: Psycholinguistics, Corpus Linguistics, and TESOL". *TESOL Quarterly* 42(3), 375-396.
- Ellis, N. C. / Simpson-Vlach R. (2009): "Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education". *Corpus Linguistics and Linguistic Theory* 5, 61–78.
- Elspaß, S. (2002): "Alter Wein und neue Schläuche? Briefe der Wende zum 20. Jahrhundert und Texte der neuen Medien – ein Vergleich". In U. Schmitz / E. L. Wyss (Hrsg.) *Briefkultur im 20. Jahrhundert*. Oldenburg (= OBST 64), 7-31.
- Erman, B. (2009): "Formulaic language from a learner perspective". In R. Corrigan / E. Moravcsik / H. Ouai (Hrsg.) *Formulaic Language Vol. 2. Acquisition, Loss, Psychological Reality, and Functional Explanations*. Amsterdam: John Benjamins, 323-346.
- Erman, B. / Warren B. (2000): "The idiom principle and the open choice principle". *Text* 20(1), 29-62.
- Esser, J. (1993): *English Linguistics Stylistics*. Tübingen: Max Niemeyer.

- Esser, J. (2000): "Medium-transferability and presentation structure in speech and writing". *Journal of Pragmatics* 32, 1523-1538.
- Esser, J. (2002): "Is netspeak a 'third medium'? Presenting speech and writing in old and new media". In S. Scholz / M. Klages/ E. Hantson / U. Römer (Hrsg.) *Language: Context and Cognition: Papers in Honour of Wolf-Dietrich Bald's 60th Birthday*. München: Langenscheidt-Longman, 83-97.
- Esser, J. (2006): *Presentation in Language: Rethinking Speech and Writing*. Tübingen: Max Niemeyer.
- Ferguson, G. (2009): "Issues in researching English as a lingua franca: a conceptual enquiry". *International Journal of Applied Linguistics* 19(2), 117-135.
- Fillmore, C. / Kay, P. / O' Connor, M. (1988): "Regularity and idiomaticity in grammatical constructions. The case of let alone". *Language* 64, 501-538.
- Firth, A. (1996): "The discursive accomplishment of normality: on 'lingua franca' English and conversation analysis". *Journal of Pragmatics* 26, 237-259.
- Fittkau & Maaß Consulting (2008): "Second Life – die aktiven Nutzer". <http://www.w3b.org/unterhaltung/second-life-die-aktiven-nutzer.html> (30.12.2012).
- Fletcher, W. (2003-2012): *PIE: Phrases in English*. [Database.]. Online verfügbar: <http://phrasesinenglish.org> (30.12.2012).
- Forchini, P. / Murphy, A. (2008): "N-grams in comparable specialized corpora". *International Journal of Corpus Linguistics* 13(3), 351-367.
- Foster, P. (2001): "Rules & routines: a consideration of their role in the task-based language production of native and non-native speakers". In M. Bygate / P. Skehan / M. Swain (Hrsg.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. London, New York: Longman, 75-94.
- Fritz, G. (2011): "Texttypen in wissenschaftlichen Blogs. Eine exemplarische Analyse am Beispiel des *Language Log*". In T. Gloning / G. Fritz (Hrsg.) *Digitale Wissenschaftskommunikation – Formate und ihre Nutzung*. Gießen: Gießener Elektronische Bibliothek, 205-286.
- Giltrow, J. / Stein, D. (2009): *Genres in the Internet: Issues in the Theory of Genre*. Amsterdam: John Benjamins.
- Gnutzmann, C. / Intemann, F. (2008): "Introduction: The globalisation of English. Language, politics and the English language classroom". In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom* (2. Edition). Tübingen: Narr, 14-29.
- Götz, S. / Schilk, M. (2011): "Formulaic sequences in spoken ENL, ESL and EFL: focus on British English, Indian English and learner English of advanced German learners". In J. Mukherjee / M. Hundt (Hrsg.) *Exploring Second-Language Varieties of English: Bridging a Paradigm Gap*. Amsterdam: John Benjamins, 79-100.
- Goldberg, A. (1995): *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A. (2006): *Constructions at Work. The Nature of Generalization in Language*. Oxford: OUP.

- Goldberg, A. / Jackendoff, R. (2004): "The English resultative as a family of constructions". *Language* 80(3), 32–568.
- Graddol, M. (1997): *The Future of English*. Online verfügbar: www.britishcouncil.org/learning-elt-future.pdf (30.12.2012).
- Graddol, M. (2006): *English Next*. Online verfügbar: www.britishcouncil.org/learning-research-english-next.pdf (30.12.2012).
- Granger, S. (1998): "Prefabricated patterns in advanced EFL writing: Collocations and formulae". In: A.P. Cowie (Hrsg.) *Phraseology: Theory, analysis and application*. Oxford: Oxford University Press, 145-160.
- Gries, S. T. (2010): "Bigrams in registers, domains, and varieties: a bigram gravity approach to the homogeneity of corpora". In *Proceedings of Corpus Linguistics 2009*, University of Liverpool.
- Gries, S. T. / Newman, J. / Cyrus, S. (2011): "N-grams and the clustering of registers". *Empirical Language Research* 5 (1).
- Günther, U. / Wyss, E. L. (1996): "E-Mail-Briefe – eine neue Textsorte zwischen Mündlichkeit und Schriftlichkeit". In E. W. B. Hess-Lüttich / W. Holly / U. Püschel (Hrsg.) *Textstrukturen im Medienwandel*. Frankfurt am Main / Berlin: Peter Lang, 61-86.
- Häcki Buhofer, A. (2000): "Mediale Voraussetzungen: Bedingungen von Schriftlichkeit allgemein". In K. Brinker / G. Antos / W. Herinemann / S. Sager (Hrsg.) *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin, New York: Walter de Gruyter, 251-261.
- Haegeman, P. (2002): "Foreigner talk in lingua franca business telephone calls". In K. Knapp / C. Meierkord (Hrsg.) *Lingua Franca Communication*. Frankfurt am Main: Peter Lang, 137-162.
- Hakuta, K. (1974): "Prefabricated patterns and the emergence of structure in second language acquisition". *Language Learning* 24, 287-297.
- Halliday, M. A. K. (1987): "Spoken and written modes of meaning". In R. Horowitz / S. J. Samuels (Hrsg.) *Comprehending Oral and Written Language*. San Diego: Academic Press, 55-82.
- Halliday, M. A. K. (1989): *Spoken and Written Language*. Oxford: OUP.
- Halliday, M. A. K. (2001): "Literacy and linguistics: relationships between spoken and written language". In A. Burns / C. Coffin (Hrsg.) *Analysing English in a Global Context: A Reader*. London: Routledge, 181-193.
- Heller, D. B. (2013): *Automatisierte Analyse von Mehrworteinheiten in englischsprachigen Korpora: eine korpuslinguistische Studie*. Unveröffentlichte Magisterarbeit, Justus-Liebig-Universität Gießen.
- Herring, S. C. (1996): *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives. Pragmatics and Beyond series*. Amsterdam: John Benjamins.
- Herring, S. C. / Kouper, I. / Paolillo, J. C. / Scheidt, L. A. / Tyworth, M. / Welsch, P. (2005): "Conversations in the blogosphere: an analysis 'from the bottom up'". In *Proceedings of the 38th Hawai'i International Conference on System Sciences*. Los Alamitos: IEEE

- Press, 1-11. Online verfügbar: www.blogninja.com/hicss05.blogconv.pdf (30.12.2012).
- Höflich, J. R. (1997): "Zwischen massenmedialer und technisch vermittelter interpersonalen Kommunikation – der Computer als Hybridmedium und was die Menschen damit machen". In K. Beck / G. Voew (Hrsg.) *Computernetze – ein Medium öffentlicher Kommunikation?*. Berlin: Spiess, 85-104.
- Hoey, M. (2005): *Lexical Priming: A New Theory of Words and Language*. London, New York: Routledge.
- Hoffmann, S. (2007): "Processing Internet-derived text – creating a corpus of Usenet messages". *Literary & Linguistic Computing* 22(2), 151-165.
- Hoffmann, S. / Evert, S. / Smith, N. / Lee, D. / Berglund Prytz, Y. (2008): *Corpus Linguistics with BNCweb - A Practical Guide*. Frankfurt am Main: Peter Lang.
- House, J. (1999): "Misunderstanding in intercultural communication: interactions in English as lingua franca and the myth of mutual intelligibility". In C. Gnutzmann (Hrsg.) *Teaching and Learning English as a Global Language*. Tübingen: Stauffenburg, 73-93.
- House, J. (2006): "Unity in diversity: English as a lingua franca for Europe". In C. Leung / J. Jenkins (Hrsg.) *Reconfiguring Europe: the Contributions of Applied Linguistics*. London: Equinox, 87-103.
- Hüllen, W. (1987): "Englisch als Fremdsprache und als lingua franca". In W. Hüllen (Hrsg.) *Englisch als Fremdsprache. Beiträge zur Theorie des Englischunterrichts an deutschen Schulen*. Tübingen: Narr, 54-60.
- Hunston, S. / Francis, G. (2000): *Pattern grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.
- Hyland, K. (2006): "A tribute to John Swales". *Journal of English for Academic Purposes* 5(3), 171-173.
- Hyland, K. (2008): "Academic clusters: text patterning in published and postgraduate writing". *International Journal of Applied Linguistics* 1(1), 41-62.
- Hymes, D. H. (1968): *The Ethnography of Speaking*. Washington, DC: Anthropological Society of Washington, 13-53. Nachdruck in J. A. Fishman (Hrsg.) (1968) *Readings in the Sociology of Language*. The Hague, Paris: Mouton, 99-138.
- James, A. (2008): "The challenges of the lingua franca: English in the world and types of variety". In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom* (2. Edition). Tübingen: Narr, 133-144.
- Jenkins, J. (2000): *The phonology of English as an International Language*. Oxford: OUP.
- Jenkins, J. (2003): *World Englishes – A Resource Book for Students*. London: Routledge.
- Jenkins, J. (2007): *English as a Lingua Franca. Attitude and Identity*. Oxford: OUP.
- Jenkins, J. (2009): *World Englishes – A Resource Book for Students*. 2nd ed. London: Routledge.
- Jespersen, O. (1924 / 1976): "Living grammar". Reprinted in D. D. Bornstein (Hrsg.) *The Theory of Grammar*. Cambridge, MA: Winthrop Publishers, 82-93.

- Jiang, N. / Nekrasova, T. (2007): "The processing of formulaic sequences by Second Language Speakers". *The Modern Language Journal* 91, 433-445.
- Jucker, A. (2004): "Gutenberg und das Internet. Der Einfluss von Informationsmedien auf Sprache und Sprachwissenschaft". *Sprache@web* 40. Online verfügbar: <http://www.mediensprache.net/networx/networx-40.pdf> (30.12.2012).
- Kachru, B. B. (1985): "Standards, codification and sociolinguistic realism: the English language in the outer circle." In R. Quirk (Hrsg.) *Englishes in the World*. Cambridge: CUP, 11-34.
- Kachru, B. B. (1988): "The sacred cows of English". *English Today* 16, 3-8.
- Kachru, B. B. (1992): "The Second Diaspora of English". In T. W. Machen / C. T. Scott (Hrsg.) *English in its Social Contexts: Essays in Historical Sociolinguistics*, Oxford: OUP, 230-252.
- Kaplan, R. B. (2001): "English - the Accidental language of science?" In U. Ammon (Hrsg.) *The Dominance of English as a Language of Science. Effects on Other Languages and Language Communities*, 3-26.
- Kirkpatrick, A. (2006): "Which model of English: native-speaker, nativized or lingua franca?". In R. Rubdy / M. Saraceni (Hrsg.) *English in the World. Global Rules, Global Roles*. London: Continuum, 71-83.
- Kirkpatrick, A. (2007): *World Englishes: Implications for International Communication and English Language Teaching*. Cambridge: CUP.
- Knapp, K. (1991): "Zurück vor Babel? Zur Zukunft der Weltsprache Englisch". *Zeitschrift für Literaturwissenschaft und Linguistik* 79, 18-42.
- Knapp, K. (2002): "The fading out of the non-native speakers: native speaker dominance in lingua-franca-situations". In K. Knapp / C. Meierkord (Hrsg.) *Lingua Franca Communication*. Frankfurt am Main: Peter Lang, 217-244.
- Koch, P. / Oesterreicher, W. (1985): "Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte". *Romanistisches Jahrbuch* 36, 15-43.
- Koch, P. / Oesterreicher, W. (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübingen: Niemeyer.
- Koch, P. / Oesterreicher, W. (1994): "Schriftlichkeit und Sprache". In H. Günther / O. Ludwig (Hrsg.) *Schrift und Schriftlichkeit. Writing and Its Use. Ein interdisziplinäres Handbuch internationaler Forschung. An Interdisciplinary Handbook of International Research* (1. Halbband). Berlin, New York: de Gruyter, 587-604.
- Koch, P. / Oesterreicher, W. (2007): "Schriftlichkeit und kommunikative Distanz". *Zeitschrift für Germanistische Linguistik* 35, 346-375.
- Kopaczyk, J. (2012): "Applications of the lexical bundle method in historical corpus research." In P. Pezik (Hrsg.) *Corpus Data across Languages and Disciplines*. Frankfurt am Main: Peter Lang, 83-95.
- Künstler, V. / Mukherjee, J. / Saage, S. (2012): *Manual for the Web – English as a lingua franca (Web-ELF) corpus*.

- Künstler, V. / Valsecchi, M. / Saage, S. / Gegenfurtner, K. / Mukherjee, J. (in Vorb.): "An eye-tracking study of processing lexical bundles by native and non-native speakers", 1-21.
- Kuiper, K. (1996): *Smooth talkers: the linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Lawrence Erlbaum.
- Langacker, R. W. (1987): *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. (1991): *Foundations of Cognitive Grammar, Volume 2: Descriptive Application*. Stanford: Stanford University Press.
- Langacker, R. W. (2008): *Cognitive Grammar: A Basic Introduction*. New York: OUP.
- Lave, J. / Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*.
- Leech, G. (2005): "Adding linguistic annotation." In M. Wynne, M. (Hrsg.) *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow.
- Li, D. C. S. (2010): "When does an unconventional form become an innovation?" In A. Kirkpatrick (Hrsg.) *The Routledge handbook of world Englishes*. Abingdon: Routledge, 617-633.
- Lyons, J. (1968): *An Introduction to Theoretical Linguistics*. Cambridge: CUP.
- Lyons, J. (1981): *Language and Linguistics: An Introduction*. Cambridge: CUP.
- Mair, C. (1997): "The spread of the *going-to*-future in written English: a corpus-based investigation into language change in progress." In R. Hickey / S. Puppel (Hrsg.) *Language History and Linguistic Modelling: A Festschrift for Jacek Fisiak*. Berlin: Mouton de Gruyter, 1537-1543.
- Maley, A. (2009): "ELF: a teacher's perspective". *Language and Intercultural Communication* 9(3), 187-200.
- Maley, A. (2010): "The reality of EIL and the myth of EFL". In C. Gagliardi / A. Maley (Hrsg.) *EIL, EFL, Global English: Teaching and Learning Issues*. Peter Lang: Frankfurt am Main, 25-44.
- Mauranen, A. (2003): "The corpus of English as a lingua franca in academic settings". *TESOL Quarterly* 27:3, 113-127.
- Mauranen, A. (2005): "English as a lingua franca – an unknown language?". In G. Cortese / A. Duszak (Hrsg.) *Identity, Community, Discourse: English in Intercultural Settings*. Frankfurt am Main: Peter Lang, 269-293.
- McArthur, T. (1987): "The English languages?". *English Today* 11(3:3), 9-13.
- McArthur, T. (1998): *The English Languages*. Cambridge: CUP.
- McCrone, J. (1999): *States of mind*. New Scientist 2178, 20th March: 30-33.
- McEnery, T. / Hardie, A. (2011): *Corpus Linguistics: Method, Theory and Practice*. Cambridge: CUP.
- McLuhan, M. (1962): *The Gutenberg Galaxy*. Toronto: University of Toronto Press.
- McLuhan, M. / Gordon, W. T. (1964 / 2003): *Understanding Media. The Extensions of Man*. Corte Madera: Gingko Press.
- McMahon, A. M. S. (1994): *Understanding Language Change*. Cambridge: CUP.

- McNeill, L. (2005): "Genre under construction: the diary on the Internet".
Language@Internet 1/2005. Online verfügbar:
http://www.languageatinternet.de/articles/2005/120/Genre10_06DOULOS.rtf.pdf/
 (30.12.2012).
- Meierkord, C. (1996): *Englisch als Medium der interkulturellen Kommunikation: Untersuchungen zum Non-Native/Non-Native-Sprecher-Diskurs*. Frankfurt: Peter Lang.
- Meierkord, C. (2004): "Syntactic variation in interactions across international Englishes".
English World-Wide 25(1), 109-132.
- Meierkord, C. (2005): "Interactions across Englishes and their lexicon". In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom*. Tübingen: Narr, 89-104.
- Meise-Kuhn, K. (1998): "Zwischen Mündlichkeit und Schriftlichkeit: Sprachliche und konversationelle Verfahren in der Computerkommunikation". In A. Brock / M. Hartung (Hrsg.) *Neuere Entwicklungen in der Gesprächsforschung. Vorträge der 3. Arbeitstagung des Pragmatischen Kolloquiums Freiburg*. Tübingen: Narr, 213-235.
- Meißner, F. J. (2005): "Soll das Mammut Elefanten und Mücken fressen? Zur Weiterentwicklung des Sprachenunterrichts in der Europäischen Union". In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom*. Tübingen: Narr, 129-146.
- Mesthrie, R. / Bhatt, R. (2008): *World Englishes: The Study of New Linguistic Varieties*. Cambridge: CUP.
- Meyerhoff, M. (2004): "Communities of practice". In J. K. Chambers / P. Trudgill / N. Schilling-Estes (Hrsg.) *The Handbook of Language Variation and Change*. Oxford: Wiley-Blackwell, 526-548.
- Modiano, M. (1999): "Standard English(es) and educational practices for the world's lingua franca". *English Today* 15(4), 3-13.
- Mollin, S. (2006a): "English as a Lingua Franca: A new variety in the new expanding circle?". *Nordic Journal of English Studies* 5:2, 41-57.
- Mollin, S. (2006b): *Euro-English. Assessing Variety Status*. Tübingen: Narr.
- Moon, R. (1998): *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Mukherjee, J. (2005): "English as a global pidgin (EGP) in academia: some prolegomena". In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom*. Tübingen: Narr, 107-115.
- Mukherjee, J. (2008): "Tradition und Innovation im englischen Internet Sprachgebrauch: konzeptionelle Überlegungen zu *Netspeak* versus *Webwrite*". In T. Gloning / I. Bons / D. Kaltwasser (Hrsg.) *Festplatte für Gerd Fritz*. Online verfügbar:
http://www.festschrift-gerd-fritz.de/files/mukherjee_2008_tradition-und-innovation-im-englischen-internet-sprachgebrauch.pdf (30.12.2012).
- Mukherjee, J. / Gries, S. T. (2010): "Lexical gravity across varieties of English: an ICE-based study of *n*-grams in Asian Englishes". *International Journal of Corpus Linguistics* 15(4). 520-548.

- Myles, F. / Mitchell R. / Hooper J. (1999): "Interrogative chunks in French L2: a basis for creative construction?". *Studies in Second Language Acquisition* 21, 49-80.
- Nattinger, J. / DeCarrico, J. (1992): *Lexical Phrases and Language Teaching*. Oxford: OUP.
- Nesi, H. / Basturkmen H. (2006): "Lexical bundles and discourse signalling in academic lectures". *International Journal of Corpus Linguistics* 11(3), 147-168.
- Nielsen, J. (1997): "How users read on the Web". Online verfügbar: <http://www.useit.com/alertbox/9710a.html> (30.12.2012).
- Oehmen, X. (2011): "Keilschriften: Zehntausend Freunde Mesopotamiens". *Frankfurter Allgemeine Zeitung* 15.06.2011. Online verfügbar: <http://www.faz.net/artikel/C31399/keilschriften-zehntausend-freunde-mesopotamiens-30439723.html> (30.12.2012).
- O'Reilly, T. (2007): "What is Web 2.0?" Online verfügbar: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (30.12.2012).
- Paolillo, J. (1999): "The virtual speech community: social network and language variation on IRC". Online verfügbar: <http://jcmc.indiana.edu/vol4/issue4/paolillo.html> (30.12.2012).
- Paton, G. (2008): "University students cannot spell". *Daily Telegraph*. Online verfügbar: <http://www.telegraph.co.uk/news/2510704/University-students-cannotspell.html> (30.12.2012).
- Pawley, A. / Syder F. (1983): "Two puzzles for linguistic theory: nativelike selection and nativelike fluency". In J. Richards / R. Schmidt (Hrsg.) *Language and Communication*. London: Longman, 191-225.
- Pawley, A. / Syder F. (2000): "The one-clause-at-a-time hypothesis". In H. Riggenbach (Hrsg.) *Perspectives on fluency*. Ann Arbor: University of Michigan Press, 163-199.
- Peng, F. / Schuurmans, D. / Wang, S. (2003): "Language and task independent text categorisation with simple language models". In M. Hearst / M. Ostendorf (Hrsg.) *HLT-NAACL 2003: Main proceedings*. Edmonton, Canada: Association for Computational Linguistics, 189-196.
- Peters, A. M. (1977): "Language learning strategies: does the whole equal the sum of the parts?". *Language* 53(3), 560-573.
- Peters, A. M. (1983): *The Units of Language Acquisition*. Cambridge: CUP.
- Pennycook, A. (2011): "Global Englishes". In B. Johnstone / R. Wodak / P. Kerswill (Hrsg.) *The SAGE Handbook of Sociolinguistics*. London: SAGE publications, 513-526.
- Peyawary, A. S. (1999): *The Core Vocabulary of International English: A Corpus Approach*. Bergen: The humanities information technologies research programme.
- Philipson, R. (1992): *Linguistic Imperialism*. Oxford: OUP.
- Poole, M. / Field, T. (1976): "A comparison of oral and written code elaboration". *Language and Speech* 19, 305-311.
- Prodromou, L. (2007): "Is ELF a variety of English". *English Today* 90, 23(2), 47-53.
- Prodromou, L. (2008): *English as a Lingua Franca. A Corpus-based Analysis*. London, New York: Continuum.

- Quirk, R. / Leech, G. / Svartvik, J. (1972): *A Grammar of Contemporary English*. Cambridge: CUP.
- Raichle, M. E. (1998): "The neural correlates of consciousness: an analysis of cognitive skill learning". *Philosophical Transactions of the Royal Society of London, Series B* 353, 1889-1901.
- Renouf, A. / Sinclair, J. M. (1991): "Collocational frameworks in English". In K. Ajimer / B. Altenberg (Hrsg) *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. London: Longman, 128-143.
- Rettberg, J. W. (2008): *Blogging*. Cambridge: Polity Press.
- Richardson, W. (2006): *Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms*. Thousand Oaks: Corwin Press.
- Rosenzweig, R. (2006): "Can history be open source? Wikipedia and the future of the past". *The Journal of American history* 93(1), 117-146.
- Rowe, C. / Wyss, E. L. (2009): *Language and New Media: Linguistic, Cultural, and Technological Evolutions*. Creskill, NJ: Hampton Press.
- Saage, S. / Künstler, V. / Mukherjee, J. (2012): "Korpuslinguistische Analysen von Mehrworteinheiten in englischsprachiger Internetkommunikation". In J. Klawitter / H. Lobin / T. Schmidt (Hrsg.) *Kulturwissenschaften digital. Neue Forschungsfragen und Methoden. Interaktiva: Schriftenreihe des Zentrums für Medien und Interaktivität, Gießen* (Band 12). Frankfurt am Main: Campus, 135-154.
- Sacks, H. / Schegloff, E. A. / Jefferson, G. (1974): "A simplest systematic for the organization of turn-taking for conversation". *Language* 50, 696-735.
- Sapir, E. (1921): *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and company.
- Saussure, F. de (1916 / 1966): *Course in General Linguistics*. New York: McGraw-Hill.
- Schäpers, U. K. E. (2009): *Nominal versus Clausal Complexity in Spoken and Written English. Theory and Description*. Frankfurt am Main: Peter Lang.
- Schlobinski, P. (2000): "Chatten im Cyberspace". In K. Eichhoff-Cyrus / R. Hoberg (Hrsg.) *Die deutsche Sprache zur Jahrtausendwende. Sprachkultur oder Sprachverfall?*. Mannheim, Leipzig: Dudenverlag, 63-79.
- Schmidt, T. / Hagelmoser, R. / Saage, S. (2010): "To tweet or not to tweet: Das Kommunikationstool Twitter im Fremdsprachenunterricht". *Praxis Fremdsprachenunterricht* 4-2010, 7-11.
- Schmitt, N. (2004): *Formulaic Sequences. Acquisition, Processing and Use*. Amsterdam: John Benjamins.
- Schneider, E. (2003): "The dynamics of New Englishes: from identity construction to dialect birth". *Language* 79(2), 233-281.
- Schneider, E. W. (2007): *Postcolonial English – Varieties Around the World*. Cambridge: CUP.
- Schneider, E. W. (2012): "Exploring the interface between World Englishes and Second language acquisition – and implications for English as a lingua franca". *Journal of English as a Lingua Franca* 1(1), 57-91.

- Schuchardt, H. (1909): "Die Lingua Franca". *Zeitschrift für Romanische Philologie* 33, 441-461.
- Scott, M. (2008): *WordSmith Tools Version 5*. Liverpool: Lexical Analysis Software.
- Seidlhofer, B. (2001): "Closing a conceptional gap: The case for a description of English as a lingua franca". *International Journal of Applied Linguistics* 11(2), 133-158.
- Seidlhofer, B. (2003): "A concept of International English and related issues: from 'Real English' to 'Realistic English'?". *Council of Europe. Language Policy Division*. Strasbourg: Council of Europe.
- Seidlhofer, B. (2004): "The VOICE of ELF – English as a lingua franca". *What's New? Autumn/Winter 2004*, 8-9.
- Seidlhofer, B. (2005): "English as a Lingua Franca". *Oxford Advanced Learner's Dictionary*. Oxford: OUP, 92.
- Seidlhofer, B. (2009): "Common ground and different realities: World Englishes and English as a lingua franca". *World Englishes* 28(2), 236-245.
- Seidlhofer, B. (2010): "Giving VOICE to English as a lingua franca". In R. Facchinetti / D. Crystal / B. Seidlhofer (Hrsg.) *From International to Local English – and Back Again*. Frankfurt am Main: Peter Lang, 147-163.
- Seidlhofer, B. / Berns M. (2009): "Perspectives on English as a lingua franca: introduction". *World Englishes* 28(2), 190-191.
- Shirky, C. (2003): "Power laws, web logs, and inequality". Online verfügbar: www.shirky.com/writings/powerlaw_weblog.html (30.12.2012).
- Sieber, P. (1998): *Parlando in Texten. Zur Veränderung kommunikativer Grundmuster in der Schriftlichkeit*. Tübingen: Niemeyer.
- Simpson, R. (2004): "Stylistic features of academic speech: The role of formulaic expressions". In U. Connor / T. A. Upton (Hrsg.) *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 37-64.
- Simpson-Vlach, R. / Ellis, N. C. (2010): "An academic formulas list: new methods in phraseology research". *Applied Linguistics Online*. Online verfügbar: <http://appliedjournals.org/content/31/4/487.full> (30.12.2012).
- Sinclair, J. M. (1987): "Collocation: a progress report". In R. Steele / T. Treadgold (Hrsg.) *Essays in honour of Michael Halliday*. Amsterdam: John Benjamins, 319-331.
- Sinclair, J. M. (1991): *Corpus, Concordance, Collocation*. Oxford: OUP.
- Siyanova-Chanturia, A. / Conklin K., Schmitt N. (2011a) "Adding more fuel to the fire: An eye-tracking study of idiom processing". *Second Language Research* 27 (2): 251-272.
- Siyanova-Chanturia, A. / Conklin, K. / van Heuven, W. (2011b): "Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multiword sequences". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(3), 776-784.
- Spöttl, C. / McCarthy, M. (2004): "Comparing knowledge of formulaic sequences across L1, L2, L3, and L4". In N. Schmitt (Hrsg.) *Formulaic Sequences. Acquisition, Processing and Use*. Amsterdam: John Benjamins, 191-225.

- Stamatatos, E. / Fakotakis, N. / Kokkinakis, G. (2001): "Automatic text categorisation in terms of genre and author". *Computational Linguistics* 26, 471-495.
- Stefanowitsch, A. / Gries, S. T. (2003): "Collostructions: investigating the interaction between words and constructions". *International Journal of Corpus Linguistics* 8(2), 209-243.
- Stein, D. (2006): "The website as a domain-specific genre". *Language@Internet* 3, Article 6, 1-9.
- Storrer, A. (2001): "Getippte Gespräche oder dialogische Texte? Zur kommunikationstheoretischen Einordnung der Chat-Kommunikation". In A. Lehr / M. Kammerer / K-P. Konerding / A. Storrer / C. Thimm / W. Wolski (Hrsg.) *Sprache des Alltags. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*. Berlin, New York: Walter de Gruyter, 439-465.
- Storrer, A. (2009): "Rhetorisch-stilistische Eigenschaften der Sprache des Internets". In U. Fix / A. Gardt / J. Knappe (Hrsg.) *Rhetorik und Stilistik. Ein internationales Handbuch historischer und systematischer Forschung*. Berlin, New York: Walter de Gruyter, 2211-2226.
- Stubbs, M. (2007): "An example of frequent English phraseology: distributions, structures and functions". In R. Facchinetti (Hrsg.) *Corpus Linguistics 25 years on*. Amsterdam, New York: Rodopi, 89-106.
- Svartvik, J. / Ekedahl, O. (1995): "Verbs in public and private speaking". In B. Aarts / C. F. Meyer (Hrsg.) *The Verbs in Contemporary English: Theory and Description*. Cambridge: CUP, 273-289.
- Swales, J. (1990): *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.
- Tagliamonte, S. (2011): "Teen language in the virtual speech community: building and analyzing a corpus of Internet media". *Präsentation auf der International Computer Archive of Modern and Medieval English (ICAME) Konferenz* in Oslo, Norwegen, 1.-5. Juni 2011.
- Taylor, J. (2002): *Cognitive Grammar*. Oxford: Oxford University Press.
- Thaler, V. (2007): "Mündlichkeit, Schriftlichkeit, Synchronizität. Eine Analyse alter und neuer Konzepte zur Klassifizierung neuer Kommunikationsformen". *Zeitschrift für Germanistische Linguistik* 35, 146-181.
- Thurlow, C. (2006): "From statistical panic to moral panic: the metadiscursive construction and popular exaggeration of new media language in the print media". *Journal of Computer-Mediated Communication* 11(3), Article 1. Online verfügbar: <http://jcmc.indiana.edu/vol11/issue3/thurlow.html> (30.12.2012).
- Thurlow, C. / Mroczek, K. (2011): *Digital Discourse: Language in the New Media*. New York, London: OUP.
- Thurlow, C. / Mroczek, K. (2011): "Introduction: fresh perspectives on New Media Sociolinguistics". In C. Thurlow / K. Mroczek (Hrsg.) *Digital Discourse: Language in the New Media*. New York, London: OUP, ix-xliv.
- Tomasello, M. (2003): *Constructing a Language. A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

- Trier, J. (1931): *Der deutsche Wortschatz im Sinnbezirk des Verstandes. Geschichte eines sprachlichen Feldes*. Heidelberg: Winter.
- Trudgill, P. / Hannah J. (2008): *International English: A Guide to the Varieties of Standard English*. London: Hodder Education.
- Underwood, G. / Schmitt, N. / Galpin, A. (2004): "The eyes have it: an eye-movement study into the processing of formulaic sequences". In N. Schmitt (Hrsg.) *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam, Philadelphia: John Benjamins, 153-172.
- Van Lancker, D. (1987): "Nonpropositional speech: neurolinguistic studies". In: Ellis, A. W. (Hrsg.) *Progress in the Psychology of Language Vol. 3*. Hillsdale, NJ: Lawrence Erlbaum, 49-118.
- Van Lancker, D. (1988): "Nonpropositional speech: neurolinguistic studies". In A. E. Ellis (Hrsg.) *Progress in the Psychology of Language Vol. III*. Hillsdale, NJ: Lawrence Erlbaum: 49-118.
- Verma, M. K. (2008): "English as an economic investment: who will earn the dividends?" In C. Gnutzmann / F. Intemann (Hrsg.) *The Globalisation of English and the English Language Classroom* (2. Edition). Tübingen: Narr, 41-54.
- Walker, J. (2005): "Weblog". In D. Herman / M. Jahn / M. L. Ryan (Hrsg.) *Routledge Encyclopedia of Narrative Theory*. London, New York: Routledge, 45.
- Warschauer, M. (2000): "Language, identity and the internet". In B. E. Kolko / L. Nakamura / G. B. Rodman (Hrsg.) *Race in Cyberspace*. New York: Routledge, 151-170.
- Warschauer, M. / Grimes, D. (2007): "Audience, authorship, and artifact: the emergent semiotics of Web 2.0". *Annual Review of Applied Linguistics* 27, 1-23.
- Warschauer, M. / Black, R. / Chou, Y-L. (2010): "Online Englishes". In A. Kirkpatrick (Hrsg.) *The Routledge Handbook of World Englishes*. New York: Routledge, 490-505.
- Weinert, R. (1995): "The role of formulaic language in second language acquisition: A review". *Applied Linguistics* 16, 180-205.
- Widdowson, H. (1997): "EIL, ESL, EFL: global issues and local interests". *World Englishes* 16(1), 135-146.
- Widdowson, H. (1998): "EIL: squaring the circles. A reply". *World Englishes*, 17(3), 397-401.
- Wilkinson, D. M. / Huberman, B. A. (2007): "Assessing the value of cooperation in Wikipedia". *First Monday* 12(4). Online verfügbar: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1763/1642> (30.12.2012).
- Wong-Fillmore, L. (1976): *The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition*. Unpublished doctoral dissertation, Stanford University.
- Wray, A. (2002): *Formulaic Language and the Lexicon*. Cambridge: CUP.
- Wray, A. (2008): *Formulaic Language. Pushing the Boundaries*. Oxford: OUP.
- Wray, A. / Perkins, M. R. (2000): "The functions of formulaic language. An integrated model". *Language & Communication* 20, 1-28. Online verfügbar: <http://www.cardiff.ac.uk/encap/resources/wrayperkins.pdf> (30.12.2012).
- Wynne, M. (2005): *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow.

- Yano, Y. (2001): "World Englishes in 2000 and beyond". *World Englishes* 20, 119–31.
- Yano, Y. (2009): "English as an international lingua franca: from societal to individual". *World Englishes* 28(2), 246-255.
- Ziegler, A. (2003): "E-Mail: Textsorte oder Kommunikationsform?". In A. Ziegler / C. Dürscheid (Hrsg.) *Kommunikationsform E-Mail*. Tübingen: Stauffenburg, 9-32.

Webseiten:

- Berliner Kurier: <http://www.berliner-kurier.de/digital/erste-studie-belegt-560-000-deutsche-sind-internetsuechtig,7168826,10900900.html> (30.12.2012).
- Blog Boing Boing: <http://boingboing.net> (30.12.2012).
- Blog Causes of Conflict: http://causesofconflict.vision.org/causes_of_conflict (30.12.2012).
- Blog Coping with Childhood Cancer: <http://www.copingwithchildhoodcancer.blogspot.com> (30.12.2012).
- BlogPulse: <http://www.blogpulse.com> (30.12.2012).
- Blog Willy Sudiarto Raharjo: <http://willysr.blogspot.com> (30.12.2012).
- Facebook: <http://www.facebook.com> (30.12.2012).
- Facebook Statistiken: <http://www.facebook.com/press/info.php?statistics> (30.12.2012).
- Massively.Joystiq.com: Second Life Statistiken:
<http://massively.joystiq.com/2009/09/08/second-life-median-concurrency-declines-as-bots-campers-progress/> (30.12.2012)
- Miniwatts Marketing Research (2011): "Top ten languages used in the Web", Internet World Stats. <http://www.internetworldstats.com/stats.htm> (30.12.2012).
- Netcraft: <http://news.netcraft.com> (30.12.2012).
- Netz-Tipp: <http://www.netz-tipp.de/languages.html> (30.12.2012).
- Second Life: <http://secondlife.com/?lang=de-DE> (30.12.2012).
- Technorati: <http://technorati.com> (30.12.2012).
- The Daily Post: <http://dailypost.wordpress.com> (30.12.2012).
- The Well (Whole Earth 'Lectronic Link): <http://www.well.com> (30.12.2012).
- Twitter: <http://twitter.com> (30.12.2012).
- Verein Deutsche Sprache: <http://www.vds-ev.de/denglisch> (30.12.2012).
- Wikipedia: "Wikipedia": <http://de.wikipedia.org/wiki/Wikipedia> (30.12.2012).
- Wikipedia Statistiken: http://meta.wikimedia.org/wiki/List_of_Wikipedias (30.12.2012)
- Wolfram Alpha: <http://www.wolframalpha.com/about.html> (30.12.2012).

Appendix

A. Liste der 100 am häufigsten rekurrierenden 4-grams für jedes Web-Genre (*pmw*)

Foren		
1	I would like to	120
2	if you want to	97
3	I would love to	79
4	to be able to	78
5	I don't want to	74
6	when it comes to	71
7	at the same time	64
8	had a nose job	64
9	I don't know if	52
10	she looks like a	57
11	at the end of	54
12	is one of the	54
13	a lot of people	53
14	to get rid of	52
15	I am going to	50
16	I had my surgery	50
17	send a message via	50
18	the end of the	50
19	if you have any	49
20	this message has been	48
21	message has been edited	48
22	I don't know what	46
23	I can tell you	45
24	I don't think she	44
25	have a lot of	44
26	good luck to you	43
27	the rest of the	43
28	I just want to	41
29	had a tummy tuck	41
30	will be able to	41
31	on the other hand	40
32	I think she looks	39
33	I don't know how	39
34	one of the most	38
35	in the first place	38
36	if you have a	37
37	I want to know	37
38	you have to be	37
39	you have to do	36
40	I am not sure	35

41	would like to know	34
42	I don't know why	34
43	for the rest of	34
44	for the first time	33
45	to have the surgery	33
46	one of the best	33
47	to be on the	33
48	nothing to do with	33
49	I think it is	32
50	I would have to	32
51	thank you for your	32
52	is going to be	32
53	you don't have to	31
54	I do not know	31
55	be on the show	30
56	would like to see	30
57	I don't think it	30
58	I just wanted to	30
59	you will have to	30
60	I don' think that	30
61	I have to say	29
62	as well as the	29
63	the easy way out	29
64	are a lot of	28
65	as far as I	28
66	you need to be	28
67	she used to be	28
68	you have any questions	27
69	there are a lot	27
70	thank you so much	27
71	I was able to	27
72	before and after pictures	27
73	to go to a	27
74	I don't know about	27
75	I want to get	27
76	I don't think I	26
77	I love the show	26
78	you are going to	26
79	I agree with you	26
80	I have a friend	26
81	I had the surgery	26
82	As far as the	26
83	I have had a	26
84	a lot of the	26

85	the rest of my	25
86	thank you very much	25
87	the only thing that	25
88	I would do it	25
89	thank you for the	24
90	think it would be	24
91	but I don't think	24
92	should be able to	24
93	there is no way	24
94	had a breast reduction	24
95	I have heard that	24
96	look in the mirror	24
97	I don't think so	24
98	I don't think she's	24
99	to look like a	24
100	there are plenty of	24

Blogs		
1	in the united states	358
2	is one of the	151
3	one of the most	132
4	at the university of	99
5	in the field of	80
6	as well as the	80
7	when it comes to	75
8	a fraction of the	70
9	it is important to	70
10	the medical tourism industry	66
11	medical tourism in india	61
12	the United States and	61
13	on top of the	61
14	fraction of the cost	61
15	the surface of the	61
16	of the most common	56
17	will be able to	56
18	of the inner ear	56
19	at the same time	56
20	to make sure that	56
21	in the number of	56
22	cost in panama cost	51
23	cost in the usa	51
24	in panama cost in	51
25	panama cost in the	51

26	of lasik eye surgery	51
27	as one of the	47
28	medical tourism in Kerala	47
29	some of the most	47
30	on the right side	47
31	from around the world	47
32	the centers for disease	47
33	medical council of India	47
34	procedure cost in Panama	47
35	make sure that you	42
36	for disease control and	42
37	are more likely to	42
38	more and more people	42
39	centers for disease control	42
40	in front of the	42
41	the shape of the	42
42	in the medical tourism	42
43	the food and drug	42
44	food and drug administration	42
45	at a fraction of	37
46	of Lasik eye surgery	37
47	at the age of	37
48	anti aging skin care	37
49	some of the best	37
50	issue of the journal	37
51	as a case of	37
52	are some of the	37
53	of medical tourism in	37
54	all over the world	37
55	your health care provider	37
56	as a result of	37
57	can be used to	37
58	in a news release	37
59	the high cost of	37
60	disease control and prevention	37
61	the medical council of	33
62	knee and hip replacement	33
63	said in a news	33
64	a wide range of	33
65	you may want to	33
66	the laser eye surgery	33
67	the dissection of the	33
68	if you are considering	33
69	and more people are	33

70	minimize the risk of	33
71	to put in contact	33
72	it comes to your	33
73	is a surgical procedure	33
74	on the left side	33
75	put in contact lenses	33
76	how to put in	33
77	on the other hand	33
78	if you are a	33
79	the left side of	33
80	of oral and maxillofacial	33
81	in the form of	33
82	in such a way	33
83	is a type of	33
84	shape of the cornea	33
85	Lasik eye surgery is	33
86	is the most common	33
87	the best way to	33
88	of the eye and	33
89	this type of surgery	33
90	reduce the risk of	33
91	total knee replacement surgery	33
92	corrective laser eye surgery	28
93	for the first time	28
94	want to make sure	28
95	Kodiyeri Balakrishnan minister of	28
96	you need to take	28
97	you will have to	28
98	the University of California	28
99	the process of organ	28
100	would I need to	28

kommerzielle Homepages		
1	at the same time	160
2	one of the most	158
3	is one of the	154
4	the best candidates for	118
5	it is important to	118
6	will be able to	109
7	as well as the	100
8	as a result of	95
9	for a few days	93
10	the size of the	91
11	it is possible to	89

12	surgery in cape town	73
13	you will be able	71
14	on the day of	70
15	is a surgical procedure	70
16	plastic and reconstructive surgery	68
17	a day or two	66
18	the extent of the	66
19	on an outpatient basis	66
20	on the extent of	66
21	a good candidate for	64
22	surgery in south africa	64
23	in the area of	64
24	should be able to	62
25	size and shape of	61
26	he or she will	61
27	the area of the	61
28	if you have any	59
29	in the united states	59
30	the first few days	59
31	can be used to	59
32	you will need to	59
33	excess skin and fat	55
34	in the field of	55
35	the size and shape	55
36	you should be able	53
37	of plastic and reconstructive	53
38	the swelling and bruising	53
39	in the case of	53
40	the back of the	53
41	within a few days	50
42	depending on the extent	50
43	at the time of	50
44	a week or two	48
45	for the first few	48
46	able to return to	46
47	it is necessary to	46
48	a basic understanding of	46
49	surgery will be performed	44
50	before and after surgery	44
51	of aesthetic plastic surgery	44
52	it is recommended to	44
53	stay in the clinic	44
54	you will be given	44
55	how to prepare for	44

56	for a week or	43
57	for a day or	43
58	enhance your appearance and	43
59	the rest of the	43
60	on how to prepare	43
61	a member of the	43
62	with the help of	43
63	performed by a qualified	41
64	days after the surgery	41
65	two to three weeks	41
66	be performed under local	41
67	plastic and reconstructive surgeons	41
68	society of aesthetic plastic	41
69	in front of the	41
70	surgeon will give you	41
71	society of plastic surgery	39
72	you can expect to	39
73	may be performed in	39
74	the duration of the	39
75	the first few weeks	39
76	international society of aesthetic	39
77	for two to three	39
78	plastic and hand surgery	39
79	for the treatment of	39
80	is a procedure that	39
81	can be performed under	39
82	to be able to	39
83	weeks after the surgery	39
84	go back to work	39
85	before and after the	39
86	a qualified plastic surgeon	37
87	will be required to	37
88	the procedure is performed	37
89	by a qualified plastic	37
90	the end of the	37
91	if you have a	37
92	the surgery will be	37
93	Gomez-Mendoza and his team	37
94	reduce the size of	37
95	be able to return	37
96	all over the world	37
97	it is very important	37
98	at the university of	37
99	done on an outpatient	37

100	the skin of the	37
-----	-----------------	----

Wikipedia-Artikel		
1	in the united states	319
2	is a surgical procedure	120
3	the royal college of	119
4	as a result of	101
5	is one of the	90
6	at the university of	83
7	royal college of surgeons	82
8	for the treatment of	82
9	in the treatment of	76
10	one of the most	76
11	in the case of	76
12	operations and other procedures	76
13	can be used to	72
14	of the royal college	72
15	as well as the	66
16	at the time of	61
17	in the united kingdom	59
18	has been shown to	58
19	and other procedures on	58
20	other procedures on the	56
21	the use of a	56
22	the surgical removal of	55
23	the american academy of	51
24	is the most common	50
25	surgical procedure in which	50
26	the size of the	48
27	food and drug administration	48
28	the end of the	47
29	oral and maxillofacial surgery	44
30	the base of the	44
31	is a type of	44
32	in the field of	43
33	can also be used	43
34	at the same time	41
35	the removal of the	41
36	college of surgeons of	41
37	the use of the	41
38	at the end of	40
39	a part of the	40
40	the top of the	40
41	a portion of the	40

42	the rest of the	38
43	the location of the	38
44	a surgical procedure in	38
45	head and neck cancer	38
46	other parts of the	37
47	studies have shown that	37
48	is the surgical removal	37
49	in the absence of	36
50	on the musculoskeletal system	36
51	the surface of the	36
52	is a form of	36
53	the head of the	36
54	procedures on the musculoskeletal	36
55	the back of the	36
56	been shown to be	36
57	reduce the risk of	36
58	an increased risk of	34
59	the united states and	34
60	surgical removal of the	34
61	in the form of	34
62	is inserted into the	33
63	may be used to	33
64	procedure in which the	33
65	in some cases the	33
66	the american college of	33
67	the tip of the	31
68	the head and neck	31
69	the american board of	31
70	part of the body	31
71	with the use of	31
72	a member of the	31
73	and drug administration FDA	30
74	of surgeons of england	30
75	sometimes referred to as	30
76	it is important to	30
77	is used to treat	30
78	as an alternative to	29
79	can be used for	29
80	also known as a	29
81	of one or more	29
82	can be caused by	29
83	college of physicians and	27
84	of physicians and surgeons	27
85	also be used to	27

86	is the removal of	27
87	referred to as a	27
88	the american society of	27
89	of the university of	27
90	Royal College of physicians	26
91	an incision is made	26
92	as a treatment for	26
93	the American Medical Association	26
94	college of surgeons in	26
95	the part of the	26
96	which can lead to	26
97	is thought to be	26
98	it is used to	26
99	is referred to as	26
100	parts of the body	26

B. Chi-Quadrat-Tests für die einzelnen Kategorien

Kategorie 1	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
Conversation	---	0,0001	0,0001	0,35	0,68	0,0001
Academic Prose	0,0001	---	0,0001	0,0001	0,0001	---
Foren	0,0001	0,0001	---	0,0001	0,0001	0,0001
Blogs	0,35	0,0001	0,0001	---	0,27	0,0001
Kommerzielle Homepages	0,68	0,0001	0,0001	0,27	---	0,0001
Wikipedia-Artikel	0,0001	---	0,0001	0,0001	0,0001	---

Kategorie 2	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
Conversation	---	0,0001	0,0001	0,0001	0,0001	0,89
Academic Prose	0,0001	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	1	0,49	0,0001
Blogs	0,0001	0,0001	1	---	0,82	0,001
Kommerzielle Homepages	0,0001	0,0001	0,49	0,82	---	0,0001
Wikipedia-Artikel	0,89	0,0001	0,0001	0,001	0,0001	---

Kategorie 3	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
Conversation	---	0,0001	0,0001	0,0001	0,0001	0,35
Academic Prose	0,0001	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	0,37	0,23	0,0001
Blogs	0,0001	0,0001	0,37	---	0,11	0,0001
Kommerzielle Homepages	0,0001	0,0001	0,23	0,11	---	0,0001
Wikipedia-Artikel	0,35	0,0001	0,0001	0,0001	0,0001	---

Kategorie 4	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia-Artikel
Conversation	---	0,0001	0,31	0,78	0,05	0,01
Academic Prose	0,0001	---	0,0001	0,0001	0,01	---
Foren	0,31	0,0001	---	0,42	0,01	0,001
Blogs	0,78	0,0001	0,42	---	0,13	0,01
Kommerzielle Homepages	0,05	0,01	0,01	0,13	---	0,25
Wikipedia-Artikel	0,01	---	0,001	0,01	0,25	---

Kategorie 5	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,0001	0,0001	0,056	0,001	1
Academic Prose	0,0001	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	0,34	0,33	0,001
Blogs	0,056	0,0001	0,34	---	0,82	0,13
Kommerzielle Homepages	0,001	0,0001	0,33	0,82	---	0,05
Wikipedia- Artikel	1	0,0001	0,001	0,13	0,05	---

Kategorie 6	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,05	0,0001	0,0001	0,0001	0,55
Academic Prose	0,05	---	0,0001	0,0001	0,0001	0,05
Foren	0,0001	0,0001	---	1	0,19	0,0001
Blogs	0,0001	0,0001	1	---	0,42	0,001
Kommerzielle Homepages	0,0001	0,0001	0,19	0,42	---	0,01
Wikipedia- Artikel	0,55	0,05	0,0001	0,001	0,01	---

Kategorie 7	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,0001	0,0001	0,0001	0,0001	0,0001
Academic Prose	0,0001	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	0,0001	0,0001	0,0001
Blogs	0,0001	0,0001	0,0001	---	0,19	0,05
Kommerzielle Homepages	0,0001	0,0001	0,0001	0,19	---	0,0001
Wikipedia- Artikel	0,0001	0,0001	0,0001	0,05	0,0001	---

Kategorie 8	Conversation	Academic Prose	Foren	Blogs	Kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,0001	0,0001	0,0001	0,0001	0,0001
Academic Prose	0,0001	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	0,0001	0,0001	0,0001
Blogs	0,0001	0,0001	0,0001	---	0,46	0,05
Kommerzielle Homepages	0,0001	0,0001	0,0001	0,46	---	0,0001
Wikipedia- Artikel	0,0001	0,0001	0,0001	0,05	0,0001	---

Kategorie 9	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,0001	0,0001	0,0001	0,0001	0,0001
Academic Prose	0,0001	---	0,0001	0,06	0,0001	0,22
Foren	0,0001	0,0001	---	0,81	0,05	0,1
Blogs	0,0001	0,06	0,81	---	0,16	0,51
Kommerzielle Homepages	0,0001	0,0001	0,05	0,16	---	0,01
Wikipedia- Artikel	0,0001	0,22	0,1	0,51	0,01	---

Kategorie 10	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,001	0,0001	---	0,0001	0,0001
Academic Prose	0,001	---	0,16	0,48	0,0001	0,0001
Foren	0,0001	0,16	---	0,3	0,0001	0,0001
Blogs	---	0,48	0,3	---	0,0001	0,01
Kommerzielle Homepages	0,0001	0,0001	0,0001	0,0001	---	0,01
Wikipedia- Artikel	0,0001	0,0001	0,0001	0,01	0,01	---

Kategorie 11	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,06	0,0001	0,79	0,001	0,05
Academic Prose	0,06	---	0,0001	0,54	0,06	0,49
Foren	0,0001	0,0001	---	0,08	0,19	0,05
Blogs	0,79	0,54	0,08	---	0,25	0,4
Kommerzielle Homepages	0,001	0,06	0,19	0,25	---	0,53
Wikipedia- Artikel	0,05	0,49	0,05	0,4	0,53	---

Kategorie 12	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,79	0,0001	0,0001	0,0001	0,0001
Academic Prose	0,79	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	0,3	0,42	0,0001
Blogs	0,0001	0,0001	0,3	---	0,13	0,14
Kommerzielle Homepages	0,0001	0,0001	0,42	0,13	---	0,0001
Wikipedia- Artikel	0,0001	0,0001	0,0001	0,14	0,0001	---

Kategorie 13	Conversation	Academic Prose	Foren	Blogs	kommerzielle Homepages	Wikipedia- Artikel
Conversation	---	0,78	0,0001	0,0001	0,0001	0,0001
Academic Prose	0,78	---	0,0001	0,0001	0,0001	0,0001
Foren	0,0001	0,0001	---	0,39	0,0001	0,0001
Blogs	0,0001	0,0001	0,39	---	0,0001	0,0001
Kommerzielle Homepages	0,0001	0,0001	0,0001	0,0001	---	0,0001
Wikipedia- Artikel	0,0001	0,0001	0,0001	0,0001	0,0001	---

C. Funktionale Klassifikation von gemeinsamen 4-grams der vier Web Genres

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
I. Stance Expressions				
A. Hedges				
I have to say	29			
it looks like she	22			
it would be a	22			
you may want to		32	23	

B. Epistemic stance				
I don't know if	55			
I think she looks	39			
I don't know what	37			
I don't think she	37			
I am not sure	35			
I do not know	31			
I don't know how	31			
I don't know why	31			
I don't think that	27			
I don't think it	26			
I agree with you	25			
I would do it	25			
think it would be	25			
I don't know about	24			
I was wondering if	23			
I think it was	22			
but I don't think	22			
I don't think she's	22			
I don't think so	21			
I don't think it	20			
I think this is	20			
according to a new		18		

C. Obligation and directive

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
you have to be	37			
you have to do	36			
I would have to	32			
you will have to	30	28		
please let me know	28			
you need to be	28			
you don't have to	27			
you are going to	26	32		
but you have to	21			
you don't want to	20			
you want to be	20			
to make sure that		55		
make sure that you		42		
want to make sure		28		
you need to take		28		
you need to know		23		
you need to make		23		
you will need to			58	
you will be given			44	
you will be encouraged			25	
that you will be			26	
will be asked to			25	
you will be asked			23	
it is recommended to			44	
will be required to			37	
you will be required			33	
and you will be			32	
will need to be			25	
you will also be			25	

D. Ability and possibility

to be able to	77	23	39	19
will be able to	41	55	107	
I was able to	27			
should be able to	24		62	
there is no way	24			
are more likely to		42		
can be used to		37	58	72
all you need to		23		
may be able to		23		

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
not be able to		23		
will allow you to		23		
you may not be		23		
be more likely to		18		
it is possible to			88	22
you should be able			53	
you can expect to			39	
you can reduce your			32	
there is no need			30	
can also be used			26	32
may be used to				32
can be used for				28
also be used to				27
are more likely to				23
it can also be				22
it is used to				19
may or may not				23
more likely to be				22

E. Evaluation

it is important to		69	118	31
it is necessary to			46	
it is very important			37	
is very important to			28	

F. Intention / volition

I would like to	117			
if you want to	96			
I would love to	78			
I don't want to	53			
I am going to	50	37		
I just want to	41	23	23	
I want to know	37			
would like to know	34			
I think it is	32			
I just wanted to	30			
would like to see	30			
I want to get	27			
I am trying to	22			
and I want to	21			

II. Discourse organising expressions

A. Metadiscourse and textual reference

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
if you look at	21			
the fact that the		23		
keep in mind that			35	
has been shown to				57
studies have shown that				36
have been shown to				24

B. Topic introduction and focus

if you have any	48		58	
I can tell you	45			
I was going to	23			
in the first place	39	23		
if you have a	37		37	
if you are a		32		
if you are considering		32		
if you have a		28		

C. Topic elaboration

when it comes to	71	74	35	
for the first time	33	28		24
in the long run	25			
I have heard that	24			
for the most part	22			
as a result of		37	93	99
is due to the		28		
so that you can		28		
the reason for this		23		
in the area of			63	
depending on the extent			49	
and as a result			30	
with the results of			28	
in most cases the			26	20
in some cases the				35
which can lead to				26
can be caused by				24
the most common cause				23
a result of the				19

D. Discourse markers

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
at the same time	64	55	157	41
thank you for your	32			
thank you so much	26			
thank you very much	25			
thank you for the	24			
in conjunction with a			33	
in conjunction with other			28	

III. Referential expressions**A. Specifications of attributes**

the rest of the	42		42	38
one of the most	37	129	155	74
for the rest of	34			
one of the best	32	23	23	
are a lot of	28			
a lot of the	26			
the only thing that	25			
the rest of my	25			
there are plenty of	24			
there are so many	24			
had a lot of	23			
in the middle of	22			
in the process of	22			
is a lot of	21			
a lot of money	21			
a lot of weight	21			
there are a lot	27			
the end of the	49		37	46
a lot of people	55	18		
in the field of		79	55	43
a fraction of the		69		
the surface of the		60	26	35
in the number of		60		
of the most common		55		24
as one of the		46		
some of the most		46		
the shape of the		42	35	31
are some of the		37		
as a case of		37		

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
at a fraction of		37		
some of the best		37		
the high cost of		37		
a wide range of		32	23	
in the form of		32	35	34
is the most common		32		55
the best way to		32		
a small amount of		28	28	24
of the cost of		28		
both sides of the		28		
the development of the		28		
the quality of the		28	32	
a great deal of		23		
a member of the		23	42	32
a number of different		23		
amount of time for		23		
on both sides of		23		
the edge of the		23		
the extent of the		23	65	
and some of the		23		
are one of the		23		
in some of the		23		
a wide variety of		18	28	
are a number of		18		18
at high risk for		18		
the size of the			92	46
on the day of			69	
the area of the			60	24
the back of the			53	35
with the help of			42	19
a basic understanding of			46	
the day of the			33	
the duration of the			39	
for a number of			33	
the member of the			33	
with the use of			33	27
the results of the			32	
a result of the			28	
for a couple of			28	
in the treatment of			28	74
on the type of			28	20
the length of the			26	

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
the outcome of the			26	
the shape of your			26	
of some of our			25	
over a period of			25	23
in the heart of			23	
the head of the			23	35
the nature of the			23	20
the position of the			23	
the size of your			23	
for the treatment of				80
the use of a				49
the base of the				42
the use of the				41
a part of the				39
a portion of the				39
the top of the				39
the location of the				38
in the absence of				35
an increased risk of				34
the tip of the				31
as a treatment for				26
the part of the				26
in the development of				24
one of the first				22
on the basis of				23
to the development of				23
one or more of				20
the center of the				20
as part of the				20
of part of the				20
or part of the				20
at the site of				19
in the context of				19
in the presence of				19
one side of the				19
a small number of				18

B. Identification and focus

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
is one of the	53	148	152	88
have a lot of	43			
is going to be	32			
i am not a	22			
if you are a	20			
that there is a	20			
there are a few	20			
this type of surgery		32		
in the case of		28		
is a type of		32		
was found to be		28		
it is not a		23		
was one of the		23		
in the case of			53	
if you are a			28	
in some cases the			28	
is located in the			28	
in the case of				74
is a type of				43
is a form of				35
sometimes referred to as				30
also known as a				28
as an alternative to				27
referred to as a				27
is referred to as				26
is thought to be				26
also referred to as				20
is also known as				22
has been found to				20
there is also a				20
was one of the				20

C. Contrast and comparison

on the other hand	40	32	32	20
as well as the	29	79	99	65
as well as a			26	24
as well as in			23	

D. Deictics and locatives

	Foren	Blogs	kommerzielle Homepages	Wikipedia Artikel
at the end of	53			
in the united states	22	351	58	
at the university of		97		
on top of the		60		
from around the world		46		
in front of the		42		
on the right side		42		
on the left side		32		
the left side of		32		
all over the world		37		
at the age of		37		24
this is the first		28		
left side of the		28		
the university of california		28		
in new york city		23		
the right side of		23		
this is one of		23		
as soon as possible		23		
at the time of		18		
at the time of			49	
prior to your surgery			26	
in front of the			41	
all over the world			37	
at the university of			37	
as soon as possible			32	
at the end of			32	
for a long time			28	
at the university of				80
in the united kingdom				59
at the time of				58
at the end of				38
in front of the				23
the front of the				19

E. Vagueness markers

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit einschließlich Tabellen und Abbildungen, die anderen Werken im Wortlaut oder Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe. Ferner versichere ich, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat, dass sie noch nicht veröffentlicht worden ist und dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt.

Ort, Datum:

Unterschrift: