
Theorie und Empirie des Phonästhemis

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

vorgelegt von
Claudia Magdalena Mucha
aus
München
2015

Erstgutachter: Prof. Dr. Dietmar Zaefferer

Zweitgutachter: PD Dr. Wilhelm Oppenrieder

Tag der mündlichen Prüfung: 21.01.2014

Vorwort

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Dietmar Zaefferer für seine wertvollen Hinweise und ausgezeichnete Betreuung im gesamten Verlauf meiner Promotion.

Auch bei Herrn Priv.-Doz. Dr. Wilhelm Oppenrieder möchte ich mich für seine sehr gute Betreuung bedanken.

Vielmals danken möchte ich auch Herrn Dr. Thomas Wimmer für seine Unterstützung bei der statistischen Analyse.

Nicht zuletzt gilt mein Dank folgenden Institutionen für ihre finanzielle Förderung: der Universität Bayern e. V., die mich mit einem Graduiertenstipendium geehrt hat, der Fazit-Stiftung für ihre Unterstützung in der Anfangsphase meiner Promotion, sowie der Graduiertenschule Sprache und Literatur München für die Abschlussfinanzierung.

Inhaltsverzeichnis

1	Einleitung	7
2	Theorie des Phonästhem	11
2.1	Eingrenzung des Gegenstands	11
2.1.1	Definition des Phonästhem nach Rupert Firth	11
2.1.2	Deskriptive Erfassung von Phonästhem	13
2.2	Zur kognitiven Repräsentation von Phonästhem	34
3	Aktueller empirischer Forschungsstand zum Phonästhem	49
3.1	Zum Hintergrund der experimentellen Phonästhemforschung	49
3.1.1	Zugrundeliegende Theorien lexikalischer Wissensrepräsentation	50
3.1.2	Methodische und methodologische Probleme morphologischer Primingstudien	53
3.1.3	Rekonstruktion der Annahmen und Hypothesen in Feldman (2000)	63
3.1.4	Zur Methodologie von Replikationsversuchen	76
3.2	Zum Urexperiment von Bergen (2004)	79
3.2.1	Motivation	79
3.2.2	Aufbau	81
3.2.3	Defizite	84
4	Empirie des Phonästhem	97
4.1	Erster Replikationsversuch von Bergen (2004)	97
4.1.1	Teilnehmer	97
4.1.2	Experimentaldesign, Aufgabenstellung und Ablauf	99
4.1.3	Ergebnisse	102
4.1.4	Diskussion	112
4.2	Zweiter Replikationsversuch: Design-, Lern- und Ermüdungseffekte	114
4.2.1	Teilnehmer	114
4.2.2	Experimentaldesign, Aufgabenstellung und Ablauf	115

4.2.3	Ergebnisse	115
4.2.4	Diskussion	117
4.3	Dritter Replikationsversuch mit systematischer Verzerrung der Token-Frequenz	121
4.3.1	Teilnehmer	121
4.3.2	Experimentaldesign, Aufgabenstellung und Ablauf	122
4.3.3	Ergebnisse	124
4.3.4	Diskussion	125
4.4	Ranking lexikalischer Zielreize	127
4.4.1	Teilnehmer	127
4.4.2	Experimentaldesign, Aufgabenstellung und Ablauf	128
4.4.3	Diskussion der Ergebnisse	130
4.4.4	Schlussfolgerungen für den dritten Replikationsversuch	133
4.5	Replikationsversuch von Bergen (2010)	141
4.5.1	Teilnehmer	141
4.5.2	Experimentaldesign, Aufgabenstellung und Ablauf	141
4.5.3	Ergebnisse	142
4.5.4	Diskussion	147
4.6	Replikationsversuch der eigenen Ergebnisse mit auditivem Priming	148
4.6.1	Teilnehmer	148
4.6.2	Experimentaldesign, Aufgabenstellung und Ablauf	148
4.6.3	Ergebnisse	149
5	Zusammenfassung	155

1 Einleitung

„Phonaesthemes [...] are frequently recurring sound-meaning pairings that are not clearly contrastive morphemes.“ (Bergen 2004:290)

Obgleich ein solches Unterfangen aus systemlinguistischer Sicht von vornherein zum Scheitern verurteilt war, berichtet Benjamin Bergen in seinem bemerkenswerten Aufsatz *The psychological reality of phonaesthemes*¹. (2004), dass es ihm gelungen ist, Effekte, die bisher im Experiment nur für Morpheme und Phoneme beobachtet worden waren, auch für Phonästhemene nachzuweisen. Aus dieser Beobachtung folgt für ihn der Schluss, dass Phonästhemene mehr sein müssen als bloße statistische Auffälligkeiten (Bergen 2004:296) im Lexikon einer Sprache. Da diese Annahme bei Bergen nicht aus einer Theorie abgeleitet ist, sondern aus empirischer Beobachtung, ergeben sich aus dieser Herangehensweise nicht nur aus wissenschaftstheoretischer, sondern auch aus praktischer Sicht Problemstellungen, die im Verlauf dieser Arbeit Schritt für Schritt gelöst werden sollten.

Wie zu Beginn angedeutet, hätte die meisten konservativen Theoretiker die Frage nach der kognitiven Repräsentation von Phonästhemene wohl relativ kalt gelassen, da für sie grammatische Funktionalität bereits psychologische Realität impliziert, und wenn das Phonästhemene ersteres nicht vorweisen kann, gibt es auch keinerlei Grund letzteres anzunehmen. So ist letztendlich die Motivation von Sprachforschern wie Benjamin Bergen, Margarete Magnus und Asa Abelin², nichtsdestotrotz Zeit und Mühen in Phonästhemene zu investieren, auf eine skeptische Einstellung gegenüber der althergebrachten, rein sprachlichen Beweisführung zurückzuführen. Dieser von einem neu aufkeimende Tatendrang begleitete Skeptizismus war für Chomsky bereits in den frühen Jahren des Paradigmenwechsels von der Generativen Grammatik hin zu den Gebrauchs-basierten Ansätzen der Kognitiven Linguistik spürbar. Mit der empirischen Trendwende zerfiel der Begriff der Evidenz in zwei erkenntnistheoretische Kategorien, näm-

¹Mit psychologischer Realität ist gemeint, dass eine Funktionalität auf kognitiver Ebene unterstellt wird. Zur psychologischen Realität von Phonemen respektive Morphemen vgl. Kempley und Morton (1982), Grainger et al. (1991), Feldman & Soltano (1999), Meunier & Segui (1999), Rastle et al. (2000); Dumay et al. (2001), Monsell & Hirsh (1998), Norris et al. (2002), Radeau et al. (1995, 1998), Slowiaczek et al. (2000) sowie Spinelli et al. (2001).

²Alle drei haben ihre Dissertationen zu Phonästhemene verfasst.

lich der „Evidenz bezüglich psychologischer Realität“ und „Evidenz, die nur zur Bestätigung einer guten Theorie dient“ (Chomsky 1981:112). Allein durch die Wortwahl wird deutlich, dass Chomsky eine solche Unterteilung widerstrebt, da Theorien, mit denen sich Sprachdaten erklären lassen, ihre Tauglichkeit nicht mehr ausschließlich mit ihrer Erklärungsadäquatheit begründen können. Stattdessen sollen sie sich nun an einem neuen, psychologischen, Gütekriterium messen lassen.

Deutlich wird hier eine divergierende Auffassung des Begriffs *Erklärung*. Während für Chomsky die beschreibende Analyse sprachlicher Auffälligkeiten bereits eine hinreichende Erklärung darstellt, so sind für Empiriker deskriptive Ansätze bestenfalls notwendige Voraussetzungen für empirisch basierte, d.h. aus dem Experiment folgende, Schlussfolgerungen.

Dabei stellt sich die Frage, welchen Wert diesem neuen Gütekriterium, der empirischen Evidenz, zugesprochen werden kann. Mit der Erforschung der „Kognisierung“, Chomskys (1981:96) Alternativbegriff zur „Kompetenz“, ist die Beweislast aufgrund des darin enthaltenen strengeren Wahrheitsanspruchs jedenfalls um ein Vielfaches gestiegen.

Ist die psychologische Evidenz der linguistischen nun überlegen, weil letztere grundsätzlich „zu schwach“ (Chomsky *ibid.*) ist? Oder sind auch Szenarien denkbar, in denen „die ‚linguistische Evidenz‘ viel überzeugender sein [kann] als Sapirs ‚psychologische Evidenz‘.“ (Chomsky *ibid.*)? Kaum ein Phänomen scheint so ideal dazu geeignet zu sein die Brisanz dieser Fragestellung zu veranschaulichen wie das Phonästhem.

Wenn man bedenkt, dass für Phonästhem kein systematisches, sprachenübergreifendes Erklärungsmodell existiert, obgleich sich zahlreiche deskriptive und psycholinguistische Befunde immer wieder aufs Neue von der Idee eines kategorialen Status faszinieren lassen (z.B. Abelin 2012), so erscheint die linguistische Evidenz definitiv zu schwach zu sein. Andererseits hängt der Wahrheitsgehalt von Erklärungen, die aus experimentellen Studien folgen, von einer Vielzahl von personenabhängigen und designabhängigen Variablen ab – bereits kleinste Fehler im Versuchsplan können hier empfindliche Effekte umkehren und zu abwegigen Schlussfolgerungen führen. Da diese Variablen aufgrund der Unabhängigkeit der meisten experimentellen Studien nicht immer vollständig kontrolliert werden können, erschweren zahlreiche widersprüchliche Befunde die Erforschung z.B. des mentalen Lexikons. Deswegen hat sich – seit Bloomfield (1933:246) dies zum ersten Mal bemängelt hat – im Hinblick auf eine standardisierte, d.h. wiederholbare, Operationalisierbarkeit des Phonästhem bisher nicht viel getan.

Zum einen besteht der Sinn und Zweck der Phonästhemforschung also darin zu klären, ob „die ‚linguistische Evidenz‘ zur Begründung ‚psychologischer Realität‘ (Chomsky 1981: 111f.) ausreichen muss oder nicht. Ausreichen müsste sie dann, wenn die psychologische Be-

weisführung als gescheitert betrachtet werden kann. Das wäre für empirische Arbeiten der Fall, die keinen Bezug zu bestehenden Theorien besitzen, Replikationen verhindern oder Resultate hervorbringen, die sich aufgrund von unberücksichtigten Störfaktoren nicht interpretieren lassen. Damit wird deutlich, dass das Erreichen dieses Forschungsziels mit dem Ausmaß an Akribie bei der Auswahl von Methodologie und Versuchsmethoden und nicht zuletzt mit dem Ausmaß an Akribie bei der statistischen Analyse steht und fällt. Deswegen ist für diese Arbeit neben dem engen Theoriebezug von besonderer Bedeutung, dass Methoden, die sich innerhalb der humanwissenschaftlichen Nachbardisziplinen bewährt zu haben scheinen, nicht nur blind angewandt, sondern auch belastet, d.h. kritisch überprüft, werden. Der Sinn und Zweck der Phonästhemforschung besteht also auch darin zu untersuchen, unter welchen Bedingungen die psychologische Evidenz dazu dienen kann die linguistische Evidenz zu bestätigen, zu ergänzen oder gegebenenfalls zu berichtigen.

Fest steht, dass psychologische Evidenz und linguistische Evidenz in einem gegenseitigen Abhängigkeitsverhältnis stehen. Die Linguistik braucht die empirische Evidenz bereits dann, wenn sie ihre Unabhängigkeit von empirischer Evidenz unter Beweis stellen möchte. Umgekehrt bilden linguistische Theorien nach wie vor die Voraussetzung für die Anwendbarkeit von Methoden (z.B. aus der Psychologie oder Soziologie) für Hypothesentests.

Während im Kapitel *Theorie des Phonästhem*s wesentliche Merkmale des Gegenstands auf der Basis alter, aber auch zahlreicher neuer Belege zusammengefasst und Phonästhemer kognitiven Theorien zuspield werden, dokumentiert *Empirie des Phonästhem*s fünf teils enge, teils zwingend modifizierte Replikationsversuche des Urexperiments von Bergen (2004). Das dazwischenliegende Kapitel *Aktueller empirischer Forschungsstand zum Phonästhem* bildet das zentrale Bindeglied zwischen dem theoretischen und dem empirischen Teil. Nicht nur werden hier Studien aufgearbeitet, die in direktem Zusammenhang mit Bergen (2004) stehen, es werden darüber hinaus auch die eigenen Methoden und die eigene Methodologie begründet. Bei näherer Betrachtung der empirischen Studien wird sich herausstellen, dass eine fehlende Differenzierung zwischen dem Konzept von *Methode* und dem Konzept von *Methodologie* nicht nur wissenschaftstheoretische, sondern letztendlich auch praktische Probleme nach sich zieht.

Charakteristisch für diese Arbeit und auch aus methodologischer Sicht von besonderer Bedeutung ist der Fokus auf Replikationen. Obgleich sie aus der experimentelle Sprachforschung nicht wegzudenken sein sollten, erfreuen sie sich keiner großen Beliebtheit. Der Grund liegt auf der Hand: Replikationen scheinen nicht originell zu sein, und da potentielle Publikationen nicht selten im Hinblick auf ihre Originalität bewertet werden, werden die allerwenigsten Befunde jemals auf ihre Validität hin überprüft. Mit dieser Arbeit soll gezeigt werden, dass sich „methodische Artefakte und eine damit einhergehende Kumulation von Irrtümern“ (Diekmann

2001:164f) unter Umständen erst dann aufspüren lassen, wenn die Replikationsversuche auch kontrolliert modifiziert werden.

2 Theorie des Phonästhem

2.1 Eingrenzung des Gegenstands

2.1.1 Definition des Phonästhem nach Rupert Firth

„Nur ein Teil der Zeichen ist völlig beliebig; bei anderen kommt eine Erscheinung hinzu, die es möglich macht, Grade der Beliebigkeit zu unterscheiden, wodurch diese doch nicht aufgehoben wird: das Zeichen kann relativ motiviert sein.“
(Ferdinand de Saussure 2001:156)

Obgleich de Saussure (2001) sich bewusst war, dass Arbitrarität kein absolutes Prinzip sein kann, da Wortbildungsprodukte unbestreitbar teil-motiviert bzw. „relativ motiviert“ sind¹, verwundert es nicht, dass das Phonästhem bis dato keinen Platz innerhalb der Strukturlinguistik gefunden hat: „It is not easy to determine what are the units of speech.“ (Firth 1930:48).

Die Bezeichnung „phonestheme“ geht auf Rupert Firth (1930) zurück und wurde u.a. von Householder (1946) und Bolinger (1965) wieder aufgegriffen. Mit der Beobachtung, dass Form und Bedeutung in einer Vielzahl englischer Simplizia systematisch zusammenhängen (vgl. auch Bolinger 1949), erinnert Firth (1930) an die Künstlichkeit sprachlicher Dichotomien. Für Firth (1930:45ff) sind Wörter „[t]he most important habitual acts of man“ und Phonästheme entsprechend „phonetic habits [...] spread from context to context in the past“. Seine ätiologisch ausgerichtete Definition von Phonästhemem beinhaltet bereits 3 Annahmen über deren Verwendung und Verbreitung:

1. Phonästhem-Lexeme werden gebraucht, d.h. gesprochen und gehört („phonetic“). Bereits 1928 hat Jespersen (1928:313) amalgamierte PÄ-Lexeme vorwiegend im „slang“ verortet. Sprachen, die vorwiegend gesprochen werden und deren Verschriftlichung noch verhältnismäßig jung ist, wie z.B. das Quechua, sind deswegen relativ reich an ikonischen Zeichen (vgl. Kohn 2005).

¹„So ist *elf* unmotiviert, aber *drei-zehn* ist es nicht im selben Grade[.]“ (de Saussure 2001:156)

2. Phonästhem-Lexeme werden regelmäßig, d.h. wiederholt, gebraucht („habit“). Daraus folgt, dass sie eher in kleinen Gemeinschaften, d.h. in Idiolekten und Dialekten, entstehen und diffundieren („spread“) als in größeren Gesellschaften. Bereits 1925 hat Joseph Vendryes (1925:183) die Konstruktion eines Form-Inhalts-Zusammenhangs als eine kognitive Gewohnheit aufgefasst: „In establishing a harmony between a thing and its name, we conform to a psychic habit as old as humanity.“
3. Die Diffusion von Phonästhem-Lexemen ist bereits abgeschlossen („in the past“).

Nicht nur kann Firth als Vordenker gebrauchsbasierter Sprachtheorien bezeichnet werden, er verblüfft darüber hinaus mit Vorüberlegungen zu einem kognitiven Erklärungsmodell für Phonästhemie, die bereits Grundgedanken der Embodied Cognition-Bewegung und des Konnektionismus (vgl. Abschnitt 2.1.2) vorwegnehmen:

„[J]ust as phones are grouped into phonetic habits, so „words“ group themselves into families of linked words or related habits. The fewer derivatives and analogues a word has, the weaker the background of bodily habit, the lonlier it is. Anything may happen to it. *Horse, folk, veal, lamp, car* are lonlier than *cat, strain, scream, crumple, clatter, slacker*. Take the phone sequence *slack*. It is not an isolated habit. It is a part of a network of grouped habits. [...] Semantically it must always be considered in an actual context of situation or experience. [...] The network of phonetic habits is therefore part of the pattern of experience. Words do not in any sense hold or contain or express the meaning shown against their written forms in a dictionary.“ (Firth 1930:48f.)

Von besondere Bedeutung sind hier

1. das Konzept phonästhemischer Analogiebildung („analogue“, vgl. auch Blevins & Blevins 2009),
2. die Vorstellung, dass Kognition physisch motiviert ist („bodily habit“),
3. die Netzwerk-Metapher als Organisationsprinzip des mentalen Lexikons („network of grouped habits“),
4. ein dynamischer enzyklopädischer Bedeutungsbegriff, der sich vom traditionellen Bedeutungsbegriff nach Wörterbuch-Schema abgrenzt („Semantically [...] experience“, „The network [...] experience“, vgl. Evans & Greene 2006:215ff), sowie
5. das Konzept lexikalischer Autonomie („lonlier [...] Anything can happen to it.“).

Da, wo Firths (1930) Überlegungen aufhören, setzt Bybee (2007) wieder an:

„The autonomy hypothesis is, then, that words will have their own lexical representation if they have one or more of the following properties: low semantic complexity, high frequency, and morphophonemic irregularity or opacity.“

(Bybee 2007:52f.)

2.1.2 Deskriptive Erfassung von Phonästhemem

Induktiv können damit erste generische Eigenschaften von Phonästhemem zusammengefasst werden:

Phonotaktik

Formal treten Phonästhemem einmal als Silbenonsetkluster (SCHNupfen, SCHNüffeln; STeif, STift, STahl), aber auch als Silbenreime in Erscheinung (sURRen, klIRRen, schnURRen; bUMMELn, tUMMELn, bIMMELn). Es handelt sich dabei jedoch um kein universelles Muster, wie folgende Beispiele aus dem Tuwinischen, einer Turksprache, die in Süd-Sibirien gesprochen wird, veranschaulichen (Quelle: Harrison 2004). Hier werden die Konzepte FORTBEWEGUNG und REIBENDER KLANG mit den Silbenskeletten [d]_[l] bzw. [k]_[j] kodiert, wobei die Zungenposition bei der Artikulation des Nukleus systematisch mit dem GEWICHT bzw. dem VOLUMEN des Agenten oder des jeweiligen Objekts variiert. Auf diese Weise fungieren die Silben D[V]L- und K[V]J als komplexe Phonästhemem²:

1a)	DILdir	[vorn]	‚trippeln‘(leichten Schritts laufen)
	DOLdur	[hinten]	‚schlurfen/trotten‘ (schweren Schritts)
	DALdur	[hinten]	‚schwer aufschlagen/umfallen‘(z.B. Pferd)
	DuLdur	[hinten]	‚laut flattern‘ (z.B. schwerer Vogel)
1b)	KIJirt/KEJirt	[vorn]	‚Quietschen neuen Leders‘ (z.B. Schuhe), ‚Knirschen von Tritten auf dünner Schneeschiicht‘
	QAJurt/QOJurt	[hinten]	‚Klang von Tritten in tiefem Schnee‘
	QuJurt	[hinten]	‚Klang von Tritten in tiefem Schnee‘, ‚Quietschen‘ (z.B. neuer Ledersattel)
	KøJyrt	[vorn]	‚Kauen‘(z.B. krachend von Karotten, knirschend von Bonbons, quietschend von Käserinde)

²Bei [u] handelt es sich um einen ungerundeten geschlossenen Hinterzungenvokal.

Während im Deutschen die Silbenkoda -TSCH- mit FEUCHTIGKEIT korreliert (c), kann auch ein monokonsonantischer³ Onset wie CZ- [tʃ] im Polnischen (b) oder die silbische Sequenz (-)R[U]-⁴ (a) lautsymbolisch interpretiert werden. Während (-)R[U]- für negativ-emotive Inhalte mit dem semantischen Nenner DERBHEIT steht, kodieren CZ-initiale Lexeme oftmals bedrohliche Konzepte mit einem Fokus auf MAGIE. Im Neologismenwörterbuch der polnischen Sprache (Worbs et al. 2007) finden sich für den Kandidaten CZ- [tʃ] einige analoge expressive Neubildungen aus der Jugendsprache (u.a. *czad* ‚ultimativer Kick‘), die ihren Ursprung dem Eintrag zufolge in der Rockmusik haben, was als (wenn auch loser) Bezug zur Semantik der CZ-Clique gewertet werden könnte:

- 2a) bRUdzic (‚beschmutzen‘) RUbasznosc (‚Derbheit‘) CZarci (‚teuflisch‘)
 gRUbas (‚Fettwanst‘) RUgac (‚herunterputzen‘) mRUk (‚Sonderling‘)
 RUina (‚Ruine‘) RUja (‚Brunft‘) tRUp (‚Leiche‘)
 kRUpic sie (‚verklumpen‘) kRUpiok (‚Blutwurst‘) kRUcac (‚schreien‘)
 kRUszyc sie (‚zerfallen‘) RUNac (‚scheitern‘) RUBacha (‚Grobian‘)
 bRUKac (‚besudeln‘) tRUdzic (‚abmühen‘) RUpiec (‚Gerümpel‘)
 maRUdzic (‚nörgeln‘) mRUczec (‚brummen‘) kRUKi (‚Geier/Hyänen‘)
 kRUciata (‚Kreuzzug‘) RUra (vulg. weibl. ‚Geschlechtsorgan‘)
- 2b) CZar (‚(Zauber-)Bann‘) CZarno (‚schwarz‘) CZyrak (‚Furunkel‘)
 CZaszka (‚Totenschädel‘) CZort (‚Teufel‘) CZuch (‚Geruch‘)
 CZarCZaf (‚Schleier‘) CZubki (‚Irrenanstalt‘) CZołg (‚Panzer‘)
 CZelusc (‚Abgrund‘) CZerada (‚Gang, Meute‘) CZerw (‚Wurm, Made‘)
 CZop (‚Eiterbläschen‘) CZosnek (‚Knoblauch‘) CZad (‚Rauchschwaden‘)
 CZador (‚Kutte‘)
- 2c) ruTSCH¹en fluTSCH¹en fleTSCH¹en planTSCH¹en luTSCH¹en
 gliTSCH¹ig MaTSCH¹

Auch Silbenkerne und sogar einzelne Artikulationsmerkmale wie die Stimmhaftigkeit können phonästhemisch gedeutet werden. Dabei sind in der einschlägigen Literatur Angaben zur Type-Frequenz der lautsymbolischen Elemente in der Regel nicht gegeben (für eine Ausnahme vgl. Blust 2003), wären jedoch notwendig, um die Verbreitung der lautsymbolischen Erscheinung zu bewerten. So korrelieren Fukuda (2003) zufolge stimmlose Konsonanten in ja-

³Wobei eine Affrikate natürlich auch als bikonsonantisch aufgefasst werden könnte.

⁴Interessanterweise scheint es sich bei der homophonen Sequenz -RÓ- [rɔ] nicht um ein Phonästhem zu handeln, obgleich *smród* [smrɔd] (‚Gestank‘) sich in das Schema fügt, was jedoch ein Einzelfall zu sein scheint. Im Übrigen scheinen bis dato weder deskriptiv noch explanatorisch ausgerichtete Untersuchungen zu polnischen Phonästhemen zu existieren, so dass sämtliche in dieser Arbeit genannten Beispiele lediglich Phonästhemkandidaten darstellen, so lange keine Klarheit über ihre Type-Frequenz herrscht. Zur Quantifizierung von Phonästhem-Kandidaten siehe Abschnitt 2.1.2.

panischen Ideophonen mit geringem Volumen/geringer Dichte und stimmhafte Kononanten mit höherem Volumen/großer Dichte (a), ein Kontrast, der sprachenübergreifend im Sinne des sogenannten Tonhöhen-Codes („Frequency-Code“, vgl. Ohala 1994) durch die Alternation hoher und tiefer Vokale ausgedrückt werden kann (vgl. auch das klassische Experiment von Sapir 1929 und Yorkston & Menon 2004 und Klink 2003 zu vokalischer Vorn-hinten- bzw. Hoch-tief-Variation).

- | | | |
|----|---------------------|---|
| 3) | sakusaku - zakuzaku | ‚Klang des Mischens feiner, sandiger/grober Stoffe‘ |
| | tonton - dondon | ‚Klang eines leichten/schweren Schlags‘ |
| | torotto - dorotto | ‚Erscheinung einer scheußlichen/zähen Flüssigkeit‘ |
| | kaka - gaga | ‚Schrei eines Vogels/Menschen/Maschine‘ |

Zu bedenken gilt es jedoch, dass bei dieser scheinbaren Korrelation zwei unterschiedliche Skalenniveaus aufeinanderprallen: Volumen und Dichte sind verhältnisskaliert, während die Stimmhaftigkeit nominalskaliert ist. Die Binarität der formalen Mittel scheint hier die natürliche Relativität der Ausprägungen auf der Inhaltsebene einzuschränken.

Dabei existieren sogar für die symbolische Verwendung der Tonhöhe Gegenbeispiele im Koreanischen (Kim 1977, Shin 2005, Winter & Grawunder 2010). Während im Tuwinischen Phonästhem im Sinne des Tonhöhen-Codes alternieren (vgl. Harrison 2004), liegt bei Diminutiva im Koreanischen die umgekehrte Tendenz vor (vgl. Kim 1977): so verweist Shin (2005) darauf, dass Koreanische Frauen die Tonhöhe senken, um Unterwürfigkeit zu signalisieren.

Produktivität

Im strukturalistischen Sinne handelt es sich bei dem Phonästhem um keine grammatische Kategorie, da eine Kombinierbarkeit zwar gegeben ist, aber (zumindest im Deutschen) synchron nicht produktiv genutzt wird. Anders sieht dies z.B. im Japanischen aus, wo adverbiale Ideophone das fehlende Phänomen der metaphorischen Bedeutungsübertragung zu kompensieren scheinen: „For example, describing the sound of your shoes and how you ran for class when you were late as, „I slap-slap ran higgledy-piggledy“ sounds unnatural and unnecessary. This is not the case in Japanese[.]“ (Sharlin 2009:4).

Beispiel 4 zeigt in (sub)morphemische Elemente dekomponierte phonästhemische Lexeme⁵ (a) (Phonästhem erscheinen in Majuskeln) und lexikalische Lücken (b) im Deutschen und Englischen:

⁵Das deutsche Wortmaterial stammt aus COSMAS II (1991-2010), dem Recherchesystem zum Deutschen Referenzkorpus am Institut für Deutsche Sprache in Mannheim, das englische aus dem deutsch-englischen Online-Wörterbuch LEO (2006-12).

- 4a) SCHM + ATZ + en, KR + ATZ + en, PL + ATZ + en, m + URR + en
 z + URR + en, SCHN + URR + en, p + URR + en, KN + URR + en
 g + URR + en, PL + UMP (,plumpsen‘), SL + UMP (,absinken‘)
 CL + UMP (,verklumpen‘), ST + UMP (,stapfen‘)
- 4b) *SCHL + URR + en, *GL + URR + en, *GL + ATZ + en
 *SCHL + ATZ + en, *GL + UMP, *FL + UMP

Wo phonästhemische Reime mit monokonsonantischen Onsets zusammentreffen (wie bei *mURRen*, *zURRen*), können letztere artikulatorisch und/oder akustisch motiviert erscheinen. Möglicherweise handelt es sich bei diesen Fällen nicht durchwegs um direkte Urschöpfungen, die phonästhemisch getriggert worden sind, sondern um Produkte aus Amalgamierungen mit bedeutungsähnlichen Lexemen (vgl. *meckern* und *ziehen*):

„Blendings of synonyms play a much greater role in the development of language than is generally recognized. [...] Such blends are especially frequent in words expressive of sounds or in some other way symbolical.“ (Jespersen 1928:312f.)

In der Dichtung (Carroll 2003) und Kindersprache (Spilling 2006-7) sind Phonästhemie im spielerischen Sinn produktiv (a), wobei direkte Ad hoc-Bildungen wie *pfitzen*, die auf ihren Initiator zurückverfolgt werden können⁶ anders als phonästhemische Neubildungen, die aus der Amalgamierung zweier phonästhemischer Lexeme hervorgehen (*shockumentary*, *buildering*), selten nennenswerte Diffusion zu zeigen scheinen. Obgleich die Amalgamierung ökonomische (vgl. Steinmetz & Kipfer 2006), saliente und damit medienwirksame Wortneubildungen hervorbringt (b), so ist sie kurzlebig (vgl. Adams 2001) im Vergleich zu derivierten und zusammengesetzten Wortneuschöpfungen (c), die ein deutlich höheres Lexikalisierungspotential zu besitzen scheinen.

- | | | | |
|-----|-----------------------|--|--------------------------------|
| 5a) | *slithy | slimy + lithe | (Carroll 2003) |
| | *gliptschick | glibberig + glitschig | (Spilling 2006-7) |
| | *krummelig | krumm + kringelig | (Spilling 2006-7) |
| | *pfitzen | pf + itz + en (‘mit e.
Gummi schießen’) | (Spilling 2006-7) |
| 5b) | Tor + Orgasmus | Torgasmus | (Seidel et al. 2012) |
| | documentary + shock | shockumentary | (Wikipedia contributors 2012a) |
| | Netz + Etikette | Netikette | (Herberg et al. 2004) |
| | citizen + network | netizen | (Herberg et al. 2004) |
| | bouldering + building | buildering | (Wikipedia contributors 2012b) |

⁶Diese Differenzierung des Begriffs „Wortneubildung“ in „Ad-hoc-Bildung“ und „Neologismus“ ist an Hohenhaus (1996) angelehnt.

5c) Doppel + gäng + er	Ge + fühl	
Doppel + klick	ge + fühl + t	(Herberg et al. 2004)
Butter + brezel	Kult + film	
auf + brezel + n	kult + ig	(Herberg et al. 2004)

In der elektronischen Version des Neologismenwörterbuchs von Herberg et al. (2004), die durch das Portal für korpusbasierte Lexikografie (OWID) des Instituts für Deutsche Sprache in Mannheim (IDS) zugänglich ist, finden sich weder direkte phonästhemische Neubildungen noch solche, die auf dem Umweg der Amalgamierung entstanden wären. Insofern lässt sich auf dieser Grundlage der Verdacht, dass es sich bei Reim-Phonästhemien um sogenannte „splinters“ (Berman 1961, Adams 1973) handeln könnte, die diachron durch Reanalyse zu einem Morphem-ähnlichen Baustein umfunktioniert worden sind⁷, nicht erhärten.

Ob und inwieweit sich phonästhemische Lexeme diachron durch Amalgamierung (wie möglicherweise bei dem nicht-phonästhemischen *Kringel* = *Ringel* + *Kreis*) oder durch (ablaufende) Analogiebildung vermehrt haben (*Kringel*~*Ringel*, *TWITTER*~*chITTER*⁸~*chATTER*), kann trotz Berücksichtigung der etymologischen Erstbelege (im Folgenden aus Harper 2001-12) nicht mit absoluter Zuverlässigkeit rekonstruiert werden. Hierfür werden plausibel erscheinende Belege aus Jespersen (1928)⁹ (a) um eigene Beispiele für hypothetische phonästhemische Amalgamierungen erweitert (b):

- 6a) ? *SLing* (,niederschlagen‘, 13. Jhd.) + *dASH* (,schlagen‘, 13. Jhd.)
= *SLASH* (,aufschlitzen‘, 14. Jhd.)
? *FLow* (,fließen‘, 14. Jhd.) + *hurry* (,eilen‘, 16. Jhd.)
= *FLurry* (,Aufruhr‘, 17. Jhd.)
? *TWist* (,verdrehen‘, 14. Jhd.) + *whIRL* (,sich drehen‘, 13. Jhd.)
= *TWIRL* (,drehen/quirlen‘, 16. Jhd.)
- b) ? *SMack* (,schmecken‘, 14. Jhd.) + *mASH* (,vermischen‘, 13. Jhd.)
= *SMASH* (,zerbrechen‘, 17. Jhd.)
? *SLack* (,lose‘, 13. Jhd.) + *dump* (,werfen/fallen‘, 14. Jhd.)
= *SLump* (,(in den Schmutz) fallen‘, 16. Jhd.)
? *fiddle* (,Geige‘, 14. Jhd.) + *TWITTER* (,zwitschern‘, 14. Jhd.)

⁷Ein prominentes Beispiel hierfür wäre das Suffixoid *-(o)holic*, das bereits den Lexikalisierungsgrad erreicht hat (vgl. Haspelmath 2002).

⁸Den Angaben auf Merriam-Webster.com (2012) zufolge erscheint es etymologisch unplausibel, dass *chITTER* sich aus *chATTER* + *TWITTER* ergibt, da *chITTER* und *chATTER* etwa ein Jahrhundert vor *TWITTER* erstbelegt sind.

⁹Jespersen's Belege sind jedoch nicht durchwegs überzeugend. So ist es fraglich, ob *slender* (,schlank‘, 13. Jhd.) auf eine Kreuzung aus *slight* (,flach‘, frühes 14. Jhd.) + *tender* (,zart‘, 13. Jhd.) zurückzugeschrieben oder schlichtweg eine Analogiebildung zu *tender* darstellt.

= TWiddle (,fideln‘, 15. Jhd.)

?SCHWingen + STengel = SCHWengel (vulg. ‚männl. Geschlechtsorgan‘)

Semantik

Expressivität¹⁰ kann lexikalisch, aber auch prosodisch ausgedrückt werden, z.B. durch die Vokalquantität: *Das Eis schmeckt sooooo gut!*. Hier hängt die Dauer des Vokals mit der Intensität der positiven Geschmacksempfindung zusammen¹¹. Bei Phonästhemen signalisieren formale Merkmale¹² expressive sensorische und kinästhetische Inhalte: *gehen* vs *SCHLendern*, *essen* vs *SCHMausen*, *übergeben* vs *kOTZen*, *schauen* vs *gLOTZen*.

Im Japanischen (vgl. Sharlin 2009) existiert eine Klassifizierung der lautsymbolischen Inhaltsseite, die auch für Beispiele aus der indoeuropäischen Sprachfamilie sinnvoll erscheint: a) Giseigo (der Klang von Tieren, z.B. einer Katze), b) Giongo (der Klang von Objekten, z.B. tropfendem Wasser) und c) Gitaigo (der Klang von physischen und emotionalen Zuständen, z.B. Trägheit¹³).

Während sich phonästhemische Inhalte oft nur unzureichend sprachlich spezifizieren lassen, ist ihre Visualisierung verhältnismäßig unproblematisch¹⁴. Um die Inhaltsseite einer phonästhemischen Lexemclique zu erfassen, wird sie an dieser Stelle vorerst auf ihr gemeinsames semantisches Merkmal reduziert und dieses als provisorisches Hyperonym eingesetzt. Dass ein und dasselbe Phonästhem in vielfacher inhaltlicher Ausprägung auftreten kann, bemerkt auch Argoud (2008:47): „Nous avons observé que 45 % des ‚mots en bl-‘ont au moins un sens renvoyant à la notion de FLOT / COURANT, et 32,5 % à celle de VISION / LUMINOSITE [...]“. Das Phonästhem SCHL- beispielsweise weist vier Lesarten auf: einmal mit dem Bedeutungsner DERBHEIT (a), einmal mit dem Inhaltskern MUND (b), dann im Sinne von WINDEN/GEWUNDEN (c) und FEHLEN VON SPANNUNG (d):

- 7a) SCHL¹ampe SCHL¹acke SCHL¹achten SCHL¹agen SCHL¹udern
SCHL¹eim SCHL¹acke SCHL¹ick

¹⁰Das heißt, der Informationsgehalt, aber auch die Form einer Äußerung, übersteigt ein für die Verständlichkeit erforderliches Minimum (Detges 2001).

¹¹Vgl. auch Kohn (2005) für Beispiele zur ikonischen Verwendung prosodischer Merkmale bei den Quechua-Sprechern in Ecuador.

¹²Abschnitt 2.1.2 nimmt hierauf ausführlicher Bezug.

¹³Während im Japanischen vorwiegend reduplikative Adverbien lautsymbolisch verwendet werden, lassen sich auch einzelne submorphemische Phonästhem identifizieren, z.B. die More [no], die für TRÄGHEIT zu stehen scheint: *NOso-NOso*, *NOro-NOro* (sich langsam, schwerfällig bewegen), *NObi-NObi* (relaxt).

¹⁴Für eine theoretisch fundierte inhaltliche Spezifizierung von Phonästhemen s. Abschnitt 2.2.

- 7b) SCHL²abbern SCHL²ucken SCHL²emmen SCHL²ingen SCHL²otzen
SCHL²ürfen
- 7c) SCHL³endern SCHL³enkern SCHL³ängeln SCHL³eife SCHL³auch
SCHL³aufe SCHL³inge
- 7d) SCHL⁴app SCHL⁴aff SCHL⁴acksig SCHL⁴abbern SCHL⁴endern
SCHL⁴enkern SCHL⁴ottern

Auch KN- kann zum einen RUND bezeichnen (a), aber auch auf GERÄUSCHE (b) referieren, während -TSCH- neben FEUCHTIGKEIT (a) ebenfalls ein GERÄUSCH-Konzept kodiert (b):

- 8a) KN¹ie KN¹auf KN¹olle KN¹üllen KN¹öchel KN¹opf KN¹äuel KN¹ödel
KN¹oten KN¹ospe ruTSCH¹en fluTSCH¹en fleTSCH¹en luTSCH¹en
gliTSCH¹ig MaTSCH¹ planTSCH¹en
- 8b) KN¹istern KN¹abbern KN¹attern KN¹irschen KN¹acken KN¹arren KN¹urren
KN¹usperrn KN¹ipsen klATSCH²en planTSCH²en manTSCH²en peiTSCH²en
laTSCH²en waTSCH²eln fleTSCH²en täTSCH²eln

Obgleich homonyme Phonästhemie in diesem Beispiel auf unterschiedliche Lexeme verteilt sind, können sie auch in ein und demselben (homonymen) Lexem auftreten, wie Argoud (2008) am Beispiel des englischen BL-Phonästhemens veranschaulicht:

„Certains lexèmes ont la particularité d’être polysémiques selon le schéma suivant : l’un des sens renvoie à la notion <Flot/Courant>, alors que l’autre (ou d’autres) renvoie(nt) à <Vision/Luminosité>. Par exemple, plusieurs formes, en général obsolètes, désignent à la fois une rafale de vent, un souffle, ou le jaillement d’un liquide ; et, d’autre part, un éclair, une lueur, l’éclat du feu etc. [...] De plus, à ces cas de polysémie s’ajoutent plusieurs mots traités comme des homonymes par le lexicographe (blaze v1 et blaze v2 OED).“ (Argoud 2008:49)

Beispiel 7 zeigt zudem, dass die Ausprägungen einer phonästhemischen Clique semantisch überlappen können. So sind einige MUND-Lexeme pejorativ (*SCHLabbern*, *SCHLotzen*, *SCHLürfen*), während Lexeme des FEHLENS VON SPANNUNG ebenfalls durch negative Konnotationen gekennzeichnet sind (*SCHLaff*, *SCHLottern*). Nichtsdestotrotz erscheint es nicht zweckdienlich, schlichtweg alle Belege mit einer negativen Komponente unter dem sehr

allgemeinen Hyperonym NEGATIV zu subsumieren (für eine vertiefte Diskussion der Spezifizierungsproblematik phonästhemischer Semantik s. Abschnitt 2.2). Als Konsequenz daraus müsste für SCHL-, SCHM¹⁻¹⁵ aber auch -OTZ- (*kOTZen*, *rOTZen*, *mOTZen*, *schmOTZen*, *schlOTZen*, *prOTZen*) eine Synonymiebeziehung auf Kosten der Besonderheiten der jeweiligen Phonästhemie unterstellt werden. Andererseits ist der Obergriff DERB für SCHL¹- gezwungenermaßen künstlich und könnte genauso gut durch SCHLECHT substituiert werden, was im Grunde keinen Fortschritt zu NEGATIV darstellt.

Im Fall der polysemen Phonästhemie SCHL- (*SCHL²abbern* *SCHL²ucken* *SCHL²emmen* *SCHL²ingen* *SCHL²otzen* *SCHL²ürfen*) und SCHM- (*SCHM²ecken* *SCHM²ausen* *SCHM²unzeln* *SCHM²atzen* *SCHM²ollen*) erscheint es angemessen Synonymie in der Ausprägung hinsichtlich des Konzepts MUND anzunehmen. Im Sinne der These von der Verkörperlichten Kognition („Embodied Cognition“, vgl. Lakoff 1987) könnte argumentiert werden, dass das Konzept des Mundes aufgrund des direkten körperlichen Bezugs zur universellen menschlichen Praxis des Essens von einem kohärenteren lexikalisches Netzwerk repräsentiert ist als das relativ abstrakte Konzept von DERB/SCHLECHT, und so mag es nicht verwundern, dass die Mehrheit dieser Lexeme Formen der Nahrungsaufnahme beschreibt.

Dabei ließe sich die Kohärenz dieser Cliquen auf formaler Ebene auf ihre augenscheinliche artikulatorische (d.h. oral-sensorische) Motiviertheit zurückführen: während ein klarer Zusammenhang zum einen zwischen dem bilabialen Nasal (SCHM²-) und dem Konzept MUND und zum anderen zwischen dem Lateral (SCHL²-) und dem Konzept ZUNGE besteht (womit SCHL²- als Meronym zu SCHM²- aufzufassen wäre), ist dies zwischen Nasal bzw. Lateral und dem Konzept DERB/SCHLECHT (*SCHL¹ampe* *SCHL¹acke* vs *SCHM¹utz*, *SCHM¹iere*) nicht gegeben.

Eine gemeinsame artikulatorische Komponente reicht jedoch nicht aus, um phonästhemische Synonymie zu begründen:

- 9a) PLatzen PLärren PLaudern PLausch PLumpsen klaTSCH¹en planTSCH¹en
manTSCH¹en peiTSCH¹en laTSCH¹en waTSCH¹eln fleTSCH¹en täTSCH¹eln
- 9b) KNallen KNattern KNistern KNarren KNurren KNuspern KNirschen KNabbern
schnATtern flATtern stOTtern brETtern wETtern rATtern knATtern

Die GERÄUSCH-Lexeme unter a) scheinen einmal artikulatorisch (PL-), einmal akustisch (-TSCH-) motiviert, während die GERÄUSCH-Belege unter b) weder artikulatorisch noch akustisch begründbar sind.

¹⁵Für einen Vergleich zwischen der Semantik von SCHM¹ und der Semantik von SCHM² und SCHM³ kann bei Bedarf zu Bsp. 14 vorgegriffen werden.

Unterstellt man also, dass bestimmte Phonästhemata sensorisch motiviert sind, lässt sich das in den meisten Fällen nur rechtfertigen, indem einzelnen Phonemen symbolische – oder in Peirces (1998) Terminologie „ikonische“ – Eigenschaften zugeschrieben werden. Es müsste demnach argumentiert werden, dass FL- als FL¹- das Merkmal LUFTSTROM und als FL²- das Merkmal FEUCHTIGKEIT einmal mit dem Reibelaut F- (**FL**etschen, **FL**ießen, ein**FL**ößen) und einmal mit dem Liquid -L- (**FL**iegen, **FL**üstern, **FL**attern, **FL**uchen, **FL**ackern, **FL**öte, **FL**amme) kodiert. Bei SCHM²- und SCHN- könnten -M- und -N- jeweils für MUND und NASE stehen und -L- in SCHL- für ZUNGE. Doch welchen Beitrag leistet dann der Frikativ SCH- ? Müsste dem zufolge KN- nicht nasal motiviert sein¹⁶? Und welche Funktion erfüllt -[V]TT-?

Sensorische Motiviertheit stellt demnach kein generisches Merkmal von Phonästhemata dar, sondern tritt innerhalb der Sprachen und sprachenübergreifend betrachtet sporadisch auf, z.B. in folgenden Lexemcliquen der austronesischen Sprachen Amis (a), Toba Batak (b) und Kadazan (c): Der anlautende phonästhematische velare Nasal [ŋ] steht dabei für MUND/NASE, d.h. für die an der Bildung des velaren Nasals beteiligten Resonanzräume. In Klammern ist die jeweilige Type Cue Validity¹⁷ (TyCV), d.h. das Verhältnis phonästhemischer ŋ-initialer Basen zur Gesamtheit der ŋ-initialen Basen im jeweiligen Lexikon, angegeben¹⁸ (Quelle: Blust 2003).

- | | | |
|------|--|------------|
| 10a) | ŋoso (,Nase‘) ŋiha (,Stimme‘) ŋoyos (,Mund‘) | 32.4% TyCV |
| 10b) | ŋajaŋ (,gähnen‘) ŋjul (,flüstern‘) ŋaur (,brüllen‘) | 31.3% TyCV |
| 10c) | ŋiŋis (,lächeln‘) ŋsap (,schmatzen‘) ŋia (,gröhlen‘) | 27.3% TyCV |

Andererseits signalisiert die More [no] in japanischen Ideophonen nicht etwa oro-nasale Funktionen, sondern steht für das Konzept TRÄGHEIT (vgl. Fukuda 2003): *NOso-NOso*, *NOro-NOro* (,sich langsam, träge bewegen‘), *NObi-NObi* (,relaxt‘).

Zwar nimmt mit zunehmender Motiviertheit die Arbitrarität ab (a,b), jedoch folgt aus eingeschränkter Arbitrarität nicht zwingend Motiviertheit (c,d).

- | | |
|------|---|
| 11a) | Hund -> Köt-er/Kläff-er/Töl-e -> Wauz-i -> Wau-wau |
| 11b) | GLitzern, STEhen -> SCHNeuzen, SCHL ² ecken -> knURRen -> peiTSCHE ¹ en |

¹⁶Die dazugehörigen Lexeme haben jedoch keinen nasalen Bezug.

¹⁷Dieser Terminus stammt aus Bergen (2010).

¹⁸Für eine ausführliche Frequenz-Betrachtung s. Abschnitt 3.1.2.

11c) ghel- -> GL- / sta- -> ST- / sneu-(pneu-?) -> SCHN- / sler- -> SCHL- /
er-/or- -> -URR-

11d) STa- (stehen, stellen) STeb- (Pfoften, stützen) STEig- (stechen) STel- (stellen)
STer- (Strich, Strahl) SMeg(h)- (S)Mei-/SMeu- SMek- (Bart, Kinn)

Es wird deutlich, dass Arbitrarität als Kontinuum¹⁹ aufzufassen ist (vgl. Chung 2009:11): Während die geläufige Bezeichnung *Hund* noch vollarbiträr ist, sind *Köter*, *Kläffer* und *Töle* durch charakteristische Attribute des Hundes motiviert, nämlich das Bellen, Wohnen in einer Kote²⁰, Herumtollen und/oder Tollwütig-Sein. Bei *Wauzi* ist das Wurzelmorphem bereits lautimitativ (*wauzen*), was bei *Wauwau*, einer meronymischen Interjektion, die das Hundebellen nachahmt und in der Kindersprache zur Gattungsbezeichnung konvertiert ist²¹, durch die Reduplikation eine klare Steigerung erfährt.

Analog hierzu verhält es sich auf phonästhemischer Ebene: die Phonästhemie GL- [gl] (*GLänzen*, *GLimmen*, *GLitzern*) und ST- [ʃtʰ] (*STehen*, *STürzen* *STützen*) scheinen weder artikulatorisch noch akustisch motiviert, wohingegen SCHN- und SCHL- partiell kinästhetische²² und -URR- lautimitative Motiviertheit unterstellt werden kann.

Da die in Beispiel 11 gelisteten Assonanzen ursprünglich protoindoeuropäische Morpheme gewesen sind (Quelle: Pokorny 1959-69), kann zumindest für letztere ein Bedeutungszusammenhang der entsprechenden Lexeme sprachhistorisch nachgewiesen werden, sofern diese eindeutig auf den Bedeutungskern der jeweiligen protoindoeuropäischen Wurzel referieren²³. Dies bedeutet jedoch nicht, dass damit alle auf protoindoeuropäische Wurzeln zurückführbaren Phonästhemie vollarbiträr sein müssen (d).

Zumal auch Morpheme existieren, die keine Bedeutung tragen, d.h. nicht lexialisiert sind, wie *-mit*, *-ceive*, *-sehr-*, *-lier-* und *zimper-* (Fromkin et al. 2003, O'Grady 2001) in *sub-mit*, *re-ceive*, *un-ver-sehr-t*, *ver-lier-en*, *zimper-lich*, genügt offenbar weder das semantische Kri-

¹⁹Für die Anregung zu dieser Sichtweise soll an dieser Stelle Prof. Detges gedankt werden.

²⁰Mittelniederdeutsch *Kötte/r* für ‚Hütte/Tagelöhner‘, (s. Schiller & Lübben 2010); ide. *kot-* ‚Wohnraum‘ (s. Pokorny 1959-69). Bei *Töle* handelt es sich vermutlich um eine Ableitung von *toll*. Im *Digitalen Grimm* (2004) verweist *Töle* auf *Dohle* (‚Straßenhure‘) und diese wiederum auf *toll*.

²¹Im COSMAS II-Korpus finden sich zu *Wauwau* 173 Belege.

²²Wie weiter oben bereits erwähnt, beschränkt sich diese Funktion jeweils auf die Nasale und den Lateral. Nichtsdestotrotz kann man argumentieren, dass SCHL- an das Sammeln von Mundspeichel erinnert. Das würde auch die pejorative Färbung des SCHL- Phonästhemis erklären, die noch stärker in der Lexemclique der englischen Kognate SL- spürbar ist (vgl. Firth 1930). SCHN- kann auch bereits lautimitativ ausgelegt werden: die Kombination von oralem Reibelaut und Nasal scheint geräuschvolle Funktionen des oro-nasalen Trakts (wie das *Schneuzen* und *Schnupfen*) nachzuahmen.

²³Argoud (2008) zeigt, dass beide Ausprägungen von BL- auf entsprechende Lesarten der protoindoeuropäischen Wurzel *bhel-* zurückgehen.

terium (vgl. Katamba 1993) noch das Kriterium der Arbitrarität (vgl. Aronoff 1981), um Phonästhe-men hinreichend von Morphemen abzugrenzen. Denn selbst bei den scheinbar eindeutig motivierten Phonästhemem wie dem austronesischen η kann nicht geklärt werden, ob bei der Neubildung von Lexemen tatsächlich der Inhalt die Form bedingt hat oder nicht vielmehr reine Analogiebildung vorliegt (und der Zusammenhang zwischen η und NASE/MUND somit eine auf Zufall beruhende Scheinkorrelation darstellt). Darüber hinaus wäre es denkbar, dass durch analoge Vermehrung z.B. SCHL-initialer Lexeme, eine Reanalyse des jeweiligen Submorphems als Phonästhem angeregt worden ist, was die variierende semantische Kohärenz innerhalb der Cliques erklären würde.

Es spricht also zunächst nichts gegen Leonard Bloomfields (1933:254) Auffassung der Phonästhe-men als „root-forming morphemes“. Obgleich ein Teil der Phonästhe-men als fossilisierte Morpheme aufgefasst werden kann²⁴, erlaubt dies keine Schlussfolgerungen auf den synchro-nen kognitiven Status des Phonästhemem.

Gebrauch

Fokussiert man Gesellschaften mit einer sehr hohen Alphabetisierungsquote²⁵ wie z. B. Deutschland oder Großbritannien, so kann angesichts der formalen und semantischen Expres-sivität phonästhemischer Lexeme in Anlehnung an Firth (1930) davon ausgegangen werden, dass diese in kleineren Gemeinschaften, d.h. in Idiolekt und Dialekten, initiiert, kultiviert und verbreitet werden (vgl. Tabelle 2.1).

Die mit einem Asterisk (*) gekennzeichneten landschaftlichen Formen sind im Duden (Bibliographisches Institut GmbH 2000) und Duden Online (Bibliographisches Institut GmbH 2012) verzeichnet, die mit zwei Asterisken (**) lediglich im Duden Online (Bibliographisches Institut GmbH 2012). Die Kriterien, die zur Aufnahme der mit einem Asterisk markierten re-gionalen Lexeme in den „für die deutsche Sprachgemeinschaft bedeutsamen Wortschatz des Deutschen“ (Bibliographisches Institut GmbH 2012) geführt haben (und damit zugleich zum Ausschluss anderer SCHN-initialer regionaler Lexeme), werden nicht offengelegt²⁶. Eine Su-che im COSMAS II-Korpus des IDS (am 12.08.2012) ergibt jedenfalls lediglich 5 Belege für *SCHNofeln* (wobei Komposita hier nicht berücksichtigt sind) und *SCHNökern*, 1 für *SCHNo-*

²⁴Eine ähnliche Ansicht findet sich bei Knowles (1987).

²⁵Diese Fokussierung erübrigt sich selbstverständlich bei Völkern, die weitgehend ohne Schriftsprache auskom-men wie beispielsweise die am Amazonas in Ecuador beheimateten Napo Runa (vgl. Uzendoski 2012, Kohn 2005).

²⁶Darüber hinaus ist *SCHNofeln* in der Online-Version des *Duden* nicht als landschaftliche (österreichische) Va-riante ausgewiesen (ebensowenig wie *SCHNopern* und *SCHNofel*). Was die Form *SCHNökern* angeht, so findet sich im *Deutsch-Plattdeutsch-Wörterbuch* (2012) lediglich *snöken*, deren Schreibvariante dann *SCHNöken* lauten würde.

Tabelle 2.1: Im *Duden* (Bibliographisches Institut GmbH 2012) sowie im *Digitalen Grimm* (DG) (2004) lexikalisierte überregionale und regionale Lexeme des Phonästhem SCHN-

PÄ-Lexem	Duden	DG	PÄ-Lexem	Duden	DG
SCHNüffeln	✓	✓	SCHNofel** ‚beleidigte Miene‘	✓	
SCHNäuzen	✓	✓	SCHNäken ‚naschen‘	✓	
SCHNiefen	✓	✓	SCHNuckern ‚naschen‘		✓
SCHNauben	✓	✓	SCHNützen ‚schneuzen‘		✓
SCHNupfern	✓	✓	SCHNuppen ‚schnupfen‘		✓
SCHNupfen	✓	✓	SCHNörzen/SCHNurzen ‚schmatzen‘		✓
SCHNalzen	✓	✓	SCHNacke(l)n* ‚schnalzen/schwatzen‘	✓	✓
SCHNurren	✓	✓	SCHNädel ‚Nasenschleim‘		✓
SCHNaufen	✓	✓	SCHNuck ‚Atemzug‘		✓
SCHNarren	✓	✓	SCHNodel/SCHNott(er) ‚Nasenschleim‘		✓
SCHNarchen	✓	✓	SCHNollen ‚schmollen‘		✓
SCHNattern	✓	✓	SCHNick ‚Geplauder‘		✓
SCHNuller	✓	✓	SCHNiffing ‚Schnupftabakdose‘		✓
SCHNabel	✓	✓	SCHNodeln ‚durch verstopfte Nase atmen‘		✓
SCHNute	✓	✓	SCHNorgeln ‚Rotz hochziehen‘		✓
SCHNauze(r)	✓	✓	SCHNorbeln ‚Luft durch Nase ziehen‘		✓
SCHNorchel	✓	✓	SCHNudeln ‚Rotz hochziehen‘		✓
SCHNodderig	✓	✓	SCHNuff ‚Nase‘		✓
SCHNaps	✓	✓	SCHNorgel ‚Nase/Mund‘		✓
SCHNieben ‚schnauben‘	✓	✓	SCHNuddeln** ‚Nase hochziehen‘	✓	
SCHNopern** ‚schnupfern‘	✓	✓	SCHNodder* ‚Nasenschleim‘	✓	✓
SCHNökern** ‚naschen‘	✓	✓	SCHNofeln* ‚schnüffeln/näseln(DWB)‘	✓	✓
SCHNadern ‚schnattern‘	✓	✓			

fel und SCHNieben, 6 für SCHNuddeln, 11 für SCHNodder und 0 für SCHNope(r)n, dagegen 103 für SCHNacken und 7 für SCHNackeln. Wenn so geringe Token-Häufigkeiten für einen Eintrag genügen, sollte nichts dagegen sprechen auch SCHNasseln und SCHNörzen (je 1 Angabe) mitaufzunehmen.

Wie die folgenden Belege aus dem *Rheinischen MitmachWörterbuch* (RMW) (2007-), dem *Rheinischen Wörterbuch* (RW) (1928-71) und dem *Deutsch-Plattdeutsch-Wörterbuch* (2012)²⁷ veranschaulichen, spiegelt der dialektale Wortschatz denselben systematischen Zusammenhang zwischen Form und Bedeutung für das Phonästhem SCHN- wider wie der Standardwortschatz, wobei insbesondere Dubletten wie SCHNützen-SCHNösen ‚naschen‘ und GLeuden - GLoöggen ‚glühen‘ gegen eine zufällige Korrelation sprechen.

²⁷Dieses Online-Wörterbuch versteht sich selbst ebenfalls als Mitmach-Wörterbuch, zu dessen Erweiterung jede/r Platt-Sprecher/in seinen/ihren Beitrag leisten kann.

Tabelle 2.2: Wortbelege für das Phonästhem SCHN- mit der Bedeutung NASE/MUND aus dem *Rheinischen MitmachWörterbuch (RMW)* (2007-) und dem *Rheinischen Wörterbuch (RW)* (1928-71) samt ihrer Standarddeutschen Entsprechungen in der jeweiligen Quelle.

Lexem RMW	Lexem RW	Standardentsprechung RMW	Standardentsprechung RW
SCHNüss	SCHNüsse	‚Mund/Schnauze‘	‚Mund/Schnauze‘
SCHNäbbeln	SCHNäbbeln	‚schwätzen‘	‚schwätzen‘
SCHNäsen	SCHNäsen	‚schnüffeln (suchen)‘	naschen
SCHNüven	SCHNüven	‚schnüffeln (suchen)‘	‚schnüffeln (suchen)‘
SCHNasseln	SCHNasseln	‚Alkohol trinken‘	‚Alkohol trinken‘
SCHNeip	SCHNeipe	‚Zigarettenstummel‘	‚Zigarettenstummel‘
SCHNitt	SCHNitt	‚billiger Tabak‘	‚billiger Tabak‘
–	SCHNotter*	–	Nasenschleim
SCHNötter	–	Nasenschleim	–
SCHNüter	SCHNüter	‚Nasenschleim‘	‚Nasenschleim‘
SCHNorksen	SCHNorksen	‚grunzen/schnarchen‘	‚grunzen/schnarchen‘
SCHNörres	–	‚Schnurrbart‘	–
SCHNuckern*	SCHNuckern	‚naschen‘	‚schluchzen‘
SCHNützen*	SCHNützen	‚naschen‘	‚naschen‘
SCHNösen	SCHNösen	‚naschen‘	‚naschen‘
SCHNuppen*	SCHNuppen	‚naschen‘	‚naschen‘
SCHNörzen*	SCHNörzen	‚Süßigkeiten sammeln‘	‚Gaben sammeln‘
SCHNüff	–	‚Lust/Appetit‘	–
SCHNuven	SCHNuven	‚schneuzen‘	‚schnauben‘

Hinweis: Hier markiert der Asterisk Einträge, die auch im *Digitalen Grimm* (2004) aufgeführt sind.

- 12a) GLupen (Leer/Niedersachsen, Lippe/NRW): ‚lauernd ansehen‘;
anglüüstern (Bremen²⁸): ‚anstarren‘
- 12b) GLeien (Leer), GLämern: ‚glänzen‘
GLirrert (Parchim/Meckl.-Vorp.), GLinstern: ‚glitzern‘
GLau (ohne Ortsangabe): ‚wohlaussehend‘
- 12c) GLojen/GLeuen (Leer, Diepholz/Niedersachsen), GLennen (Leer): ‚glühen‘
GLEuden (Parchim/Meckl.-Vorp.), GLoöggen (Gütersloh/NRW): ‚glühen‘
GLent (o. O.), GLönig (Schleswig/Schlesw.-Holst.) ‚glühend‘
GLööhentich/GLöhntig (Wuppertal/NRW): ‚glühend‘
GLüüsterGLEunig (Bremen): ‚intensiv glühend‘
GLummen (Meckl.-Vorp.): ‚funkeln‘
GLösen, GLööstern (o. O.): ‚glimmen‘;
GLEinig/GLeunig (Leer, Bremen): ‚gleißend‘

Dass die prinzipiell gegebene Kombinierbarkeit von Phonästhemem bei der Wortneubildung

nicht genutzt wird, mag an dem höheren kreativen Aufwand gegenüber den klassischen Wortbildungsverfahren liegen, und kann schlussendlich als Hinweis darauf gewertet werden, dass Phonästhemene keine psychologische Realität besitzen. Auch wenn es zunächst so scheint, als könnte dies bei Dialektsprechern angesichts der höheren Anzahl lexikalisierten phonästhemischer Lexeme anders sein, so ist bei dieser Herangehensweise Vorsicht geboten. Vorschnell könnte von der Vielzahl phonästhemischer Einträge in Dialektwörterbüchern auf eine routinierte Verwendung phonästhemischer Lexeme unter Dialektsprechern geschlossen werden oder gar darauf, dass das Landleben sich positiv auf die Differenziertheit der sensorischen Empfindungen auswirkt. Aus den Dialektwörterbüchern geht jedoch nicht hervor, ob alle Dialektsprecher über ein vergleichbar vielfältiges Inventar an sensorischen Verben verfügen, oder aber die Beiträge nicht vielmehr auf wenige Sprecher mit verhältnismäßig reichem lexikalischem Inventar zurückgehen²⁹). Ebenso ist eine Assoziation von Landleben und Kontaktfreudigkeit – als Umkehrschluss zu dem Vorurteil, dass in Großstädten aufgrund der Bevölkerungsdichte und Mobilität auf nachbarschaftliche Beziehungen weniger Wert gelegt wird –, die in direkten Zusammenhang mit Mündlichkeit, Dialektalität und Expressivität gebracht werden kann, skeptisch zu betrachten.

„Stadtbewohner haben also nicht nur gleichviele soziale Kontakte wie Landbewohner, ihnen wird auch im Bedarfsfalle von der gleichen Anzahl Personen Hilfe und Unterstützung zuteil. Somit unterscheiden sich die persönlichen Netzwerke von Stadt- und Landbewohnern lediglich in der Netzwerkdichte. [...] Damit haben Landbewohner dichtere soziale Netzwerke als Stadtbewohner, d.h. die Netzwerkpersonen von Bewohnern aus Landgemeinden kennen sich untereinander häufiger als die Netzwerkpersonen von Großstädtern.“ (Petermann 2001:35ff)

Einerseits erscheint es plausibel, dass „[...] Städte in der Regel ein höheres Maß an Individualität und Selbstverwirklichung [ermöglichen] als dörfliche Gemeinschaften, die nicht nur an traditionellen Vorgegebenheiten orientiert waren und sind, sondern durch die Nachbarschaft auch mehr Verhaltenskontrolle ausüben können.“ (Bertram 1994:3) Andererseits kommt es darauf an, wo auf dem Land Daten erhoben werden. Denn es wäre nicht unplausibel anzunehmen, dass sich in spärlich bevölkerten ländlichen Gebieten (wie z.B. in Rockern, ein Stadtteil der niederbayerischen Kleinstadt Pfarrkirchen, oder in den Streusiedlungen des Münsterlands)

²⁹Das *Rheinische Mitmachwörterbuch* (2007-) gibt eigenen Angaben zufolge den aktuellen Stand der dialektalen Umgangssprache an. Es basiert dabei auf Beiträgen von Sprechern des Rheinischen, wobei unklar ist, ob die Zahl der Einträge pro Sprecher mit erhoben wird; auf den entsprechenden Internetseiten ist jedenfalls nichts dergleichen vermerkt. Die Einträge des *Rheinischen Wörterbuchs* basieren neben dialektalen Schriften auf Fragebögen, die an Dialektsprecher versandt worden sind, sowie auf Rückfragen.

der mündliche Sprachkontakt lediglich auf Familienmitglieder und die nächsten Nachbarn beschränkt.

Die Diffusion phonästhemischer Lexeme kann also nicht ausschließlich von Sprechercharakteristika, der Varietät und/oder dem Register abhängen, sondern darüber hinaus von der Netzwerkdicke und der regionalen und sozialen Mobilität der Sprecher. Es muss demnach weder so sein, dass alle Dialektsprecher alle Varianten kennen, noch dass wenige Sprecher sehr viele kennen, sondern dass die Variation innerhalb der Expressiva dadurch zu erklären ist, dass in schwach mobilen, aber dicht vernetzten Gemeinden eigene Varianten, beispielsweise für das Knarren von neuen Schuhen (a) oder das Schlittschuhfahren (b), kultiviert werden (Quelle: Sprechender Sprachatlas von Bayern 2006-2008), die dann außerhalb dieser Gemeinschaft keine Verbreitung finden.

- 13a) garETZn (Miesbach) knarATZn (Markt Schwaben) grAUZn (Andechs)
rOUZn (Mittenwald)
kwiiggETZn (Ramsau) gnAZn (Kinding) goarETZn (Altreichenau)
qwiigATZn (Aidenbach) zwiagATZn (Lalling)

- 13b) SCHL⁵ipfetzen (Babenheim) SCHL⁵eifen (Rennertshofen) SCHL⁵effen
SCHL⁵effen (Bodenmais) SCHL⁵enzen (Ramsau)

Daraus ergeben sich schwerwiegende Folgen für die Operationalisierung des Phonästhem: Auf der Makroebene kann der Konventionalisierungsgrad des Phonästhem zwar mit der Größe der lexikalisierten Phonästhem-Lexem-Clique (d.h. der Type-Frequenz) begründet werden, jedoch kann von letzterer nicht auf die psychologische Realität der Phonästhemie auf der Mikroebene geschlossen werden (wie bei Bergen 2004), da der Gebrauch phonästhemischer Lexeme (d.h. die Token-Frequenz innerhalb des jeweiligen Lexikons) mit dem Sprecher und der Sprechergemeinschaft variiert.

Soziologisch betrachtet ist der „phonetic habit“ (Firth 1930) ein Teil des Habitus einer Sprechergemeinschaft (vgl. Bourdieu 1987). D.h. „phonetic habits“ sind Verhaltenscharakteristika, die Aufschluss über die Herkunft, den sozialen Status bis hin zur sexuellen Orientierung eines Sprechers geben können (vgl. Labov 1963, 2006; Pierrehumbert et al. 2004). Die Hyponasalität beispielsweise ist ein Merkmal, das den Sprechstil vieler homosexueller Männer charakterisiert (und Pierrehumbert et al. 2004 in ihrer Studie zum hypothetischen „homosexuellen Register“ erstaunlicherweise entgeht). Darüber hinaus könnte angenommen werden, dass die Verwendung von Redensarten wie *etwas gebacken kriegen*, *aus dem Ärmel schütteln*, *den Faden verlieren* oder *kein Blatt vor den Mund nehmen* mit Größe und fehlender Dichte des sozialen Netzwerks steigt, während ein expressiver Wortschatz (d.h. ein Wortschatz, der

reich an phonsäthemischen Lexmen ist) in vergleichsweise kleinen und dichten Netzwerken dominiert.

Werden die Einflüsse der Sprechergemeinschaft (d.h. des sozialen Netzwerks und der Varietät) in einem psycholinguistischen Experiment zum kognitiven Status des Phonästhem, wie bei der lexikalischen Erkennungsaufgabe in Bergen (2004), über die Teilnehmer hinweg nicht homogenisiert bzw. kontrolliert (z.B. indem Gruppe A aus Dialektsprechern mit dichtem Netzwerk besteht und Gruppe B aus Großstädtern mit losem Netzwerk), so wäre bei den üblichen Samplegrößen und Teilnehmercharakteristika ein signifikanter Effekt des Phonästhem aus soziolinguistischer Sicht nicht zu begründen und damit auch nicht zu erwarten.

Ein weiterer Faktor für die relativ niedrige Gebrauchsfrequenz phonästhemischer Lexeme in den Standardvarietäten könnte der Einfluss der Mobilfunknutzung sein. D.h. neben der Tatsache, dass die Fähigkeiten differenziert wahrzunehmen und diese Wahrnehmungen situationsadäquat zu kodieren, unter den Sprechern variieren, muss zunächst einmal die Notwendigkeit einer Versprachlichung subtiler Erfahrungen gegeben sein. Folglich könnte man annehmen, dass die Kommunikationskultur auf dem Land aufgrund der dortigen Arbeitsbedingungen³⁰, aber auch hinsichtlich des schwächer ausgebauten Mobilfunknetzes etwas anders ausfallen mag als in der Stadt. Es überrascht nicht, dass Sonnemann (2004) zufolge durch letzteren Faktor zwar die Quantität der sozialen Kontakte steigt, jedoch zu Lasten der Gesprächsqualität, d.h. des Informationsgehaltes. Während das expressive phonästhemische Vokabular im Dialekt ein Phänomen des Mündlichen ist, so ist es in der Standardsprache ein Phänomen der Schriftsprache. Und auch hier scheint es auf bestimmte Textformen (Literatur, Poesie, Zeitungstexte) beschränkt zu sein. Die gerade von jungen Menschen immer häufiger genutzte Websprache dagegen ist eher durch Knappheit als Elaboriertheit gekennzeichnet. So können beim Chatten oder Mailen sogenannte Emoticons expressive Inhalte kodieren und somit für die Distanz zum Kommunikationspartner kompensieren. Letztendlich scheint aber der Einfluss des Fernsehens und der Schulbildung für die Breite und die Verwendung des Wortschatzes, und damit auch für die Vorkommenshäufigkeit phonästhemischer Lexeme, zentral zu sein. Angesichts der neuesten Entwicklungen in der an Internet-Kommunikation wäre es nicht verwunderlich, wenn sich phonästhemische Lexemcliquen schleichend verkleinern würden.

„Dass die mobile Internetkommunikation unsere Kommunikationspraxis ändern wird, scheint somit unbestritten[...]. Wird sich dadurch aber auch unser Schreibstil ändern? Wenn Emails tatsächlich jederzeit in Sekundenschnelle ausgetauscht werden können [...] dann wird es vermutlich mehr emails mit diskursivem Charakter

³⁰Bei Sprechern aus Stadtrandbezirken beispielsweise könnten längere Anfahrtszeiten zum Arbeitsplatz mit öffentlichen Verkehrsmitteln eine verstärkte Mobilfunknutzung begünstigen.

ter geben. Auch der Textumfang der Emails wird geringer werden, da Anfragen über Emails schrittweise entwickelt werden können. Daraus aber zu schließen, dass dies auch für das nicht-elektronische Schreiben gelten würde, halte ich für falsch. [...] Ob es künftig aber noch nicht-elektronische Briefe geben wird, ist die Frage.“ (Dürscheid 2005:96)

Das Schwinden phonästhemischer Lexeme (wie am Beispiel der -URR- Clique zu erkennen ist) kann als Symptom eines fortschreitenden Dialektverfalls gewertet werden (vgl. Stickel 1997, Seifart 2000, Raith 2012), der auf einen zunehmenden Sprachkontakt von Dialektsprechern mit Sprechern der Standardvarietäten zurückgeführt werden könnte. Während anzunehmen ist, dass demographische Faktoren wie eine relativ dünne Besiedlung ländlicher Gebiete (z.B. der niederbayerische Landkreis Regen mit 81 Ew/km²) sowie geringe Mobilität idiolektalen Sprachgebrauch begünstigen und damit die Varianz sensorischer Verben erhöhen, spricht eine hohe Bevölkerungsdichte und hohe Mobilität (wie etwa im deutschen Ballungsraum München) für eine Vereinheitlichung des Sprachgebrauchs.

Tabelle 2.3: Wortbelege für das Phonästhem -URR mit lautimitativer Bedeutung aus dem Rheinischen Wörterbuch (1928-71)

Lexem RMW	Lexem RW	Standardentsprechung RMW	Standardentsprechung RW
–	IURRen	–	‚brüllen‘
–	SCHMURRen	–	‚schwätzen‘
–	bURRen	–	‚purren‘
–	hURRen	–	‚schnurren/raufen‘
–	schURRen	–	‚scharren‘
–	tURRen	–	‚surrend fliegen‘
–	URren	–	‚murmeln‘
–	spURRen	–	‚hervorspritzen‘
–	flURRen	–	‚prügeln‘
–	kURRen	–	‚girren/wiehern‘

Das Quantifizierungsproblem als Ausgangspunkt für eine Theorie des Phonästhem

Tabelle 2.4 veranschaulicht, dass es sich rechnerisch nicht rechtfertigen lässt, SCHM- zu den Phonästhemem zu zählen, nicht jedoch M-. Auch wenn numerische Unterschiede klar vorhanden sind (wobei keine Aussagen über Signifikanzen möglich sind³¹), wird deutlich, dass sie

³¹Das liegt zum einen daran, dass aus einer bekannten endlichen Grundgesamtheit ausgezählt wird, und zum anderen daran, dass nicht davon ausgegangen werden kann, dass phonästhemische Lexeme unabhängig voneinander initiiert wurden.

unterschiedlich groß sind und die phonästhemische Phonotaktik sich nur ungenügend präzise darstellen lässt. Phonästhemische Eigenschaften sind letztendlich ausschließlich definitiv an eine polysegmentale Erscheinungsform geknüpft. Das heißt, dass man sich zwar darauf festlegen kann, monosegmentale Phone aus dem Definitionsbereich des Phonästhem auszuscheiden, diese Festlegung jedoch nicht rechnerisch motiviert sein kann.

Insgesamt enthält Tabelle 2.4 quantitative Argumente für drei deutsche Phonästhemkandidaten. Die phonästhemische Type-Frequenz ergibt sich dabei aus der Anzahl der phonästhemischen Lexeme in Verhältnis zur Anzahl aller Lexeme, die mit demselben Phonästhemkandidaten anlauten. Von besonderer Bedeutung ist hier das Argument, dass die Type-Frequenz nur beurteilt werden kann, wenn mit einer entsprechenden „Kontrollgruppe“ aus demselben semantischen Feld verglichen wird³² (Quelle zur Erhebung der Frequenzwerte: Weis 2002). Es gilt also zu zeigen, dass ausgeschlossen werden kann, in einer beliebigen (mit ein und demselben Phonem) anlautenden Lexemclique einen Zusammenhang zu finden, wie er für das jeweilige Phonästhem unterstellt wird: Auch hier zeigt sich, dass Arbitrarität als ein Kontinuum aufgefasst werden kann. Während [k]- und [ʃ]- offensichtlich nur sehr schwach (3,2 %, 1,8 %) mit dem semantischen Feld MUND/NASE bzw. SEHEN verknüpft sind, sieht es bei dem bilabialen Nasal aufgrund der artikulatorischen Eigenschaften bereits etwas anders aus (13,6%). Mit Phonetik allein lässt sich jedoch nicht erklären, weshalb [ʃm]- eine Type-Frequenz von 18,7% erreicht, [ʃn]- dagegen eine Type-Frequenz von 56,7%. Damit wird deutlich, dass das strukturelle Problem des Phonästhem auf ein starres Verständnis von Kategorialität und Arbitrarität zurückzuführen ist, so dass es sich im Grunde genommen um ein Scheinproblem handelt.

Sinnvoller als der Versuch einen Bewertungsschlüssel für diese relativen Zahlenwerte zu entwickeln erscheint es einen erweiterten Blickwinkel einzustellen, um übergeordnete Zusammenhänge zu erkennen. Ein absolutes Bewertungskriterium für Phonästhemie lässt sich so zwar auch nicht finden, dafür kann mit Monaghan et al. (2005) statt einem quantitativen ein strukturelles Argument beige-steuert werden, durch das das Phonästhem als spontane Ordnung begreifbar wird. Bevor auf dieses Argument näher eingegangen wird, soll zunächst das Konzept der Spontanen Ordnung erläutert werden.

Synchron betrachtet bieten einige Type-Frequenz-Werte Grund zu der Annahme, dass die dazugehörigen Phonästhemie (z.B. SCHN-) konventionalisiert und damit auch bei den Spre-

³²Bei diesem kontrollierten Quantifizierungsverfahren scheint es sich um eine innovative Herangehensweise zur Beurteilung einzelner Phonästhemkandidaten zu handeln. Konkret wurden dabei aus Weis (2002) die relevanten phonästhemischen Lexeme ausgezählt und in Relation zu der ausgezählten Gesamtzahl der entsprechenden konsonantenklusterinitialen Lexeme gesetzt (d.h. z.B. alle SCHL-initialen phonästhemischen Lexeme in Relation zur Gesamtzahl aller schl-initialen Lexeme. Derivationen ein und desselben Lexems wurden definitionsgemäß unter selbigem zusammengefasst und ausgeklammert.

Tabelle 2.4: Type-Frequenzen der Phonästhematik GL-, SCHN-, SCHM- mit Kontrollwerten

Bedeutung	Lexeme	TY-Frequenz	PÄ-Lexeme	TY-Frequenz
SEHEN	scheinen sheel schillern schielen schimmern schauen	1,8 % (6/336)	GLitzern GLimmen GLühen GLeißend GLänzen GLotzen	18,2 % (6/33)
NASE/ MUND ^a	Nase niesen Napf nörgeln nuckeln nagen naschen nähren Nüster nölen nuscheln nippen	9,9 % (12/121)	SCHNüffeln SCHNäuzen SCHNute SCHNaps SCHNauben SCHNupfern SCHNalzen SCHNurren SCHNarren SCHNarchen SCHNiefen SCHNupfen SCHNaufen SCHNattern SCHNnuller SCHNabel SCHNauze SCHNorchel SCHNieben SCHNodderig	56,7 % (21/37)
MUND ^b	miauen Mine Mucks Mund/en Mensa mäkeln motzen mosern murren Maul Mimik Mumps mampfen murmeln meckern	13,6 % (15/110)	SCHMatzen SCHMausen SCHMunzeln SCHMauchen SCHMollen SCHMecken	18,7 % (6/32)
*MUND ^c	kauen keifen kichern Kippe klagen klönen knirschen knuspern kosten kotzen krähen krakeelen küssen keuchen kläffen knabbern knutschen krächzen kreischen	3,2 % (19/576)		

chern in Form einer konstanten Größe mental verankert sind. Der Soziolinguist Keller (2009) nennt Konventionen spontane Ordnungen, die wie von unsichtbarer Hand entstehen. Die Idee von emergenten Strukturen als sich selbst erzeugende, sozusagen naturwüchsige Ordnungen hatte bereits der österreichische Nobelpreisträger und Ökonom Friedrich August von Hayek (1952). D.h. auch die Konventionalisierung von Phonästhematiken kann nicht auf die Eigenschaften und Intentionen einzelner Personen zurückgeführt, sondern ausschließlich mit allgemeinen Prinzipien begründet werden, die dem Handeln dieser Individuen zugrunde liegen. Deswegen

kann ein Erklärungsmodell für die Funktionalität des Phonästhemas auch nur sehr allgemein ausfallen. Eine Konventionalisierung bestimmter Phonästhemata lässt sich also nicht mit einer kognitiven Verankerung auf der Sprecherebene begründen (so wie bei Bergen 2004), sondern nur mit Prinzipien, die eine solche Verankerung bewirken.

Fasst man also sprachliche Irregularitäten wie das Phonästhemata als spontane Ordnungen in einem komplexen System auf, so lassen sie sich z.B. im Rahmen eines Künstlichen Neuronalen Netzwerks modellieren. Da Regelwidrigkeiten traditionellen Theorien der Morphologie (Hockett 1954, Lieber 1992) Schwierigkeiten bereiten – dies wird z.B. an dem Konzept des Nullmorphems deutlich – wären Erklärungsansätze für eine Funktionalität des Phonästhemas eher von den gebrauchsbasierten Theorien (Bybee 1988, Langacker 2002, Hopper & Bybee 2001) zu erwarten. Letztendlich ist es dann auch das konnektionistisch geprägte Netzwerkmodell von Bybee (2004), das eine sinnvolle Einbettung von Phonästhemata in das mentale Lexikon verspricht³³, wobei Phonästhemata – ähnlich wie Morpheme – sich als Generalisierungen (oder auch „Schemas“³⁴) auffassen lassen, die aus komplexen Beziehungssystemen emergieren.

Vor diesem Hintergrund lässt sich festhalten, dass für eine Erklärung der diachronen Produktivität von Phonästhemata auf das allgemeine Verhaltensmuster der Analogiebildung zurückgegriffen werden kann. Für eine synchrone Erklärung der Funktionalität des Phonästhemas (und damit für eine Begründung sowohl der eigenen als auch der zugrundeliegenden Arbeit von Bergen 2004) erscheint es naheliegend einmal vom Ökonomieprinzip³⁵ auszugehen und einmal vom Prinzip der Spontanen Ordnung.

Komplexe Systeme als Beziehungsnetzwerke (sogenannte ANNe „Assoziative Neuronale Netzwerke“) zu modellieren, aus denen Generalisierungen als spontane Ordnungen hervorgehen, findet schon länger Anwendung in der Statistik. Bei diesen Modellen kann es sich um Heuristiken, aber auch um Prognoseinstrumente handeln, die idealerweise bei großem Datensatz, Drittvariablen (d.h. Störfaktoren) und fehlenden Vorannahmen über zugrundeliegende Prinzipien zum Einsatz kommen können. Mit ANNen lassen sich aus diesen Datenmengen Zusammenhänge abstrahieren, aus denen ex post Hypothesen abgeleitet werden können (vgl. Zhang et al. 1998). Einfacher und eleganter ist es jedoch Hypothesen aus vorhandenen Theorien abzuleiten als Theorien auf Datenmaterial zurechtzuschneiden. Von praktischem und spielerischem Nutzen sind ANNe beispielsweise für die Prognose von Wechselkursen (Hann & Steurer 1996) und Durchfallquoten von Studenten (Gorr et al. 1994). Auch Frühwarnsysteme (Karunanithi et al. 1994) und die Gesichtserkennung (Latha et al. 2009) basieren auf solchen

³³Eine ausführliche Modellierung folgt im nächsten Abschnitt.

³⁴Das Schema-Konzept findet sich im Übrigen bereits bei Hayek (1952).

³⁵Eine ausführlichere Erklärung dieses Prinzips folgt im nächsten Abschnitt.

Netzwerkmodellen³⁶.

Wie ist nun die Brücke von Spontanen Ordnungen zurück zu dem weiter oben angekündigten strukturellen Argument von Monaghan et al. (2005) zu schlagen? Auf dem Gebiet der für die Soziologie zentralen Rational-Choice-Theorie haben Coleman (1991) und Young (2001) bedeutende konkurrierende Arbeiten verfasst. Obgleich sie unterschiedlicher Ansicht über die Bedingungen sind, unter denen Makrophänomene hervorgebracht werden, so betrachten beide soziale Konventionen als Gleichgewichtszustände³⁷. Ähnlich verhält es sich bei Monaghan et al. (2007). Ohne zu bemerken wie weitreichend ihre Phonological-Distributional-Coherence-Hypothese (Monaghan et al. 2007, 2005) tatsächlich ist, gehen auch diese Autoren in ihrem Grammatik-Verständnis (auf einem völlig anderen Fachgebiet als Coleman und Young, vgl. auch Christiansen 2010) von einem Gleichgewichtszustand aus, der auf komplementären Mustern beruht: So ist der Zusammenhang zwischen Form und Bedeutung zwar prinzipiell arbiträr, Form und syntaktische Kategorie jedoch sind systematisch verknüpft. Z.B. sind Nomina tendentiell länger als Verben, werden eher auf der ersten Silbe betont und tragen tiefe Vokale und Nasale, während sich Verben komplementär dazu verhalten³⁸. Beim Lernen von Kategorien findet dabei eine „Arbeitsteilung“ (Christiansen 2010) zwischen formalen Kriterien und Verteilungskriterien statt:

„We found an interaction between phonological and distributional cues for all four languages indicating that when distributional cues were less reliable, phonological cues were stronger. This provides converging evidence that language is structured such that language learning benefits from the integration of information about category from contextual and sound-based sources.“

(Monaghan et al. 2005:259)

In diesem System ausgewogener Beziehungen lassen sich teilmotiviert Zeichen wie phonäthemische Lexeme als Komplement zu den arbiträren Zeichen auffassen. Damit stehen sie also nicht außen vor, sondern ergänzen das mentale Lexikon. Diese Ausgewogenheit von Systematizität und Arbitrarität³⁹ kann auch mit der Sprecherabsicht begründet werden: Arbiträre Zeichen gehören ins formelle Register und vermitteln sachliche Information, während expressive

³⁶Für Geisteswissenschaftler ist dabei vielleicht besonders interessant, dass so die Bestimmung der Autorenschaft und auch die Datierung bestimmter Schriftstücke möglich wird, wie z.B. im Nachlass von Schelling (Holthausen & Ziche 2007).

³⁷Auch Marktgleichgewichte können mit Coleman (1991) und Young (2001) als spontane Ordnungen aufgefasst werden.

³⁸Monaghan et al. (2007) zeigen dies für das Niederländische, Englische, Japanische und Französische.

³⁹Christiansen (2010) spricht hier lediglich von Arbeitsteilung. Der Bezug zu den weiter reichenden Gleichgewichtszuständen wird auch hier noch nicht scharf erkannt.

Zeichen sachliche Informationen präzisieren, indem sie sie um emotive Komponenten bereichern. Einen klaren Hinweis darauf, dass es sich dabei tatsächlich um eine komplementäre und damit im Sinne eines Gleichgewichtszustandes ausgewogene Beziehung handelt, gibt die Verarbeitungshäufigkeit phonästhemischer Lexeme: letztere sind in den allermeisten Fällen⁴⁰ weniger gebräuchlich als ihre Ausnahmen (d.h. vollarbiträre Lexeme), z.B. *GLimmen*, *GLänzen* (ide. ghel-) vs *glauben* (ide. leubh-); *sIRRen*, *klIRRen* vs *irren*; *SCHNurren*, *SCHNiefen* (ide. sneu-) vs *schneiden* (ide. sker-). Unter Berücksichtigung der Beobachtungen von Monaghan et al. (2005) lässt sich festhalten, dass Phonästhemie die Signalstärke der „phonologischen Information“ ihrer Lexeme erhöhen und so einem gebrauchsbedingten Abbau dieser Lexeme aus dem Wortschatz entgegenwirken.

„We found confirmation for our hypothesis that phonological and distributional information contributed differentially towards categorisation. At points where distributional information was better for classification—the high-frequency items phonological cues were found to be of less value. Conversely, *for the lower-frequency items, where distributional information was less useful, phonological information contributed towards more accurate classification*. It is possible that the absence of quality distributional information for lower-frequency items pressures the phonological forms to remain distinct and true to the grammatical categories [...]. Because there is no other information available for these items, phonological information becomes particularly useful, and therefore preservation of any such information to assist in categorisation would be encouraged. However, this interaction between effectiveness of cues may be more subtle than just interacting with frequency. For example, Christiansen and Monaghan (in press) found that *verbs are better classified with phonological information* than distributional information, whereas nouns demonstrate the reverse pattern.“

(Monaghan et al. 2005:176, eigene Kursivierung)

2.2 Zur kognitiven Repräsentation von Phonästhemem

Während im letzten Abschnitt generische Eigenschaften des Phonästhemem induktiv zusammengetragen worden sind, wird im Folgenden deduktiv, d.h. ausgehend von theoretischen Ansätzen über die mentale Repräsentation von Sprache, aber auch von Resultaten explorativer

⁴⁰Ein Beispiel für einen Ausreißer, der im Verhältnis zum Prototypen des gleichlautenden Phonästhemem eine niedrige Frequenz zeigt, wäre *grummeln* (Phonästhemem -[V]MMEL- mit dem dazugehörigen Schema SPATIAL MOTION_PATH_*TO-FRO wie bei *tUMMELn*, *bUMMELn*).

empirischen Studien, auf die kognitiven Funktionen des Phonästhem geschlossen.

Grundsätzlich geht die Auffassung des Phonästhem als submorphemisches Element an Firths Verständnis dieser Erscheinung vorbei. Für Firth (1930:52f.) war nicht etwa die systematische Form-Bedeutungsbeziehung ein typisches Merkmal der Phonästhem, sondern deren konzeptualisierende Funktion.

Phonästhem funktionieren „in linked contexts of experience“. Sie sind „phonetically and experientially linked“. Genau genommen sind sie demnach keine Sub-Morpheme, sondern vielmehr Supra-Morpheme oder richtiger, im Sinne Fillmores (1977, 1985), lexikalische Einheiten konzeptueller Frames. Als „habits“ verankern sie kognitiv durch ihre routinierte Verwendung in der Alltagssprache.

„[A] word’s meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning. Speakers can be said to know the meaning of the word only by first understanding the background frames that motivate the concept that the word encodes. Within such an approach, words or word senses are not related to each other directly, word to word, but only by way of their links to common background frames and indications of the manner in which their meanings highlight particular elements of such frames.“

(Fillmore & Atkins 1992:76f.)

Das Besondere an Phonästhem-Lexemen ist, dass sie Erfahrungsrahmen aktivieren, von denen nicht ausgegangen werden kann, dass sie innerhalb einer Sprachgemeinschaft in allen Akteuren kognitiv verankert sind, wie z.B. im Fall von HANDEL (mit den lexikalischen Einheiten *kaufen*, *verkaufen*, etc.). So ist es sehr wahrscheinlich, dass bei den allermeisten Briten/Deutschen/Schweden das Konzept einer kommerziellen Transaktion (HANDEL) kognitiv verinnerlicht ist⁴¹, während es fraglich ist, ob alle diese Sprecher das Konzept von TIERGERÄUSCHE, das von der phonästhemischen Clique -URR- (*sURRen*, *gURRen*, *knURRen*, *schnURRen*) kodiert wird, in ein und derselben Form teilen. Nichtsdestotrotz können diese Frames so lange als konventionalisiert gelten, wie die dazugehörigen Lexeme im jeweiligen Gebrauchswortschatz aufgeführt sind.

⁴¹Die exemplarische Beschränkung auf Briten, Deutsche und Schweden impliziert, dass es sich bei Frames um keine universellen Konzepte handelt, auch wenn dies bei HANDEL zunächst intuitiv so erscheinen mag. Allerdings werden die Begriffe Verkäufer, Käufer und Währung in einer Gesellschaft, die lediglich Tauschhandel kennt (z.B. bei den Pygmäen), vermutlich keine Bedeutung haben. Grundsätzlich gilt es bei der Annahme von Universalität zu berücksichtigen, dass bisher nicht alle Völker erforscht worden sind (z.B. in Papua Neu Guinea).

Die Künstlichkeit der sprachlichen Erfassung phonästhemischer Konzepte, auf die an anderer Stelle bereits aufmerksam gemacht worden ist, lässt sich darauf zurückführen, dass sie in der Regel nicht vollständig durch das Inventar abstrakter Bildschemas nach Johnson (1987), Lakoff (1987), Gibbs et al. (1994, der das Schema RESISTANCE vorschlägt), Cienki (1998, der STRAIGHT beiträgt) und Clausner & Croft (1999, die das Inventar um LEFT-RIGHT ergänzen) zu erfassen sind. Stattdessen müssen diese klassischen Schemata um eine körperliche (d.h. akustische, artikulatorische, visuelle und/oder haptische) und/oder emotive Dimension (d.h. um „Mimetische Schemata“, vgl. Zlatev 2007) erweitert werden: „If you treat an image schema as merely an abstract formal cognitive structure, then you leave out its embodied origin and arena of operation.“ (Johnson 2005:23)

Wie Bsp. 14 zeigt, lassen sich phonästhemische Lexemcliquen durch Extensionen der klassischen Schemata OBJECT und AGENT semantisch spezifizieren: „As they are grounded in spatial experience, most image schemas discussed in the literature can be characterized in visually perceivable, spatial terms; however, they should not be thought of as tied to any single perceptual modality.“ (Cienki 1998:108) Die hier eingeführten Extensionen sind mit einem Asterisk gekennzeichnet und basieren konzeptuell auf den Mimetischen Schemata nach Zlatev (2007). In ihrer formalen Notation sind sie jedoch unabhängig von Zlatev (2007), was nicht zuletzt dem Entwurfscharakter der Mimetischen Schemata geschuldet ist:

„Prime examples of mimetic schemas with more or less universal status are those which categorize activities such as EATING, CRYING, RUNNING, CRAWLING and FLYING and actions (i.e. goal-directed acts) such as GRASP-X, PUSH-X, KICK-X and PUNCH-X. Other mimetic schemas will be more or less culturally specific such as ICE-SKATING and SHOOT-X.“ (Zlatev, in press)

Unterschieden wird im Folgenden zwischen solchen Cliquen, deren Semantik primäre akustische (*SOUND), visuelle (*LIGHT), orale (*MOUTH) und haptische (*TOUCH[*HARD, *SOFT]) Mimetische Schemata umfasst (a), und solchen, deren Inhalt zudem noch formal akustisch (-[E]SH) und artikulatorisch (-UPF-, SPR-, GR-) vermittelt wird (b⁴²).

14a)	AGT_*TOUCH+SPATIAL MOTION+ SURFACE+ITERATION	-[V]BB-	knIBBeln rUBBeln grABBeln schrUBBen verstrUBBeln krIBBeln ⁴³ krABBeln drIBBeln
	AGT+OBJ_*TOUCH_HARD+	-ITZ-	spITZen rITZen

⁴²In Beispiel 15) b) meint COUNTERFORCE bei -UPF- eine Gegenkraft zur Schwerkraft, im Gegensatz zur Semantik von -OPF-.

	+SPATIAL MOTION_CONTACT+ +SURFACE+FORCE_RESISTANCE		krITZeln schnITZen aufschlITZen
	AGT+OBJ_*TOUCH_HARD	ST ³ -	STatue STachel STab STock STuhl STein STift STahl STelze STEif STamm STirn STRamm STarr
	AGT/OBJ_*SOUND+ITERATION	[V]TTER-	rATTern knATTern flATTern stOTTern schnATTern wETTern breTTern
	AGT_*MOUTH+OBJ+SPATIAL MOTION_CONTACT+ITERATION	-UNCH	crUNCH (,knuspern‘) mUNCH (,mampfen‘) scrUNCH (,zerkauen‘) IUNCH brUNCH
	OBJ_*LIGHT+FORCE_ATTRACTION ⁴⁴	GL-	GLänzen GLühen GLimmen GLitzern GLEißend
	*SUBSTANCE_*TOUCH_SOFT	SCHM ¹ -	SCHMieren SCHMalz SCHMinke SCHMant SCHMelzen ⁴⁵
14b)	AGT+SPATIAL MOTION_PATH_UP+ +ITERATION+FORCE_COUNTERFORCE	-UPF-	hUPFen rUPFen schnUPFen lUPFen zUPFen
	AGT_*MOUTH+SPATIAL MOTION	SCHM ³ -	SCHMausen SCHMAuchen SCHMunzeln SCHMollen SCHMatzen
	AGT_*MOUTH_*SENSATION	SCHM ² -	SCHMecken SCHMAuchen SCHMatzen SCHMausen
	AGT_*EMOTION_*RAGE+AGT/ OBJ_*SOUND+SPATIAL MOTION_ _CONTACT+FORCE	-[E]SH	dASH (,zerschmettern‘) lASH (,auspeitschen) bASH (,schlagen‘) gASH (,einschneiden‘) hASH (,zerhacken‘) quASH (,vernichten‘)

		thrASH (,verhauen‘) trASH (,demolieren‘) thrESH (,verdreschen‘) slASH (,aufschlitzen‘) crASH (,zus.stoßen‘) gnASH (,knirschen‘)
*SUBSTANCE ⁴⁶ +SPATIAL MOTION_ _CONTAINMENT_OUT+SPATIAL MOTION_PATH_*DISPERSION	SPR-	SPRießen SPRühen SPRenkeln SPRengen SPRudeln
AGT_*EMOTION_*ANGER	GR ¹ -	GRantig GRobian GRam GRoll GRimmig GRummeln GRunzen GRausam
*EXP_*EMOTION_*FEAR	GR ² -	GRuseln GRauen GRäuel GRab GRuft GRotte

Auch wenn es naheliegend erscheint bei dem Entwurf dieser primären Schemata von der kindlichen Psyche auszugehen⁴⁷, so sind Zlatevs Schemata dennoch ad hoc. Außerdem lassen sich Schemas wie PUSH-X und PUNCH-X unter FORCE-COUNTERFORCE (PUSH-X) und FORCE_COMPULSION (PUNCH-X) einordnen. Stattdessen könnte man das FORCE-Schema aber auch etwas erweitern, z.B. mit FORCE_*CONSTRUCTION und FORCE_*DESTRUCTION, oder noch abstrakter mit FORCE_*POSITIVE/NEGATIVE. Gerade am Beispiel von GR- wird deutlich, dass die konzeptuelle Repräsentation phonästhemischer Lexemcliquen nicht ohne eine Erweiterung der universellen Bildschemata nach Johnson (1987) und Lakoff & Turner (1989) um sensorische (*SOUND, *LIGHT, *MOUTH, *TOUCH) und emotive (*EMOTION) Komponenten beschreibbar ist. Da die Liste der Bildschemata kein geschlossenes Inventar darstellt⁴⁸(vgl. Hampe 2005), erfordert die Analyse in Bsp. 14 neben den sensorischen Schemas Erweiterungen in der EXISTENCE-Gruppe (*SUBSTANCE) und in der SPATIAL MOTION-Gruppe (*DISPERSION). Es wird deutlich, dass die terminologischen Konzepte von Bildschemas und Mimetischen Schemas einander ergänzen können, was schlussendlich der ineffizienten Vervielfachung von theoretischen Ansätzen und Termini entgegengewirkt.

⁴⁷ „[Mimetic schemas] constitute body-based, pre-linguistic, consciously accessible representations that serve as the child’s first concepts“ (Zlatev 2007:245).

⁴⁸Im Übrigen ist diese Liste auch nicht eindeutig hierarchisch geordnet, da es keinen Konsens über eine solche hypothetische Ordnung gibt (vgl. Hampe 2005).

Neben den sensorischen phonästhemischen Lexemcliquen existieren auch solche, die ohne sensorische Komponente auskommen (a) und deren Expressivitätsgrad mit dem Abstraktheitsgrad ihrer Bildschemas variiert:

15a)	SPATIAL MOTION_PATH_ _*TO-FRO	-[V]MMEL-	tUMMELn bUMMELn fUMMELn bIMMELn PIMMEL rAMMELn BOMMEL RUMMEL wIMMELn trOMMELn (Bammel)
	VERTICAL ORIENTATION	ST ¹ -	ST ¹ ehen ST ¹ olz ST ¹ ützen ST ¹ eigen ST ¹ ufe ST ¹ upsnase ST ¹ amm ST ¹ ilett(o) ST ¹ ürzen ST ¹ ellen ST ¹ elze ST ¹ iel
15b)	OBJ+FORCE_SPLITTING	SP ¹ -	SPreizen SPalten SPrengen (Sprung) (speisen sprechen)
	OBJ+SPATIAL MOTION_ _PATH_STRAIGHT	-OPF-	PfrOPFen stOPFen kIOPFen trOPFen
	SPATIAL MOTION_PATH_ _*TO-FRO+ITERATION	SCHW-	SCHWingen SCHWabbeln SCHWirren SCHWenken SCHWappen SCHWanken
	SPATIAL MOTION_ _CONTAINMENT_OUT+ +SPATIAL MOTION_PATH_ _*DISPERSION	SPR-	SPRießen SPRühen SPRenkeln SPRitzen SPRudeln
15c)	AGT+SPATIAL MOTION_ _(UPDOWN)*TO-FRO+ITERATION	[V]CKEL-	dACKELn ⁴⁹ wACKELn zOCKELn nUCKELn rUCKELn frICKELn
	AGT+SPATIAL MOTION_ _PATH_STRAIGHT+FORCE_ _COUNTERFORCE	-UPF-	hUPFen rUPFen zUPFen schnUPFen IUPFen
	AGT+SPATIAL MOTION_	ST ² -	ST ² echen

	_PATH_STRAIGHT+CONTACT		ST ² opfen ST ² icken ST ² emmen anST ² upsen ST ² ochern ST ² ecken ST ² ippen ST ² reifen ST ² oßen
	OBJ+FORCE_SPLITTING+ +MULTIPLICITY_PART-WHOLE	SP ² -	SPan SPreu SPore SPreißel SPLiss SPplitter SPröde
15d)	AGT_*EMOTION_*RAGE+ +AGT/OBJ_*SOUND+SPATIAL MOTION_CONTACT+FORCE	-[E]SH	dASH (,zerschmettern‘) lASH (,auspeitschen) bASH (,schlagen‘) gASH (,einschneiden‘)

Anzumerken wäre hier zunächst, dass das Pfad-Schema im Hinblick auf die Bewegungsrichtung für -[V]MMEL- unterspezifiziert ist, so dass die Erweiterung *PATH_TO-FRO vorgeschlagen wird. Von _ITERATION wird hier deswegen abgesehen, da ein Teil der Instanzen des Phonästhem-Lexem-Paradigmas ein wenig kalkulierbares Fortbewegungsmuster impliziert, das eher Zickzack-artig oder atomar zu visualisieren wäre als geradlinig-iterativ wie bei *trOMMELn* und *bIMMELn*. Natürlich erscheint es sinnvoll eine weitere Unterteilung dieser Clique in Erwägung zu ziehen.

Was den Eintrag *Bammel* anbelangt, so beschreibt er einen psychischen Zustand, der nicht notwendigerweise eine körperliche Erschütterung im Sinne eines Zitterns und damit eine räumlichen Hin-und Her-Bewegung voraussetzt, weswegen er streng genommen nicht zur Clique von -[V]MMEL- gehört und an dieser Stelle als interessanter Ausreißer gelistet ist, um das wiederkehrende Problem der Festlegung von Zuordnungs- bzw. Ausschlusskriterien von formal und inhaltlich analog erscheinenden Lexemen hinsichtlich ihrer Zugehörigkeit zu phonästhemischen Cliquen zu veranschaulichen. Stattdessen wäre *Bammel* dem Schema AGENT_*EMOTION_*FEAR zuzuordnen. Ähnlich schwierig gestaltet sich der Ausschluss von *sprechen*, *speisen* und *Sprung* aus dem Paradigma von SP- (OBJ+FORCE_SPLITTING), denn dies hängt letztendlich von der Definitionsmenge von FORCE ab, d.h. von der Art der physikalischen Kraft und (im Fall von *sprechen* und *speisen*) von ihrer Intensität. Erfolgt eine Beschränkung auf kinetische Energie, so wäre *Sprung* keine lexemische Instanz von SP-, denn streng genommen können Sprünge auch durch die Einwirkung von Wärmeenergie entstehen (z.B. im Asphalt).

Der Beobachtung, dass die abstrakteren Cliquen, die ein bis zwei Bildschemata teilen, auf Analogie beruhen, da sie keine expressive Komponente zeigen, während bei den weniger ab-

strakten Cliques die Expressivität mit der Komplexität der aktivierten Schemata steigt, wird besonders gut im Vergleich der Extreme deutlich, d.h. zwischen dem sensorisch unmotivierten -[V]MMEL- vs dem sensorisch motivierten -[E]SH.

In Hinblick auf die lexematische Verarbeitungshäufigkeit scheint eine Aufstellung der Phonästhemie nach Schemakomponenten wiederum sinnvoll. Mit einer Reduktion der Schemakomplexität nimmt naturgemäß die inhaltliche Spezifität des Lexems ab, was sich letztendlich in einer erweiterten Anwendbarkeit und damit auch in einer erhöhten Verarbeitungshäufigkeit widerspiegelt. Anschaulich wird dies, wenn nochmals das Schema AGENT/OBJECT_ *SOUND + *ITERATION für das Phonästhem -[V]TTER- (wie in *rATtern* (1604) *knATtern* (733), *schnATtern* (401), *flATtern* (5836), *stOTtern* (796) aus Bsp. 14 aufgegriffen wird. Wird das Schema unter Beibehaltung der formalen Einschränkung -[V]TTER- um *SOUND reduziert, so fällt nur noch das hochfrequente *zittern* darunter, das mit 20814 Einträgen im COSMAS II Korpus gebräuchlicher ist als alle Instanzen des erweiterten -[V]TTER-Paradigmas zusammengenommen. Dem gleichen Prinzip folgend ist *ST¹upsnase*(302) innerhalb des einfachen Schemas VERTICAL ORIENTATION gebräuchlicher als *anST²upsen*(48) innerhalb des erweiterten Schemas von *ST²-AGENT + SPATIAL MOTION_PATH_STRAIGHT + CONTACT*.

Warum also verschwinden die wenig gebräuchlichen Phonästhem-Lexeme nicht aus der Sprache?⁵⁰ Als erstes Argument kann das Ökonomieprinzip herangezogen werden, das auch der phonästhemischen Analogiebildung zugrunde liegt: Phonästhemie kodieren subtile semantische Information auf einem relativ geringen formalen Raum (vgl. Ohala 1994). So ist es viel naheliegender zu sagen: *Die Vase dort drüben SCHWankt schon* als *Die Vase dort drüben beginnt gerade nicht mehr fest auf dem Untergrund zu stehen*, und situationsabhängig⁵¹ ist *Er ist heute etwas wIRR* ökonomischer als *Er ist heute etwas durcheinander*⁵².

Bei denjenigen phonästhemischen Lexemen, für die es kein phonästhemisches oder nicht-phonästhemisches Synonym zu geben scheint, wie z.B. *SCHNattern*, ist die Antwort unmittelbar gegeben. Wie verhält es sich jedoch mit *SCHWanken*, *wACKELn*, *schLOTTERn*? Auch hier ist die Synonymie sofern vorhanden nur vordergründig, d.h. partiell. In gewisser Weise scheint sogar ein komparatives Verhältnis zu bestehen: von *SCHWanken* über *wACKELn* hin

⁵⁰Dass ein solcher Abbau zumindest nicht synchron beobachtet werden kann, lässt sich aus einem eigenen Ranking-Versuch schließen, der sich jedoch in erster Linie aus einer anderen Fragestellung ergibt. Diskutiert wird dieser Versuch in Abschnitt 4.4.

⁵¹Aufgrund des pejorativen Nebensinns kommt es natürlich darauf an, wo und wem gegenüber man einen solchen Satz äußert. Sitzt man gerade in einem wissenschaftlichen Vortrag neben der Frau des referierenden Professors und meint dessen konfuse Stil kommentieren zu müssen, kann „wirr“ unter Umständen kostspieliger sein als ein „durcheinander“.

⁵²Dieses Beispielpaar für partielle Synonymie geht auf die Wikipedia contributors (2013) zurück.

zu *schlOTTERn* scheint eine Steigerung in der Frequenz der Bewegungskomponente impliziert zu sein.

Doch wie sieht es bei *SCHL⁴endern*, *TRotten*, *zOCKELn* aus? Es liegt zwar kein komparatives Verhältnis vor, jedoch sind auch hier die Lexeme nicht austauschbar. Das zweite Argument könnte lauten: In Situationen, wo die Notwendigkeit zu einer bildhaften Sprache entsteht, zum Beispiel im humoristischen Register, besteht ein hoher Bedarf an phonästhemischen Lexemen. Der Satz: „Das Kind lief die Straße entlang.“ enthält weniger anschauliche Information als der Satz: „Das Kind zOCKELte die Straße entlang.“ Letzterer beschreibt eine ganz bestimmte Art der Fortbewegung, die von einer gewissen Unbeholfenheit und Ignoranz im Hinblick auf die eigene Außenwirkung⁵³ gekennzeichnet zu sein scheint. Auch klingt ein obstruierender Faktor mit: Ein Kind, dass die Straße entlangzockelt, wird entweder sehr jung sein und eventuell durch eine Windel obstruiert, oder es trägt einen großen Tornister auf dem Rücken.

Fokussiert man, als drittes Argument, die sensorische und kinästhetische Semantik der Phonästhemie, lässt sich die Persistenz der niedrigfrequenten phonästhemischen Lexeme mit dem entwicklungspsychologischen Aspekt von Zlatev (2007:268) erklären: „[Mimetic schemas] constitute the first form of (conscious) internal representation,, d.h. konventionalisierte Phonästhem-Lexeme kodieren elementare körperliche Konzepte.

Das vierte Argument ist ein strukturelles. Offenbar kann man hier beobachten, dass relativ niedrigfrequente (phonästhemische) Lexeme einmal formal mit anderen niedrigfrequenten Lexemen verknüpft zu sein scheinen und einmal zusätzlich inhaltlich. Allerdings sind diese Verknüpfungen dabei nicht auf das jeweilige Paradigma beschränkt, sondern erstreckt sich auf Paradigmen anderer niedrigfrequenter (phonästhemischer) Lexeme: Bei einem Lexem, das zwei Phonästhemie enthält, z.B. *SPRITZen*, könnte die Aktivierung über formale und inhaltliche Verknüpfungen einen quasi konservierenden Effekt sowohl auf *SPRießen*, *SPRenkeln* usw. ausüben als auch auf das Paradigma von -ITZ-, z.B. die phonologischen Nachbarn *BLITZen* und *FL¹ITZen*. Zudem könnte eine zusätzliche inhaltliche Verknüpfung das Phonästhem BL- aktivieren und dieses wiederum das synonyme GL-. Im Sinne einer Kettenreaktion wäre es demnach denkbar, dass von der Aktivierung des phonästhemischen SPR-Paradigmas die Paradigmen von -ITZ- und BL-, und indirekt auch GL- und FL-, profitieren.

Im Falle des mehrdeutigen SCHL-Phonästhemis (*SCHL¹eim SCHL²ürfen SCHL³eife SCHL⁴ottern*) erscheint es nicht unplausibel, dass über eine semantische Verbindung zu *SCHL²-* nicht nur das *SCHM²-* (und auch *SCHM¹-*) Paradigma gestärkt würde (denn *SCHL²* und *SCHM²-* gehören ja beide zum Erfahrungsrahmen oraler Tätigkeiten), sondern über die forma-

⁵³Wobei dieser Aspekt eher bei einem zockelnden Erwachsenen von Bedeutung wäre.

le Ambiguität alle 4 Ausprägungen⁵⁴ von SCHL- sowie eventuell vorhandene mit SCHL¹, 3, 4-inhaltlich verwandte Phonästhemparadigmen. Durch einen stärkenden Effekt der Synonymie von SCHM²- und SCHL²- ließe sich somit das Ungleichgewicht zwischen der Type-Frequenz von SCHM²- (18,7%) und der Type-Frequenz von SCHN- (56,7%) abmildern (7/57-> 12,3%, 31% gesamt).

Da diese theoretische Überlegungen sich ex post ergeben⁵⁵, wäre eine empirische Überprüfung für künftige Forschung interessant. Dennoch sprechen die Ergebnisse der in Kapitel 4 beschriebenen Versuchsreihe – obgleich sie nicht auf diese hypothetischen Prozesse zugeschnitten ist – eher für eine solche Erklärung als dagegen. Zumindest ließe sich auf diese Weise verstehen, weshalb phonästhemische Lexeme trotz ihrer eingeschränkten Verwendbarkeit konventionalisiert sind und nicht aus dem standardsprachlichen Lexikon verschwinden. Legt man die Gedächtnistheorie Frederic Bartletts (1932), des Begründers der Frame-Semantik⁵⁶, zugrunde, bei dem Konventionen als kognitive Schemata aufgefasst werden, vgl. Busse (2012, 2007⁵⁷), kann die Schemaeinordnung in den Beispielen 14 und 15 dabei als eine Heuristik dieser Konventionalisierung dienen. Darüber hinaus wird deutlich, dass sich der Frame-semantische und konnektionistische Ansatz nicht gegenseitig ausschließen müssen.

Vor dem oben dargestellten theoretischen Hintergrund scheint es also, als bestünde die Funktion des Phonästhems darin, für die Spezifität und damit für einen geringen Routinisierungsgrad des Frames, an das es gebunden ist (wie z.B. TIERGERÄUSCHE: gURRen, schnURRen, knURRen oder EMISSION VON LICHT: *GLitzern*, *GLimmen*, *GLänzen*), zu kompensieren, indem es durch formale und semantische Verknüpfung zu phonästhemisch-lexikalischen Einheiten desselben Frames und anderer (sensorischer) Frames, selbsterhaltend,

⁵⁴Genau genommen finden sich in dieser Arbeit 5 Ausprägungen von SCHL-, wobei SCHL⁵- rein dialektal ist (s. Bsp. 13b).

⁵⁵Damit ist gemeint, dass der empirische Teil vorliegender Arbeit zum Zeitpunkt dieser Argumentation bereits abgeschlossen ist.

⁵⁶„Als erster Begründer des Frame-Gedankens wird meistens Marvin Minsky wahrgenommen und rezipiert. Es mag dahingestellt bleiben, ob diese Wahrnehmung den historischen Tatsachen entspricht, da zumindest in der Linguistik Fillmore bereits 1968 erste Fassungen eines Prädikations-Rahmen-Modells formuliert hat, und auch innerhalb der Kognitionswissenschaft auch nach Meinung von Minsky selbst seine Frame-Konzeption nichts anderes sei als eine neue Durchdekliniation des älteren Schema-Konzepts von Bartlett (1932).“ (Busse 2012:20)

⁵⁷„Nach dieser Überlegung konvergieren also die Prinzipien von Sprache und Kognition, von Sprache und Episteme in einem wichtigen Punkt, der diesen gemeinsam (als notwendige Voraussetzung, als *conditio sine qua non* ihrer jeweiligen Möglichkeit) zugrunde liegt: in den Prinzipien des menschlichen Gedächtnisses.[...] Nach dieser Konventionstheorie spielen also die Präzedenzfälle bisheriger erfolgreicher Zeichenverwendungen in ähnlichen / identischen Kontexten die tragende Rolle für die Existenz gesellschaftlicher Konventionen. Es ist dabei für das Entstehen und die individuelle Speicherung und Beherrschung von Konventionen unerheblich, ob in diesen vergangenen Präzedenzfällen die Zeichenbenutzer aktiv oder rezipierend tätig waren. Deutlich wird, dass Konventionalisierung hier als eine Art Kategorisierungsleistung aufgefasst wird [...]“ (Busse 2007:5f.)

die Salienz dieses Erfahrungsrahmens erhöht.

Abschließend bleibt noch anzumerken, dass das Phonästhem vielmehr stellvertretend für einen Frame zu stehen scheint, als dass es an ihn gebunden wäre, obgleich im Sinne der zu Beginn dieser Arbeit zitierten Autonomiehypothese⁵⁸ von Bybee (2007:52f.) phonästhemische Lexeme alles andere als autonom wären. Dies versteht sich im Hinblick auf die Regularität ihrer Ableitung, impliziert aber nicht, dass sie nur durch die Vorbedingung ihres Frames interpretierbar wären. Umgekehrt sind es gerade die autonomen, relativ abstrakten Verben (wie z.B. *gehen*), die von ihren Frames abhängig zu sein scheinen. In dieser Hinsicht weicht der Ansatz dieser Arbeit zur mentalen Repräsentation phonästhemischer Lexeme von Fillmore & Atkins (1992) ab, die offensichtlich von nicht-phonästhemischen Lexemen ausgehen. So gesehen zeigt das phonästhemische Lexem eine bemerkenswerte Autonomie gegenüber seinem Erfahrungsrahmen.

Abbildung 2.1 liefert einen Ausschnitt aus einem Netzwerkmodell des Phonästhem -IRR-. Dabei kann -IRR- als -URR- übergeordnet aufgefasst werden. Zugrundegelegt sind Bybees (2004) Kriterien für lexikalische Stärke:

„Some forms of a paradigm are more closely related to one another than to other forms, depending upon the following factors: semantic similarity, phonological similarity, and the token frequency of the forms. [...] Two forms will be considered more closely related if they share features of meaning than if they share features of form. The degree of semantic relatedness is determined both by the number and the nature of shared features. [...] [I]nflected forms of the same paradigm, which share the same stem and thus lexical meaning, are very closely related to one another, since inflectional affixes tend to make very small meaning contributions. [...] Different degrees of relatedness obtain among derivationally related forms. In fact, derivationally related forms tend to move away from one another semantically to a greater extent than inflectionally related forms do.“

(Bybee 2004:117f.)

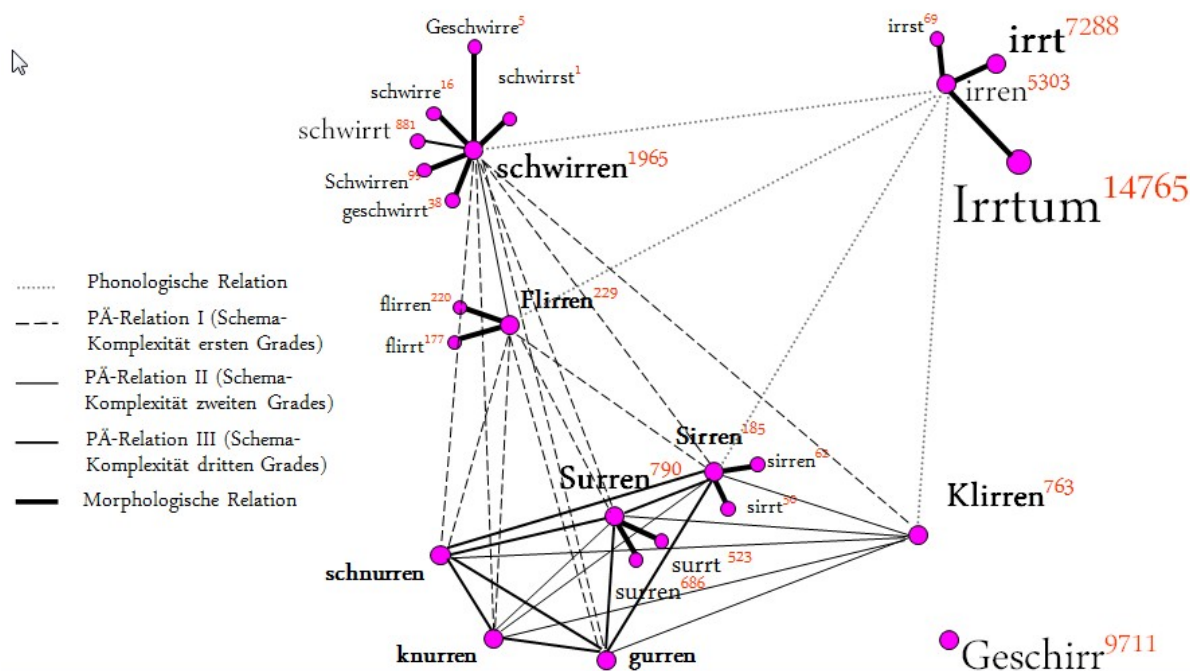
Nach diesem Modell werden in der Abbildung das Ausmaß der lexikalischen Stärke⁵⁹ durch die Schriftgröße angedeutet (d.h. *Irrtum* ist das stärkste Lexem, *schwirrst* das schwächste) und die semantischen Beziehungen⁶⁰ durch die Länge und Stärke der Verbindungslinien (so dass

⁵⁸„The autonomy hypothesis is, then, that words will have their own lexical representation if they have one or more of the following properties: low semantic complexity, high frequency, and morphophonemic irregularity or opacity.“ (Bybee 2007:52f.)

⁵⁹Diese steigt mit der Verarbeitungshäufigkeit des jeweiligen Lexems, d.h. mit der Anzahl der lexematischen Tokens. Für eine detaillierte Erläuterung des Begriffs vgl. Bybee (2004:113ff).

⁶⁰Hohe semantische Überlappung bildet kürzere Linien aus, relativ niedrige Überlappung längere Linien.

Abbildung 2.1: Vereinfachte Netzwerkdarstellung des Phonästhem -[V]RR- nach dem Modell von Bybee (2004)



innerhalb des morphologischen Paradigmas von bspw. *schwirren* die Verknüpfungen kürzer und stärker sind als zwischen *schwIRRen* und *klIRRen*). Dabei stellen die gepunkteten Linien reine Formverknüpfungen dar.

Die gestrichelten Linien bilden die phonästhemische Verknüpfung ersten Grades ab, d.h. sie gehören zum übergeordneten Phonästhem -[V]RR- mit dem groben Schema einer Hin- und Her-Bewegung (SPATIAL MOTION_PATH_*TO-FRO + ITERATION), für die noch zusätzlich eine Erweiterung von hoher Frequenz⁶¹ erarbeitet werden müsste (z.B. *_SPEED_FAST*).

Die dünnsten durchgezogenen Linien beschreiben die phonästhemische Verknüpfung zweiten Grades, die den etwas komplexeren Schemas -IRR¹- einerseits und -IRR²- und -URR- andererseits zugrundeliegt: *schwIRR¹en* und *flIRR¹en* gehören zu SPATIAL MOTION_PATH_*TO-FRO_*SPEED_*FAST + ITERATION mit der Erweiterung einer visuellen Wahrnehmung, während -IRR²- und -URR- (z.B. *klIRR²en* und *sURRen*) SPATIAL MOTION_PATH_*TO-FRO_*SPEED_*FAST + ITERATION mit der Erweiterung einer auditiven Wahrnehmung kodieren. In Langackers (2002) Terminologie wäre *schwIRR¹en* hier der Prototyp von -IRR¹-, während *flIRR¹en* eine Extension dieses Prototyps darstellen würde (vgl. auch Taylor 2002).

Die nächstbreitere durchgezogene Linie symbolisiert die phonästhemische Verbindung drit-

⁶¹Hier ist zunächst die physikalische Größe gemeint, nicht die Tonhöhe.

ten Grades zwischen den Lexemen *sIRR²en*, *sURRen*, *knURRen*, *schnURRen*, *gURRen*. Das zuletzt genannte klangliche Schema wird nun um eine weitere Komponente spezifiziert, nämlich den AGENT: alle Klänge sind in erster Linie tierischen Ursprungs (auch wenn sie von Menschen imitiert werden können). In der Peripherie dieses synchronen Netzwerks ist außerdem das Lexem *Geschirr* aufgeführt. Ähnlich wie *Irrtum* genießt es eine gewisse Autonomie, obgleich bei beiden Lexemen inhaltliche Bezüge zu den letztgenannten phonästhemischen Schemas konstruiert werden können⁶².

Abschließend soll betont werden, dass es sich bei dieser Darstellung vorerst nur um eine Heuristik und nicht um ein Erklärungsmodell für kognitive Funktionalität des Phonästhemis handelt⁶³. Was es so wertvoll macht, sind folgende Gegebenheiten: Erstens ist es aus empirischen Beobachtungen abgeleitet worden⁶⁴ und zweitens nimmt es auf eine allgemeine naturwissenschaftliche Gesetzmäßigkeit Bezug, nämlich auf das bereits angesprochene Ökonomieprinzip⁶⁵: Lexeme, die Form und Bedeutung teilen, sind stärker verknüpft als Lexeme, die nur Form oder nur Bedeutung teilen. Dahinter steckt die Logik eines zweifaktoriellen Algorithmus, wobei der Verarbeitungsvorteil gegenüber einer einfaktoriellen Suchfunktion auf einer Reduktion der zu vergleichenden Elemente beruht: in einem einfaktoriellen Suchalgorithmus muss jeder neu abgerufene Lexikoneintrag mit allen anderen gespeicherten Einträgen abgeglichen werden, während dieser Abgleich in einer zweifaktoriellen Suche auf der Basis einer Vorauswahl durch den zweiten Faktor beschleunigt wird⁶⁶. Drittens berücksichtigt es den zentralen Faktor der Type- und Tokenhäufigkeit von Lexemen. Die deskriptiven Erkenntnisse zu Phonästhemis scheinen also mit diesem theoretischen Modell kompatibel zu sein: niedrige Typefrequenz, wie im Fall der Phonästhemis (wenn man deren Type-Frequenz mit der Type-Frequenz von Derivations- und Flexionsmorphemen vergleicht) schwächt die Bildschemas,

⁶²So lässt sich ein Fehler oder Irrtum mit einer Hin- und Her-Bewegung des Kopfes signalisieren. Darüber hinaus liegt *Irrtum* die Proto-Wurzel ER- zugrunde, die „in Bewegung setzen“ bedeutet (Pokorny 1959-69), womit ein Bezug zum Basis-Schema SPATIAL MOTION_PATH unterstellt werden könnte. Daneben ist ein Hantieren mit *Geschirr* üblicherweise ziemlich geräuschvoll. Interessant ist hierbei, dass *Geschirr* diachron betrachtet wohl zum inzwischen ungebräuchlichen Lexem *anschirren* gehört hat (*Digitaler Grimm* (2004), womit auf Pferdegeschirr Bezug genommen wurde. Auditiv erfahrbar wird das Pferdegeschirr dabei durch das Kopfschütteln des Tieres. Man könnte also spekulieren, dass *anschirren* und *Geschirr* einmal zum Schema -IRR²- gehört haben könnten, auch wenn das synchron nicht mehr ganz so transparent erscheint.

⁶³Ob es sich tatsächlich dazu eignet die kognitive Funktionalität des Phonästhemis zu begründen, kann ausschließlich auf der Grundlage von Experimenten beantwortet werden.

⁶⁴Bybee (2004:136) beobachtet, dass hochfrequente Lexeme zwar den Ausgangspunkt von Lautwandel darstellen, selbst aber widerstandsfähig gegenüber Lautwandel sind.

⁶⁵Auf Ökonomie wird in Abschnitt 3.2.3 nochmals Bezug genommen.

⁶⁶Dieses Prinzip kann mit einer Spielkisten-Analogie veranschaulicht werden: es ist anzunehmen, dass bei der Suche mit geschlossenen Augen in einer Spielkiste, die mit diversen geometrischen Objekten wie Pyramiden, Zylindern, Quadern, Würfeln und Kugeln angefüllt ist, das manuelle Aussortieren aller Würfel (Faktor 1) länger dauern wird als wenn noch der Umstand hinzukommt, dass alle Würfel klebrig sind (Faktor 2).

d.h. sie zeigen niedrige Produktivität (vgl. Bybee 2004:28); niedrige Tokenfrequenz wiederum stärkt die Verknüpfung zu benachbarten Formen und macht sie anfällig für analogische Veränderungen (Bybee 2004:136).

Alles in allem lassen sich aus dem theoretischen Modell von Bybee (1985, 2004) Annahmen über die kognitive Repräsentation und Verarbeitung von Phonästhemen ableiten, die dann innerhalb eines psycholinguistischen Experiments überprüft werden können. Zudem können die Ergebnisse von Bergens (2004) Empirie zur kognitiven Realität von Phonästhemen, der das theoretische Ausgangsmodell fehlt, damit neu bewertet werden⁶⁷.

⁶⁷Eine genauere Ausführung folgt in Abschnitt 3.2.3.

3 Aktueller empirischer Forschungsstand zum Phonästhem

Den Kern dieser Arbeit bildet eine Versuchsreihe zur kognitiven Repräsentation des Phonästhem, die im Jahre 2009 mit zwei engen Replikationsversuchen von Bergen (2004) ihren Anfang genommen hat und mit einem finalen Replikationsversuch der eigenen Ergebnisse im Herbst 2011 zum Abschluss gekommen ist. Vom ersten bis zum letzten Replikationsversuch hat das ursprüngliche Experiment systematisch Modifikationen erfahren, die sich entweder aus der kritischen Diskussion des jeweils vorangegangenen Replikationsversuchs ergeben haben oder aber auf den Autor des Urexperiments selbst zurückzuführen sind, so dass der Versuchsaufbau schrittweise optimiert werden konnte. Bevor diese Experimente im folgenden Kapitel eingehend diskutiert werden, soll ein kritischer Überblick über die zugrundeliegenden Versuche von Feldman (2000) und Bergen (2004)¹ gegeben werden. Hierfür ist es zunächst notwendig auf zentrale Theorien lexikalischer Wissensrepräsentation einzugehen, aber auch auf wiederkehrende methodische und methodologische Probleme morphologischer Primingstudien, ohne die eine vollständige Bewertung der Ergebnisse von Feldman (2000) und Bergen (2004) nicht möglich wäre.

3.1 Zum Hintergrund der experimentellen

Phonästhemforschung

„Mit Replikationen erhöht sich die Wahrscheinlichkeit drastisch, dass fehlerhafte Ergebnisse aussortiert werden. Ein statistisch-signifikanter Effekt, erstmalig in der Literatur berichtet, ist nicht viel mehr als eine Hypothese. Signifikante Effekte sollte man erst dann zum Nennwert nehmen, wenn sie in Folgestudien reproduzierbar sind. Deshalb sind methodisch sorgfältige Replikationen für ein

¹Da es sich bei Bergen (2010) nicht um ein Urexperiment, sondern um eine modifizierte Replikation von Bergen (2004) handelt, die der Autor selbst durchgeführt hat, wird diese Veröffentlichung – obgleich sie die Grundlagen für einen eigenen Replikationsversuch liefert – an dieser Stelle nicht aufgeführt.

kumulatives, empirisch orientiertes Forschungsprogramm so bedeutsam. “
(Diekmann 2008:545)

Im Frühjahr 2009 wurde aus methodologischer Sicht das Verfahren des Replikationsversuchs gewählt, um der Frage nach der kognitiven Funktionalität des Phonästhem nachzugehen². Aufgrund der experimentellen Ausrichtung und nicht zuletzt aufgrund der Eklatanz des Befundes wurde hierfür als methodische Grundlage die Veröffentlichung *The psychological reality of phonaesthemes* von Bergen (2004) herangezogen.

Abgesehen davon, dass Bergens Studie aufgrund ihrer innovativen experimentellen Herangehensweise an diesen Gegenstand für einen Replikationsversuch prädestiniert erschien, enthält sie ebenso wie die Arbeiten von Feldman (2000) und Feldman & Soltano (1999), an denen sich Bergen orientiert, diverse Unklarheiten. Der Replikationsversuch einer Studie sollte also nicht nur in einer Eins-zu-eins-Wiederholung bestehen, sondern idealerweise mit einer Weiterentwicklung einhergehen. Da Bergen (2004) in gewisser Hinsicht selbst einen modifizierten Replikationsversuch eines Teilexperiments aus Feldman (2000) darstellt, müssen die Versuche von Feldman (2000) aufgearbeitet und zu Bergen (2004) in Beziehung gesetzt werden. Zuvor jedoch soll es darum gehen, das Verfahren des morphologischen Primings theoretisch zu begründen.

3.1.1 Zugrundeliegende Theorien lexikalischer Wissensrepräsentation

Weder Bergen (2004) noch Feldman (2000) gehen bei ihrer experimentellen Arbeit explizit von Theorien aus, sondern begründen ihr Forschungsinteresse damit, dass vorangegangene Befunde zum Morphem- bzw. Phonästhempriming entweder nicht schlüssig sind (Feldman) oder methodisch fragwürdig (Bergen).

Dabei bleibt es dem Leser überlassen, durch die Lektüre älterer Morphemprimingstudien (z.B. Drews & Zwitserlood 1995, Forster et al. 1987, Frost et al. 1997, Grainger et al. 1991) die Empirie in Feldman (2000) und Bergen (2004) als Konsequenz der theoretischen Debatte über die Verarbeitung (d.h. Aktivierung und Speicherung) lexikalischer Wissensrepräsentationen zu begreifen (d.h. weder Bergen noch Feldman gehen explizit auf ihre Hintergründe ein): Der Wort-basierte sogenannte „Single-Route“-Ansatz nimmt an, dass alle einfachen und komplexen Lexeme holistisch gespeichert werden (vgl. Butterworth 1983, Bybee 1985, 1988, Stemberger 1994, Giraudo & Grainer 2000), während die Morphem-basierte Theorie unterstellt, dass komplexe Wortformen als Morphembauusteine repräsentiert sind (vgl. Taft & Forster 1975, Taft 1985, 2004). Durch experimentelle Arbeit auf der Basis von regulären und

²Eine eingehende Diskussion der Methodologie von Replikationen folgt in Abschnitt 3.1.4.

irregulären Verben (Kempler & Morton 1982) hat sich dabei als hybride Theorie der Wort- und Regel-basierte Ansatz („dual route approach“) herausgebildet, der separate Mechanismen für irreguläre und reguläre Lexeme postuliert: erstere (z.B. *singen, sang, gesungen*) werden in einem assoziativen Netzwerk holistisch im Langzeitgedächtnis gespeichert, letztere (z.B. *spielen, spielte, gespielt*) werden bei der Sprachdekodierung auf der Basis der regulären Transformationsregel und der entsprechenden Wurzellexemrepräsentation rekonstruiert (vgl. Pinker 1991, Pinker & Prince 1988, 1994, Prasada & Pinker 1993).

Die generative Single-Route-Bewegung (auch: „full decomposition model“ oder „the localist approach“) nimmt demnach eine ökonomische Wissensrepräsentation an, wohingegen die konnektionistische Single-Route-Fraktion (auch: „Parallel-Distributed-Processing-view“, vgl. Daugherty & Seidenberg 1992, 1994, Elman et al. 1996, Rumelhart & McClelland 1986, Plaut & Gonnerman 2000, Rueckl et al. 1997, Seidenberg 1992) von einer immensen Speicherkapazität³ des menschlichen Gedächtnisses ausgeht (vgl. Bybee 1985, 1988, Stemberger 1994). Dafür berücksichtigt sie die anatomischen Gegebenheiten des menschlichen Gehirns: morphologische Beziehungen zwischen den Wurzellexemen und flektierten, derivierten und zusammengesetzten Formen sollen dabei aus formalen, kontextuellen (d.h. semantischen) und aktivationsbedingten⁴ Ähnlichkeitsbeziehungen hervorgehen. Im Gegensatz zum „full decomposition model“ handelt es sich dabei (wie auch bei der hybriden zweispurigen Theorie und insbesondere ihrer „Horse race“-Variante, vgl. Alegre & Gordon 1999, Baayen et al. 2007, Hay 2001, Lehtonen et al. 2006, McCormick et al. 2009) um relativ innovative Ansätze.

Da eine Fülle von empirischer Evidenz für alle drei Ansätze, die radikalen Ansätze und den hybriden Ansatz, vorgebracht worden ist, lässt sich ihre Erklärungsadäquatheit praktisch nicht beurteilen. Theoretisch wäre es aber möglich die zugrundeliegenden Studien einer eingehenden Prüfung (und gegebenenfalls einem Replikationsversuch) zu unterziehen⁵.

³Professor Robert R. Birge, Lehrstuhlinhaber für Chemie an der Universität von Connecticut und Experte für biomolekulare Elektronik, hat auf der Basis seiner Forschung zu Protein-basierten Speichermedien in einem Radiointerview des amerikanischen Radiosenders Science Update die Speicherkapazität des menschlichen Gehirns auf etwa 30 bis 40 Terabytes geschätzt. Wenn heute ein Sony Notebook der gehobenen Preisspanne über etwa 1000 Gigabyte Festplattenspeicher verfügt, so entspricht die von Prof. Birge geschätzte Hirnspeicherkapazität 30 bis 40 solchen Notebookfestplatten. Sogar Taft & Forster (1975) – obgleich ihre Dekompositionstheorie dem Ökonomieprinzip geschuldet ist – räumen ein, dass der Ökonomieanspruch einer neurophysiologischen Grundlage entbehrt.

⁴Damit ist die gebrauchsbasierte Häufigkeit der Aktivierung eines Lexemes gemeint.

⁵Sind wissenschaftliche Veröffentlichungen, in denen die Evidenz ausschließlich (deskriptiv) argumentativ beigebracht wird, relativ transparent, so gestaltet sich die Beurteilung psycholinguistischer (explanativer) Studien um einiges zeitintensiver und damit mühsamer, zumal nicht nur theoretischer und methodische, sondern nicht zuletzt auch statistische Argumente vorgebracht werden, was Linguisten ohne eine entsprechende Zusatzausbildung für gewöhnlich erhebliche Schwierigkeiten bereitet. Linguisten, die ihr Methodenspektrum erweitern möchten, sind deswegen immer noch weitgehend auf das Lehrangebot benachbarter Disziplinen wie der Soziologie oder Psychologie angewiesen.

Auf der Basis empirischer Arbeit zugunsten der Dual-Route-Annahme argumentieren dabei Baayen et al. (1997), Clahsen et al. (1992), Jaeger et al. (1996), Friederici et al. 1993, Kuperman et al. (2008), Marcus et al. (1995), Marslen-Wilson & Tyler (1997, 1998), Marsolek & Burgund (1997), Pinker (1991, 1997), Prasada & Pinker (1993) und Pinker & Prince (1994) sowie Ullman (1999). Unterschiedliche Bahnungseffekte zwischen regulären und irregulären Lexemen zeigen auch Stanners et al. (1979), Gross et al. (1998), Schreuder & Baayen (1995), Sonnenstuhl et al. (1999). Marslen-Wilson & Tyler (1998) bspw. beobachten an aphasischen Patienten Priming irregulärer (*gave-give*), aber nicht regulärer Verb-Primes (*called-call*) (vgl. auch Tyler et al. 2002a, 2002b, Miozzo 2003).

Zugunsten der konnektionistischen Single-Route-Theorie sprechen dagegen Bybee (1995), Hahn & Nakisa (2000), Marchman (1997), McClelland & Patterson (2002a, 2002b, 2003), Nakisa et al. (1998), Plunkett & Marchman (1991, 1993), Raveh (2002), Seidenberg & Gonnerman (2000), Sereno & Jongman (1997), Smolensky (1995), Stemberger & MacWhinney (1986, 1988). Eddington (2000) bspw. gelingt es die Verhaltensdaten von Prasada & Pinker (1993) mit einer Exemplar-basierten einspurigen Simulation zu replizieren, während Joanisse & Seidenberg (1999) die von neuropsychologischen Studien veranschlagte Dissoziation irregulärer und regulärer Wortformen innerhalb eines Netzwerkmodells reproduzieren.

Da sich die empirischen Ergebnisse von Stockall & Marantz (2006) weder mithilfe des Konnektionismus noch der Dual-Route-Theorie erklären lassen, schlagen sie die radikale Dekompositionsrichtung ein (vgl. auch Badecker & Caramazza 1991, Bowers 2000). Ihren Ansatz entwickeln die Autoren basierend auf der Beobachtung, dass ein Prime-Target-Paar nach dem Muster *taught-teach* zwar magnetische Aktivität im Gehirn hervorruft (gemessen mit Hilfe der Magnetoenzephalographie, kurz MEG) hervorruft, jedoch kein Priming, das umgekehrte *teach-taught* dagegen Priming und MEG-Aktivität zeigt. Für Paare, die formal stärker überlappen wie *give-gave* respektive *gave-give* sind die Befunde sowohl im Hinblick auf den MEG-Effekt als auch die Primingrichtung robust. Unerwähnt bleiben dagegen die widersprüchlichen Befunde in Stolz & Feldman (1995), die keinen signifikanten Unterschied zwischen *marked-mark* und *spoke-speak* zeigen. Auch in Feldmann (2000) werden für Paare wie *vowed-vow* morphologische Primingeffekte berichtet. Das mögliche Argument, der Unterschied bestünde darin, dass *vowed* keine irreguläre Regel aktiviert wie *taught*, wäre dabei nicht sehr überzeugend. Denn wenn Stockall & Marantz ihr ausbleibendes Priming für *taught-teach*-Paare damit erklären, dass die Aktivierung von *teach* durch die vorhergehende Aktivierung der irregulären Transformationsregel (durch das Prime *taught*) verlangsamt würde⁶, so stellt sich die Frage,

⁶„If the irregular allomorph is the prime, the rule generating the irregular allomorph will be activated by processing the prime. This rule, and the route linking the lexical entry *teach* to the irregular rule, would then be active. It's then plausible that when the root *teach* is reactivated by the target *teach*, the system is inclined

weshalb die Aktivierung einer regulären Regel weniger Zeit benötigen sollte. Ebenso wenig (und das räumen die Autoren selbst ein) zeigt sich der veranschlagte inhibierende Effekt der irregulären Regel bei *gave-give*.

Angesichts der wenig überzeugenden Argumentation im generativen Sinne, drängt sich der Verdacht auf, dass sich die Unterschiede im Priming irregulärer vs regulärer Wortformen auf den Grad der formalen Überlappung zwischen Prime und Target zurückführen lassen (vgl. Stanners et al. 1979). Unterstützung findet diese Hypothese in Pastizzo & Feldman (2002), die bei *taught-teach*-Paaren ebenso wenig wie Stockall & Marantz (2006) einen Primingeffekt messen, Instanzen eines regulären Paradigmas (*hatched-hatch*) sowie eines irregulären mit hoher formaler Überlappung (*fell-fall*) dagegen vergleichbare Primingeffekte zeigen. Aus den Daten in Voga & Grainger (2004) geht hervor, dass mit zunehmendem Grad der formalen Ähnlichkeit bei morphologischen-Prime-Target-Paaren die Fehlerrate fällt. Bei rein formal stark überlappenden Paaren wie *board-boar* messen Münte et al. (1998) keinen ERP-Effekt, so wie auch Gonnerman et al. (2007) keinen Primingeffekt für pseudo-morphologische Paare wie *hardly-hard* berichten können; stattdessen können letztere für morphologische Paare graduelle Effekte formaler Überlappung beobachten (vgl. auch Rueckl et al. 1997 und Feldman 1995 für einen gegenteiligen Bericht).

Es wird also deutlich, dass mit Hilfe von Verhaltensdaten nicht zwischen den konkurrierenden Theorien unterschieden werden kann. In Folgenden wird eine Auswahl neuerer Morphem-Primingstudien etwas ausführlicher besprochen, um auf wiederkehrende Probleme dieser Methode aufmerksam zu machen und zu zeigen, dass Mängel in Bergen (2004) sich aus einer althergebrachten Forschungsmentalität ergeben, die neu überdacht werden muss.

3.1.2 Methodische und methodologische Probleme morphologischer Primingstudien

Vorweg sollte man vielleicht erwähnen, dass es sich bei der experimentellen Sprachforschung um ein relativ junges Fachgebiet handelt. Weder Methodenkurse noch Statistikvorlesungen sind standardmäßig in die sprachwissenschaftlichen Lehrpläne integriert, wie dies z.B. bei den Sozialwissenschaftlern der Fall ist. Das Sortiment einschlägiger Publikationen zum Thema wissenschaftstheoretische Grundlagen und quantitative Forschungsmethoden scheint dementsprechend wenig Beachtung zu finden. Der Gesamteindruck der Lektüre ist dabei von einer starken Tendenz zu Standardinstruktionen geprägt, die sich weitgehend auf die Anwendung

to follow the link to the irregular rule again.“[...] „Why the prior activation of the irregular rule should be a factor in the low overlap pairs like taught teach but not in the high overlap pairs like gave give, is not immediately clear.“ (Stockall & Marantz 2006:109)

statistischer Verfahren bei bereits vorliegenden Daten beschränken (Johnson 2008, Baayen 2008, Rietveld & van Hout 2005, für eine Ausnahme s. Gries 2008). Ausnahmen sind zwar vorhanden (z.B. Broschart 2008), allerdings sind diese in den meisten Fällen pluralistisch konzipiert und passen daher ihre Methoden sehr spezifischen Fragestellungen an (z.B. Featherstone & Sternefeld 2007, Kepser & Reis 2005, Solte-Gresser et al. 2005, Stolz & Kolbe 2003, Albert & Koster 2002, Tuldava 1995), die sich zu keiner übergeordneten Erkenntnis zusammenführen lassen. Äußerst oberflächliche Anweisungen lassen dabei die ungenügende Auseinandersetzung mit grundlegenden wissenschaftstheoretischen Fragestellungen erkennen (z.B. werden die zentralen Unterschiede zwischen Falsifikationismus und Verifikationismus, Hypothesenbildung und Definition nicht diskutiert). So schlägt Bald (1977:22) vor: „benutze deine ‚Fantasie‘“, und Vaux (1999:118): „benutze deinen [...] gesunden Menschenverstand“.

Innerhalb der psycholinguistischen Experimentalforschung sind insbesondere Primingdesigns mit Worterkennungsaufgabe problembehaftet. Das ist darauf zurückzuführen, dass Priming zum einen auf sehr simplen, aber doch ausgesprochen künstlichen Aufgabenstellungen basiert, die zwar Versuchseffekte isolieren können, die dann aber nicht unbedingt auf die interessierenden Sprachverarbeitungsprozesse generalisierbar sind; zum anderen ist es besonders anfällig für Drittvariableneffekte, wie z.B. Prime- und Target-Oberflächenfrequenz, Wurzelfrequenz bei morphologisch komplexen Stimuli (nochmals unterteilbar in Familiengröße, sprich der Type-Frequenz der Wurzeln, und deren Gebrauchs- bzw. Tokenfrequenz), Zeitintervalle wie die Stimulus-Onset-Asynchrony (SOA) und das Inter-Stimulus-Intervall (ISI) und Teilnehmereffekte⁷. Auf diese Beschränkungen prallen nun die vergleichsweise hohen Ansprüche und Erwartungen der Versuchsleiter, d.h. mit einfachen Mitteln möchte man zumeist sehr komplexe Zusammenhänge zwischen kognitiven Prozesse und Repräsentationen erklären und Generalisierungen von der Versuchsgesamtheit auf eine Grundgesamtheit ableiten. Es ist deswegen von einfachen Mitteln die Rede, da keine Gewissheit über die neuronalen Mechanismen des Primings zu herrschen scheint (Schacter et al. 1993). Bemerkenswert ist dabei, dass das Priming-Verfahren nichtsdestotrotz so fest etabliert ist, dass es kaum noch hinterfragt wird⁸. Durch den relativ geringen Kostenaufwand des lexikalischen Primings und der Vielzahl der Parameter, die innerhalb eines solchen Designs manipuliert werden können, verwundert es nicht, dass sie sich nach wie vor großer Beliebtheit erfreut (vgl. Libben & Jarema 2002).

Vielleicht sollte noch hinzugefügt werden, dass nicht nur der Versuchsleiter für gewöhnlich eine favorisierte Hypothese, die sogenannte „pet hypothesis“, hat. Orne (1962) erinnert daran,

⁷Ein Teilnehmereffekt könnte z.B. darin bestehen, dass ein bestimmter Teilnehmer ein unterdurchschnittliches Reaktionsvermögen hat, übermüdet ist o.ä.

⁸Für eine Ausnahme vergleiche die Überprüfung subliminaler Effekte in Balota (1983) und Cheesman & Merikle (1984).

dass auch der Versuchsteilnehmer bestrebt ist, dem Wissenschaftler dabei zu helfen, seine Arbeitshypothese zu verifizieren.

„The subject’s performance in an experiment might almost be conceptualized as problem-solving behavior; that is, at some level he sees it as his task to ascertain the true purpose of the experiment and respond in a manner which will support the hypotheses being tested. Viewed in this light, the totality of cues which convey an experimental hypothesis to the subject become significant determinants of subjects’ behavior.“ (Orne 1962:779)

Bedenkt man die forschungspragmatischen Vorzüge des Primings, erscheint die beinahe unüberschaubare Fülle von Priming-basierter Evidenz nur ganz natürlich. Und mit der Zahl der einschlägigen Arbeiten steigt die Wahrscheinlichkeit, neben solchen Resultaten, die den eigenen Forschungsergebnissen widersprechen, auch solche vorzufinden, die sie bestätigen.

Ein gutes Beispiel für die potentielle Überforderung, die sich aus der Vielzahl möglicher Drittvariablen und der zunehmenden Unübersichtbarkeit der Forschungslandschaft ergibt, aber auch für die Schwierigkeit, mit Hilfe von Verhaltensdaten die Erklärungsadäquatheit konkurrierender theoretischer Ansätze zu beurteilen, stellen McCormick et al. (2009) dar: Die Autoren fragen sich, ob morphologische Dekomposition alle oder nur wenig gebräuchliche (niedrigfrequente) Lexeme betrifft, wie aus dem Augmented-Addressed-Morphology-Modell von Caramazza et al. (1988) und Chialant & Caramazza (1995) folgt. Schon bei der Hypothesenbildung wird deutlich, dass eine Falsifikation der Dekomposition praktisch ausgeschlossen ist, lediglich ihre Reichweite wird zur Debatte gestellt.

„We compared masked priming for derived primes that were both high in frequency (on average 60 per million) and of a higher frequency than their stem targets (on average more than four times as frequent) with the effects observed for two types of low-frequency primes. The first type consisted of low-frequency words (on average 2 per million) that were on average more than four times less frequent than their stem targets. The second type consisted of nonwords with no lexical frequency. *If morphological decomposition is limited to unfamiliar words, as predicted by the horse-race style of dual-route models [...], then priming should be limited to the last two conditions. [...]* On the other hand, if *morphological decomposition* is a routine process that applies to all morphologically-structured stimuli, then it should be observed in all three conditions.“

(McCormick et al. 2009:12, eigene Kursivierung).

Da die Drittvariable Wortfrequenz als Maß für das Aktivierungspotential von Lexemen gelten kann – d.h. hochfrequente Lexeme werden schneller erkannt als niedrigfrequente Lexeme (Marslen-Wilson 1987, Allen et al. 1992, Davis & Lupker 2006, MacKay 1982, McClelland & Rumelhart 1981, Morton 1970, Scarborough et al. 1977, Segui & Grainger 1990; dasselbe gilt für Silben s. Levelt & Wheeldon 1994, Cholin et al. 2006) – ist zu erwarten, dass niedrigfrequente Primes mit einer relativen Verzögerung der Reaktionszeiten einhergehen, wie es z.B. beim morphologischen Priming von Giraudo & Grainger (2000) der Fall ist, die von McCormick et al. (2009) offenbar übersehen worden sind. Für formal-ähnliche Paare mit kurzer SOA und Maskierung (ähnlich wie bei McCormick et al. 2009) werden dagegen inhibitorische Effekte hochfrequenter Primes berichtet (z.B. Segui & Grainger 1990), bei Davis & Lupker (2006) ist dieser Zusammenhang lediglich numerisch, bei Radeau et al. (1995) umgekehrt als numerischer Trend niedrigfrequenter Primes. Bei Grainger & Ferrand (1994) erzeugen relativ zum Target hochfrequente (homophone) Primes einen signifikant beschleunigenden Effekt. Den Widerspruch zu Segui & Grainger (1990) begründen Grainger & Ferrand (1994) ad hoc damit, dass für homophone orthographisch ähnlichen Wortpaare, so wie sie in letztgenannter Veröffentlichung zum Einsatz kommen (z.B. *fois-foie*), andere Gesetze gelten müssen als für rein orthographisch ähnliche Paare (z.B. *chat-char*). Es müsse demnach mindestens zwei mentale Lexika geben, das phonologische und das orthographische.

Berücksichtigt man diesen unklaren Kenntnisstand zur relativen Prime-Target-Frequenz erscheinen die Resultate in McCormick et al. (2009) tatsächlich wenig vorhersehbar. Die Logik zugunsten des Full-Decomposition-Ansatzes wäre also folgende: Wie bereits Giraudo & Grainger (2000) bemerkt haben, kann Dekomposition nie eine Funktion der Oberflächenfrequenz von komplexen Lexemen sein. Die Variation dieses Faktors im Design von McCormick et al. (2009) stellt somit einen auf den ersten Blick überzeugenden Falsifikationsversuch dar (wobei keine signifikanten Unterschiede zwischen den drei Gruppen erwartet werden). Zu bemängeln bleibt zunächst, dass „der Primingeffekt“ der Nullhypothese nicht definiert wird: ist ein beschleunigender Effekt oder ein verlangsamender Effekt gemeint?

Wenn nun hochfrequente komplexe Primes holistisch gespeichert werden, wie kann es dann begründet werden, dass Dekomposition des komplexen Primes, bei der die Aktivierung der Wortwurzel erst nach ihrer Freilegung erfolgt, die Reaktionen auf das Target gegenüber dem direkten Zugriff auf das jeweilige komplexe Lexem in seiner holistischen Form beschleunigt? Zunächst einmal klingt das, als würde man behaupten, ein Apfel, der erst noch geschält werden muss, (wobei der Schälprozess in dieser Analogie dem morphologischen „parsing“ entspricht) ließe sich rascher verzehren als ein Apfel, den man direkt samt Schale verspeist. Wenn Portin et al. (2008:453) von der „faster full-form route“ sprechen, so erscheint auch ihnen der

umgekehrte Zusammenhang plausibler.

Um eine Hypothese wie die obige zu rechtfertigen, müsste zusätzlich angenommen werden, dass das mentale Lexikon ähnlich formal sortiert ist wie ein Wörterbuch. Vertreter dieser abstraktionistischen Sichtweise sind u.a. Bowers & Michita (1998), Caramazza et al. (1988), Humphreys et al. (1990), Marslen-Wilson & Warren (1994a) und Morton (1979), während die Vorstellung von Lexemen als episodische Repräsentationen von Goldinger (1998), Goldinger & Azuma (2004) und Jacoby (1983) geteilt wird (für eine Diskussion der Ansätze s. Tenpenny 1995).

Im Rahmen dieser Vorstellung hätte ein freigelegtes Wurzelmorphem (z.B. *brennen*) eines komplexen derivierten Lexems (z.B. *verbrennen*) tatsächlich einen Verarbeitungsvorteil gegenüber *verbrennen*, da angenommen werden kann, dass letzteres in einen Wettstreit mit phonologischen Nachbarn gerät: es müssten also alle Lexeme, die mit *ver-* beginnen, nach *verbrennen* durchsucht werden, während eine Suche innerhalb des *br*-Klusters aufgrund der sehr viel niedrigeren Type-Frequenz⁹ relativ zügig verlaufen könnte. Im Übrigen wird dieses Argument in McCormick et al. (2009) nicht vorgebracht, was nicht gerade zur Transparenz der Hypothesen beiträgt, sondern stammt aus Knuth (1973).

Es leuchtet unmittelbar ein, dass die Wörterbuch-Analogie nur in Fällen präfigierter Formen treffend ist, McCormick et al. (2009) implementieren jedoch nur suffigiertere Formen. Tatsächlich können Baayen et al. (2007) einen deutlichen Verarbeitungsunterschied zwischen präfigierten und suffigierten Formen verzeichnen, was gegen eine Freilegung des Wurzelmorphems und damit gegen einen Morphem-basierten Ansatz spricht, da dieser vergleichbare Reaktionszeiten für Wurzeln präfigierter und affigierter Lexeme prognostizieren würde.

Ist mit „priming“ nun inhibierendes, d.h. negatives Priming gemeint, so wäre entsprechend positives Priming für die hochfrequenten Primes (*government-govern*¹⁰) bereits im Hinblick auf die genannte Unmittelbarkeit des direkten lexikalischen Zugriffs zu erwarten. Darüber hinaus kann man die Einstufung eines Primes wie *escortment* im Paar *escortment-escort* als maximal niedrigfrequente (d.h. ungebräuchliche weil nicht lexikalisierte) Instanz eines Lexemparadigmas in Frage stellen. Genauso gut könnte man für diese Kondition eine rein formale Beziehung zwischen Prime und Target mit hoher phonologischer Überlappung unterstellen (vgl. Chéreau et al. 2007, Coltheart et al. 2001 und Harm & Seidenberg 2004, die ebenfalls eine positive Korrelation zwischen Ausmaß der Überlappung und phonologischem Effekt annehmen). Im Hinblick auf die niedrige SOA von 42 ms, mit der hier gearbeitet wird, ein Bereich, in dem zumeist reine Formeffekte berichtet werden (u.a. von Davis et al. 1998, Feldman

⁹Das heißt, sehr viel weniger Wörter fangen mit *br-* an als mit *ver-*, was auch „family size effect“ genannt wird (vgl. z.B. Bertram et al. 2000, De Jong et al. 2000).

¹⁰Dieses Beispiel stammt aus McCormick et al. (2009).

2000, Perre et al. 2009, Shen & Forster 1999) ist dieser Konditionseffekt nicht vom Formeffekt separierbar, da keine entsprechende Vergleichsgruppe existiert. Da sich die drei Konditionen (hochfrequentes, niedrigfrequentes und infrequentes Prime) nicht voneinander unterscheiden, könnte ebenso gut darauf geschlossen werden, dass es sich bei den von McCormick et al. (2009) unterstellten Morphemeffekten um reine Formeffekte handelt.

Dass der Falsifikationsversuch der Full-Decomposition-Theorie durch McCormick et al. (2009) nicht streng genug ist (vgl. Gadenne 1976, Popper 2002), lässt sich schlussendlich mit Booij (2005) untermauern:

„If the word is stored, the direct route will be faster than the indirect route if the stored complex word has a high frequency, that is, a high level of activation. [...] If the complex word has a lower frequency than the base word, this will make parsing of the complex word easier and more efficient than direct retrieval.“
(Booij 2005:250)

Auch wenn es angesichts der Linearität der menschlichen Wahrnehmung (d.h. Zeit wird als Sequenz von Ereignissen wahrgenommen) nach wie vor intuitiv plausibler erscheint zu unterstellen, dass ein direkter lexikalischer Zugang schneller erfolgt als ein geparster Zugang, so kann ein vergleichbarer Effekt aller drei Konditionen ebenso gut mit dem Frequenz-basierten Dual-Route-Ansatz (dem Horse-Race-Ansatz) erklärt werden (u.a. Hay 2001).

Ein weiteres Beispiel für das gezielte Verifizieren der Dekompositionstheorie gibt Taft (2004). So wird zu Beginn der einleitenden Zusammenfassung folgende Alternativhypothese formuliert: „If recognition of a polymorphemic word always takes place via its decomposition into stem and affix, then the higher the frequency of its stem (i.e., base frequency) the easier the lexical decision response should be when frequency of the word itself (i.e., surface frequency) is controlled.“ (Taft 2004:745) In der Schlussdiskussion lautet es dann angesichts der Resultate: „[T]he existence of a surface frequency effect with a concomitant lack of a base frequency effect need not imply that whole-word access has taken place without decomposition.“ (Taft 2004:761) Ungeachtet der Ad-hoc-Hypothesen, die konstruiert werden, um an der favorisierten Theorie festhalten zu können, wird unmittelbar deutlich, dass hier Evidenz zugunsten der Nullhypothese schlichtweg nicht akzeptiert wird. Es stellt sich letztendlich die Frage, wozu eine empirische Überprüfung durchgeführt wird, wenn von dem unerwünschten Ergebnis von vornherein angenommen wird, dass es ungültig ist. Taft's Argumentation (2004) stellt mit Sicherheit keinen Einzelfall dar. Um den Finger auf dieses weit verbreitete Vorgehen, bei dem die theoretische Erklärung an die Resultate angepasst wird, zu legen, hat Kerr (1998) den Terminus „HARKing“ eingeführt: „Hypothesizing After the Results are Known“.

Stockall & Marantz (2006), McCormick et al. (2009) und Taft (2004) dienen damit als Negativbeispiele für den Einsatz von Immunisierungsstrategien durch das Vermeiden riskanter Hypothesen, das Konstruieren von Ad-hoc-Annahmen sowie das relativ häufig anzutreffende Übergehen eigener oder fremder Teilresultate, die der neu gewonnenen Datenlage widersprechen (so wird bei Stockall & Marantz (2006) selektiv zitiert¹¹ und das in der tabellarischen Darstellung offensichtlich fehlende Priming für schwach flektierende Paare wie *jump-jumped* übergangen¹²).

Zwar bietet die Veröffentlichung von Stockall & Marantz (2006) Anlass zur Kritik, verweist aber auch auf eine wenig beachtete¹³ Studie von Rastle et al. (2000), die vielleicht als erste morphologisches Priming mit phonästhemischem Priming kombinieren, leider ohne die theoretischen Implikationen eines solchen Designs auszuschöpfen.

„Morphological effects that are greater than the combined effects of semantic and orthographic similarity can be interpreted as evidence for the lexical representation of morphological structure within the lexicon, but this outcome also is consistent with Conjoint Similarity accounts given the greater similarity along each single dimension. Both an account based on activation of a shared morpheme and an account based on the conjoint effects of shared form and shared meaning could accommodate a morphological effect that is greater than the sum of facilitation for the two separate dimensions of similarity. Crucially, one domain in which the Morpheme Activation and the distributed Conjoint Similarity accounts of morphological processing make very different predictions is when prime-target pairs are similar in form and are similar in meaning *but fail to share a base morpheme.*“ (Pastizzo & Feldman 2009:5, eigene¹⁴ Kursivierung)

Die phonästhemische Relation, die von Rastle et al. (2000) im Experimentaldesign der morphemischen Relation zwischen Prime und Targetpaaren als Kontrollgruppe von formal und semantisch ähnlichen Wörtern gegenübergestellt wird, steht nicht im Zentrum des Interesses, gilt es doch den psychologischen Status von Morphemen zu untersuchen. Wie auch den

¹¹Feldmans (2000) Überzeugung, dass morphologisches Priming nicht auf kombinierte Form- und Bedeutungsähnlichkeit zurückgeführt werden kann, wird aufgegriffen, widersprüchliche Resultate aus Stolz & Feldman (1995) dagegen werden ausgeblendet.

¹²Es kann natürlich nicht ausgeschlossen werden, dass es sich in dem Fall um einen Darstellungsfehler handelt. Das heißt, die Durchschnittswerte für die Reaktionszeiten der regulären morphologischen Relation könnten lediglich in der Darstellung, nicht jedoch in der Analyse mit denen der relationslosen Vergleichsgruppe verwechselt worden sein. Ein ähnliches Problem findet sich auch bei Bergen (2004) und wird in Abschnitt 3.2.3 diskutiert.

¹³Diese für sein Design zentrale Veröffentlichung ist von Bergen (2004) offenbar übersehen worden.

¹⁴Dies meint in vorliegender Arbeit stets eine Kursivierung durch die Autorin, die nicht im Originaltext vorhanden ist.

meisten anderen Autoren, die sich experimentell mit dieser Thematik befassen, geht es Rastle et al. (2000) dabei in erster Linie darum zu zeigen, dass sich der Morphemeffekt nicht auf einen kombinierten Form- und Bedeutungseffekt zurückführen lässt (vgl. Napps 1989, Feldman 2000, Rastle et al. 2000, Kouider & Dupoux 2009, für eine gegenteilige Ansicht z.B. Rueckl et al. 1997).

Während Rastle et al. (2000) davon absehen, ihr innovatives Design ins rechte Licht zu rücken, scheinen Stockall & Marantz (2006) bei der Stimuli-Zuordnung zur (S+O-M)-Gruppe¹⁵ – vermutlich vor dem Hintergrund der Ergebnisse von Bergen (2004) (nämlich, dass Phonästhememorpheme psychologisch real sind und nicht auf kombinierte Form- und Bedeutungseffekte zurückführbar) – bewusst von einer Implementierung von Phonästhememorphemen abzusehen, wobei sie aber nicht ganz konsequent sind. Ihre (S+O-M)-Gruppe setzt sich demnach recht heterogen aus sogenannten Pseudophonästhem¹⁶-Lexemen wie *blossom-bloom* und *crinkle-wrinkle*, Kuriositäten wie einem homophonen Minimalpaar, das eine metonymische Beziehung zeigt wie *pat-pet*, sowie aus Phonästhememorphemen wie *gl-* (hier in antonymer Ausprägung im Sinne eines FEHLENS VON LICHT: *gloom-glum*) bzw. Phonästhememorphemkandidaten wie *scr-* (UNANGENHMER KLANG: *screech-scream, scrape-scratch*) und *-[V]sh* (lautmalerisch: *squish, squash*) zusammen, was wiederum gegen eine Auseinandersetzung mit Bergen (2004) spricht.

Da Stockall & Marantz (2006) keine Effekte dieser heterogenen (S+O-M)-Kondition messen können, schließen sie darauf, dass „[t]he failure of the boil–broil items to evoke a priming response in either the neural or behavioural measures provides key support for a model of lexical organization wherein morphological relatedness can not be explained as a combination of semantic relatedness and phonological/orthographic relatedness.“ (Stockall & Marantz 2006:106) Dass ein solcher Schluss nicht ganz gerechtfertigt erscheint, lässt sich schlichtweg mit der Tatsache begründen, dass die semantische und orthographische Relation der „boil–broil items“ nicht auf Morpheme übertragen werden kann, oder anders gesagt: Die Annahme, dass die morphologische Relation nicht auf die semantischen und formalen Aspekte reduziert werden kann, sollte nicht anhand von Lexemen überprüft werden, die überhaupt keine morphologische Relation teilen.

Darüber hinaus sind Morpheme nun einmal durch gemeinsame Form- und Inhaltsmerkmale definiert. Damit stellen letztere die einzigen inhärenten Merkmale der unabhängigen Variablen (morphologische Relation) dar, die überhaupt variiert werden können, wobei sie offen-

¹⁵Damit ist die phonästhemische Kontrollgruppe gemeint. Der Term setzt sich aus einer positiven semantischen, einer positiven orthographischen und einer negativen morphologischen Beziehung zusammen.

¹⁶Der Terminus wird von den Autoren nicht verwendet, jedoch handelt es sich sinngemäß um Pseudophonästhememorpheme, eine Bezeichnung, die auf Bergen (2004) zurückgeht, der damit Phonästhememorphem-ähnliche Elemente bezeichnet, die eine verglichen mit Phonästhememorphemen relativ geringe Type-Frequenz aufweisen.

sichtlich graduelle Effekte erzeugen, die von der Art und dem Ausmaß der phonologischen und semantischen Überlappung zwischen Prime und Target abhängen (vgl. Bybee & Moder 1983, Feldman & Prostko 2002, Feldman et al. 2004, Hahn & Nakisa 2000, Gonnerman et al. 2007, Kielar & Joannisse 2011, Kotz et al. 2002, Rueckl et al. 1997). So zeigen z.B. Kotz et al. (2002) in einer Arbeit, die die funktionelle Magnetresonanztomographie (fMRT) mit Priming kombiniert, dass Wortpaare, die eine kategoriale semantische Relation teilen (z.B. *cow-dog*, aus Kotz et al. 2002) aufwendiger verarbeitet werden als Paare, die eine assoziative semantische Relation (z.B. *cat-dog* aus Kotz et al. 2002) teilen (vgl. auch Moss et al. 1995, Kotz 1998, Shelton & Martin 1992 und Neely 1991 für einen Überblick).

Zentral, und für diese Arbeit von besonderer Bedeutung, ist bei der Studie von Rastle et al. (2000) allerdings der Befund, dass die phonästhemische Relation sich bei kurzer SOA nicht von der Bedeutungsrelation unterscheidet. Leider lassen die Autoren die Quelle ihrer Frequenzwert-Erhebungen unerwähnt. Die relative Verarbeitungshäufigkeit lexikalischer Stimuli stellt nämlich ein nicht zu unterschätzendes Problem für die meisten wortbasierten Primingstudien dar, so dass sich die Validität der Resultate von Rastle et al. (2000) ohne eine unabhängige Kontrolle des Frequenzeffekts streng genommen nicht vollständig beurteilen lässt. Wie an anderer Stelle bereits erwähnt, handelt es sich dabei um einen Störfaktor, der verschiedene Ausprägungen annehmen kann:

- einmal tritt er als Oberflächenfrequenz einfacher oder komplexer Lexeme (z.B. *calculate*, *sized*) in Erscheinung, wobei *sized* eine niedrigere Oberflächenfrequenz aufweist als *size*- (Bsp. aus Chialant & Caramazza 1995)
- einmal als Base-Frequenz einfacher Lexeme; für *calculate* entspräche dies der Summe der Oberflächenfrequenzen aller inflektierten Formen von *calculate* plus der Oberflächenfrequenz des Infinitivs: *calculate* + *calculate* (1.Pers.) + *calculates* + *calculated* + *calculating* (Bsp. aus De Jong et al. 2000)
- einmal als kumulierte Wurzel-Token-Frequenz komplexer Lexeme, die sich aus der Summe aller Base-Token-Frequenzwerte der (De Jong et al. 2000 zufolge) derivierten Formen (d.h. der sogenannten „Familienmitglieder“) von *calculate* (*calculable*, *calculation*, *calculator*, *calculus*, *incalculable*, *incalculably*, *miscalculate*, *miscalculation*) ergibt (vgl. auch Giraudo & Grainger 2000, Taft 1979)
- einmal in der Ausprägung als relative Prime-Target-Tokenfrequenz; d.h. in einem *tanzentänzeln*-Trial hat das Prime eine höhere Tokenfrequenz als das Target (vgl. Segui & Grainger 1990)

- desweiteren als kumulierte Token-Frequenz von Lexemen, von denen angenommen wird, dass sie mit den Stimuli mit aufgerufen werden (vgl. Frauenfelder & Peters 1998), und als deren relative Token-Frequenz (vgl. Grainger et al. 1989). Bei letzterer wird nicht die summierte (Base-)Frequenz der Nachbarn betont, sondern eine mögliche Inhibition durch einen im Vergleich zum Stimulus höher-token-frequenten Nachbarn oder auch die Anzahl solcher höherfrequenten Nachbarn (vgl. Perea & Carreiras 1998, Alvarez et al. 2001) sowie
- als Umfang des lexematischen Paradigmas, sprich als die Type-Frequenz des Wurzelexems eines komplexen Primes respektive Targets (d.h. für *calculate* addieren De Jong et al. 2000 alle derivierten Formen). Man spricht dabei auch vom „family size effect“ (vgl. Bertram et al. 2000, De Jong et al. 2000, Pytkänen et al. 2004, Schreuder & Baayen 1997, Lüdeling & De Jong 2002, Sainz & Lazaro 2009).

Mit aufgerufene Lexeme sind unter dem etwas unscharfen Terminus „lexikalische Nachbarn“ in der Literatur verbreitet. Man versteht darunter zum einen Minimalpaar-Nachbarn mit maximaler Überlappung, z.B. *ford*, *work* und *word* (Bsp. aus Grainger et al. 2005, vgl. auch Coltheart et al. 1977, Luce & Pisoni 1998) und zum anderen „cohort competitors“ mit Betonung auf der Onset-Kohorte (z.B. Marslen-Wilson & Zwitserlood 1989 und Norris 1994): „bat will not be activated when the input is cat; but words like cab, cattle, cavern, catatonic.“ (Magnuson et al. 2007:2) Die Onset-Kohorten werden dabei im perzeptiven Zeitverlauf mit Zunahme der Überlappung reduziert. So sind Onset-Nachbarn im frühen Stadium der Worterkennung von Bedeutung, maximal überlappende Nachbarn dagegen im späten Stadium. Allerdings lässt sich der Moment der Identifizierung nicht genau bestimmen: „Using the ‚captain‘ example, how is one to know if a match has been achieved after the word CAP or if one has to continue reducing the cohort?“ (Field 2004:65)

Die Klassifizierung der relativen Verarbeitungshäufigkeit von Lexemen ist damit allerdings noch nicht abgeschlossen, denn es kursieren noch folgende Differenzierungen:

- Eine dritte Form von lexikalischem Nachbarn wäre die „transposed letter similarity“ (vgl. Andrews 1996, Chambers 1979, Taft & Van Graan 1998). Nachbarn von *calm* wären hier sowohl *call* als auch *clam* (Bsp. aus Davis 2005)
- Desweiteren ist die sublexikalische Bigram/Trigram-Type- und Token-Frequenz von Bedeutung, wobei hier für gewöhnlich nicht zwischen Type- und Tokenfrequenz unterschieden wird (Broadbent & Gregory 1968, Rice & Robinson 1975, Westbury & Buchanan 2002, vgl. negative Befunde in Andrews 1992 und Peereman & Content 1995). Carreiras et al. (1993) sprechen dabei auch vom „syllable frequency effect“

Wie man sieht, wird es praktisch unmöglich sein, alle Ausprägungen dieser Drittvariablen zu kontrollieren. Selbst in einem maximal einfachen, d.h. ungeprinten, Worterkennungsdesign, der sogenannten „single presentation technique“ (vgl. z.B. Perea & Rosa 2002), wäre an eine Homogenisierung der Oberflächenfrequenz kaum zu denken. Angenommen, die Stimuli fallen in mehrere Konditionen, so geschieht die Zuordnung der Stimuli zur Kondition für gewöhnlich nach bestimmten linguistischen Gesichtspunkten und nicht rein zufällig, so dass eine Kontrolle durch Randomisierung praktisch ausgeschlossen werden kann. Eine Homogenisierung erscheint deswegen unrealistisch, da die üblicherweise auf Frequenz kontrollierten Konditionen¹⁷ immer noch Randwerte enthalten. Würde man solchen Randwerten stets mit Beschneidung begegnen, so würde es schlussendlich überhaupt kein Versuchsmaterial mehr geben.

Es wird nun deutlich, dass in einem Primingdesign, das Primes und Targets systematisch variiert, an eine durchgreifende Kontrolle der lexematischen Type- respektive Token-Frequenz aufgrund der Komplexität dieser Variablen, aber auch angesichts der Größe und Zusammensetzung des jeweiligen Referenzkorpus, nicht zu denken ist. Vor diesem Hintergrund wird der Frequenzaspekt bei der Besprechung der direkten Vorgängerarbeiten von Bergen (2004) als auch der darauf basierenden eigenen Forschungsarbeiten von besonderer Bedeutung sein. Die Strategie besteht dabei darin, zu überprüfen, inwieweit die Störvariable Verhaltensdaten verzerren kann, und diese gegebenenfalls im zweiten Schritt durch ein spezielles Primingdesign bzw. nachträglich in der statistischen Analyse zu eliminieren, um so Aussagen über die in erster Linie interessierenden Zusammenhänge zu gewinnen.

3.1.3 Rekonstruktion der Annahmen und Hypothesen in Feldman (2000)

Methodische und methodologische Probleme finden sich auch bei Feldman (2000), deren Versuchsdesign als Vorlage für Bergen (2004) gedient hat. Wie weiter unten veranschaulicht werden wird, bestehen die Hauptmängel in Feldman (2000) im Gegensatz zu den im vorhergehenden Abschnitt diskutierten Arbeiten in der fehlenden Bezugnahme auf linguistische Theorien¹⁸, desweiteren in der fehlenden Berücksichtigung disziplinübergreifender allgemeiner Prinzipien wie des Analogie-Prinzips (vgl. Goldberger 1991 und Manski 1988 für die Statistik) und des Ökonomie-Prinzips (vgl. Raj & Chen 2011, Kirchgässner 2008 für die Soziologie, Trost 1995 für die Chemie) bei der Interpretation der Ergebnisse, als auch in der aus

¹⁷Konkret wird darauf geachtet, dass in jeder Kondition ähnlich häufige Lexeme vorhanden sind, d.h. solche, die in einem Referenzkorpus ähnliche Mittelwerte erzielen.

¹⁸Es wird nicht klar, von welchen Hypothesen ausgegangen wird und aus welchen Theorien diese unterschwelligen Hypothesen abgeleitet worden sind.

Sicht des kritischen Rationalismus (Popper 2002) zirkulären, jedoch weit verbreiteten methodologischen Spielart der Verifikation¹⁹.

Ebenso wie Napps (1989), Rastle et al. (2000), Kouider & Dupoux (2009) und später Bergen (2004), geht es Feldman (2000) darum zu zeigen, dass Morpheme die Worterkennung erleichtern, ohne dass dies auf die Form- oder Bedeutungsübereinstimmung oder die kombinierte Form- und Bedeutungsübereinstimmung zwischen Lexemen, die ein und dasselbe Morphem teilen, reduzierbar wäre. Bedenkt man die Verbreitung dieser Schlussfolgerung unter den einschlägigen Studien (vgl. auch Damian & Martin 1999, Frost et al. 2000, Garcia-Albea et al. 1998, Marslen-Wilson & Tyler 2007, für Ausnahmen vgl. Rueckl et al. 1997, und die ERP²⁰-Studie von Martin & Thierry 2008) erscheint es so, als befände sie sich auf dem Weg hin zu einem Dogma. Dies mag teils an der aus wissenschaftshistorischer Sicht berechtigten skeptischen Einstellung gegenüber der methodologischen Maxime²¹ von Occams Rasiermesser liegen (vgl. Gernert 2007), teils an einer Verwechslung von Einfachheit und Trivialität. Andererseits argumentiert Popper (2002), dass einfache Erklärungen leichter falsifizierbar sind. In erster Linie scheint die Abneigung gegen die Kombinationsannahme jedoch mit einem lückenhaften statistischen Verständnis zusammenzuhängen.

Feldman (2000) begründen letztere hauptsächlich mit Ergebnissen aus ihrem Telexperiment 2, einer lexikalischen Entscheidungsaufgabe, bei der zwischen Prime und Target Distraktoren implementiert sind („long-term lexical decision task“). Da in Experiment 1 die Verarbeitungserleichterung durch die Morphemgruppe im Vergleich zum relativ robusten Bedeutungseffekt mit Ansteigen der SOA zunimmt, wurde Experiment 2 konstruiert, nicht zuletzt um Evidenz für die Alternativhypothese zu mehren²². In Experiment 2 scheint es nun so, als könne der Morphemeffekt weder auf Form- noch auf Bedeutungseffekte zurückzuführen sein (da beide sich nicht signifikant von der relationslosen Kontrollgruppe unterscheiden). Merkwürdig dagegen ist, dass die kombinierte Verarbeitungserleichterung durch Form- und Be-

¹⁹Da jegliches Prinzip die Erfahrung/Beobachtung übersteigt, kann es nicht mit Beobachtungen verifiziert werden; es kann nur mit Beobachtungen falsifiziert werden: „no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white.“ (Popper 2002:4) Es reicht nicht, den Satz *Alle Schwäne sind weiß*. mit Beobachtungen zu rechtfertigen, da induktiv nicht alle Schwäne beobachtet werden können, und selbst wenn, dann führt die Suche nach weißen Schwänen dazu, dass schwarze Schwäne ausgeblendet werden. Beweisen lässt sich diese Hypothese nur, indem nach einer Beobachtung für seine Widerlegung gesucht wird, d.h. nach einem Schwan, der nicht weiß ist. Lässt sich keiner finden, so kann der Satz vorläufig als wahr gelten.

²⁰ERP steht für Event-Related Potential und meint eine elektrophysiologische Reaktion des Gehirns auf einen Stimulus.

²¹Um ein Prinzip handelt es sich dabei wohl eher nicht, da es davon abhängt, wie Einfachheit definiert wird.

²²Im Übrigen wird H1 nicht umsonst Alternativhypothese genannt, verweist sie doch auf die methodologische Priorität der Nullhypothese (d.h. *zuerst* muss H0 widerlegt sein, erst dann kann die Alternativhypothese beibehalten werden), vgl. Neyman & Pearson (1933).

deutungsgruppe sich nicht signifikant von der morphematischen Verarbeitungserleichterung unterscheidet. Trotz nicht-signifikanter Haupteffekte ergibt sich der Morphemeffekt demnach additiv aus der Summe der numerischen Haupteffekte.

Im Versuchsdesign 1, das die SOA (*stimulus onset asynchrony*)²³ ohne den Einsatz von Distraktoren variiert (*short-term lexical decision task*), lässt sich der Morphemeffekt bei den SOAs von 66 ms, 116 ms und 300 ms ebenfalls nicht von der Summe der Form- und Bedeutungseffekte unterscheiden. Genau gesagt beschränkt sich ein klarer Morphemeffekt auf 300 ms inklusive 50 ms Leere, wogegen sich bei 300 ms inklusive 50 ms Maske kein unterscheidbarer Effekt zeigt. Weshalb die SOA hier zusätzlich zu der Maskierung variiert wurde, die übrigens keinen signifikanten Effekt auf die Verarbeitung der Formkondition hat, wird nicht erläutert.

Feldman (2000) betrachtet die Situation jedoch im Sinne eines halb vollen Glases und hebt die für die Fehlerrate signifikante Überlegenheit der Morphemgruppe gegenüber der kombinierten Form- und Bedeutungsgruppe (in Experiment 2) hervor. Bei genauerem Hinsehen heißt dies aber nur, dass bei der Form- und Bedeutungsrelation zusammen genommen signifikant mehr Fehler gemacht wurden als bei der Morphemrelation. Erst wenn diese Fehler dadurch entstanden sind, dass die Reaktionszeit von Probanden 1500 ms überschritten hat, könnte man die Fehlerrate für die Reaktionszeiten interpretieren (die Differenz zwischen morphematischem und formal-inhaltlichem Effekt wäre dann eigentlich noch größer). Ist das nicht der Fall spricht das sogar für den Kombinationseffekt, mit dem sich die morphematische Verarbeitungserleichterung trotz der höheren Fehlerrate erklären lässt.

Eine Interpretation dieser Ergebnisse wird durch den Umstand erschwert, dass Feldman (2000) davon absieht, ihre Annahmen und Hypothesen explizit zu formulieren. Deswegen müssen diese zusammen mit den zugrundeliegenden Theorien über die mentale Repräsentation von Sprache ex post aus der vergleichsweise präzisen Dokumentation ihres methodischen Vorgehens rekonstruiert werden:

Annahme 1: Die kognitive Realität einer Kategorie ist operationalisierbar.

Annahme 2: Da bewusstes Wissen konstruiert ist, erlaubt nur eine Erforschung unbewussten Wissens Rückschlüsse auf die kognitive Realität von Kategorien.

Im Grunde genommen begründen die Annahmen 1 und 2 den Einsatz der Priming-Methode zur Messung von Kategorialität. Da Morpheme im prozeduralen Gedächtnis verortet werden,

²³Damit ist der zeitliche Abstand zwischen Einsetzen des Primes und Einsetzen des Targets gemeint, d.h. ISI + Primedauer = SOA.

wäre eine introspektive Argumentation auf Grundlage deklarativer Gedächtnisinhalte zirkulär. D.h. auch wenn eine linguistische Laiin nach Aufforderung ein komplexes Lexem wie *Freundschaft* in seine funktionalen Bestandteile zerlegen könnte, weil sie die Bedeutung des Wortes *Freund* erklären kann und das Suffix aufgrund von Analogie zu *Mannschaft* und *Gesellschaft* im Prinzip als solches erkennt, auch wenn ihr der Fachausdruck nicht geläufig ist, so heißt das noch nicht, dass ihre mentale Representation von *Freundschaft* in Wort-und-Regel-Form vorliegt. Mit deklarativem Wissen ist hier also konstruiertes Wissen gemeint.

Aus dem Versuchsaufbau lässt sich folgende Arbeitshypothese rekonstruieren:

H1: Wenn der Morphemeffekt stärker ist als der Form- und der Bedeutungseffekt (zusammen), dann ist das Morphem als Kategorie psychologisch real.

H0: Wenn der Morphemeffekt nicht stärker ist als der Form- und der Bedeutungseffekt (zusammen), dann ist das Morphem als Kategorie nicht psychologisch real.

Hier werden die Auswirkungen der unabhängigen Variablen Bedeutung, Form und Morphem auf die abhängige Variable Reaktionszeit gemessen. Der kategoriale Status des Morphems wird also relativ zu den Haupteffekten von Form und Bedeutung betrachtet. Da Feldman (2000) auf eine Spezifikation der Bedingung für Kategorialität verzichtet, ist es ungewiss, ob das Morphem bereits kategorialen Status erhält, wenn sich der Morphemeffekt sowohl vom Form, als auch vom Bedeutungseffekt unterscheidet, oder aber erst dann, wenn der Morphemeffekt nicht durch den *kombinierten* Form- und Bedeutungseffekt erklärt werden kann²⁴.

Obgleich Experimentaldesign 2 einen vom Bedeutungs- und Formeffekt scheinbar unabhängigen Morphemeffekt zeigt, verbieten folgende Teilresultate aus den Experimenten 1 und 2 eine Verwerfung von H0:

- In Experiment 2 lässt sich zumindest augenscheinlich²⁵ der Effekt der Bedeutung vom Effekt des Morphems separieren; aus dem Fehlen eines Haupteffekts lässt sich jedoch nicht auf das Fehlen eines Interaktionseffekts zwischen Haupteffekten schließen (Campbell & Stanley 1973)
- In Experiment 1 ist dagegen kein signifikanter Unterschied zwischen Morphemeffekt und Bedeutungseffekt nachweisbar (d.h. H0 wird beibehalten)
- In Experiment 1 ist für die SOAs von 32 ms und 116 ms inkl. Leere und SOA 300 ms inkl. Maske der Morphemeffekt nicht vom kombinierten Form- und Bedeutungseffekt

²⁴Dasselbe Problem tritt auch in Bergen (2004) auf.

²⁵Auf die Resultate von Experiment 2 wird weiter unten noch näher eingegangen.

differenzierbar; in Experiment 2 ist der Morphemeffekt bei fehlenden Haupteffekten für Form und Bedeutung ebensowenig von der Summe der numerischen Haupteffekte differenzierbar (d.h. H0 wird beibehalten)

Punkt 1 wird von den Autoren als Indiz dafür gewertet, dass ein (zeitabhängiger²⁶) Morphemeffekt ohne Beteiligung der Bedeutungs- und Formkomponente vorliegt, jedoch sind die Ergebnisse alles in allem zu inkonsistent.

Von fehlenden Haupteffekten auf fehlende Interaktion zu schließen, ist ein Fehler, der in einer späteren Arbeit der Autorin (s. Pastizzo & Feldman 2009) nicht mehr begangen wird: „[N]ot only can shared form and meaning facilitate target recognition in a way that cannot be predicted by simply adding the two dimensions of relatedness, but also [...] the magnitude of facilitation between prime and target seems to be related to the number of words within a rime family that share meaning as well as form.“ Es wird nun ein Kombinationseffekt eingeräumt, der von einem Additionseffekt abgegrenzt wird und abgegrenzt werden muss, zumal sich Interaktionseffekte und Additionseffekte per Definition ausschließen.

Doch gerade wenn kein Additionseffekt gemessen werden kann, ist die Wahrscheinlichkeit groß, dass kombinierte Haupteffekte den Morphemeffekt hervorbringen, indem sie miteinander interagieren. Das Prinzip der Interaktion lässt sich mit folgenden Analogien veranschaulichen:

Analogie A: Gemessen wird (bei konstant gehaltener Sendezeit)

- a) der Effekt des Moderators Thomas Gottschalk (genauer gesagt, einer von Thomas Gottschalk moderierten Sendung, z.B. *Gottschalk Live*) auf die Einschaltquoten.
- b) der Effekt der Sendeformats *Wetten dass...?* ohne Moderation durch Thomas Gottschalk (stattdessen: z.B. Wolfgang Lippert) auf die Einschaltquoten.
- c) der Effekt von *Wetten dass...?* moderiert von Thomas Gottschalk auf die Einschaltquoten.
- d) der Effekt eines alten Kriminalfilms auf die Einschaltquoten.

Nehmen wir also an, die Quoten von a) und b) unterscheiden sich nicht signifikant von den durch d) erzielten Quoten, d.h. es werden keine Haupteffekte von a) und b) gemessen. Dafür lässt sich für c) ein signifikanter Effekt auf die Einschaltquoten nachweisen. Der Effekt c) beruht auf einer wechselseitigen Abhängigkeit von a) und b). a) und b) interagieren, obgleich weder der Effekt von a) noch von b) für sich genommen signifikant vom Effekt der Kontroll-

²⁶Hier wurde die Zeit zwischen Primestart und Targetstart von SOA 32 ms auf 300 ms (Experiment 1) und weiter durch Implementierung von durchschnittlich 10 Distraktoren (Experiment 2) verlängert.

bedingung d) (Kriminalfilm) abweichen.

Analogie B: Gemessen wird (bei einem Zeitfenster von 30 Sekunden)

- a) der Effekt der Zugabe von Würfelzucker²⁷ in ein Heißgetränk, z.B. Kaffee, auf dessen Süße.
- b) der Effekt des Rührvorgangs in einem Heißgetränk auf dessen Süße.
- c) der Effekt der Zugabe von Würfelzucker in ein Heißgetränk und anschließendem Rühren auf dessen Süße.

Stellt sich unter Bedingung c) ein messbarer Effekt ein (ausgewogene Süße), bei a) und b) dagegen nicht (zu wenig oder gar kein süßer Geschmack), so wäre es unsinnig zu behaupten, die Süße des Heißgetränks könne nicht mit der Gabe von Zucker und anschließendem Rühren (zur Beschleunigung der Auflösung und Verteilung in der Flüssigkeit) erklärt werden.

Übertragen auf Feldman (2000) heißt dies, dass der Effekt des Morphemprimings in ihrem Experiment 2 durchaus auf eine Interaktion von Form und Bedeutung zurückführbar ist und nicht als Gegenbeweis für einen obskuren morphematischen Zusatzeffekt gewertet werden kann. Es sollte an diesem Punkt nichtsdestotrotz angemerkt werden, dass Interaktionen aus dem Alltag sehr viel plausibler erscheinen, weil es sich um relativ zum Morphempriming transparente und robuste (d.h. bewährte) Effekte handelt. Dass gezuckerter Kaffee süß schmeckt und Einschaltquoten die relative Beliebtheit einer Fernsehsendung widerspiegeln ist so unmittelbar einleuchtend, dass es trivial erscheint. Die Mechanismen der Interaktion formaler und semantischer kognitiver Repräsentationen sind dagegen wenig offensichtlich, und nur so lange sie das sind, mag es unangemessen erscheinen hierfür ähnlich „triviale“ Interaktionen anzunehmen wie in den beiden Analogien.

Zusammengefasst kann man sagen, dass Feldman (2000) die folgenden methodischen und methodologischen Probleme zugrunde liegen:

1. Weder wird auf eine Theorie Bezug genommen noch eine Theorie beigesteuert, aus der Hypothesen abgeleitet werden könnten. Daraus folgt, dass Variationen der Versuchsbedingungen (Maskierung, SOA, Distraktoren) entweder gar nicht oder lediglich mit Erfahrungswerten aus vergangenen Experimenten begründet werden.
2. Es werden keine klaren Arbeitshypothesen formuliert. Deswegen müssen sie aus dem Vorgehen rekonstruiert werden.

²⁷Diese Analogie ist nur dann zweckdienlich, wenn von der Würfelform des Zuckers ausgegangen wird.

3. Dort, wo es zur Arbeitshypothese über einen Effekt der morphematischen Konstruktion passt, werden für die Gesamtauswertung Werte interpretiert, die sich unterhalb des Signifikanzniveaus bewegen (d.h. die lediglich als Tendenzen²⁸ und numerische Unterschiede²⁹ vorliegen). Dort, wo es nicht zur Arbeitshypothese passt, unterbleibt eine umfassende Diskussion fehlender Signifikanzen, d.h. fehlende Haupteffekte von Form und Bedeutung in Experiment 2 werden als Indiz für die Arbeitshypothese gewertet anstelle auf einen möglichen Interaktionseffekt zu verweisen, der – wie weiter oben ausgeführt – auch bei fehlenden Haupteffekten gegeben sein kann.

Aus Punkt 1 ergeben sich nachstehende Unklarheiten:

- Es leuchtet nicht nur angesichts möglicher Interaktionseffekte, sondern auch aus theoretischer Sicht³⁰ nicht ein, weshalb der kategoriale Status des Morphems nicht durch einen Kombinationseffekt der Form- und Bedeutungsseite erklärt werden kann, sondern sich unbedingt von diesen Effekten (bzw. der Summe dieser Effekte) unterscheiden sollte (wodurch die scheinbare Paradoxität³¹ der Ergebnisse des Experiments 2 erst entsteht). Durch die Ablehnung eines Kombinationseffekts kann zumindest eine Vermutung über die (in dem Fall anti-konnektionistische) theoretische Grundeinstellung der Autorin angestellt werden.
- Desweiteren wird nicht deutlich, ob die Differenzen zwischen Morphemeffekt vs Formeffekt und Morphemeffekt vs Bedeutungseffekt bereits eine hinreichende Bedingung für morphematische Kategorialität darstellen oder nicht. Wenn ja, so bedarf es für die Gültigkeit von H1 keines Kombinationseffekts³². Wenn nicht, dann gälte es zu begründen, weshalb die Bedingung einer signifikanten Differenz zwischen Morphemeffekt-Formeffekt und Morphemeffekt-Bedeutungseffekt als Indiz für morphematische Kategorialität nur notwendig, aber nicht hinreichend ist, die Bedingung einer signifikanten Differenz des Morphemeffekts zum *addierten* Form- und Bedeutungseffekt dagegen notwendig *und* hinreichend ist.

²⁸„The magnitude of morphological facilitation tended to be greater than that of semantic facilitation.“ (Feldman 2000:1441)

²⁹„Facilitation resulting from morphological relatedness was numerically and sometimes even statistically greater than that of semantic relatedness.“ (Feldman 2000:1441)

³⁰Der hier implizit unterstellte Widerspruch zwischen einem morphembasierten mentalen Lexikon und kombinierten Effekten von Form und Bedeutung im Morphempriming lässt sich in Frage stellen.

³¹Paradox in dem Sinne, dass keine Haupteffekte von Form und Bedeutung messbar sind, der Morphemeffekt sich jedoch durch eine Aufsummierung der nicht vorhandenen Effekte begründen lässt.

³²In dem Fall würde das geklammerte „zusammen“ im Hypothesensatz entfallen.

- Schlussendlich wird nicht klar, welche Eigenschaft von Morphemen den postulierten Verarbeitungsvorteil bewirkt. Deswegen könnte man ebenso gut annehmen, dass er nicht einmal auf die Interaktion von Form und Bedeutung, sondern gänzlich auf die Bedeutungsübereinstimmung und/oder die Frequenzübereinstimmung zwischen identischen Wurzelmorphemen (z.B. *vowed-vow* vs *pledge-vow*) oder etwa auf das Fehlen eines Inhibitionseffekts durch die phonologische Abweichung des Targets zum (inhaltlich synonymen) Prime-Lexem zurückzuführen ist³³. „[M]orphological effects could be distinguished unequivocally from the effects of orthographic similarity and more tentatively from those of semantic similarity.“ (Feldman 2000:1439) stellt also ein ausgesprochen optimistisches Fazit dar, wenn man bedenkt, dass sich der Morphemeffekt in Experiment 1 nicht signifikant vom Bedeutungseffekt separieren lässt.
- Da die „psychologische Realität“ einer Kategorie nicht klar definiert wird, muss im Hinblick auf die Methode (Priming) mit Annahme 2 auf eine Differenzierung dieses Begriffs in 2 Dimensionen geschlossen werden: einmal die unbewusste psychologische Realität (R^1) und einmal die bewusste, d.h. konstruierte, psychologische Realität (R^2).

Feldman nimmt also an, dass der konstruierten Kategorie des Morphems im Unbewussten eine mentale Repräsentation entspricht, die mit dem konstruierten Begriff kongruent ist. Das Versuchsdesign stellt ja nichts anderes dar als eine Strategie, mit der die Existenz des Konstrukts nachgewiesen werden soll; immerhin wird erwartet, dass der Morphemeffekt die Summe seiner Teile übersteigt.

„To reiterate, in order to preserve an argument for the special status of morphological representations when facilitation for BOAT–FLOAT and VOWED–VOW pairs are both greater than one would predict based on effects of form and effects of meaning, one could posit that different mechanisms underlie each type of facilitation.“ (Pastizzo & Feldman 2009:8, eigene Kursivierung)

Nicht nur ist eine solche Argumentation *ad hoc*. Vielmehr lässt sich ein rein kategorial begründeter Morphemeffekt angesichts eines Teilergebnisses aus Rastle et al. (2000) gänzlich in Frage stellen: Entgegen der Annahme, dass eine Kondition, bei der die Wortpaare zwar eine morphologische, aber keine Bedeutungsrelation teilen (in Rastle et al. 2000 z.B. *rational*, *organ-organise*, *casual-casualty*, *depart-department*) einen vergleichbaren Einfluss

³³Unterstellt man für *pledge* und *vow* einen im wesentlichen identischen semantischen Gehalt, erscheint ein Trial wie *vowed-vow* im Hinblick auf Phonologie und Frequenz homogener als *pledge-vow*, so dass ein sogenannter „task effect“ (vgl. u.a. Kuiken & Vedder 2007, Dhar 1997) nicht ausgeschlossen werden kann; liegt ein solcher Effekt vor, dann wird die abhängige Variable nicht etwa von der unabhängigen Variable beeinflusst, sondern von Drittvariablen der Aufgabenstellung.

auf die Verhaltensdaten ausübt wie Wortpaare, die ein transparentes Ableitungsverhältnis aufweisen (z.B. *hunt-hunter*, *swim-swimmer*), unterscheidet sich die opake morphologische Relation nicht von der Formrelation.

Dass jede Kategorie im Zuge ihrer Konstruktion kognitive Realität R^2 erlangt, stellt noch keinen Beweis dafür dar, dass eine Kategorie (wie das Morphem oder Phonästhem) auch unabhängig von dessen Begriffsfassung real im Sinne von R^1 ist. Dementsprechend unklar ist es, welche neuronalen Strukturen einer solchen Realität R^1 entsprechen sollen.

In vorliegender Arbeit wird davon ausgegangen, dass alle Begriffsdefinitionen (bewusste) Konstrukte sind, so dass der ontologische Status einer Kategorie zunächst nicht über die bewusste Konstruktion des Begriffs Kategorie hinausgehen kann. Desweiteren wird angenommen, dass mit Verfahren wie dem Priming nur solche unbewussten Prozesse operationalisiert werden können, die der Begriffskonstruktion (und damit der Konstruktion von kategorialen Konzepten) insofern vorausgehen, als dass sie ihre Konstruktion überhaupt erst ermöglichen.

Annahme 3: Nur unter der Bedingung, dass Morpheme psychologisch real (unterbewusst gegeben) sind, sind sie bewusst konstruierbar ($R^1 \rightarrow R^2$).

Annahme 3': Nur unter der Bedingung, dass Morpheme bewusst konstruierbar sind, sind sie psychologisch real (R^2).

Da Pastizzo & Feldman (2009) sich nicht mit einer Erklärung auf der Grundlage eines Kombinationseffekt zufrieden geben wollen, muss angenommen werden, dass sie von Annahme 3 ausgehen. Damit erhält das Morphem a priori Gültigkeit. Annahme 3' ist vorsichtiger, da sie auf die Realitätsebene R^2 beschränkt ist, und sagt zunächst nichts über die aprioristischen Denkprinzipien aus, die Kategorisierung ermöglichen³⁴. Vor dem Hintergrund der erkenntnistheoretischen Überlegungen Kants (1990)³⁵ ist das Morphem, nur insofern es als begriffliches

³⁴Statt „konstruierbar“, könnte man hier auch „vorstellbar“ sagen.

³⁵An dieser Stelle könnte man einwenden, dass eine Argumentation mit Kant (1990) im Hinblick auf die dualistische Ausrichtung seiner Philosophie überholt ist. Damit wird unterstellt, dass Wissenschaft im Laufe der Zeit fortschreitet, und dass alte Ideen „veraltet“ sind. Rückschritten oder Stagnation in der Wissenschaft wird von Wissenschaftlern natürlich nicht gern nachgegangen, so dass sich kaum einschlägige Diskussionen finden lassen. Beschäftigt man sich etwas eingehender mit Alltagsproblemen wie z.B. zahnärztlichen Behandlungen, so können als Gegenargument (gegen das Axiom der fortschreitenden, d.h. sich im positiven Sinne mehrenden Wissenschaft) neben dem oft genannten Übergang von der Antike zum Mittelalter aktuelle Entwicklungen in der Zahnmedizin vorgebracht werden. So kommt der Trockenlegung zu behandelnder Zähne mittels des bereits im 19. Jahrhundert erfundenen Kofferdams heute eine Randstellung zu, obgleich sich die Methode nicht unwesentlich auf die Erfolgchancen endodontischer Eingriffe (u.a. durch die Isolierung des Zahns vom Mundspeichel und eine damit verbundene Minimierung der Keimzahl) auswirkt (Hellwig et al. 2009). Auch Goldhämmerfüllungen, die wie Schädelfunde belegen, bereits im 17. Jahrhundert zu den zahnärztlichen Leistungen gehört haben und die allen anderen Restaurationsmaßnahmen überlegen sind (s. Kamann 2001, Kamann & Gängler, 1999, Splieth et al. 2000), werden heute kaum noch durchgeführt. Doch zurück

Denkwerkzeug konstruierbar ist, real, d.h. es besitzt keine Existenz $E(R^1)$, die über seinen Status als Konstrukt $E(R^2)$ hinausgeht. Ballmer (1976) würde sagen, dass es sich hierbei um einen rein formalen Realitätsstatus handelt.

Da Annahme 3 mit der Unterscheidung von bewussten und unbewussten Denkprozessen³⁶ Bezug nimmt auf Erkenntnisse der Gedächtnisforschung (genauer gesagt der Unterscheidung zwischen deklarativem und prozeduralem Gedächtnis) könnte man annehmen, es handle sich dabei um die ‚fortschrittlichere‘ Annahme. Jedoch setzt letztere im Gegensatz zur ersteren einen strikten Dualismus voraus: Während 3' einer empirischen Erklärung über Gesetzmäßigkeiten der Gehirntätigkeit offen bleibt, setzt 3 dogmatisch eine übergeordnete Rolle der mentalen Repräsentation voraus. Es wird ja mit 3 nicht zwischen bewusster und unbewusster Konstruktion von Kategorien in dem Sinne unterschieden, dass Kategorien bewusst konstruiert werden können, weil dies von unbewussten Prinzipien bzw. Denkprozessen gesteuert wird³⁷, sondern weil sie dem menschlichen Denkapparat in Form der Universalgrammatik bzw. der Konstruktionen gegeben sind: „[C]onstructions themselves carry meaning.“ (Goldberg 1995:1) Es wird demnach der Kategorie eine Realität zugesprochen, die es überhaupt erst möglich macht, sich eine bewusste Vorstellung von ihr zu machen. Eine Paraphrase von 3 würd sich folgendermaßen lesen: *Morpheme sind mehr als Begriffe*. D.h. sie sind über ihre Begrifflichkeit hinaus als Konstruktionen real³⁸. Obgleich sich die drei Hauptströmungen der Konstruktionsgrammatik, d.h. die generativ geprägte Berkeley-Schule (Kay & Fillmore 1999), der Lakoffsche Ansatz (Lakoff & Johnson 1999) sowie die radikale Schule Crofts (2001), darin einig sind, dass Konstruktionen erlernt werden müssen, und nicht etwa angeboren sind (Fischer & Stefanowitsch 2006), so wird ihre Konstruktion durch eine (in Kants Terminologie) rein aprioristische kognitive Prädisposition ermöglicht, die allen Menschen von Natur aus gegeben sein muss. Zumindest die Konstruktionsgrammatiker gestehen sich also ein, dass es ohne gemeinsame physiologische Voraussetzungen keine kognitiv funktionalen und damit in

zu Kant: nach dessen Auffassung wären Morpheme zunächst nicht mehr als Begriffe, was für die Universalgrammatiker oder auch Konstruktionsgrammatiker wie z.B. Goldberg (1995) höchst unbefriedigend wäre.

³⁶Annahme 3' bezieht sich nur auf bewusste Prozesse.

³⁷Deswegen visualisieren die Konnektionisten Morpheme als emergierende Schemata.

³⁸An dieser Stelle wird auf das Vokabular der Konstruktionsgrammatik Bezug genommen: „[M]orphemes are clear instances of constructions in that they are pairings of meaning and form that are not predictable from anything else [...].“ (Goldberg 1995:4) Dabei „[schließt] [d]er Begriff der Konstruktion [...] also alle konventionalisierten linguistischen Ausdrücke ein, die die folgenden Bedingungen erfüllen: (i) ihre Form ist direkt mit einer bestimmten Bedeutung oder Funktion gepaart, (ii) ihre Form lässt sich nicht (bzw. nicht völlig) aus anderen Formen der Sprache ableiten, und (iii) ihre Semantik ist nicht (bzw. nicht völlig) kompositionell“. (Fischer & Stefanowitsch 2006:6). Wenn alle drei Bedingungen gelten müssen, wären Phonästheme keine Konstruktionen, denn „ihre Form ist direkt mit einer bestimmten Bedeutung oder Funktion gepaart“. Im Hinblick auf die beiden anderen Kriterien sind sie Konstruktionen. Es wird also deutlich, dass die Konstruktionsgrammatik nicht innovativ genug ist, um Phonästheme sinnvoll einzubetten.

R^1 realen Konstruktionen geben kann. Deswegen gilt es Hypothesen über eben diese physiologischen Voraussetzungen des kognitiven Prozesses der Kategorisierung zu formulieren.

Fest steht, dass kognitive Theorien wie die Konstruktionsgrammatik³⁹, wenn sie davon absehen Annahmen über die physiologischen Voraussetzungen für ihre Postulate zu formulieren, keinen explanativen Gehalt haben.

Diese Feststellung ist einer naturwissenschaftlichen Auffassung des Begriffs *explanativ* geschuldet, dem zufolge Theorien einen Sachverhalt nur dann erklären können, wenn aus ihnen *empirisch überprüfbare* Annahmen über Kausalzusammenhänge abgeleitet werden können. Ohne Hypothesen über die physiologischen Bedingungen dieser Kausalzusammenhänge bedarf es auch keines psycholinguistischen Experiments, das den Effekt einer unabhängigen sprachlichen Variable auf eine abhängige metrische Variable misst.

Im konnektionistischen Netzwerk-Modell des mentalen Lexikons (Bybee 2001) sind solche Bedingungen enthalten. Physiologisch ist dieser Ansatz von der Erkenntnis geprägt, dass Gehirntätigkeit auf die Interaktion von vernetzten Neuronen zurückzuführen ist, die mittels der Elektroenzephalographie oder auch des funktionalen Magnetresonanzverfahrens gemessen werden kann. Die Fähigkeit des Nervensystems zur Gedächtnisbildung (d.h. zur Speicherung, Ordnung und zum Abruf von Informationen) wird dabei im Hinblick auf die Dauer der Speicherung in Kurz- und Langzeitgedächtnis (Atkinson & Shiffrin 1968) sowie nach dem Inhalt in deklaratives und prozedurales Langzeitgedächtnis unterteilt (Anderson 1976, Boyland 1996). Die subkortikale Verortung des letzteren geht mit unbewusstem (Verhaltens-)Wissen einher, ersteres mit bewusstem Wissen (semantisch und episodisch). Dass eine Abgrenzung dieser Wissensareale problematisch ist, zeigt die Tatsache, dass sich semantische Analysen zwar bewusst durchführen lassen (also auf Ebene von R^2), wir uns der kognitiven Repräsentation dieser symbolischen Verbindungen (auf Ebene von R^1) zwischen Form und Zeichen (Langacker 1987) jedoch nicht bewusst sind. So bemerken Evans & Greene (2006), dass innerhalb der kognitiven Semantik Johnsons (1987) „image schemas“ im Hinblick auf Ihre Abstraktheit und Unschärfe nicht etwa mentalen Bilder darstellen, die bewusst und erschöpfend beschrieben werden können (auch wenn wir sie beschreibend zu rekonstruieren versuchen), sondern emergente Basiskonzepte. Sie sind also nicht deckungsgleich mit ihrer symbolischen Notation. Auf der Basis des Modells vom Neuronengeflecht, aus dem Generalisierungen ordnend emergieren, folgt bei Bybee (2001) die Annahme, dass lexikalische Einheiten im proze-

³⁹Auch wenn einige Vertreter der Konstruktionsgrammatik, wie z.B. Kay & Fillmore (1999), eine angeborene Universalgrammatik voraussetzen und andere emergentistisch argumentieren (Tomasello 2003, Langacker 1987, 1991) so geht es beiden Strömungen letztendlich um die Art und Weise wie Sprache kognitiv repräsentiert und organisiert ist.

duralen Gedächtnis⁴⁰ miteinander verknüpft sind, wobei die Stärke der Verknüpfung zwischen Lexemen auf 1) inhaltlicher und 2) formaler Ähnlichkeit sowie 3) deren Verarbeitungshäufigkeit basiert:

„The semantic connections are the strongest and the most important in determining the closeness of the relations among words. [...] The strongest sort of relation between forms is the morphological relation which consists of parallel semantic and phonological connections. Morphological relations can be closer or more distant, depending upon three factors: 1) the degree of semantic relatedness, which is determined by the *number* of shared features, and the *nature* of shared features. 2) The extent of phonological similarity between the items (e.g. *sing* and *sang* are more closely related than *bring* and *brought*). 3) Word frequency: high-frequency words form more distant connections than low-frequency words. In the case of morphologically complex words, this means that high-frequency words undergo less analysis, and are less dependent on their related base words than low-frequency words.“ (Bybee 1985:118).

Es handelt sich hiermit also um eine paradigmatische⁴¹ wort-basierte konnektionistische Theorie über die Struktur des Lexemspeichers⁴², aus der falsifizierbare Hypothesen abgeleitet werden können:

H1: Die Verbindung zwischen Lexemen, die Form und Bedeutung teilen, ist stärker als zwischen Lexemen, die nur Form oder nur Bedeutung miteinander teilen.

Die rein formale Beziehung zwischen Lexemen sollte in einem enzyklopädischen mentalen Lexikon eine untergeordnete Rolle spielen, da formale Familienähnlichkeit keine Voraussetzung für paradigmatische Beziehungen darstellt, wie am Beispiel von *go* und der dazugehörigen Suppletivform *went* (Bybee 1985) ersichtlich wird. Das heißt aber nicht, dass kein formales Ordnungssystem geschaffen werden kann, was in einzigartiger Weise vom TOT („Tip Of the Tongue“)-Phänomen veranschaulicht wird (vgl. die klassische Abhandlung von Brown & McNeill 1966 und Brown 1991): Soll ein relativ unroutiniertes Wort (z.B. der Name eines Schauspielers oder einer Hunderasse) erinnert werden, das dem Sprecher nicht einmal

⁴⁰Bybee (2001) ordnet lexikalisches Wissen dem prozeduralen Gedächtnis zu, ebenso wie Pawley & Syder (1983), die dafür argumentieren, dass sogenannte Satzstämme bei Muttersprachlern als „soziale Institutionen“ lexikalisiert sind. Demnach klingt *I have an ache in the head.* in den Ohren von Muttersprachlern schlichtweg deswegen sonderbar, da die Formel *I have a headache.* erwartet wird.

⁴¹Dieses Attribut geht auf Booij (2004) zurück.

⁴²Den Terminus *Lexem* könnte man hier aus konstruktionsgrammatischer Sicht auch durch den Begriff *Konstruktion* ersetzen.

notwendigerweise bereits auf der Zunge liegen muss, so können phonologische Hinweise auf Anfangs- und Endränder den Erinnerungsprozess offenbar beschleunigen⁴³.

Über das Verhältnis der Stärke der Verbindung zwischen Instanzen des Lexemparadigmas und Lexemen kann also gesagt werden, dass erstere gegenüber letzterer überwiegt. Ausgehend hiervon könnte man annehmen, dass *go* und *went* über eine kürzere Aktivierungsstrecke miteinander verbunden sind als *go* und *stand*, diese wiederum über eine kürzere Strecke als *go* und *drink*. Diesem Modell zufolge müssten Konzepte, die notwendige Bedingungen eines Frames sind, wie *Käufer*, *Verkäufer* und *Ware* im Frame *Handel* (Fillmore 1985) stärkere Verbindungen aufweisen als phonästhemische Cliques.

Feldman (2000) kann also keine Theorie, weder *ex ante* noch *ex post*, beisteuern, aus der sich überprüfbare Vorhersagen ableiten ließen. Stattdessen wird versucht das Apriori des Morphems zu verifizieren, jedoch nicht vermittelt allgemeiner disziplinübergreifender Prinzipien oder in Bezugnahme auf Erkenntnisse der Neurowissenschaften über menschliche Gehirnstrukturen, sondern – ob nun im Dienste der Transformations- oder der Konstruktionsgrammatik ist unklar – als inhärente Größe. Man muss sich also im Klaren darüber sein, dass Goldbergs Konstruktionen im allgemeinen und dem Morphem (ob nun als Tiefenstruktur oder Konstruktion) im Speziellen, so wie es in Feldman (2000) operationalisiert wird, mit Ballmer (1976) gesprochen weitläufig „materiale Realität“, d.h. stoffliche (neuronale) Realität unterstellt wird: Da das Morphem mehr sein muss als ein konnektionistisches Schema, das als Summe von Eigenschaften emergiert, die Lexeme miteinander teilen, muss es im Experiment auch mehr sein als die Summe seiner Teile. Vermutlich ohne es zu beabsichtigen entwirft Feldman (2000)⁴⁴ ein Versuchsdesign, das die Konzepte der Konstruktion und der Tiefenstruktur dem Schemakzept gegenüberstellt. Im Unterschied zur konnektionistischen Vorstellung vom Morphem (als Schema) handelt es sich bei ersteren jedoch lediglich um unüberprüfbare Postulate, denn der Mehrwert wird nicht spezifiziert. In Anbetracht ihres undurchsichtigen theoretischen Bezugs ist es nicht verwunderlich, dass es Feldman (2000) nicht gelingt, ihre spektakulären Befunde zugunsten der konnektionistischen Schemavorstellung ins rechte Licht zu rücken.

⁴³Auch episodische Hinweise, z.B. Visualisierungen einer DVD-Hülle mit dem vergessenen Namen eines bestimmten Schauspielers, scheinen zum Erfolg zu führen. Dieses Beispiel stammt aus der Selbstbeobachtung, findet sich aber auch ähnlich in Maril et al.(2001)

⁴⁴Im Übrigen stellen Feldman & Soltano (1999) in einer Kurzversion die Ergebnisse der ein Jahr später veröffentlichten ausführlichen Arbeit von Feldman (2000) vor. Bergen (2004) bezieht sich später nur auf Feldman & Soltano (1999). Es bestehen aber bis auf den Umfang keine wesentlichen inhaltlichen Unterschiede zwischen den beiden Veröffentlichungen.

3.1.4 Zur Methodologie von Replikationsversuchen

Vorliegende Arbeit möchte nicht nur durch die Evaluation vorangegangener Forschungsberichte und -ergebnisse einen Beitrag zum Fortschritt der empirischen Sprachwissenschaft leisten, sie möchte auch die Bedingungen erörtern, unter denen der Paradigmenwechsel (Kuhn 1996) der empirischen Wende auch tatsächlich mit einem Fortschritt in der Wissenschaft einhergehen kann, um schlussendlich die eigene empirische Arbeit an diesen Bedingungen auszurichten.

Mit der Etablierung der kognitiven Linguistik hat sich nicht nur die Vorstellung von Sprache erweitert, sondern zugleich auch ihr Darstellungsmittel. Der Notation der Transformationsgrammatiker steht nun das statistische Argument gegenüber, eine Entwicklung, die nicht unkritisch betrachtet werden sollte. Denn obgleich die experimentelle Datengewinnung und statistische Analyse eine Inferenz von einer Teilgesamtheit auf die Grundgesamtheit und damit die Erklärung von (sprachlichen) Phänomenen in Aussicht stellt, so hat das sprachliche Hilfswerkzeug, mit dem die Wirklichkeit erfasst werden soll, gegenüber dem statistischen den Vorteil der Transparenz. Wenn gemäß Ballmers (1976) Definition von Empirie die intersubjektive Verständigung von besonderer Bedeutung ist, dann ist das sprachliche Argument insofern empirischer, als dass es leichter zu interpretieren und zu bewerten ist als das statistische.

Zwar ist die Überprüfbarkeit des Wahrheitsgehalts erklärender Aussagen, die mit Hilfe von statistischen Verfahren getroffen worden sind, nicht unmittelbar gegeben, dafür beruht ihr hoher „Grad an Empirizität“ (Ballmer 1997:92) darin, dass sie eine nicht unwesentliche Steigerung des Erkenntnisgewinns versprechen. Um diesen in der Praxis auch zu erzielen, sind im Zuge der vorliegenden Arbeit folgende Maximen entwickelt worden:

- Das Schließen aus statistischer Evidenz erfordert eine Grundausbildung in Wissenschaftstheorie und quantitativen Messmethoden (*Kompetenz*).
- Der Umgang mit eigener und fremder statistischer Evidenz erfordert aufgrund der hohen Fehleranfälligkeit quantitativer Verfahren in methodischer *und* methodologischer Hinsicht Sorgfalt und Vorsicht (*Skepsis*)⁴⁵.
- Um Intersubjektivität zu gewährleisten, müssen Datengewinnung- und auswertung vollständig dokumentiert werden (*Transparenz*).
- Der Erkenntnisgewinn ist immer nur vorläufig (Diekmann 2008, Popper 2002), d.h. es müssen Replikationsversuche durchgeführt werden (*Konvergenz*)

⁴⁵Dieses Schlagwort ergibt sich aus dem Grundgedanken des Falsifikationismus (Popper 2002).

„From among these statements, those are selected which are not derivable from the current theory, and more especially those which the current theory contradicts. Next we seek a decision as regards these (and other) derived statements by comparing them with the results of practical applications and experiments. If this decision is positive, that is, if the singular conclusions turn out to be acceptable, or verified, then the theory has, for the time being, passed its test: we have found no reason to discard it. But if the decision is negative, or in other words, if the conclusions have been falsified, then their falsification also falsifies the theory from which they were logically deduced.

It should be noticed that a positive decision can only temporarily support the theory, for subsequent negative decisions may always overthrow it. So long as theory withstands detailed and severe tests and is not superseded by another theory in the course of scientific progress, we may say that it has ‚proved its mettle‘ or that it is ‚corroborated‘ by past experience.“ (Popper 2002:10)

Eigentlich müsste man das Kriterium der Kompetenz um einen ethischen Aspekt erweitern, der dem, was Ballmer (1976) „Immunisierungsstrategien“ nennt, zugrunde liegt. Popper dagegen setzt eine aufrichtige und gewissenhafte Herangehensweise bereits voraus:

„There are, I believe, some intuitive desiderata which cannot be satisfied by any formal definition. For example, a theory is the better confirmed the more ingenious our unsuccessful attempts at its refutation have been. My definition incorporates something of this idea—if not as much as can be formalized. But one cannot completely formalize the idea of a sincere and ingenious attempt.“ (Popper 2002:418f)

Der empirische Sozialforscher Andreas Diekmann bringt es in einem Spiegel-Interview (Evers & Grolle 2006) auf den Punkt: Um überhaupt wissenschaftlich arbeiten zu können, bedarf es eines gewissen Vertrauens in das Umfeld. Vielleicht ist es auch dieses Grundvertrauen, das die Annahme motiviert, „dass die Strenge der Prüfung nicht von psychologischen Tatsachen beim Ausführenden abhängt.“ (Musgrave 1980). Grundsätzlich darf schon vorausgesetzt werden, dass dem Forscher in erster Linie an Wahrheitsfindung gelegen ist und erst in zweiter Linie an einer Veröffentlichung in einer renommierten Fachzeitschrift. Nichtsdestotrotz muss das eigene und fremde wissenschaftliche Vorgehen einer Prüfung standhalten können. Vielleicht ist es ja auch die Implikatur des Misstrauens, die Replikationsversuche so unbeliebt macht: es mag respektlos erscheinen und peinlich sein, die Arbeit des Kollegen unter die Lupe zu nehmen. In ersterer Hinsicht nimmt man bereits vorweg, dass ihm ein böser Fehler passiert ist

(und möchte diese Vermutung durch eine Replikation falsifizieren), in letzterer Hinsicht macht man es sich ja mit der Überprüfung auch recht einfach, da dies scheinbar keine Kreativität erfordert.

Mit dieser Arbeit soll gezeigt werden, dass kreative Replikationsversuche nicht nur möglich, sondern auch notwendig sind. Was die oben genannte Implikatur anbelangt, so wird hier nichtsdestotrotz nach dem Leitsatz vorgegangen: Vertrauen ist gut, Kontrolle ist besser. Denn auch der Replikator muss grundsätzlich davon ausgehen, dass seine Arbeit überprüft werden kann, und diese möglichst vollständige Überprüfbarkeit auch gewährleisten.

Da eine quantitative Formalisierung der Prüfungsstrenge unmöglich erscheint, richtet sich die Versuchsreihe methodologisch nach Volker Gadennes qualitativer Formalisierung⁴⁶:

„U1 ist [...] genau dann eine strengere Prüfung als U2, wenn die beiden folgenden Bedingungen erfüllt sind:

1. In Untersuchung 1 (U1) wurden alle die potentiell störenden Bedingungen erster und zweiter Art kontrolliert, die in Untersuchung 2 (U2) kontrolliert wurden.

2. Es gibt mindestens eine potentielle störende Bedingung erster⁴⁷ oder zweiter Art⁴⁸, die in U1, nicht jedoch in U2 kontrolliert wurde.“

(Gadenne 1976:64)

Etwas irreführend ist hier die Nummerierung: die kontrollierte Untersuchung wäre besser U2 genannt worden und die Untersuchung, von der ausgegangen wird, U1. Dann würde es auch nicht so erscheinen als würde eine einzige Replikation genügen⁴⁹. Worum es aber hauptsächlich geht, ist der Umstand, dass mit jedem Replikationsversuch ein weiterer eliminierbarer Störfaktor beseitigt und zugleich der Grad der Prüfungsstrenge schrittweise erhöht wird. Ausgangspunkt ist dabei, wie oben bereits angedeutet, der kritische Rationalismus Poppers, vor dessen Hintergrund nur die erfolglose Falsifikation als Bewährung einer Theorie (oder Hypothese) gelten kann.

Wenn man bedenkt, dass der modernen experimentellen Sprachforschung die Annahme zugrundeliegt, dass die unbewusste Realität von Sprache⁵⁰ empirischer ist als die bewusste Konstruktion von Sprache⁵¹, ergeben sich heute ganz neue Möglichkeiten, Immunisierungsstrategien zu entwickeln als noch zur Zeit von Ballmer (1976), der mit „politischer Leidenschaft“

⁴⁶Gadennes „Prinzip der Kontrolle“ (Gadenne 1976) ist u.a. auch für die Politikwissenschaften von Bedeutung, vgl. Schmidt (1977).

⁴⁷Damit ist die interne Validität aufgehoben.

⁴⁸Damit ist die externe Validität aufgehoben.

⁴⁹Eine Benennung mit U0 wäre ja nicht sinnvoll.

⁵⁰Damit sind kognitive Strukturen gemeint, über die *Hypothesen* angestellt werden können.

⁵¹Damit sind formale Strukturen gemeint, die per *Definition* festgelegt werden. Die hervorhebende Kursivierung

(Dittmar 2005) an die wissenschaftstheoretischen Mindestanforderungen für die empirische Sprachwissenschaft erinnert hat.

„Even in light of the neologism results [...], one could still hold the position that phonaesthemes are only static, distributional facts about the lexicon, which speakers of a language can access consciously⁵². This is problematic since essentially all normal morphological processing happens unconsciously.“ (Bergen 2004:295)

Unter dieser Annahme steigt auch der Anspruch bei der Kontrolle: Nicht nur, dass das rein sprachliche Gegenargument (z.B. gegen die Behauptung, Phonästhem seien psychologisch real) nicht mehr ausreicht, Replikationsversuche können (und sollen) aufwendiger sein als die ursprünglichen Experimente, da der Replikator als solcher unter dem gesteigerten Druck eines maximalen Genauigkeitsanspruchs steht.

3.2 Zum Urexperiment von Bergen (2004)

3.2.1 Motivation

Bergens Arbeiten (2004, 2010) können als Konsequenz aus der Fülle von deskriptiven Studien zu Phonästhemem⁵³ (u.a. Firth 1930, Bloomfield 1895, Householder 1946, 1978, Debon 2001, Joseph 1991, Liberman 1990, Nuckolls 1996) sowie aus einer Vielzahl experimenteller Publikationen, die auf lexikalischen Entscheidungsaufgaben (u.a. Brown et al. 1955, Koriat & Levi 1977, Westbury 2005, Köhler 1929, Klink 2000, Magnus 2000) und auf Korpusanalyse beruhen (Drellishak 2006, Otis & Sagi 2008), aufgefasst werden.

Noch bevor Firth (1930) den Terminus Phonästhem eingeführt hat, hat Jespersen (1928) bereits bemerkt, dass manche sprachlichen Zeichen durch submorphemische Korrelationen

erfolgt hier deswegen, da die eigene Erfahrung gezeigt hat, dass Hypothesen und Definitionen weitläufig verwechselt werden. Der Unterschied zwischen beiden besteht darin, dass Definitionen nicht falsch sein können: sie sind per Definition wahr und können nicht falsifiziert werden, weswegen sich eine Überprüfung erübrigt. *Hypothesen dagegen müssen ein gewisses Risiko beinhalten falsch sein zu können.*

⁵²Bergen bezieht sich damit auf die Methoden seiner Vorgänger, u.a. Hutchins (1998) und Magnus (2000), die ihre Daten auf der Grundlage von Sprecherintuitionen erhoben haben. Wenn auch die Begriffe „bewusst“ und „unbewusst“ nicht unproblematisch sind, insbesondere wenn sie den Einsatz wissenschaftlichen Methoden begründen, so ist Bergens Kritik dennoch berechtigt. Die Fähigkeit eines Sprechers sprachspezifische Form-Inhalt-Korrelationen zu identifizieren und zu interpretieren wird kaum die Signifikanz solcher Phänomene für das Sprachsystem als Ganzes unter Beweis stellen.

⁵³Neben klassischen Abhandlungen existieren auch sprachenspezifische Studien, u.a. Abelin (1999) für das Schwedische, Austerlitz (1994) für das Finnische, Berlin (1994) für die jivaroanische Sprachenfamilie, Blust (2003) für das Austronesische, für Bengali Dimock (1957), Durand (1961) für vietnamesische Befunde, für das Japanische Hamano (1998) und Kakehi et al. (1996), für Sanskrit Hoffmann (1952), Joseph (1987, 1991, 1994) für das Griechische, für das Koreanische Kim (1977) und Martin (1962), sowie statistische Argumente in Drellishak (2006) und Otis & Sagi (2008).

zwischen Form und Inhaltsseite auffallen⁵⁴ und dass Sprecher solche Zusammenhänge intuitiv erkennen können (Jespersen 1928). Experimentell wurde diese Hypothese von Sapir (1929) und Newman (1933) überprüft und bestätigt, was wiederum als Inspiration für eine Reihe methodisch recht fragwürdiger Experimente zur Universalität von Lautsymbolik gedient hat, bei der Testpersonen Antonyme aus einer oder mehreren unbekanntem Sprachen Antonymen aus der Muttersprache zuordnen sollten, teils mit positivem Ausgang (Brown et al. 1955, Gebels 1969, Brown & Nuttall 1959, Tsuru & Fries 1933), teils mit negativem (Maltzman et al. 1956). Auch Gestalt-theoretische Studien über den Zusammenhang von visuellen Reizen und Lauten (vgl. Köhler 1947, Fox 1935, Davis 1961, Westbury 2005, Nielsen & Rendall 2011, für kontroverse Befunde Rogers & Ross 1975), die als Vorläufer der Image-Schema-Theorie (Johnson 1987) aufgefasst werden können und zu denen auch die zahlreichen Arbeiten zur Synästhesie⁵⁵ gerechnet werden können (Ramachandran & Hubbard 2001, Berlin 1994, Asher et al. 2009) deuten an, dass Lautsymbolik über eine neurologische Basis verfügt.

Bei den jüngeren Publikationen zum Thema Lautsymbolik sind einige Arbeiten hervorzuheben, mit Hilfe derer sich die Funktion des Phonästhem mit Lernprozessen im Spracherwerb begründen lässt. So erleichtern Ähnlichkeiten zwischen Formen und Inhalten das Lernen von Kategorien (Yoshida & Smith 2006, Imai et al. 2010, Christiansen 2010). Außerdem ist es denkbar, dass phonästhemische Silben die kindliche Segmentierung des akustischen Signals mittels komplexer Silbenonsets (Friederici & Wessels 1993) vereinfachen können: Durch das Zusammentreffen von expressiven Formen und expressiven Inhalten könnten erlaubte Onsets einfacher von unerlaubten unterschieden werden. Eine Übertragung dieser Erkenntnisse auf das Phonästhem wäre jedoch nur unter der Voraussetzung ihrer Kategorialität erlaubt. Eine solche Eigenschaft müsste aber erst separat (und stets vorläufig) unter Beweis gestellt werden, was sich Bergen (2004) zum Ziel gesetzt hat.

Die fehlende Konvergenz bestehender Forschungsergebnisse, aber auch die bereits erwähn-

⁵⁴In Jespersen (1928) finden sich bereits Vorüberlegungen, die Ohala (1994) zu seiner Theorie des Frequency Codes erweitert (vgl. hierauf aufbauende Arbeiten von Hays 1994, Joseph 1994, Fitch 2000, Diffloth 1994, Sereno 1994 und Tsur 2006).

⁵⁵Ogleich es es sich bei letzterem um einen seltenen neurologischen Zustand der „additiven Wahrnehmung“ handelt (Cytowic 1995), bei dem ein sensorischer Reiz unwillkürlich ein oder mehrere andere Sinnesempfindungen hervorruft, so dass beispielsweise Buchstaben als farbig empfunden werden (Brang et al. 2011, Callejas et al. 2007), scheint auch beim Nicht-Synästheten diese Kondition zumindest ansatzweise vorhanden zu sein (z.B. wenn Parfums als süß, schwer oder grün beschrieben werden) und sich kultivieren zu lassen (z.B. in der Poesie). Hier wird der Bezug zur Gestaltpsychologie besonders deutlich: da die Nase wesentlich am Geschmackempfinden beteiligt ist, können beide Sinne als Einheit empfunden werden. Die Synästhesie erhält damit eine physiologische Grundlage. Dass dies für die Konsumgüterindustrie, wo die Anpassung des Markennamens an die jeweiligen Produkteigenschaften eine entscheidende Rolle für den Verkaufserfolg spielt, von wirtschaftlicher Bedeutung ist, leuchtet unmittelbar ein (vgl. Klink 2000, Yorkston & Menon 2004, Ngo et al. 2011).

ten Mängel klassischer und jüngerer empirischen Studien zur Lautsymbolik zwingen zu einer originellen Herangehensweise. Denn während die Resultate in Magnus (2000), Hutchins (1998) und Abelin (1999) auf Sprecherintuitionen basieren, so dass zwischen Offline-Effekten einer möglicherweise signifikanten kognitiven Klasse von Phonästhemen und Online-Effekten einer analogen Ad-hoc-Derivation von salienten lexikalischen Prototypen nicht unterschieden werden kann, ist Otis & Sagis (2008) Evaluation von Phonästhem-Kandidaten auf der Grundlage ihrer Bedeutungsseite mit einer Kollokationsanalyse, der *Latent Semantic Analysis*, auf das Gutenberg-Korpus beschränkt, das sich zum Großteil aus literarischen Werken des 19. Jahrhunderts zusammensetzt und somit keinerlei Aussagen über synchrone Zustände ermöglicht.

Zwar gelingt es Bergen (2004) mit einem Primingexperiment das Phonästhem als Funktion von Reaktionszeiten zu operationalisieren, was unter kontrollierten Bedingungen erstmals Aussagen über dessen kognitive Realität erlaubt, andererseits versäumt auch er es, ein theoretisches Erklärungsmodell für die Funktionalität des Phonästhem beizusteuern. Und ohne theoriegestützte Hypothesen lässt sich sein Primingdesign schlichtweg nicht begründen.

3.2.2 Aufbau

Auf die Arbeit, an der sich Bergen (2004) in erster Linie orientiert, nämlich Feldman & Solitano (1999) respektive Feldman (2000), ist bereits in Kapitel 3.1.3 eingegangen worden. Vordergründig verfolgt er damit einen hypothesengesteuerten, explorativen Ansatz, der jedoch nicht expliziert wird. Die unterschwellige Hypothese lautet dabei: Wenn die Phonästhem systematische Form-Bedeutung-Korrelationen mit einem kategorialen Mehrwert zu dieser Korrelation sind, so müssten sie kognitiv wie Morpheme verarbeitet werden (die scheinbar dieselben Eigenschaften haben), wenn sie in ein gleichwertiges Experimentaldesign implementiert werden.

Um seine Vermutung zu überprüfen entwirft Bergen (2004) in enger Anlehnung an Feldmans (2000) Design ein phonästhemisches Priming-Experiment. Damit wird die experimentelle Forschung zu Phonästhemen fort von der Sprecherbefragung und damit dem bewussten Zugriff auf Sprache in Abhängigkeit von interpersonell variierenden Intuitionen hin zur unbewussten Sprachverarbeitung gelenkt, wobei der vage Begriff „Bewusstsein“ hier mittels der SOA operationalisiert wird. Vereinfacht lässt sich sagen: Je kürzer die SOA, desto unbewusster die Reaktionen, wobei ein natürlicher Grenzwert nach oben und unten hin angenommen werden muss. Wenn einem Teilnehmer bei einer Befragung zur Beurteilung relativ viel Zeit zur Verfügung steht (z.B. 10 Sekunden für je ein zu vergleichendes und zu bewertendes Wortpaar), fallen seine Entscheidungen bewusst aus. Es können dabei Zusammenhänge konstruiert

werden, die zuvor noch keine psychologische Realität besessen haben und deren (konstruierte) psychologische Realität vermutlich nur vorübergehender Natur sein wird. D.h. die Zusammenhänge sind dann außerhalb des Experiments nicht mehr von Nutzen (vgl. Kuiken & Vedder 2007, Dhar 1997 zum Einfluss sogenannter „task effects“).

Bergen (2004) übernimmt mit dem Design von Feldman (2000) auch die methodologischen Schwächen der Vorgängerin. Trotz einer genialen Übertragung handelt es sich bei Bergens Vorüberlegung um keine Hypothese, da sie aus einer Methode folgt und nicht aus einer Theorie abgeleitet ist:

Bergens 1. Annahme: Da es sich beim Priming um ein indirektes Verfahren zur Messung unbewusster kognitiver Prozesse handelt, haben Feldman & Soltano (1999) die Kategorialität von Morphemen gemessen.

Bergens 2. Annahme: Da Priming ein bewährtes Verfahren ist, sind Feldman & Soltanos (1999) Resultate zuverlässig.

Bergens Arbeitshypothese: Wenn sich in einem Primingdesign nach dem Vorbild von Feldman & Soltano (1999) Phonästhemeeffekte nicht additiv aus Effekten der Form- und Bedeutungsseite ergeben, wie Feldman & Soltano (1999) es in einem Telexperiment für die Morpheme zeigen konnten, hat auch das Phonästhem den Mehrwert der Kategorialität.

Dass es sich bei den gefälligen Ergebnissen in Feldman & Soltano (1999) um Teilergebnisse handelt, die sich erst nach mehrfacher Variierung der Parameter erzeugen ließen, erwähnt Bergen (2004) nicht⁵⁶.

In Bergens übernommenem Design werden die (englischsprachigen) Testpersonen mit einer lexikalischen Entscheidungsaufgabe konfrontiert, bei der es darum geht, möglichst schnell zu entscheiden, ob ein auf einem Computermonitor dargebotenes orthographisches englisches Wort (das Target) zu einem existierenden Lexem des Englischen⁵⁷ passt oder nicht. Alle implementierten Stimuli sind dabei phonologisch mögliche Wörter, jedoch ist nur die Hälfte von Ihnen lexikalisiert. Die andere Hälfte stellt phonotaktisch korrekte lexikalische Lücken dar. Die Entscheidungsaufgabe ist lediglich vordergründig; nicht um das Differenzierungsvermögen zwischen inhaltlich belegten und inhaltlich leeren Wörtern geht es, sondern um die kognitive Verarbeitung der Relationen, die zwischen einem bahnenden Wort (Prime) und

⁵⁶Das geht aus der Lektüre von Feldman & Soltano (1999) auch nicht hervor, da der Artikel lediglich eine Kurzversion von Feldman (2000) darstellt. Vielleicht wäre Bergen (2004) die schwache Aussagekraft der Ergebnisse auch nicht entgangen, wenn er seine Lektüre auf Feldman (2000) beschränkt hätte.

⁵⁷Aus der Publikation geht nicht hervor, ob es sich bei allen Teilnehmern um Muttersprachler gehandelt hat oder nicht und ob sie alle derselben Variätät des Englischen zugeordnet werden konnten.

dem zu bewertenden Wort (Target) vermutet werden. Bei dem Prime handelt es sich um einen kaum lesbaren, 150 Millisekunden auf dem Monitor erscheinenden Stimulus, der dem Target in jedem Wortpaar (Trial) vorangeht. Alle Primes sind lexikalisierte Wörter, von denen angenommen wird, dass sie bei den Versuchspersonen neuronale Verbindungen aktivieren. Die Stärke dieser Aktivierung hängt von der Relation zwischen Prime und Target ab; es wird also unterstellt, dass bestimmte Relationen eine neuronale Verarbeitung beschleunigen und andere wiederum verlangsamen wirken. Das Fehlen einer Relation würde demnach die geringste Bahnung bewirken. Folgende Relationen sind in Bergen (2004) konstruiert:

1. Der phonästhemische Zusammenhang: formale und semantische Gemeinsamkeiten zwischen Prime und Target
2. Der negative Zusammenhang (die so genannte Baseline): weder formale noch semantische Gemeinsamkeiten zwischen Prime und Target
3. Der semantische Zusammenhang: eine semantische Gemeinsamkeit zwischen Prime und Target
4. Der phonologische Zusammenhang: eine formale Gemeinsamkeit zwischen Prime und Target
5. Der pseudo-phonästhemische Zusammenhang: formale und semantische Gemeinsamkeiten zwischen Primes und Targets mit niedriger Type-Frequenz

Die Kondition der Pseudo-Phonästhemie ist dabei eine zusätzliche Neuerung zu Feldman (2000). Pseudo-Phonästhemie unterscheiden sich von den Phonästhemien im Frequenzaspekt: erstere zeigen verglichen mit letzteren eine ausgesprochen niedrige Type-Frequenz. Im Grunde genommen stellt die pseudo-phonästhemische Kondition neben der Baseline eine weitere Kontrollgruppe für die phonästhemische Kategorie dar.

„In order to test whether phonaesthemes also have a psychological reality – whether they are internalized and used by speaker-hearers – we have to answer two related questions. First, does the presence of a phonaestheme in a word affect the processing of that word? If phonaesthemes play a role in language processing, then words in which they occur should be processed differently from words in which they do not. Second, if phonaesthemes do in fact affect lexical processing, is this effect significantly different for phonaesthemes than for subparts of words that correlate some form and some meaning, but not in a statistically significant

way in the lexicon? In other words, is there a role for frequency of a form-meaning pairing in determining processing effects?“ (Bergen 2004:296)

Die Messung der durchschnittlichen Reaktionszeiten der Probanden auf die einzelnen Relationen soll nun darüber entscheiden, ob die Phonästhematik auf einen additiven Effekt von Form und Bedeutung reduzierbar sind oder einen kategorialen Mehrwert haben, d.h. signifikant stärkere neuronale Verknüpfungen mit einem dementsprechend höheren Aktivierungspotential erzeugen. Von zentralem Interesse ist dabei der Faktor der lexikalischen Verarbeitungshäufigkeit. In einem randomisierten Ablauf werden den Probanden die Stimuli (150 ms Primedauer, 300 ms ISI, 1000 ms Targetdauer) in aufeinanderfolgenden Trials präsentiert.

Bei einem Primingdesign handelt es sich also um ein Zeitmessungsverfahren (Rickheit et al. 2002), bei dem die Reaktionszeiten auf einen Reiz (das Target) im Rahmen einer simplen Aufgabenstellung (z.B. der hier verwendeten lexikalischen Entscheidungsaufgabe) in Abhängigkeit von einem bahnenden Stimulus (dem Prime) interpretiert werden. Vorausgesetzt wird dabei, dass bei der Verarbeitung des Primes durch assoziative Verbindungen (die nicht nur auf rein semantische Merkmale beschränkt sind) das Target-Lexem gebahnt, d.h. voraktiviert wird (vgl. Grimm & Engelkamp 1981). Die Differenz der Reaktionszeit der interessierenden Kondition zu den Kontrollkonditionen ergibt dabei den Primingeffekt (vgl. Neely 1991).

3.2.3 Defizite

Die Hauptdefizite in Bergen (2004) bestehen in der Versäumnis, klare Hypothesen zu formulieren und diese – und damit auch sein Experimentaldesign – aus theoretischer Sicht zu begründen.

Um mit den Replikationsversuchen nicht auch alle groben Fehler in Bergen (2004) zu replizieren, soll in vorliegender Arbeit ex post von einem bewährten metalinguistischen Prinzip ausgegangen werden, von dem angenommen wird, dass es 1) die Effekte in Bergen (2004) erzeugt und 2) replizierbar ist: Prinzipiell könnten sich die Ergebnisse seiner psycholinguistischen Studie – gesetzt dem Fall, dass sie replizierbar sind – mit Hilfe der bewährten Ökonomiemaxime erklären lassen (vgl. Martinet 1962:139, Werner 1989, Wilder et al. 1997, Wurzel 1997:305f, Ronneberger-Siebold 1997, Perkell et al. 2002; Whitney 1877). Nicht nur liegt die Ökonomiemaxime dem kommunikativen Anspruch auf expressive und interpretative Optimierung zugrunde und motiviert damit (a) die Idee bidirektionaler Beschränkungsgrammatiken (Boersma & Hamann 2008:256, Gärtner 2004), (b) die pragmatische Theorie Q- und R-basierter Implikatur (Horn 1984), (c) das H und H-Modell der Phonetik (Lindblom 1990) und (d) die Sprachwandeltheorie Rudi Kellers (Keller 2003); auch über die Sprachwissenschaften

hinaus ist sie u.a. für die Rechtswissenschaften (Friedmann 1999), die Volkswirtschaftslehre (Ramb et al. 1993), Biologie (Kokko et al. 2002), und Psychologie (McGonigle & Chalmers 1998, 2002) von Bedeutung.

Die Form des von Bergen (2004) verwendeten Experimentaldesigns, das Quasi-Experiment, wurde in dieser grundlegenden Struktur für die fortführende Arbeit beibehalten. Es ist deswegen von einem Quasi-Experiment (feststehender Terminus) die Rede, da weder die Auswahl von Probanden aus der Gesamtbevölkerung („random sampling“ Shadish et al. 2002) zufällig erfolgt ist noch Probanden zufällig der Kontrollgruppe bzw. den Versuchsgruppen zugeteilt worden sind („random assignment“ Shadish et al. 2000). Um eine maximale Generalisierbarkeit der Versuchsergebnisse zu ermöglichen, wäre es zum einen erforderlich gewesen, z.B. per Münzwurf Personen aus einer möglichst heterogenen Teilgesamtheit der Gesamtbevölkerung für das Experiment auszuwählen, wobei sich natürlich bei weitem nicht alle ausgewählten Testpersonen zu einer Teilnahme an einem wissenschaftlichen Experiment bereit erklären (Shadish et al. 2002)⁵⁸ Zum anderen hätte eine randomisierte Zuordnung die Anzahl der Probanden vervielfacht, da im idealen Experiment für jede Versuchsbedingung eine neue möglichst große Personengruppe erforderlich ist, und nicht, wie im Quasi-Experiment bei Bergen (2004), jede Kondition von ein und derselben Gruppe bearbeitet wird: anstelle der von Bergen veranschlagten 20 Personen, die fünf Konditionen ausgesetzt waren, wären demnach mindestens 20 Personen pro Kondition nötig gewesen.

Die finanziellen und zeitlichen Vorteile des Quasi-Experiments gehen also mit potentiellen Teilnehmereffekten hinsichtlich der Kondition und Teilnahmebereitschaft einher. Demnach kann es sein, dass die Probanden nur aus finanziellen Gründen teilnehmen und die Aufgabe nicht ernst nehmen, dass ihre Aufmerksamkeit, Motivation und/oder Reaktionsfähigkeit aus diversen Gründen im Verlauf des Experiments variiert. Ersterer Umstand ist jedoch hinsichtlich des menschlichen Handlungsmusters der Reziprozität⁵⁹ (Stegbauer 2002, Camerer & Fehr 2004, Braun & Gautschi 2011) weniger zu erwarten, insbesondere wenn man Reziprozität wie Hondrich (2001) als eine Norm auffasst.⁶⁰ Dennoch kann es nicht ausgeschlossen werden,

⁵⁸Deswegen kann man annehmen, dass teilnahmebereite Personen sich auch hier (d.h. im hypothetischen Szenario eines idealen Experiments) systematisch (z.B. in sozioökonomischer Hinsicht) von Teilnahmeverweigerern unterschieden hätten (Shadish et al. 2002:249).

⁵⁹Reziprozität bezeichnet „jeweils nichteigennütziges Verhalten gegenüber einem anderen Akteur, welches durch vorherige Handlungen dieses Anderen bedingt wird.“ (Braun & Gautschi 2011:264)

⁶⁰Stellt man sich die Situation zwischen Versuchsteilnehmer und Versuchsleiter als einen Tauschhandel vor, so wird der Proband im Vertrauen auf die Vergütung eine adäquate Leistung abliefern, auf die der Versuchsleiter seinerseits vertraut. Probleme ergeben sich lediglich, wenn die Vorstellungen der Probanden davon, was eine vergütungsadäquate Leistung darstellt, stark variieren. Wie genau eine potentiell abweichende Bewertung der Vergütung sich in der Leistung im Experiment umsetzen lässt, stellt eine separate Fragestellung dar, die in vorliegender Arbeit nicht bearbeitet worden ist. Aus forschungspragmatischen Gründen wurde deshalb bei der eigenen empirischen Arbeit unterstellt, dass 10 Euro einen über alle Teilnehmer hinweg gleichgewich-

dass Personen, die wissen, dass ihre Leistung nicht vergütet wird, sich dementsprechend weniger Mühe geben, während mit der Aussicht auf monetäre Entschädigung bereitwillig der Erwartung entsprochen wird, sich hierfür auch anzustrengen.

Da jeder Proband auf mehr als eine Kondition reagieren soll, können diese Teilnehmereffekte nun wesentlich schwerer vom interessierenden Effekt der unabhängigen Variablen unterschieden werden. Zu den Teilnehmereffekten sind auch mögliche Selektionseffekte zu zählen. Diese könnten darin bestehen, dass Universitätszugehörige, die Bergen für seinen Versuch rekrutiert hat, sich in ihrem Literaturkonsum quantitativ und qualitativ von universitätsferne Personen unterscheiden, was schlussendlich durch Routinisierung (Croft 2000) zum Entrenchment (für eine genauere Erläuterung dieser Prozesse vgl. Langacker 1987, Croft 2000, Schmid 2007) mentaler Repräsentationen führen könnte, von denen nicht angenommen werden kann, dass sie zum kognitiven Inventar der Allgemeinbevölkerung gehören. Plakativ gesprochen wäre es wahrscheinlicher, dass in der Kognition des Germanistikstudenten eine phonästhemische Kategorie existiert als dass dies bei einem Vermessungsingenieur oder der Kassiererin im Penny-Markt der Fall wäre⁶¹.

Dennoch ist die Aussagekraft quasi-experimenteller Studien nicht von vornherein als minderwertig einzustufen, da es sich um einen Mangel handelt, der durch den akribischen Umgang mit potentiellen Drittvariablen ausgeglichen werden kann. In Shadish et al. (2002) findet sich hierzu eine hilfreiche Anleitung: zunächst gilt es, diese Drittvariablen vorwegzunehmen, und im Idealfall in einem Pretest zu untersuchen. So könnte bspw. die Leistung von Testpersonen, die einen finanziellen Anreiz erhalten, mit der Leistung von Personen, die ohne Vergütung teilnehmen, miteinander verglichen werden.

Eine andere Möglichkeit der Kontrolle von Drittvariablen, die nicht von den Teilnehmern ausgehen, sondern in Zusammenhang mit der unabhängigen Variablen stehen, bietet das Experimentaldesign selbst. In dieser Hinsicht geht Bergen (2004) vorbildlich vor: linguistische Drittvariablen können vom interessierenden Effekt der phonästhemischen Kategorie separiert werden, indem sie diesem als Kontrollgruppen gegenübergestellt werden (vgl. die Ausführungen in Abschnitt 3.2.2). Z.B. kann durch eine Implementierung der pseudo-phonästhemischen Kondition die Bedeutung der Type-Frequenz für die mentale Verarbeitung von Phonästhemen überprüft werden.

Bergens vollständige Auflistung des Stimulisets scheint zunächst einen Vorteil für die Replikationsversuche darzustellen, der jedoch durch den Umstand, dass er doppelt so viele Stimuli

teten Wert darstellen, der den (für die Probanden gut im Vorfeld kalkulierbaren) Zeit- und Arbeitsaufwand rechtfertigt, eine bestmögliche Leistung abzuliefern.

⁶¹Der Vergleich impliziert keine Wertung der jeweiligen Berufsstände, sondern betont vielmehr die Verschiedenheit der damit einhergehenden Routinen.

aufführt als letztendlich in dem Versuch zum Einsatz gekommen sind, entscheidend gemindert wird. Da Bergen seine Auswahl nicht offenlegt, beläuft sich die Wahrscheinlichkeit mit seinem Original-Set zu arbeiten lediglich auf 252⁵. Hierbei ist folgende Erkenntnis zentral: Zu gewährleisten, dass sich die Durchschnittswerte der Token-Frequenz der jeweiligen Stimuli-Subsets in jeder der fünf Konditionen ähneln, ist hier völlig unerheblich. Seine Mittelwerte beziehen sich nämlich auf das komplette Set aus 50 lexikalischen Stimuli, von denen nur die Hälfte Verwendung gefunden hat. Im Hinblick darauf, dass die Frequenzwerte innerhalb der Konditionen nicht homogen sind (d.h. es existieren extreme Randwerte, über die gemittelt wird), wäre es also theoretisch möglich, unbeabsichtigt das Wortmaterial so zusammenzustellen, dass die interessierende phonästhemische Kondition (Token-Frequenz-bedingt) die kürzesten Reaktionszeiten hervorruft. Im Hinblick auf die Tokenfrequenz führt Bergen die Angleichung der Konditionen einmal auf Basis des Brown-Korpus (Kucera & Francis 1967, ca. 1 Million Einträge) und Thorndike-Lorge-Korpus (Thorndike & Lorge 1944, ca. 18 Millionen Einträge) und einmal auf Grundlage des English Lexicon Projects (Balota et al. 2007) durch, das gemittelte Reaktionszeiten auf isolierte (d.h. ungeprimte) Stimuli bereitstellt. Dass die Daten der Korpora textbasiert und die des English Lexicon Projects verhaltensbasiert sind, sollte Bergen nicht entgangen sein. Vielmehr zeigt seine Argumentation, aber auch die fehlende Reflektion des Frequenzbegriffs, dass hier größere Zusammenhänge nicht erkannt werden. Man muss kein Soziologe sein, um zu erkennen, dass die Elizitierung von Verhaltensdaten sich auf der Mikroebene menschlicher Individualhandlungen abspielt und Korpusdaten auf der Makroebene anzusiedeln sind. Dementsprechend ist lexikalische Token-Frequenz einmal als Verarbeitungshäufigkeit und einmal als Vorkommenshäufigkeit aufzufassen. Nun ist es zwar legitim sich im Rahmen einer Verhaltensstudie an sekundären Verhaltensdaten zu orientieren, nicht legitim ist es dagegen, anhand dieser Daten die Validität textbasierter Daten zu beurteilen: „Fortunately, there exist measures of lexical decision times for words in isolation, which help us determine how serious the frequency differences described above are.“ (Bergen 2004:298)

Der Eindruck entsteht als würde Bergens sich darum bemühen von den textbasierten Frequenzwerten durch Verhaltenswerte abzulenken, denen – anders als bei den Korpusdaten – augenscheinlich nicht unterstellt werden kann, dass sie den Effekt der unabhängigen Variablen (d.h. der Konditionen) erzeugt haben könnten. Letztere scheinen nämlich den tatsächlichen Reaktionszeitgewinn der Phonästhemie in Bergens Versuch vorwegzunehmen: in beiden Korpora handelt es sich bei den Phonästhemien um die hochfrequenteste Gruppe. Zwar kann nicht von vornherein unterstellt werden kann, dass die Verarbeitungshäufigkeit der Instanzen eines Lexems bei einzelnen Sprechern auch die Vorkommenshäufigkeit auf der Makroebene widerspiegeln, andererseits sind Makrophänomene unweigerlich auf Mikroprozesse zurück-

zuführen. Der Übergang zwischen Mikro- und Makroebene lässt sich jedoch kaum über ein Invisible-Hand-Phänomen (Keller 2003) hinaus spezifizieren. Erschwerend kommt noch hinzu, dass die Dynamik zwischen Mikro- und Makroebene nicht unidirektional verläuft, sondern dass „[d]ie Übermittlung von Informationen *von der Makroebene zu individuellen Akteuren* [...] deren Handlungen und somit auch das Systemverhalten stark beeinflussen [kann.]“ (Coleman 1991:26, eigene Kursivierung)

Während es unmittelbar einleuchtet, dass die Vorkommenshäufigkeit von Lexemen die Verarbeitungshäufigkeit beeinflusst und umgekehrt, so bedarf es einer eigenen Empirie zu untersuchen, ob Sprecher intuitiv zwischen beiden differenzieren oder nicht bzw. wie gut sie Vorkommensverteilungen prognostizieren können. Erst auf der Basis einer solchen Studie können die Ergebnisse von Bergen (2004) endgültig beurteilt werden. Können die Probanden nämlich zwischen den beiden Ausprägungen von Token-Frequenz differenzieren, stellt die Vorkommensfrequenz trotz abweichender Verhaltensdaten nach wie vor eine ernst zu nehmende Drittvariable dar, durch die die Gültigkeit der Versuchsergebnisse in Frage gestellt wird. Im Rahmen der vorliegenden Arbeit ist ein solcher Versuch ergänzend durchgeführt worden (s. Abschnitt 4.4).

Abschließend muss noch auf einige unschöne rechnerische Probleme in Bergen (2004) aufmerksam gemacht werden. Eine Überprüfung der aufgeführten Kucera-Francis- und Thorndike-Lorge-Tokenfrequenz-Durchschnittswerte auf der Basis der MRC-Datenbank⁶² (Wilson 1988) hat das erstaunliche Resultat zu Tage gebracht, dass lediglich zwei der Durchschnittswerte bis auf die erste Nullkommastelle mit denen Bergens übereinstimmen. Dass es bei den anderen Werten zu erheblichen Abweichungen kam, liegt schlichtweg daran, dass Bergen für die Konditionen, in denen Stimuli mit Frequenzwert 0 vorkamen (z.B. Kondition 1 = Phonästhem) durch entsprechend weniger Fälle geteilt hat. Hatten also zwei von 10 Wörtern den Korpusertrag 0, wurde die Summe dieser Konditionswerte nicht etwa durch 10 geteilt, sondern durch 8!

Für die Phonästhem-Kondition hat sich der Gesamtwert 126 ergeben. 126 geteilt durch 8 ergibt 15,75, was genau dem in Bergen (2004) angegebenen Wert entspricht. Teilt man jedoch richtig, nämlich durch 10, ergibt sich 12,6. Ähnlich verhält es sich mit dem Durchschnittswert 136,56 aus dem Thorndike-Lorge. Das Wort *blob* liefert hier 0 Einträge. Wird die Summe der Werte (1229) durch 9 geteilt (10 minus *blob*) ergibt sich gerundet 136,5, was dem in Bergen angegebenen Wert entspricht. Teilt man jedoch korrekt durch 10 ergibt sich 122,9.

⁶²Es handelt sich um dieselbe Datenbank, die Bergen (2004) verwendet hat. Im Übrigen fehlt letztere in seiner Bibliographie. Noch ein Hinweis: am 07.11.2013 hat die Abfrage des Brown-Korpus (anders als bei Thorndike-Lorge) über die MRC-Datenbank nicht funktioniert. Alternativ wurden die eigenen Tabellenwerte mit Hilfe einer alternativen Datenbank, dem Concordancer auf www.lexutor.ca, überprüft.

Dass die Wörter mit Frequenzwert 0 nicht als nichtexistent behandelt werden dürfen, ist einleuchtend, zumal sie ja lexikalisiert sind. Bei einem umfangreicheren Korpus als dem Brown-Korpus (Kucera & Francis 1967), beispielsweise im BNC⁶³ (The British National Corpus 2007, ca. 100 Millionen Einträge) oder COCA (Davies 2008-, ca. 450 Millionen Einträge), ist für diese Fälle jedoch ein Ertrag zu verzeichnen: Während das Brown-Korpus und Thorndike-Lorge 0 Einträge für *sneeze* zeigen, enthält das COCA bereits 632. Zum Vergleich: das COCA zeigt 18980 Einträge für *smell*, Brown dagegen nur 280 Einträge.

Extrem unschön kommt hinzu, dass Bergen die gemittelten Frequenzwerte den falschen Konditionen zuordnet. In Bergen (2004) steht der Mittelwert der Form-Kondition (111.5) bei den Pseudophonästhemem! Der Wert der Pseudo-Phonästhemem (136.5) steht dafür bei der Form-Kondition, wobei hier wieder der systematische Teilungsfehler hinzukommt. Teilt man die Summe der Pseudophonästhememwerte durch 9 erhält man 136.5, teilt man korrekt durch 10, ergibt das 122.9. Dasselbe gilt auch für Bergens Angaben aus dem Brown-Korpus: Die Form- und Pseudo-Phonästhemem-Konditionen sind vertauscht!

Auch bei der Begründung seines Designs scheint Bergen ein Fehler unterlaufen zu sein: Die Festlegung seines ISIs (Inter-Stimulus-Intervall, d.h. des Zeitfensters zwischen Prime und Target) auf 300 ms erklärt er folgendermaßen: „This inter-stimulus latency of 300ms was chosen because that delay has been reported to most effectively separate morphological from nonmorphological priming effects (Feldman & Soltano 1999).“ (Bergen 2004: 296f) Allerdings ist in Feldman & Soltano (1999) keine Rede von einem ISI von 300 ms, lediglich in der späteren Langfassung Feldman (2000) wird ein Morphemeffekt bei einer Stimulus-Onset-Asynchrony von 300ms berichtet⁶⁴. ISI und SOA sind aber nicht dasselbe.

Vielleicht sollte zu guter Letzt noch erwähnt werden, dass grobe Fehler nicht nur bei Bergen, sondern auch bei Feldman (2000) vorgekommen sind. Das Verhältnis von Prime-Frequenz zu Token-Frequenz begründet sie in der Beschreibung ihres Experimentaldesigns folgendermaßen: „Following Grainger (1990) all primes were of lower frequency than their targets.“ (Feldman 2000:1435) Bei Grainger (1990) heißt es aber:

„When the target-word is a higher frequency orthographic neighbour of the prime then, during prime-word processing competition from the activated target representation must be eliminated in some way. The way in which such competition is removed produces interference in the subsequent processing of the target.“
(Grainger 1990:240)

⁶³ „Data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.“

⁶⁴ „In Experiment 1C, primes appeared for 250 ms and were followed by a blank of 50 ms before onset of the target (total SOA = 300 ms).“ (Feldman 2000:1435)

Tabelle 3.1: Token-Frequenzwerte der lexikalischen Targets in Bergen (2004)

	Thorndike- Lorge	Kucera- Francis	Letters	Phones	Syllables
Phonaestheme	181,4	15,75	5,2	4	1
Form	136,56	12,7	5	4,5	1,1
Meaning	160,5	13,1	4,7	4,1	1,4
Pseudo-Phonaestheme	111,5	5,5	5,1	4,5	1,1
Baseline	96,3	12,22	4,8	3,8	1,2

Damit hat Feldman (2000) ihren Versuch so aufgebaut, dass die Konstellation von niedrigerfrequenten Primes und höherfrequenten Targets einen Störfaktor darstellt, von dem mit Granger (1990) angenommen werden kann, dass er maßgeblich die Effekte in ihrer Form-Kondition beeinflusst. Bei der Orientierung an bestehenden Experimentaldesigns ist deren eingehende Überprüfung demnach unentbehrlich.

Tabelle 3.2: Token-Frequenzwerte des eigenen Subsets (E) aus dem gesamten Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus Thorndike-Lorge written (TLwr) und Kucera-Francis written (KFwr)

TAR	KOND	TO_KFwr	TO_TLwr	DS_KFwr(E)	DS_TLwr(E)
gleam	1	4	196	12,6	181,4
sniff	1	2	65	(:8=15,75)	(181,4)
smell	1	34	280		
glare	1	7	164		
sneeze	1	-	18		
glimpse	1	16	186		
glow	1	16	254		
smoke	1	41	545		
snack	1	6	15		
snarl	1	-	91		
nest	2	20	113	11	96,3
hymn	2	9	49	(:9=12,22)	(96,3)
shiver	2	4	132		
twin	2	7	100		
barn	2	29	176		
ugly	2	21	178		
quill	2	9	8		
wrangle	2	-	15		

Fortsetzung nächste Seite

Tabelle 3.2: Token-Frequenzwerte des eigenen Subsets (E) aus dem Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus TLwr und KFwr

(Forts.)

TAR	KOND	TO_KFwr	TO_TLwr	DS_KFwr(E)	DS_TLwr(E)
roast	2	10	156		
yearn	2	1	36		
button	3	10	196	13,1	160,5
quit	3	15	250	(13,1)	(160,5)
wax	3	14	100		
blossom	3	7	170		
bass	3	16	28		
rope	3	15	244		
clever	3	17	225		
dim	3	19	156		
bandit	3	3	33		
lawn	3	15	203		
glove	4	9	167	5,5	111,5
flag	4	16	135	(12,7)	(:9=136,5)
plunge	4	5	227		
stab	4	3	38		
drip	4	1	94		
dread	4	9	181		
flake	4	1	59		
frost	4	6	95		
grunt	4	2	91		
provoke	4	3	28		
blast	5	15	113	12,7	122,9
cradle	5	7	45	(5,5)	(111,5)
skiff	5	9	2		
throat	5	51	438		
blot	5	6	72		
blank	5	14	166		
blob	5	2	-		
blink	5	4	83		
crook	5	3	95		
sketch	5	16	215		

Tabelle 3.3: Token-Frequenzwerte d. eigenen Subsets (E) aus d. Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus dem BNC u. COCA

TAR	KOND	TO_BNC	TO_COCA	DS_BNC(E)	DS_COCA(E)
gleam	1	336	1052	953	4632,2
sniff	1	299	1254		
smell	1	3509	17249		
glare	1	533	3027		
sneeze	1	88	579		
glimpse	1	982	5588		
glow	1	910	6573		
smoke	1	3814	21202		
snack	1	341	3318		
snarl	1	108	533		
nest	2	1426	5512	1042,8	4443,6
hymn	2	332	934		
shiver	2	328	1310		
twin	2	1832	8291		
barn	2	1296	6171		
ugly	2	1299	9315		
quill	2	70	408		
wrangle	2	72	189		
roast	2	514	4147		
yearn	2	81	610		
button	3	1629	9041	1094,4	7007
quit	3	893	11735		
wax	3	697	3470		
blossom	3	356	1600		
bass	3	1897	9189		
rope	3	1501	7447		
clever	3	2239	5219		
dim	3	743	4806		
bandit	3	78	643		
lawn	3	1031	9614		
glove	4	375	3919	620,6	3970,8
flag	4	1488	10276		
plunge	4	549	2458		
stab	4	409	1553		
drip	4	282	1648		

Fortsetzung nächste Seite

Tabelle 3.3: Token-Frequenzwerte d. eigenen Subsets (E) aus d. Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus dem BNC u. COCA

(Forts.)

TAR	KOND	TO_BNC	TO_COCA	DS_BNC(E)	DS_COCA(E)
dread	4	556	2548		
flake	4	267	781		
frost	4	981	4233		
grunt	4	184	1208		
provoke	4	589	1957		
blast	5	1132	5893	959,6	4574,2
cradle	5	358	1757		
skiff	5	26	582		
throat	5	2984	14082		
blot	5	298	557		
blank	5	1596	6995		
blob	5	124	516		
blink	5	217	2247		
crook	5	375	1285		
sketch	5	749	3690		

Tabelle 3.4: Token-Frequenzwerte d. eigenen Subsets (E) aus d. Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus dem BNC u. COCA unter besonderer Berücksichtigung der Jahre 2000-2001

TAR	KOND	DS_BNC(B)	DS_COCA(B)	COCA_00-01	DS_COCA_00-01(B)
gleam	1	1092	6037,5	103	576,3
sniff	1			121	
smell	1			1525	
glare	1			309	
sneeze	1			64	
glimpse	1			559	
glow	1			648	
smoke	1			2015	
snack	1			370	
snarl	1			49	
nest	2	725	3688,7	516	369,9
hymn	2			83	
shiver	2			108	

Fortsetzung nächste Seite

Tabelle 3.4: Token-Frequenzwerte d. eigenen Subsets (E) aus d. Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus dem BNC u. COCA unter besonderer Berücksichtigung der Jahre 2000-2001

(Forts.)

TAR	KOND	DS_BNC(B)	DS_COCA(B)	COCA_00-01	DS_COCA_00-01(B)
twin	2			1098	
barn	2			623	
ugly	2			851	
quill	2			23	
wrangle	2			12	
roast	2			328	
yearn	2			57	
button	3	1106,4	6276,4	845	576,4
quit	3			1098	
wax	3			328	
blossom	3			160	
bass	3			708	
rope	3			639	
clever	3			506	
dim	3			465	
bandit	3			46	
lawn	3			969	
glove	4	568	3058,1	324	300,9
flag	4			1249	
plunge	4			259	
stab	4			138	
drip	4			207	
dread	4			236	
flake	4			47	
frost	4			310	
grunt	4			86	
provoke	4			153	
blast	5	785,9	3760,4	529	362,3
cradle	5			189	
skiff	5			54	
throat	5			1293	
blot	5			54	
blank	5			802	

Fortsetzung nächste Seite

Tabelle 3.4: Token-Frequenzwerte d. eigenen Subsets (E) aus d. Targetset von Bergen (2004) mit den entsprechenden Durchschnittswerten (DS) aus dem BNC u. COCA unter besonderer Berücksichtigung der Jahre 2000-2001

(Forts.)

TAR	KOND	DS_BNC(B)	DS_COCA(B)	COCA_00-01	DS_COCA_00-01(B)
blob	5			42	
blink	5			200	
crook	5			120	
sketch	5			340	

4 Empirie des Phonästhemis

4.1 Erster Replikationsversuch von Bergen (2004)

Der erste Replikationsversuch von Bergen (2004) ist im Laufe des Wintersemesters 2008/09 am Institut für Theoretische Linguistik der Ludwig Maximilians Universität im Rahmen eines Methoden-Seminars von Prof. Dr. Zaeferrer von einer dreiköpfigen Experimentalgruppe unter eigener Leitung im Sinne einer Übung durchgeführt worden. Es verwundert also nicht, dass es hier, trotz der gewissenhaften Bemühung Bergens Studie möglichst 1:1 zu wiederholen, aufgrund der fehlenden praktischen Erfahrung der Arbeitsgruppe mit psychologischen Experimenten zu einem vermeidbaren methodischen Fehler gekommen ist. Wie weiter unten noch ausgeführt wird, kann er im Nachhinein als eine erste, wenn auch unbeabsichtigte Modifikation des zugrunde liegenden Experiments aufgefasst werden, die eine erste Bewertung der Resultate Bergens (2004) erlaubt.

4.1.1 Teilnehmer

Was das statistische Problem der Stichprobengröße anbelangt, so ist mit Diekmann (2001:189f.) davon auszugehen, dass letztere sich umgekehrt proportional zur erwarteten Effektgröße verhält. D.h. seltene Ereignisse bzw. kleine Effekte erfordern relativ umfangreiche Stichproben, während große Effekte bereits mit relativ kleinen Stichproben gemessen werden können. Natürlich setzt diese Faustregel die Einschätzbarkeit der Effektstärke voraus, welche idealerweise aus der theoretischen Fundierung folgt. Im Fall des Phonästhemis handelt es sich um eine Größe, von der im Sinne einer Alternativhypothese unterstellt werden kann, dass sie zum kognitiven Inventar jeden einzelnen Sprechers einer Sprache gehört, in deren Wortschatz phonästhemische Lexeme statistisch auffällig sind. Da Bergens Versuch in Analogie zum Morphempriming von Feldman (2000) bzw. Feldman & Soltano (1999) entworfen worden ist und auch ähnliche Effekte unterstellt, kann man annehmen, dass ebensoviele Probanden wie bei Feldman (2000) anzusetzen sind, die mit 95-96 (Experiment 1) und 120 Probanden (Experiment 2) gearbeitet hat. Erstaunlicherweise kann Bergen (2004) signifikante Ergebnisse mit

bereits nur 20 Teilnehmern zeigen. Konkret bedeutet dies, dass Bergen angenommen haben muss, dass das Phonästhem stärkere Effekte erzielt als das Morphem, was sich theoretisch nicht unmittelbar begründen lässt. Dennoch hat sich der Replikationsversuch sinngemäß an der Probandenzahl des Vorgängers zu orientieren, so dass für die eigene Versuchsreihe ebenfalls 20 Teilnehmer angesetzt wurden. Letzendlich gelang es lediglich 19 Teilnehmer zeitnah zu rekrutieren, von denen eine/r Teilnehmer/in weit unterdurchschnittlich langsame Reaktionszeiten aufwies und aus der Berechnung ausgeschlossen werden musste.

Wie bereits im vorhergehenden Abschnitt angesprochen, stellte der Umstand, dass Bergen sein Experiment unvollständig dokumentiert, ein unüberwindbares Problem für einen möglichst engen Replikationsversuch dar. Zwar sind seine Angaben hinsichtlich der Anzahl der Probanden und der linguistischen Stimuli sehr genau, Details zum Vorgehen bei der Auswahl der Stimuli und bei der Stichproben- und Datengewinnung fehlen dagegen völlig. Diese fehlenden Informationen ließen sich trotz Nachfrage bei Bergen nicht rekonstruieren¹. Auch wenn an dieser Stelle nicht davon ausgegangen werden darf, dass Bergen (2004) bei der Dokumentation seines Experiments gezielt selektiv vorgegangen ist, muss dennoch darauf hingewiesen werden, dass die jeweiligen Lücken eine Immunisierungsstrategie ermöglichen (vgl. Ballmer 1976): Abweichende Resultate künftiger Replikationsversuche können so stets mit Abweichungen vom ursprünglichen Stimuliset begründet werden.

Obgleich Bergen mit Muttersprachlern des Englischen gearbeitet hat (wobei keine Angaben zur Varietät gemacht wurden), die sich teils aus Studenten, teils aus anderen nicht näher spezifizierten Universitätsangehörigen² zusammengesetzt haben, sollte sich für den Replikationsversuch die Rekrutierung der Testpersonen nicht auf den Campus der Ludwig Maximilians Universität München beschränken, da eine solche Einschränkung nicht aus der Theorie des Phonästhem folgen kann³. Außerdem war anzunehmen, dass der monetäre Anreiz von 5 Euro Aufwandsentschädigung⁴ nicht ausreichend Studierende mit der Muttersprache Englisch zeitnah zu einer Teilnahme würde motivieren können, selbst wenn diese nur wenige Minuten in Anspruch genommen hätte. Tatsächlich hielten sich die Rückmeldungen in Gren-

¹Auf Anfrage per Email gab Benjamin Bergen an, über den Entwurf des Experimentaldesigns nicht mehr zu verfügen. Er begründete dies damit, dass das Experiment zum Zeitpunkt der Nachfrage bereits 9 Jahre zurücklag. Aus seiner Erinnerung konnte er lediglich einige unwesentliche Angaben zu den Probanden und zum Versuchsdesign machen. Um die statistische Analyse auf mögliche Zahlendreher hin zu überprüfen wurde um Einsicht in die ursprünglichen Sprachdaten gebeten, doch auch diese schienen ihm im Laufe der Zeit leider abhanden gekommen zu sein.

²Quelle dieser Angaben ist die persönlichen Kommunikation mit Herrn Bergen per Email.

³Es kann ja nicht angenommen werden, dass Phonästhem lediglich bei einer bestimmten Bildungsschicht entrencht sind, auch wenn dies eine interessante Hypothese darstellt, der man in einem separaten, auf diese Fragestellung abgestimmten Experimentaldesign nachgehen könnte.

⁴Bei Feldman (2000) stellte die Teilnahme entweder die Zugangsvoraussetzung für ein Seminar dar oder wurde mit 5 Dollar vergütet. Bergen (2004) dagegen macht keine Angaben zur Vergütung.

zen, obgleich auch auf dem Campus der nahegelegenen Technischen Universität München ein Aufruf ausgehängt worden ist. Das Rekrutierungsterrain wurde daraufhin erweitert. Die fehlenden Probanden wurden schlussendlich über das Internetforum www.toytowngermany.com angeworben, eine Kommunikationsplattform für englischsprachige⁵ Personen in Deutschland.

4.1.2 Experimentaldesign, Aufgabenstellung und Ablauf

Das Design von Bergen (2004) wurde übernommen, so weit es seine Dokumentation gestattete. In die lexikalische Entscheidungsaufgabe fielen 50 Wortpaare (z.B. *SNiffle* – *SNeeze*), die für jeden Teilnehmer identisch waren und in derselben Reihenfolge am PC-Monitor dargestellt wurden. Bei der Auswahl der Stimuli mussten aus Bergens 10-er Set lexikalisierter Targets (pro Kondition) 5 Sets (pro Kondition) ausgesucht werden. Insgesamt war von Bergen also ein Pool aus 50 Sets aufgelistet, von denen jedoch nur 25 Anwendung fanden. Um welche es sich genau gehandelt hat, konnte aufgrund fehlender Angaben nicht ermittelt werden, so dass die Wahrscheinlichkeit mit Bergens Originalset zu arbeiten 25/2⁵. Also wurde aus dem jeweiligen Konditionsset von 10 lexikalisierten Targets der Reihe nach abwechselnd ein Target übernommen und eines verworfen. Ergänzend dazu kamen pro Kondition 5 Sets mit nicht-lexikalisierten Targets (lexikalische Lücken), die den phonotaktischen Regeln des Englischen entsprechen. Pro Kondition wurden demnach 10 Wortpaare bestehend aus Prime und Target präsentiert, von denen nur die Hälfte der Targets lexikalisiert ist⁶. Wie bereits erwähnt handelt es sich um insgesamt 5 Testkonditionen:

1. Der phonästhemischen Zusammenhang: formale und semantische Gemeinsamkeiten zwischen Prime und Target (Kondition 1⁷)
2. Der negative Zusammenhang (die so genannte Baseline): weder formale noch semantische Gemeinsamkeiten zwischen Prime und Target (Kondition 2)
3. Der semantische Zusammenhang: eine semantische Gemeinsamkeit zwischen Prime und Target (Kondition 3)

⁵Die Varietäten wurden zweckmäßig auf Britisches Englisch, Amerikanisches Englisch, Australisches und Kanadisches Englisch beschränkt.

⁶Insgesamt setzten sich für jeden Teilnehmer die Wort-Trials aus 5 Phonästhem-, 5 Pseudo-Phonästhem-, 5 Bedeutungs-, 5 Form- und 5 relationslosen Prime-Target-Paaren zusammen, die um 25 Pseudowort-Trials ergänzt worden sind. Erwähnenswert ist, dass Bergen das Wort *bling* zu den Pseudowörtern gerechnet hat; allerdings ist es mittlerweile als Modewort geläufig (268 Einträge im COCA (Davies 2008-))

⁷Die Nummerierung der Konditionen wird für die eigene Empirie im Weiteren beibehalten. Mit „Kondition 3“ wird also stets auf die Bedeutungsrelation Bezug genommen (z.B. in Tabellen und im PXLab-Dokument).

4. Der phonologische Zusammenhang: eine formale Gemeinsamkeit zwischen Prime und Target (Kondition 4)
5. Der pseudo-phonästhemische Zusammenhang: formale und semantische Gemeinsamkeiten zwischen Primes und Targets mit relativ⁸ niedriger Type-Frequenz (Kondition 5)

Der Replikationsversuch wurde mit PXLab (Irtel 2007) geschrieben, einer Sammlung von Java-Klassen und Anwendungen für computergestützte psychologische Experimente. Das Design umfasste eine Anleitung, einen Testlauf zur Einstimmung auf die Aufgabenstellung sowie das linguistische Material. Die Primedauer belief sich wie bei Bergen auf 150 ms, das ISI auf 300 ms (SOA 450) und die Targetdauer auf 1000 ms, wobei Reaktionszeiten, die dieses Limit überschritten haben, als Fehler gewertet wurden. An dieser Stelle muss nochmals darauf hingewiesen werden, dass Bergen seine ISI mit der Begründung festlegt, Feldman & Solitano (1999) hätten bei dieser ISI einen Morphemeffekt gemessen (Bergen 2004:296f.), was so nicht stimmt. Dieser Effekt wurde bei einer SOA von 300 ms und einer Primedauer von 250 ms ermittelt, während sich Bergens SOA auf 450 ms beläuft. Messbare Phonästhem-Effekte wären in diesem Bereich demnach gar nicht zu erwarten. Da der Primingeffekt für die Morphemkondition in Feldman (2000) jedoch mit zunehmender SOA an Stärke gewonnen hat, könnte man diese Erhöhung der SOA in Bergen (2004) nachträglich mit einer dem Morphem nachgestellten Verarbeitungsphase des Phonästhems begründen⁹. Doch auch das wäre nur eine unzulängliche Rechtfertigung, wenn man bedenkt, dass die Erhöhung der SOA hier auf Kosten der Primedauer geht¹⁰. Obgleich Bergens SOA von 450 ms also weder theoretisch noch mit vorangegangenen Erfahrungswerten begründbar ist, musste sie für den Replikationsversuch übernommen werden.

Wo Informationen zu der Präsentationsform des Sprachmaterials auf dem Computerbildschirm bei Bergen gefehlt haben, wurde nach eigenem Ermessen vorgegangen. Z.B. wurden die Targets zweckmäßig in Grün (vor schwarzem Hintergrund) hervorgehoben und zentriert, während eine graue Schriftfarbe von den Primes ablenken sollte, die (wie bei Bergen) etwas oberhalb der Bildschirmmitte erschienen sind. Die Instruktion liest sich wie folgt:

„Your task in this experiment is to decide whether a GREEN sequence of letters appearing in the center of the screen constitutes a valid word of English or not.

⁸„Relativ“ meint hier relativ zur phonästhemischen Kondition.

⁹Das heißt, man könnte behaupten, dass die Aktivierung phonästhemischer Links in einer späteren Verarbeitungsphase, also nach der formalen, semantischen und morphemischen Verarbeitung, stattfindet, so dass das SOA-Zeitfenster erweitert werden muss.

¹⁰Bei Feldman (2000) beträgt die ISI 50 ms und bei Bergen (2004) 300 ms. Von einer 450 ms-SOA bei Bergen (2004) fallen also bereits 300 ms nur für die Pause weg.

Immediately before the sequence of GREEN letters will be shown, you will see a grey sequence of letters slightly above the center of the screen. Try to ignore these grey letters and only attend to the GREEN letter sequence. If the GREEN sequence which is shown is a word of English, then press the RIGHT arrow key on your keyboard. If the sequence is NOT a word of English, then press the LEFT arrow key on your keyboard. Please use the fingers of both hands to respond. Whether your response is right or wrong, will NOT be shown to you. Please try nevertheless to respond correctly and as quickly as possible. Please press the UPWARD arrow key to start the short training phase of the experiment. After this short training, the monitor will advise you to start the real experiment.“

Der Trainings-Block umfasste 10 Trials (z.B. *roar* – **spess*) und wurde vom nächsten Block durch eine Pause getrennt. Die Pause wurde von folgender Bildschirmanweisung signalisiert: „PAUSE – Relax for a while. To go on, please press the UPWARD arrow key!“ Auf diese Weise konnte sich jeder Teilnehmer die Pause nehmen, die er zur Konzentrationssteigerung benötigte. Bei der Begrüßung des Teilnehmers wurde jedoch von der Versuchsleitung darauf hingewiesen, dass nur soviel Zeit wie absolut nötig zum Ruhen aufgewandt werden sollte. Keiner der Probanden fiel durch wesentliche Abweichungen in der Bearbeitungsdauer auf. Die Pausen-Option stellt die einzige willkürliche Abweichung zum Vorgänger dar und sollte das bei einzelnen Teilnehmern, aber auch zwischen den Teilnehmern variierende Ermüdungslevel kontrollieren bzw. die Aufmerksamkeitsleistung möglichst konstant halten. Hierfür wurde auch auf eine ruhige Arbeitsumgebung geachtet.

Der Versuchsablauf gestaltete sich folgendermaßen: Die Probanden wurden nach Terminvereinbarung einzeln in einem angemieteten Seminarraum auf dem Campus der Ludwig Maximilians Universität München empfangen, wobei ihnen bereits mit dem Aufrufplakat eine Aufwandsentschädigung von 10 Euro in Aussicht gestellt worden war. Alle Fenster des Seminarraums waren geschlossen, um den Geräuschpegel möglichst niedrig zu halten, und es wurde darauf geachtet, dass das Notebook, auf dem die Aufgabenstellung bearbeitet werden sollte, so platziert war, dass die Versuchsperson nicht von Lichtreflexen auf dem Display abgelenkt werden konnte. Die Versuchsleitung wies jeden Teilnehmer darauf hin, dass das Experiment a) aus der Bearbeitung einer sehr einfachen Aufgabe besteht, b) in einem sehr raschen Tempo abläuft und deswegen hohe Konzentration erfordert, c) es aber eine Trainingsphase und auch selbst regulierbare Pausen gibt, d) die Pausen jedoch nicht länger als unbedingt nötig in Anspruch genommen werden sollen, e) alles in allem möglichst zügige und möglichst richtige Reaktionen erwünscht sind und f) die Trainingsphase zudem den Zweck hat, bei Unklarheiten noch einmal aus dem Raum treten und die Versuchsleitung fragen zu können ohne

negative Auswirkung auf das Versuchsergebnis. Für die Dauer des Experiments befand sich der Proband demnach alleine in dem Seminarraum, während die Versuchsleitung vor der Tür wartete. Lediglich eine einzige Person hat über alle Replikationsversuche hinweg die Möglichkeit nutzen müssen noch einmal herauszutreten und Fragen zu stellen. Nach Beendigung der Aufgabe wurden die Probanden gebeten Angaben zum Schulabschluss, Alter, Händigkeit und der Zeit, die sie in einem englischsprachigen Land verbracht hatten, zu machen. Da dieser Bitte nur sehr zurückhaltend entsprochen worden ist, konnte auf keine zusätzlichen potentiellen Teilnehmereffekte hin kontrolliert werden.

4.1.3 Ergebnisse

Deutlich wird in Tabelle 4.1, dass erwartungsgemäß bei nicht-lexikalisierten Targets die meisten Fehler gemacht werden und die gemittelten Reaktionszeiten am längsten sind. Keine Fehler und auch die schnellsten Reaktionszeiten sind dagegen für die Kondition 3 (Bedeutungsrelation) charakteristisch. Obwohl somit für die vorliegende Stichprobe gilt, dass ein Bedeu-

Tabelle 4.1: Fehlerraten, Fallzahlen, durchschnittliche Reaktionszeiten und Standardabweichungen nach Konditionen

Kondition	Fehlerrate	N (richtig)	MW Reaktionszeit	Standardabw.
Pseudo-Target	0,910	355	726,205	118,760
Phonästhem	1	82	688,816	123,574
Keine Relation	0,963	79	683,501	127,139
Bedeutung	1	86	640,024	109,474
Form	0,977	84	645,020	98,467
Pseudo-Phonästhem	0,988	82	641,983	166,310

Hinweis: Für alle Konditionen wurden nur die Reaktionszeiten der richtigen Antworten für die Berechnung der durchschnittlichen Reaktionszeit in ms verwendet. Das Label *Pseudo-Target* bezieht sich auf die Gruppe der phonotaktisch möglichen, aber nicht-lexikalisierten Targets.

tungszusammenhang zwischen Prime und Target sowohl die kleinsten Fehlerraten als auch die kürzesten Reaktionszeiten nach sich zieht, kann dieser Zusammenhang nur unter bestimmten Voraussetzungen auf die Grundgesamtheit verallgemeinert werden. Deutlich wird dies an den Standardabweichungen der Reaktionszeiten in den einzelnen Konditionen. Hierbei wäre z.B. ein Wert von 109,474 so zu interpretieren, dass – eine Normalverteilung der Reaktionszeiten vorausgesetzt – 95% der Reaktionszeiten in der Bedeutungs-Kondition zwischen 425 ms und 855 ms liegen. Somit könnte eine Reaktionszeit von beispielsweise 700 ms in jeder der 5

Konditionen beobachtet werden. Um nun herauszufinden, ob wesentliche Unterschiede zwischen den durchschnittlichen Reaktionszeiten der einzelnen Konditionen existieren, müssen geeignete Signifikanztests durchgeführt werden, mit denen sich Wahrscheinlichkeiten für die Gültigkeit von Nullhypothesen berechnen lassen. Idealerweise wird hier eine Nullhypothese über die Gleichheit der durchschnittlichen Reaktionszeiten angesetzt, wobei z.B. bei einer einfaktoriellen¹¹ Varianzanalyse (ANOVA) ein F-Test durchgeführt wird. Hierbei ergibt sich bei einer Beschränkung auf lexikalisierte Targets¹² für die Fälle aus Tabelle 4.1 ein F-Wert von 2,98, wobei der p-Wert der korrespondierenden F-Verteilung ($F(4, 408)$) 0,0192 beträgt. Bei einem Signifikanzniveau α von 0,05 bzw. 5% kann die Nullhypothese somit abgelehnt und die Alternativhypothese angenommen werden, oder anders gesagt: Bei den durchschnittlichen Reaktionszeiten von mindestens zwei der 5 Konditionen gibt es einen signifikanten Unterschied. Die Logik dieses Signifikanztests lässt sich dabei anhand der theoretischen F-Verteilung veranschaulichen (Abbildung 4.1): Die Freiheitsgrade im Zähler und Nenner der theoretischen Verteilung der Teststatistik ergeben sich aus der Streuungszerlegung der Reaktionszeiten (insgesamt 413 Fälle, 5 Faktorstufen) bei der Varianzanalyse. Der kritische Wert der Teststatistik wird durch das Signifikanzniveau α von 0,05 festgelegt und liegt bei 2,39 (senkrechte Linie). Somit liegen 95% der Werte der Teststatistik unter der Nullhypothese (H_0 : Es gibt keine wesentlichen Unterschiede zwischen den Konditionen.) zwischen 0 und 2,39. Größere Werte sind dagegen unter H_0 sehr unwahrscheinlich und sprechen für die Alternativhypothese (H_1 : Es gibt wesentliche Unterschiede zwischen den Konditionen.) Der p-Wert bezieht sich jeweils auf die Fläche unterhalb der Kurve der F-Verteilung, die rechts vom beobachteten F-Wert liegt, wobei der p-Wert bei $F = 2,39$ genau 0,05 beträgt. Gemeint ist hiermit, dass sich die durchschnittlichen Reaktionszeiten nicht nur in der jeweiligen Stichprobe, sondern mit sehr großer Wahrscheinlichkeit auch in der Grundgesamtheit voneinander unterscheiden, so dass ein Inferenzschluss¹³ möglich wird.

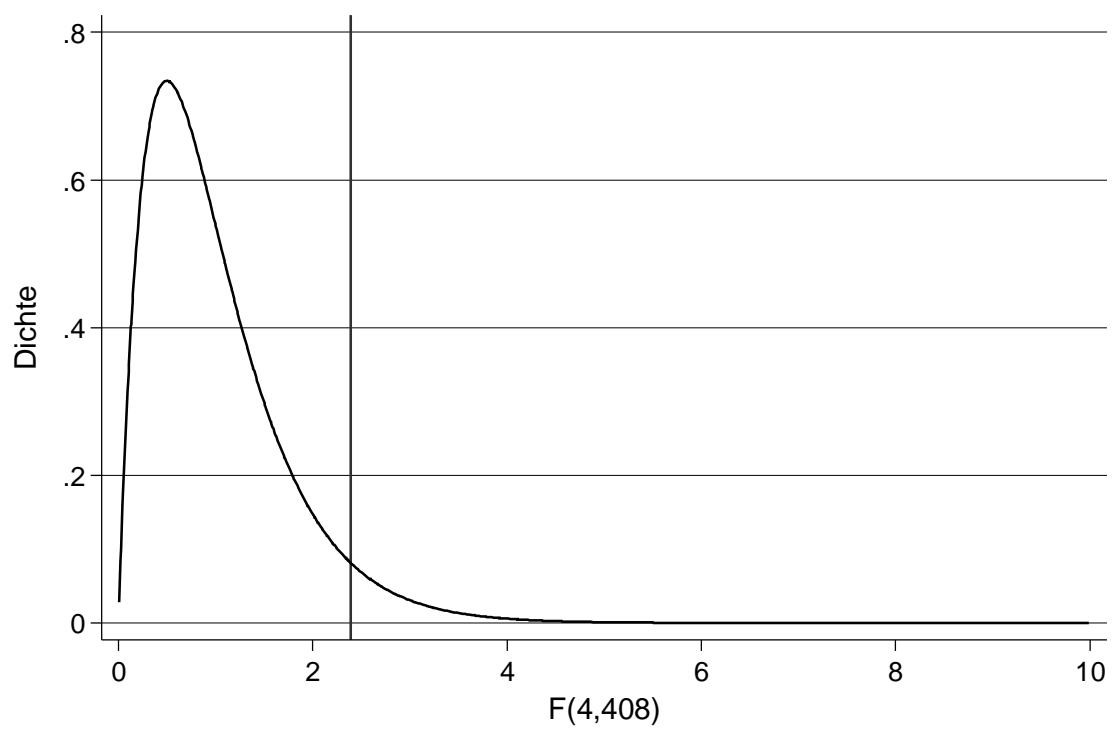
Wie bereits besprochen gilt es zudem zu berücksichtigen, dass nicht nur die unabhängige Variable (d.h. der Zusammenhang zwischen Prime und Target), sondern auch zahlreiche Drittvariablen die Reaktionszeiten beeinflussen können. Deshalb werden bei vergleichbaren Experimenten (Feldman 2000, Rastle et. al. 2000) zwar ein Relationseffekt („relatedness“) und ein Konditions-Effekt („prime type“, „condition“) berichtet, für weitere potentielle Störfaktoren (z.B. Vorkommenshäufigkeit der Targets, Reihenfolge der Trials im Experiment, Wortlänge)

¹¹Einfaktoriell bezieht sich hier auf nur einen Faktor, der jedoch abgestuft sein kann, z.B. auf den Faktor „Relation“, der aus 5 Faktorstufen besteht: 1. phonästhemische Relation, 2. keine Relation, 3. semantische Relation, 4. phonologische Relation, 5. pseudo-phonästhemische Relation.

¹²Die Beschränkung ist notwendig, da bei der einfaktoriellen Varianzanalyse auf den Effekt der lexikalisierten Targets (gegenüber den nicht-lexikalisierten) verzichtet wird.

¹³Damit ist der Schluss von einer Stichprobe auf die Grundgesamtheit gemeint.

Abbildung 4.1: Theoretische F-Verteilung mit 4 Freiheitsgraden im Zähler und 408 Freiheitsgraden im Nenner



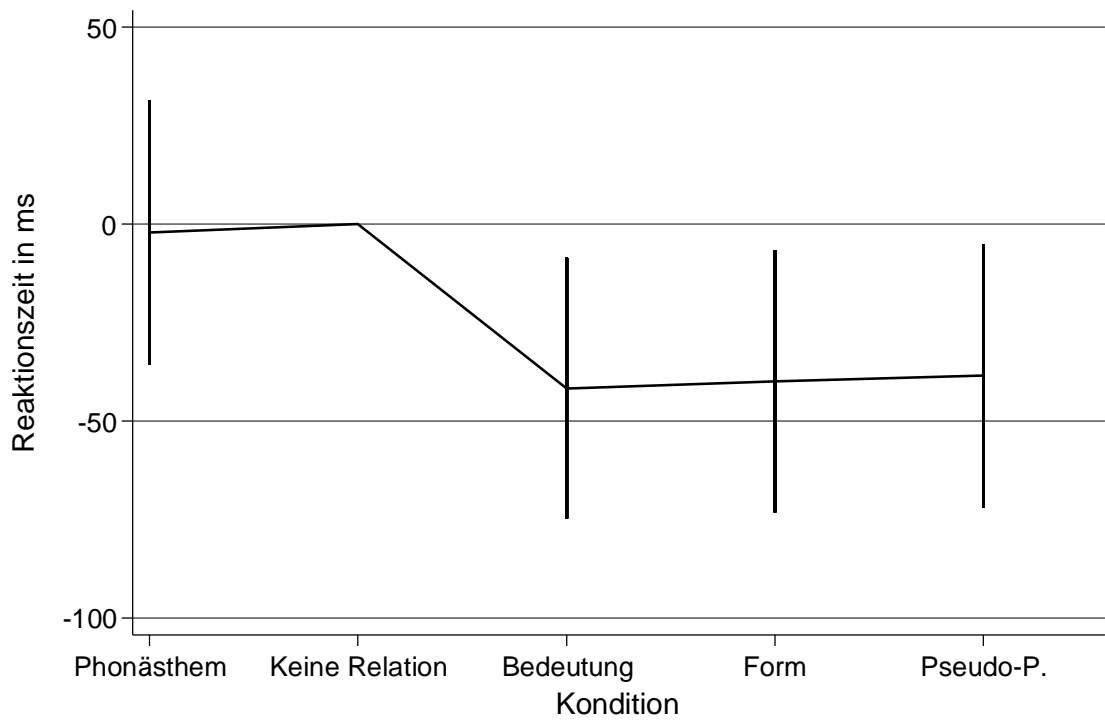
wird dagegen schlichtweg Homogenität bzw. Kontrolle durch Randomisierung unterstellt. So bewegt sich der Häufigkeitsdurchschnitt bei Feldman (2000) zwischen 20 (Formgruppe) und 34 (Bedeutungsgruppe), die Silbenlänge variiert bei Bergen (2004) zwischen 1 (Phonästhemgruppe) und 1,4 (Bedeutungsgruppe). Hierbei ergeben sich für die Fälle aus Tabelle 4.1 inkl. der Pseudo-Wort-Targets p-Werte von 0,0063 (Target-Typ-Effekt, d.h. Wort-Targets werden von Pseudo-Wort-Targets unterschieden), 0,0238 (Relationseffekt, d.h. die Konditionen mit Relation werden von der relationslosen Kondition 2 unterschieden), 0,0000¹⁴ (Teilnehmereffekt, d.h. dass die Reaktionsfähigkeit der Teilnehmer stark variiert) und 0,0130 (Unterschiede zwischen den Konditionen). Die Daten von einem Teilnehmer mussten aus der Statistik entfernt werden, weil seine Reaktionszeiten systematisch die 1000 ms-Marke überschritten haben. Nur solche Teilnehmer zu selektieren, die über ein vergleichbares Reaktionsvermögen verfügen, ist deswegen problematisch, da damit die Sample-Größe reduziert würde. Ein Sample von 20 Teilnehmern wie bei Bergen (2004) ist allerdings zu knapp bemessen, um sich eine Korrektur für Teilnehmereffekte leisten zu können. Somit können Effekte sowohl für den Unterschied a) zwischen Wort-Targets und Pseudo-Wort-Targets, b) zwischen der relationslosen Baseline und den Konditionen mit Relation, c) bei der Reaktionsfähigkeit der Teilnehmer (Drittvariable) als auch d) zwischen den einzelnen Konditionen mit Relation (z.B. zwischen Form- und Bedeutungskondition) nachgewiesen werden.

An dieser Stelle noch eine grundsätzliche Bemerkung zur Varianzanalyse: es ist etwas un schön, dass mit diesem Rechenverfahren standardmäßig (d.h. u.a. auch in Bergen 2004, Feldman 2000, Rastle et al. 2000, McCormick et al. 2009) nur F-Tests berechnet werden, die für sich genommen keine Aussagen über die Richtung und Stärke der getesteten Effekte zulassen. Man muss sich also klar machen, dass mit einer Varianzanalyse die Fragen, a) ob die durchschnittliche Reaktionszeit bei den Pseudo-Wort-Targets kleiner ist als bei den Wort-Targets oder nicht etwa umgekehrt oder b) um wieviele Millisekunden die durchschnittliche Reaktionszeit bei der Bedeutungs-Kondition kleiner ist als bei der relationslosen Kondition, nicht direkt beantwortet werden können. Um dieses Problem zu lösen, werden üblicherweise im Anschluss daran Tabellen und Grafiken erstellt, aus denen die fehlenden Informationen abgelesen werden können. Mit einem Marginsplot lassen sich die Effekte der einzelnen Konditionen darstellen (Abbildung 4.2). Grundlage hierfür sind Regressionsmodelle, die in vielen Statistikprogrammen¹⁵ zur Berechnung der Varianzanalyse verwendet werden. Im Gegensatz zur ANOVA-Darstellung lassen sich damit die Richtungen und Signifikanzen der interessierenden Unterschiede (zwischen den Konditionen) bei Spezifikation des zugrundeliegenden Regres-

¹⁴P-Werte mit $p = 0,0000$ ergeben sich dann, wenn die Wahrscheinlichkeit für die Nullhypothese kleiner als 0,001 % ist.

¹⁵So auch bei Stata, das für die vorliegende Arbeit verwendet wurde.

Abbildung 4.2: Marginsplot mit 95 % Konfidenzintervall für die Effekte der Konditionen



Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht, eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde für die Erstellung des Plots eine ANOVA ohne Relationseffekt geschätzt.

sionsmodells dann unmittelbar aus der Regressionstabelle (Tabelle 4.2) ablesen. Hier wird

Tabelle 4.2: Einflussfaktoren auf die Reaktionszeit

	Reaktionszeit	
Target/Pseudo-Target	-37,12**	(-2,76)
Phonästhem	-2,113	(-0,12)
Bedeutung	-41,64*	(-2,47)
Form	-39,85*	(-2,35)
Pseudo-Phonästhem	-38,45*	(-2,25)
Konstante	807,5***	(45,92)
Beobachtungen	768	
R^2	0,310	

t-Werte in Klammern

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Relationseffekt geschätzt. Auf die Darstellung des Teilnehmer-Effekts (17 Dummyvariablen) wurde aus Platzgründen verzichtet.

sofort deutlich, dass die Reaktionszeiten in fast allen Konditionen schneller (negative Vorzeichen der Effekte) sind als in der Referenzkategorie „keine Relation“. Nur die Phonästhem unterscheiden sich nicht signifikant von der Referenzkategorie. Aufgrund der Linearität des Modells (d.h. Regressionen und Varianzanalysen sind lineare Modelle) stimmen die Regressionskoeffizienten zudem mit den Marginalen Effekten aus Tabelle 4.2 überein. Zu interpretieren sind sie so, dass z. B. die Reaktionen in der Bedeutungs-Kondition durchschnittlich 41,64 ms schneller erfolgen als in der Referenzkategorie „keine Relation“. Genauso ist auch der Target-Typ-Effekt zu interpretieren, wobei die Reaktionszeiten bei Wort-Targets durchschnittlich 37,12 ms kürzer sind. Die Konstante steht dagegen für die durchschnittliche Reaktionszeit in den Referenzkategorien, also „Pseudo-Wort-Targets“ und „Keine Relation“, wobei der im Vergleich zu Tabelle 4.1 relativ hohe Wert darauf zurückzuführen ist, dass für den Teilnehmer-Effekt der langsamste Teilnehmer (Teilnehmer 11) als Referenzwert spezifiziert wurde.

Bei der Modellierung musste auf den Relationseffekt verzichtet werden, da eine Aufnahme der entsprechenden Dummyvariablen in das Modell zum Ausschluss einer der relationsbehafteten Konditionen geführt hätte. Die Ursache hierfür ist das Verfahren zur Bestimmung der Regressionskoeffizienten bzw. Gruppenmittelwerte (Varianzanalyse), bei dem ein voller Spaltenrang der Designmatrix (d.h. des Datensatzes, der in den Zeilen die Beobachtungen und in den Spalten die Einträge für die einzelnen Variablen enthält) vorausgesetzt wird. Letztendlich

handelt es sich hierbei jedoch nur um zwei Spezifikationsvarianten, die beide ein und dasselbe Modell implizieren, so dass der Relationseffekt durch einen entsprechenden F-Test auch im Rahmen des Modells in Tabelle 4.1 bestimmt werden kann ($F(4, 745) = 3,19$, mit einem p-Wert von 0,0130). Interessant ist zudem die Anpassungsgüte R^2 , die erkennen lässt, dass etwa 31% der Varianz der Reaktionszeit durch das Modell erklärt werden kann.

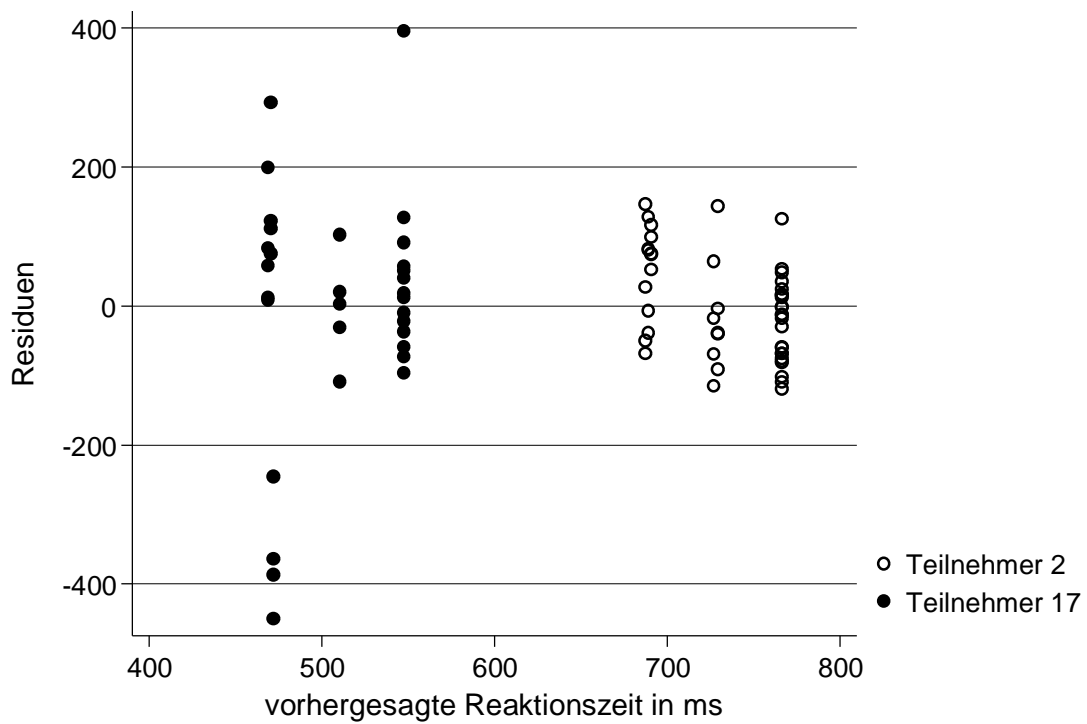
Stellt man diese relativ einfache und übersichtliche Zusammenfassung von Ergebnissen den oftmals kryptischen Darstellungen und Interpretationen von ANOVA-Outputs in einschlägigen Veröffentlichungen (z.B. Feldman 2000) gegenüber, wird deutlich, dass Regressionstabellen bei den Experimenten, die im Rahmen dieser Arbeit durchgeführt wurden, im Vergleich zu ANOVA-Erträgen nur Vorteile mit sich bringen¹⁶.

Besonders hervorzuheben sind an dieser Stelle sogenannte regressionsdiagnostische Verfahren, mit denen statistische Annahmen geprüft werden können, die bei der Varianzanalyse meist ungeprüft vorausgesetzt werden. Eine solche Annahme, mit deren Verletzung eine Verfälschung von Signifikanztests einhergehen kann, ist die Homoskedastizitätsannahme (Fahrmeir et al. 2004). Diese besagt, dass die Streuung der Reaktionszeiten für jede Faktorkombination (in Varianzanalysen) bzw. an jeder Stelle der Regressionsgeraden (in Regressionsmodellen) dieselbe sein sollte. Diese Annahme lässt sich nun mit einem signifikanten Breusch-Pagan-Test für Heteroskedastizität widerlegen (Hackl 2005), wobei die Ursache für Heteroskedastizität auf teilnehmerspezifische Varianzen der Reaktionszeiten zurückzuführen ist. Darstellen lässt sich dieses Phänomen mit einem sogenannten Residual-versus-Fitted-Plot (Hamilton 2009), bei dem die Residuen (d.h. die Differenzen zwischen der Regressionsgeraden und den Messwerten) gegen die vorhergesagten Werte geplottet werden (Abbildung 4.3). Ebenfalls als problematisch erweist sich die Normalverteilungsannahme (Fahrmeir et al. 2004), die sich im Hinblick auf die Varianzanalyse so interpretieren lässt, dass die Reaktionszeiten für alle Faktorkombinationen der Gaußschen Normalverteilungskurve folgen sollten. Geht man vom Regressionsmodell aus kann diese Annahme überprüft werden, indem die Residuen des Modells einer korrespondierenden (theoretischen) Normalverteilung gegenübergestellt werden (Abbildung 4.4). Deutlich wird hierbei, dass die Residualverteilung leicht asymmetrisch ist, was schlussendlich auf zu viele relativ hohe Reaktionszeiten (d.h. langsame Reaktionen) zurückzuführen ist. Das liegt daran, dass es sich bei der Reaktionszeit um eine Größe handelt, die einseitig beschränkt ist (d.h. eine Reaktionszeit < 0 ms ist ausgeschlossen) und nach oben hin offen¹⁷. Somit kann eine Normalverteilung nur als eine mehr oder weniger gute Annäherung

¹⁶Im Allgemeinen eignen sich Varianzanalysen nur für Experimente, in denen keine metrischen unabhängigen Variablen verwendet werden. Da im Laufe der Versuche jedoch viele metrische Störfaktoren mit aufgenommen werden, würde die Darstellung mit Varianzanalysen ausgesprochen unübersichtlich ausfallen.

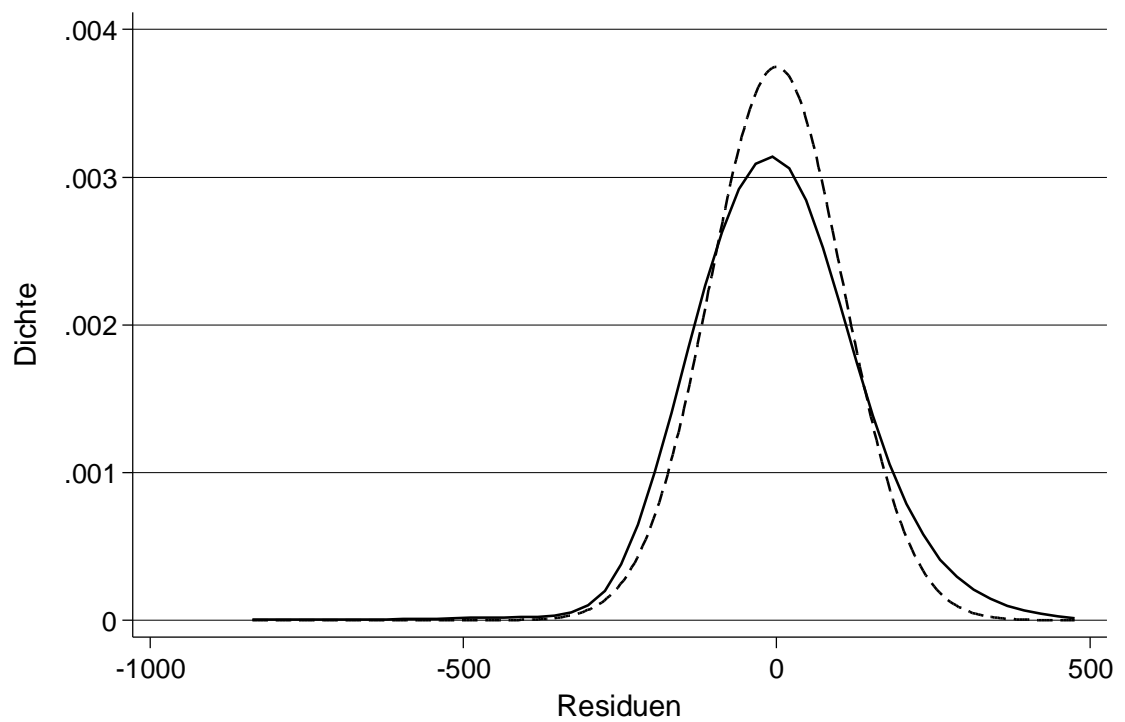
¹⁷Dabei handelt es sich um eine ganz allgemeingültige Aussage, die nicht auf die einseitige Zensur bei 1000 ms

Abbildung 4.3: Test der Homoskedastizitätsannahme – Residual-versus-Fitted-Plot



Hinweis: Aus darstellungstechnischen Gründen ist eine Beschränkung auf zwei Teilnehmer erfolgt.

Abbildung 4.4: Test der Normalverteilungsannahme – Kerndichteschätzer der Residualverteilung



Hinweis: Die durchgezogene Linie bezieht sich auf die Verteilung der Residuen, die gestrichelte Linie auf die dazugehörige theoretische Normalverteilung.

an die tatsächliche Verteilung der abhängigen Variablen gelten.

Aus den Ergebnissen der Regressionsdiagnostik lässt sich schließen, dass einfache Regressionsmodelle bzw. Varianzanalysen, wie sie gewöhnlich in der Psycholinguistik zum Einsatz kommen, für die spezifischen Fragestellungen dieser Arbeiten nur bedingt geeignet sind. Gerade weil die zu schätzenden Modelle nicht deskriptiv sind, sondern für Hypothesentests verwendet werden, kann nicht darauf verzichtet werden, eine nicht erfüllte Homoskedastizitätsannahme bereits bei der Modellierung zu berücksichtigen, da sonst eine sinnvolle Interpretation von Signifikanztests nicht möglich wäre. Dieses Problem lässt sich lösen, indem bei der Modellierung unterstellt wird, dass die Kovarianzmatrix über eine flexiblere Struktur verfügt, als dies tatsächlich der Fall ist (vgl. Fahrmeir et al. 2004). Zudem wären kleinere Varianzen der Schätzer und damit bessere Signifikanztests zu erwarten, wenn man für den Teilnehmer-Effekt eine Random-Effects-Struktur annimmt (Woolridge 2003). Das wäre insofern sinnvoll, weil es sich aufgrund der Messwiederholungen (50 Wortpaare pro Teilnehmer) trotz des experimentellen Designs um ein klassisches Paneldatenszenario¹⁸ handelt (Fahrmeir et al. 2004). Die (theoretische) Random-Effects-Struktur geht davon aus, dass alle Teilnehmer sich ähnlich verhalten, d.h. beispielsweise, dass keiner der Probanden die Ja- und Nein-Tasten verwechselt hat (was praktisch nicht ausgeschlossen werden kann)¹⁹. Grundsätzlich kann damit das Problem der Abhängigkeit von Beobachtungen einzelner Teilnehmer gelöst werden, wodurch sich ein schlichtes Experimentaldesign mit wenigen Teilnehmern und vielen Messwiederholungen mit einem aufwendigen Design mit vielen Teilnehmern und jeweils nur einer Messung (d.h. ein Wortpaar pro Teilnehmer) gleichsetzen lässt.

Darüber hinaus wäre es grundsätzlich möglich, entsprechende Modelle für die Fehlerraten (z.B. Logit-Modelle für Paneldaten) zu spezifizieren (Woolridge 2003), jedoch wird darauf im Rahmen dieser Arbeit verzichtet, da die Fehlerraten durchwegs sehr niedrig sind. Dies spricht zwar für die Qualität des Designs und der Instruktion, erlaubt aber aufgrund der geringen Varianz der abhängigen Variablen (d.h. im Fall der Fehlerrate) keine hinreichend präzise Schätzung von Effekten. Darüber hinaus sollte noch erwähnt werden, dass es sinnvoll wäre, insbesondere dann, wenn sehr langsame Reaktionen nicht aus der Analyse ausgeschlossen werden – wie es bei der eigenen Empirie der Fall ist –, bereits bei der Modellierung *keine* Normalverteilung der Residuen bzw. der abhängigen Variablen zu unterstellen. Da die für derartige Situationen geeigneten Modelle zur Ereigniszeitanalyse relativ komplex sind und

im Experiment bezogen ist.

¹⁸Bei einer Panelstudie handelt es sich um eine Form von Längsschnittstudien, bei der dieselben Personen über einen gewissen Zeitraum hinweg beobachtet werden.

¹⁹Im Übrigen lässt sich die Annahme über eine Ähnlichkeit der Verhaltensmuster auch durch entsprechende Tests überprüfen (Woolridge 2003).

zudem keinen Vergleich mit den in der Literatur üblichen Varianzanalysen erlauben, wird im Folgenden darauf verzichtet, diesen Ansatz weiterzuverfolgen.

Obwohl Abweichungen von der Normalverteilung dazu führen können, dass theoretische Verteilungen von Teststatistiken (z.B. F-Tests) nicht mehr korrekt sind, ist dies eher für kleine Stichproben problematisch. Angesichts der designbedingt relativ großen Fallzahlen²⁰ kann somit davon ausgegangen werden, dass die beobachtete Abweichung von der Normalverteilung bei den im Rahmen dieser Arbeit zu schätzenden Modellen nicht zu fehlerhaften Tests führen sollte.

Zusammengefasst ist also ein lineares Regressionsmodell mit einer Random-Effects-Struktur des Teilnehmereffekts, bei dem Clusterbildungen in der Kovarianzmatrix²¹ unterstellt werden, das Modell der Wahl, welches für die Auswertung aller in der Beschreibung noch folgenden Replikationsversuche verwendet wird. Dabei wird deutlich, dass sich numerisch gesehen nur geringfügige Unterschiede zum Ausgangsmodell (Tabelle 4.2) ergeben, die nichtsdestotrotz für die Interpretation (im Hinblick auf theoretische Vorgaben) sehr bedeutsam sein können. Dies gilt insbesondere für den Effekt der Pseudo-Phonästhem-Kondition, die im Ausgangsmodell im Gegensatz zur Phonästhem-Kondition signifikant schnellere Reaktionen als die relationslose Referenzkategorie aufweist. Trotz eines nach wie vor negativen Vorzeichens ist dies nun nicht mehr der Fall, wobei der Verlust an Signifikanz vermutlich auf das Heteroskedastizitäts-Problem im ursprünglichen Modell zurückgeführt werden kann. Deutlich wird dies auch im Marginsplot (Abbildung 4.5).

4.1.4 Diskussion

Mit diesem ersten Versuch konnte das Ergebnis von Bergen (2004), der für die Phonästhem-Kondition die signifikant kürzesten Reaktionszeiten gegenüber den anderen Relationen beobachten kann, trotz nahezu vollständiger Übernahme aller bekannten Parameter nicht repliziert werden. Weder die Phonästhem noch die Pseudo-Phonästhem unterscheiden sich signifikant von der relationslosen Referenzkategorie, während in der Bedeutungsgruppe die

²⁰Fallzahlen sind hier nicht mit Teilnehmerzahlen zu verwechseln, sondern beziehen sich auf das Produkt aus Teilnehmern und Beobachtungen. Wenn also 20 Teilnehmer auf (immer dieselben) 10 Trials, z.B. in der Phonästhem-Kondition, reagieren, beträgt die Fallzahl für diese Kondition idealerweise 200. Idealerweise, da es natürlich sein kann, dass einige Reaktionen die obere Zensur übersteigen (d.h. > 1000 ms, korrekterweise müsste man hier das „Größer-gleich“-Symbol verwenden) und deswegen für die Reaktionszeiten nicht direkt gewertet werden. Indirekt lässt sich aber sagen, dass die Reaktionen in einer Kondition mit weniger Fällen im Durchschnitt tatsächlich noch langsamer erfolgt sind. Wenn also 19 Teilnehmer innerhalb des Experiments je 50 Trials bearbeiten, würden das im Idealfall 950 Fälle ergeben. Sind es aber nur 768 Beobachtungen, bedeutet dies, dass die Differenz auf Reaktionszeiten > 1000 ms zurückgeführt werden kann. Diese wurden im Übrigen nicht für die Fehlerraten interpretiert.

²¹Gemeint sind die teilnehmerspezifischen Varianzen bei den Reaktionszeiten.

Tabelle 4.3: Einflussfaktoren auf die Reaktionszeit – Random-Effects-Modell

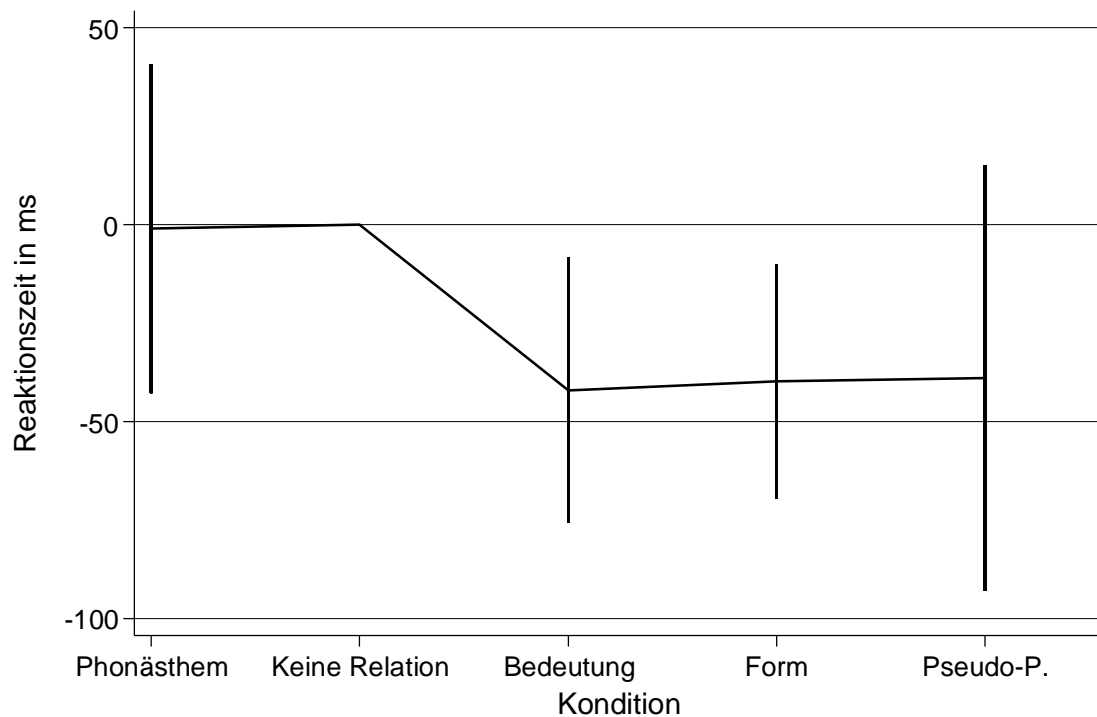
	Reaktionszeit	
Target/Pseudo-Target	-38,12*	(-2,37)
Phonästhem	-0,997	(-0,05)
Bedeutung	-42,01*	(-2,45)
Form	-39,78**	(-2,62)
Pseudo-Phonästhem	-38,94	(-1,41)
Konstante	721,3***	(48,20)
Beobachtungen	768	
R^2		

z-Werte in Klammern

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Relationseffekt geschätzt. Die Anpassungsgüte R^2 wird nicht angegeben, da sie bei Paneldatenmodellen nicht sinnvoll interpretiert werden kann.

Abbildung 4.5: Marginsplot mit 95 % Konfidenzintervall für die Effekte der Konditionen



signifikant schnellsten Reaktionen verzeichnet werden können. Ein möglicher Grund hierfür könnte in der fehlenden Randomisierung des Trialablaufs liegen. Da im Kontakt mit Bergen erst im Nachhinein in Erfahrung gebracht wurde, dass er ein randomisiertes Design verwendet hat (was im Grunde genommen auch das naheliegende ist), so ist es auf die zu diesem Zeitpunkt fehlende eigene Erfahrung mit psychologischen Experimenten zurückzuführen, dass diese leicht vermeidbare Drittvariable nicht berücksichtigt worden ist. Damit würde sich auch die relative große Differenz zwischen den Reaktionszeiten der Phonästhem-Relation und der Pseudo-Phonästhem-Relation erklären lassen: Im Experimentaldesign 1 (s. Anhang) folgen die Konditionen blockweise aufeinander, so dass die Phonästhem-Gruppe (Kondition 1) von der relationslosen Gruppe (Kondition 2) gefolgt wird, diese wiederum von der Bedeutungsgruppe (Kondition 3), diese von der Formgruppe (Kondition 4) und diese schlussendlich von der Pseudo-Phonästhem-Gruppe (Kondition 5). Man könnte nun unterstellen, dass zu Beginn des Trialablaufs die Reaktionen der Testpersonen mit der Eingewöhnung an die Aufgabenstellung systematisch langsamer ausfallen als in der späteren Phase, für die eine gesteigerte Trainings- und Aufmerksamkeitssituation angenommen werden kann. Darüber hinaus fällt auf, dass das Leistungspk (in der signifikanten mittleren Kondition 3) in der zweiten Hälfte des Experiments keinen wesentlichen Rückgang zeigt, was vermutlich daran liegt, dass die theoretische Mindestbearbeitungszeit (d.h. abzüglich des Trainings und der Pausen) der Aufgabenstellung weniger als 1 Minute betragen hat.

Insgesamt wäre festzuhalten, dass die Drittvariable *Nicht-randomisiertes Design* den interessierenden Effekt verzerrt und die Signifikanz der Bedeutungsgruppe erzeugt haben könnte. Folgerichtig sollte dieser Störfaktor im nachfolgenden Replikationsversuch kontrolliert werden.

4.2 Zweiter Replikationsversuch: Design-, Lern- und Ermüdungseffekte

4.2.1 Teilnehmer

Für den zweiten Replikationsversuch von Bergen (2004) konnte über die Internet-Plattform *Toytowngermany* ein neues Sample von 21 Englisch-Muttersprachlern angeworben werden, die unter denselben Bedingungen wie im ersten Replikationsversuch (Instruktion durch die Versuchsleitung und am Monitor, Vergütung, Arbeitsumgebung usw.) dieselbe Aufgabenstellung bearbeiten sollten (lexikalische Entscheidungsaufgabe, jeweils 50 Wortpaare in 5 Konditionen).

4.2.2 Experimentaldesign, Aufgabenstellung und Ablauf

Alle Teilnehmern wurden derselben Trial-Abfolge ausgesetzt, wobei der einzige Unterschied zum Replikationsversuch Nummer 1²² in der Randomisierung der Konditionen bestand, um so den weiter oben diskutierten potentiellen Trainingseffekt (bestehend aus Anfangsträgheit und Leistungspeak) auf alle Konditionen zu verteilen. Das heißt, dass wenn in Block 1 (zehn Trials, gefolgt von der Bildschirminstruktion eine Pause einzulegen) eine allgemeine Anfangsträgheit der Probanden die Reaktionszeiten beeinflusst, durch eine Mischung der Konditionen ein möglicher Störeffekt dieser Größe für alle Konditionen unterstellt werden kann.

4.2.3 Ergebnisse

Tabelle 4.4 zeigt, dass bei Pseudo-Worten (wie im ersten Replikationsversuch) und in der Pseudo-Phonästhem-Kondition die meisten Fehler gemacht werden. Anders als im ersten Replikationsversuch sind jetzt die Reaktionen der Teilnehmer in der Phonästhem-Kondition am schnellsten. Wie bereits beim ersten Experiment wird auch hier ein lineares Regressionsmodell

Tabelle 4.4: Fehlerraten, Fallzahlen, durchschnittliche Reaktionszeiten und Standardabweichungen nach Konditionen

Kondition	Fehlerrate	N (richtig)	MW Reaktionszeit	Standardabw.
Pseudo-Target	0,917	441	717,699	115,590
Phonästhem	0,990	103	627,602	141,234
Keine Relation	0,980	100	646,581	112,959
Bedeutung	0,980	98	645,109	131,477
Form	0,990	103	632,647	131,050
Pseudo-Phonästhem	0,788	78	647,437	120,104

Hinweis: Bei den Konditionen wurden nur die Reaktionszeiten der richtigen Antworten für die Berechnung der Reaktionszeit in ms verwendet.

mit einer Random-Effects-Struktur des Teilnehmereffekts, bei dem Clusterbildungen in der Kovarianzmatrix (d.h. teilnehmerspezifische Varianzen der Reaktionszeiten) unterstellt werden, geschätzt (Tabelle 4.5). Durch den Verzicht auf die Abfolge der Konditionen in separaten Blöcken mit jeweils 10 Wortpaaren, ist es nun aber auch möglich, den Einfluss zweier potentieller Störfaktoren, die im vorhergehenden Abschnitt bereits angesprochen worden sind,

²²Auch hier setzten sich wie in Replikationsversuch Nummer 1 für jeden Teilnehmer die Wort-Trials aus 5 Phonästhem-, 5 Pseudo-Phonästhem-, 5 Bedeutungs-, 5 Form- und 5 relationslosen Prime-Target-Paaren zusammen, die um 25 Pseudowort-Trials ergänzt worden sind.

Tabelle 4.5: Einflussfaktoren auf die Reaktionszeit – Random-Effects-Modell

	Reaktionszeit	
Target/Pseudo-Target	-69,15***	(-5,67)
Trialnummer	1,771*	(1,97)
Trialnummer (quad.)	-0,0351*	(-2,17)
Phonästhem	-20,09	(-1,21)
Bedeutung	-9,596	(-0,69)
Form	-13,27	(-0,93)
Pseudo-Phonästhem	2,143	(0,14)
Konstante	702,8***	(45,80)
Beobachtungen	923	

z-Werte in Klammern

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Relationseffekt geschätzt. Die Anpassungsgüte R^2 kann bei Paneldatenmodellen nicht sinnvoll interpretiert werden und wird deshalb nicht berichtet.

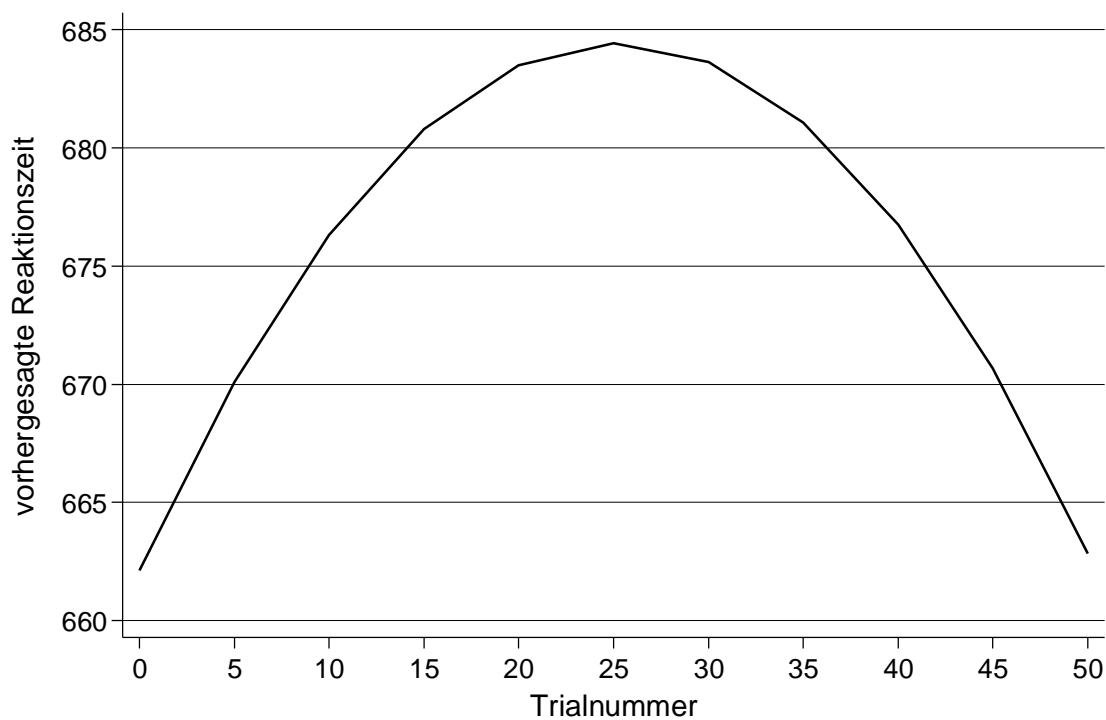
zu untersuchen, nämlich von Lern- und Ermüdungseffekten. Nach Vorbereitung der Teilnehmer durch die Trainingstrials (10 Trials) wäre es auch denkbar anstelle von einer Anfangsträgheit, wie bei Replikationsversuch Nummer 1 vermutet, von maximaler Aufmerksamkeit auszugehen, d.h. den Leistungsspeak für die Anfangsphase zu unterstellen und eine abfallende Leistungskurve (Ermüdungseffekt) zum Schluss des Experiments hin. Es stehen sich hier also zwei gegenläufige Faktoren gegenüber: a) die im Zeitverlauf (zunächst²³) abnehmende Anfangsträgheit, die man auch umgekehrt als Lerneffekt auffassen kann, für den aufgrund der zunehmenden Vertrautheit mit der Aufgabe eine Leistungsverbesserung zum Experimentende hin angenommen werden könnte und b) ein Ermüdungseffekt.

Veranschaulichen lässt sich dies beispielsweise mit der Akkordarbeit an der Spülmaschine einer Großküche: Nimmt im Laufe der Zeit mit der Routinisierung der Arbeitsbläufe (Essensreste vom Tablett nehmen und in den Müll werfen, Besteck in den Besteckkorb stellen, Teller und Tablett auf das Fließband sortieren) der Arbeitsfluss und damit die Schnelligkeit zu, so lässt sich diese Leistung aus physikalischen Gründen (u.a. aufgrund des Zweiten Hauptsatzes der Thermodynamik) weder beliebig steigern noch dauerhaft aufrecht erhalten (d.h. der Spüler wird unvermeidlich erschöpft).

²³Es kann davon ausgegangen werden, dass sich dieser Zustand nur eine begrenzte Zeit lang aufrecht erhalten lässt.

Um beide Effekte zu berücksichtigen erscheint es somit sinnvoll, einen linearen und einen quadratischen Effekt der Trialnummer in das Modell aufzunehmen. Beide Effekte sind signifikant und können in ihrem Verlauf mit einem Conditional-Effects-Plot (Hamilton 2009) dargestellt werden (Abbildung 4.6) Sofern es die Unterschiede zwischen den Konditionen betrifft,

Abbildung 4.6: Conditional-Effects-Plot für den Lerneffekt und den Ermüdungseffekt



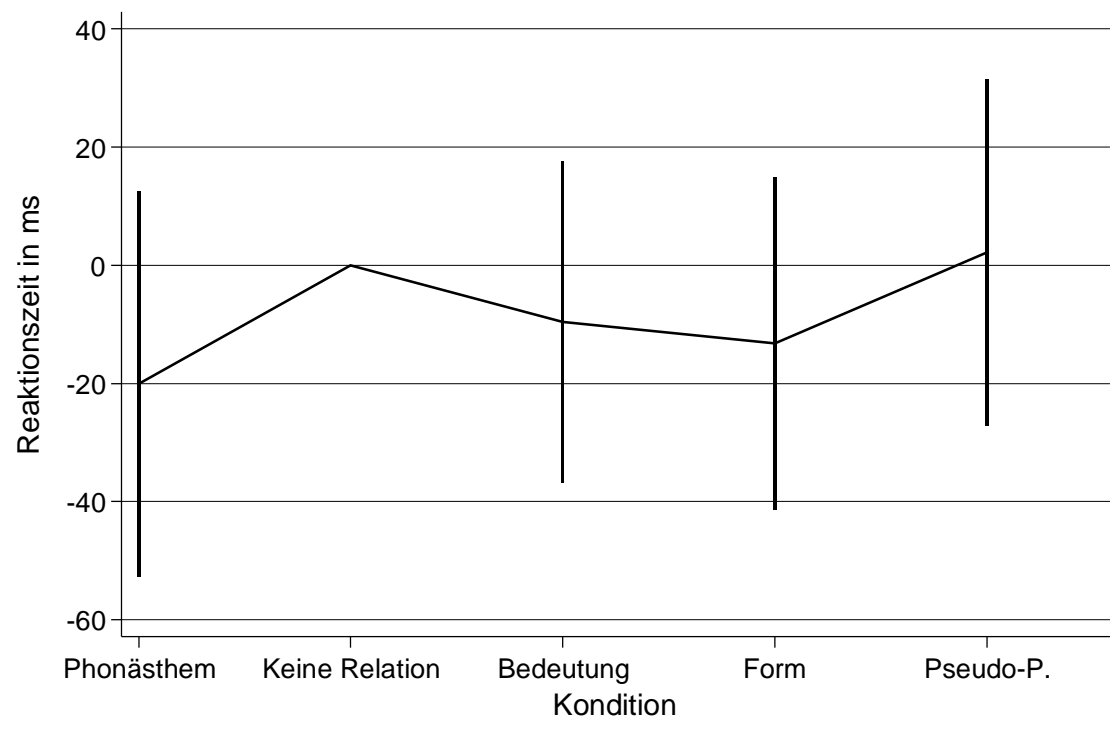
Hinweis: Die vorhergesagten Reaktionszeiten werden berechnet, indem die jeweiligen Trialnummern zusammen mit den tatsächlichen Werten für die übrigen Variablen in die Regressionsgleichung eingesetzt werden. Anschließend werden die berechneten Werte über alle Beobachtungen gemittelt.

lassen sich diesmal weder ein Relationseffekt noch Unterschiede zwischen den Konditionen nachweisen. Dargestellt werden kann dieses Ergebnis mit einem Marginsplot (Abbildung 4.7).

4.2.4 Diskussion

Da im Replikationsversuch Nummer 2 weder ein signifikanter Relationseffekt noch Effekte zwischen den relationsbehafteten Konditionen gemessen werden können, lässt sich unmittelbar darauf schließen, dass die im ersten Replikationsversuch berichteten Effekte vollständig

Abbildung 4.7: Marginsplot mit 95 % Konfidenzintervall für die Effekte der Konditionen



von dem nicht-randomisierten Design erzeugt werden²⁴.

Die schlechtere Leistung bei den Phonästhemen im Vergleich zu den Pseudo-Phonästhemen – denn es hätte eigentlich aufgrund der niedrigeren Type-Frequenz der Pseudo-Phonästhematome umgekehrt sein müssen – hat in der deskriptiven Statistik des ersten Replikationsversuchs die Vermutung motiviert, dass dies auf einen positiven Lerneffekt für Pseudo-Phonästhematome, die im nicht-randomisierten Design zum Experimentende hin präsentiert worden sind, zurückzuführen ist. Obgleich ein solcher Lerneffekt mit dem zweiten Replikationsversuch bestätigt werden kann, bleibt der Widerspruch bestehen: Mit einem signifikanten Lern- und einem signifikanten Ermüdungseffekt ergibt sich ein Szenario, bei dem zu Beginn (fehlende Ermüdung) und zum Schluss (Lerneffekt) des Experiments ein Leistungshoch zu verzeichnen ist. Zwar profitieren die Pseudo-Phonästhematome in Replikationsversuch Nummer 1 von ihrer Endpositionierung, allerdings gilt dasselbe für die Phonästhematome, die durch ihre Anfangspositionierung vom fehlenden Ermüdungseffekt profitieren.

Interpretiert man die Rangfolge der Reaktionszeiten in der deskriptiven Statistik des ersten Replikationsversuchs unter Berücksichtigung dieser Drittvariablen, würde das bedeuten, dass die Phonästhematome-Kondition, die aufgrund der hier noch fehlenden Randomisierung der Konditionen den Teilnehmern als erster Block präsentiert wurde unter optimalen Bedingungen die längsten Reaktionszeiten erzielt, während die Bedeutungsgruppe unter schlechtesten Bedingungen (da sie als Trialblock mittig positioniert ist) mit den kürzesten Reaktionszeiten einhergeht und tatsächlich noch schnellere Reaktionen hervorruft als es rein numerisch betrachtet den Anschein macht.

Auch wenn die nicht-randomisierte Trialabfolge zu größeren Verzerrungen führt als die Drittvariablen Ermüdungs- und Lerneffekte, so sind letztere trotzdem nicht zu vernachlässigen. Denn wie in der Diskussion des ersten Replikationsversuchs bereits angesprochen worden ist, lassen sie sich durch Randomisierung zwar besser verteilen, jedoch nicht einfach herauskürzen. Da die Randomisierung keine gleichmäßige Verteilung der Konditionen auf die Trialblöcke gewährleistet, kann es nämlich immer noch sein, dass sich im letzten Block 3 Trials der Kondition 1 und nur 1 Trial der Kondition 3 befinden. Da Bergen (2004) keine Angaben zu diesen beiden Variablen macht (ebensowenig wie zu den Fehlerraten), lässt sich nicht beurteilen, ob der von ihm behauptete signifikante Effekt der Phonästhematome-Kondition nicht etwa auf eine günstige Konstellation von hochfrequenten Stimuli-Set-Exemplaren²⁵ und einer zufälligen Verdichtung phonästhematischer Trials an den Anfangs- und Endrändern des Designs,

²⁴Das heißt, dass die Unterscheidung der Konditionen durch deren blockweise Abfolge erleichtert wird.

²⁵An dieser Stelle soll daran erinnert werden, dass die Wahrscheinlichkeit mit Bergens (2004) Original-Set zu arbeiten nur 25²⁵ beträgt, da er zwar sein gesamtes Set dokumentiert, seine konkrete Trialauswahl jedoch nicht.

auf die trotz Randomisierung hin kontrolliert werden sollte, zurückzuführen ist.

Da trotz fehlender signifikanter Effekte²⁶ der Konditionen im zweiten Replikationsversuch zu beobachten ist, dass die Phonästhem-Kondition tendentiell die kürzesten Reaktionszeiten erzielt, so liegt es nahe eine gezielte Modifikation (u.a. eine Fallzahlerweiterung) des Urexperiments von Bergen (2004) in Erwägung zu ziehen. Da die Replikationsversuche Nummer 1 und 2 aufgrund der Datenlage nicht hätten durchgeführt werden können, ohne die genannten Abweichungen zum Original hinzunehmen, konnte ein Replikationsversuch, der frei von Modifikationen ist, von Anfang an nicht realisiert werden. Weitere Abwandlungen sollen dabei nicht in erster Linie dazu dienen, die Gültigkeit von Bergens (2004) Befunden in Frage zu stellen, sondern vielmehr dazu sich ihnen anzunähern²⁷. Ausdrücklich abzugrenzen ist dieses Vorgehen vom Verifikationsversuch, bei dem es darum geht Beweise für eine favorisierte Hypothese zu erbringen. Es geht also darum, die interessierenden Effekte schrittweise von Störvariablen zu befreien, von denen unter H₀ angenommen wird, dass sie erstere überdecken. Das heißt, dass der Beweis einer psychologische Realität von Phonästhemem nicht dadurch erzwungen werden soll, dass so lange Replikationsversuche durchgeführt werden, bis der gewünschte Effekt zu verzeichnen ist. Vielmehr soll ein solcher (vorläufiger) Beweis indirekt²⁸ erbracht werden, und zwar indem unvoreingenommen versucht wird zu zeigen, dass naheliegende Störfaktoren die beobachteten Zusammenhänge *nicht* hervorgebracht haben können.

Dass sich Bergens Resultate auch mit diesem Versuch nicht replizieren ließen, kann zum einen an den unumgänglichen Abweichungen zu seinem Urversuch liegen²⁹, es kann aber auch daran liegen, dass Bergen (2004) noch gravierendere Fehler unterlaufen sind als bereits im Abschnitt zu den Schwachstellen seiner Arbeit angesprochen worden ist. Und auch wenn das der Fall sein sollte, muss dies noch nicht bedeuten, dass die interessierenden Effekte nicht messbar sind. Es geht im Folgenden also nicht darum, Bergen weitere Fehler nachzuweisen, sondern im Sinn der Sache selbständig zu überprüfen, *unter welchen Bedingungen phonästhemische Effekte – falls überhaupt – messbar sind*.

Mit dem zweiten Replikationsversuch konnte demnach gezeigt werden, dass a) das nicht-

²⁶Aus dem Fehlen von Effekten können sich natürlich ebenso wertvolle Aussagen ableiten lassen wie aus dem Vorhandensein von Effekten.

²⁷Damit wird zunächst von der Validität der Beobachtungen Bergens (2004) ausgegangen.

²⁸Vergleiche hierzu die wissenschaftstheoretischen Ausführungen in Abschnitt 3.1.4.

²⁹Vergleiche hierzu die Ausführungen in Abschnitt 4.1. Zu den dort genannten zwingenden Abweichungen könnte auch noch ein systematischer Unterschied zwischen Bergens (2004) Teilnehmern und den für die Replikationsversuche Nummer 1 und 2 rekrutierten englischsprachigen Personen bestehen. Obgleich alle Probanden einen gebildeten Eindruck gemacht haben und im Gespräch erkennen ließen, dass sie vorwiegend in der Muttersprache kommunizieren, können Selektionseffekte nicht ausgeschlossen werden. Beispielsweise hätte es sich bei Bergens Teilnehmern durchwegs um Linguistikstudenten handeln können, deren mentales Lexikon sich von Nicht-Linguisten unterscheiden könnte. Bergen hat leider keine genauen Angaben zur Zusammensetzung seiner Stichprobe gemacht.

randomisierte Design als Störgröße tatsächlich nicht zu unterschätzen ist und dass b) Lern- und Ermüdungseffekte unter ungünstigen Zusatzbedingungen die interessierenden Effekte überdecken können.

Darüber hinaus fungiert das anfangs als vermeidbar gehandelte nicht-randomisierte Design hervorragend als Umkehrprobe für die Zuverlässigkeit der statistischen Analyse: Sind im randomisierten Experiment keine Effekte zwischen den Konditionen messbar (wie in Replikationsversuch Nummer 2), sollten sie in der nicht-randomisierten Version dieses Experiments (Replikationsversuch Nummer 1) messbar sein. Dass Umkehrproben in der psycholinguistischen Experimentalforschung anders als in den angewandten Naturwissenschaften (wie z.B. der Geodäsie) unüblich sind, ist ausgesprochen bedauerlich, zumal sie dazu beitragen können die Verlässlichkeit von Forschungsergebnissen nicht unwesentlich zu erhöhen.

Wie weiter oben bereits angesprochen, wurde im nächsten Schritt eine gezielte Modifikation des Urexperiments vorgenommen. Zum einen wurde im Hinblick auf die aus der deskriptiven Statistik hervorgehenden Tendenzen die Fallzahl drastisch erhöht und zum anderen wurde der Einfluss der vielleicht bedeutendsten Drittvariablen für sprachbasierte psycholinguistische Primingexperimente unter die Lupe genommen, nämlich den der Vorkommenshäufigkeit lexikalischer Stimuli (Token-Frequency). Während der Einfluss von Type-Frequency bei Bergen einen interessierenden Faktor darstellt, der dazu dient Phonästhem von Pseudo-Phonästhem zu unterscheiden, lässt ein Blick auf die Token-Frequenzen seiner Stimuli-Sets im Thorndike-Lorge-Korpus (und auch in den ergänzenden Korpora in Abschnitt 3.2) einen gleichgerichteten Zusammenhang zwischen Reaktionszeiten und Frequenzeffekten vermuten. Ob es bei einem augenscheinlichen Effekt bleibt oder nicht, soll mit Versuch Nummer 3 überprüft werden.

4.3 Dritter Replikationsversuch mit systematischer Verzerrung der Token-Frequenz

4.3.1 Teilnehmer

Die Idee, Bergens (2004) Design auf das Deutsche zu übertragen, war zunächst von rein forschungspragmatischen Überlegungen³⁰ geprägt. Da angesichts der Terminkoordination und Raumsituation die Durchführung der ersten beiden Replikationsversuche bereits mit erheblichen Schwierigkeiten verbunden gewesen war, wurde von einer Erhöhung der Fallzahl durch

³⁰Deutsche Probanden waren in dem Fall leichter zu rekrutieren.

eine Sampleerweiterung abgesehen und stattdessen die Trialzahl verfünffacht³¹. Die Testpersonen setzen sich aus 18 monolingualen Studenten mit der Muttersprache Deutsch zusammen, die über Aushänge auf dem Campus der Ludwig-Maximilians-Universität München angeworben worden sind. Für die etwa 10-minütige Teilnahme erhielten sie jeweils 10 Euro Aufwandsentschädigung. Die Fachzugehörigkeit rangierte dabei von Lehramtsstudenten bis hin zum Maschinenbau-Studenten, wobei die sprachenbezogene Fachzugehörigkeit unter den Teilnehmern dominiert hat.

4.3.2 Experimentaldesign, Aufgabenstellung und Ablauf

Da für die Überprüfung des Effekts der lexikalischen Target-Token-Häufigkeit die Unterscheidung zwischen Phonästhemen und Pseudo-Phonästhemen nicht von Bedeutung ist, wurde die Pseudo-PÄ-Kondition nicht in das deutsche Design integriert (d.h. es verbleiben 4 Relationen à 50 Trials, davon die Hälfte lexikalische Targets, die andere Hälfte lexikalische Lücken). Darüber hinaus hat das den Vorteil, dass die Bearbeitungszeit und damit auch der Einfluss des Ermüdungseffekts reduziert wird, von der angenommen werden kann, dass er mit steigender Bearbeitungszeit zunimmt.

Bei der Zusammenstellung der Trials wurde darauf geachtet, dass möglichst alle Target-Onsets aus Konsonantenklustern bestehen, um den Einfluss der potentiellen Drittvariable Phonäthem-Typ gering zu halten. Dieser Vorsatz konnte nur so weit realisiert werden wie es die Grenzen des deutschen Onset-Phonäthem-Inventars zugelassen haben. Mit dessen Erschöpfung musste folglich auf Reim-Phonätheme zurückgegriffen werden. Dabei wurde darauf geachtet den Reim-Phonäthem-Targets Primes mit monokonsonantischen Onsets gegenüberzustellen, z.B. *sIRRen* und *gURRen*, und entsprechend viele monokonsonantische Targets auf alle anderen Relationen zu verteilen (wie z.B. das Trial *weinen* – *heulen* in der Bedeutungsrelation).

Um den inhibierenden Effekt der phonologischen Nachbarn gering zu halten, wurden Primes und Targets durchgehend so angeordnet, dass die Primes eine höhere Frequenz aufweisen als die Targets. Wie bereits erwähnt worden ist, hat Feldman (2000) diesen aus Grainger

³¹Dieser Faktor ist aus der persönlichen telefonischen Kommunikation mit Prof. Kotz vom Max Planck Institut für Kognitions- und Neurowissenschaften in Leipzig hervorgegangen und lässt sich mit Erfahrungswerten begründen. So hat Feldman (2000) bei 95-96 und 120 Testpersonen Signifikanzen zu verzeichnen, Rastle et al. (2000) arbeiten mit 72 Teilnehmern, McCormick et al. (2009) mit 60. Außerdem wurde zu diesem Zeitpunkt eine Zusammenarbeit an einer kombinierten Priming-EEG-Studie zur Messung der N400/P300-Komponenten als Anhaltspunkt für den Einfluss der Token-Frequenz ins Auge gefasst. Hierfür wäre eine Stichprobenerweiterung auch aus methodisch Gründen unumgänglich gewesen. Die Idee zu dieser Kooperation wurde verworfen, nachdem sich eine kostengünstigere Alternative gefunden hat, auf die weiter unten noch näher eingegangen wird.

(1990) hervorgehenden Störfaktor genau anders herum umgesetzt. Bei der Erkennung niedrigerfrequenter Primes werden laut der Theorie von Grainger (1990) höherfrequente phonologische Nachbarn dieser Primes mitaufgerufen und müssen unterdrückt werden. Problematisch dabei ist, dass auch das (höherfrequente) Target mit unterdrückt wird, wodurch es bei der Verarbeitung des Targets zu Verzögerungen kommt. Für die Phonästhem-Gruppe wurden Verben folgender Phonästhem- und Phonästhemkandidaten³² eingesetzt: SCHN- (z.B. *schneuzen* – *schniefen*), GL- (z.B. *glitzern* – *glimmen*), SCHM- (z.B. *schmecken* – *schmatzen*), SCHL- (z.B. *schlucken* – *schlemmen*), -[V]TZ- (z.B. *klotzen* – *protzen*), -TSCH- (z.B. *quietschen* – *plantschen*), KN- (z.B. *knoten* – *knüllen*), SPR- (z.B. *sprudeln* – *sprießen*), ST- (z.B. *starren* – *stieren*), -[V]DDEL- (z.B. *paddeln* – *buddeln*), -[V]BBEL- (z.B. *krabbeln* – *brabbeln*), FL- (z.B. *fliegen* – *flattern*), -[V]RR- (z.B. *schnurren* – *knurren*). Die Pseudo-Targets wurde durch Ersetzen des Vokals aus lexikalisierten Targets abgeleitet, z.B. *pflücken* – **rapfen* in Kondition 3 aus *pflücken* – *rupfen*, *nehmen* – **frenen* in Kondition 2A³³) aus *nehmen* – *freuen*, und für die Phonästhem-Gruppe (Kondition 1) beispielsweise *knarren* – **schmirren* aus *knarren* – *schnarren*. Wie bei dem zweiten Replikationsversuch war die Abfolge der Wortpaare für alle Teilnehmern identisch und erfolgte in zufälliger Reihenfolge.

Abschließend wurde die relationslose Kondition (2) im Hinblick auf Target-Token-Frequenz in zwei Gruppen aufgeteilt: in der Testgruppe A (9 Teilnehmer) enthielt diese Kondition hochfrequente Verb-Targets wie *brauchen*, *stehen*, während für Gruppe B (9 Teilnehmer) in dieser Kondition niedrigfrequente Targets wie *pfänden*, *gröhlen* eingesetzt wurden³⁴. Die Token-Frequenzwerte wurden aus dem COSMAS II-Korpus des IDS Mannheim erhoben (COSMAS II 1991-2010). Keiner der Teilnehmer wusste von der Gruppenaufteilung, wobei die Zuordnung der einzelnen Teilnehmer zu einer der beiden Gruppen, der Reihe nach abwechselnd, von der Versuchsleitung am PC erfolgt ist, noch bevor der jeweilige Teilnehmer den Raum betrat.

³²Phonästhemkandidaten unterscheiden sich von Phonästhem- dadurch, dass sie mit dem in Abschnitt 2.1.2 beschriebenen Quantifizierungsverfahren nicht bewertet worden sind. Um die Stichprobenerweiterung zu gewährleisten, konnte auf die Kandidaten leider nicht verzichtet werden.

³³Das „A“ wird im nächsten Absatz erklärt.

³⁴Bei der Anordnung der lexikalischen Stimuli ist es zu vermeidbaren, aber unschädlichen Fehlern gekommen: das Target *stutzen* taucht im Set von Gruppe A und B jeweils doppelt auf (*stemmen* – *stutzen*) (Gruppe 4) und (*kürzen* – *stutzen*) (Gruppe 3). Auch war die Trialzusammensetzung in Gruppe A und B nicht ganz ausgewogen. Während das Design für Gruppe A 26 Form-Paare enthielt, waren es in Gruppe B 27 Paare. Für Gruppe A gab es pro Teilnehmer insgesamt 95 Pseudoworte, 105 Worte, davon 26 Bedeutungs-, 26 Form-, 27 Phonästhem- und 26 relationslose Trials; in Gruppe B erhielt jeder Teilnehmer 95 Pseudoworte, 106 Worte, davon 26 Bedeutungs-, 27 Form-, 27 Phonästhem- und 26 relationslose Trials. Da diese Asymmetrien bei der statistischen Analyse sofort auffielen und berücksichtigt wurden, konnte eine Beeinträchtigung der Zuverlässigkeit des Gesamtergebnisses verhindert werden. Darüber hinaus hat sich bei näherer Betrachtung des Wiederholungswerts in Übereinstimmung mit den Erfahrungen aus Studien zu repetitiven Primingeffekten (u.a. Ellis & Ellis 1998) gezeigt, dass sich bei der Wiederholung die Reaktionszeiten verkürzen und die Fehleranfälligkeit sinkt.

Die Versuchsbedingungen waren dieselben wie bei den Replikationsversuchen Nummer 1 und 2. Zu diesem Zweck waren die englischen Instruktionstexte (Versuchsbeschreibung, Pausenhinweise, Verabschiedung) der ersten beiden Versuche vorab ins Deutsche übersetzt worden.

4.3.3 Ergebnisse

Deutlich wird in Tabelle 4.6, dass in den Konditionen „Bedeutung“ und „Form“ die wenigsten Fehler gemacht werden. Anders als bei den vorangegangenen Experimenten kann die relationslose Kondition (2) die schnellsten Reaktionen verzeichnen, was auf die ausgesprochen kurzen Reaktionszeiten in der Gruppe A zurückzuführen ist, für die in dieser Kondition systematisch hochfrequente Targets eingebaut worden sind. Wie bei Replikationsversuch Nummer

Tabelle 4.6: Fehlerraten, Fallzahlen, durchschnittliche Reaktionszeiten und Standardabweichungen nach Konditionen

Kondition	Fehlerrate	N (richtig)	MW Reaktionszeit	Standardabw.
Pseudo-Target	0,920	1297	762,358	112,167
Keine Relation	0,913	389	665,725	131,623
Keine Relation A	0,974	223	606,685	107,420
Keine Relation B	0,843	166	745,037	119,057
Phonästhem	0,870	380	711,474	130,382
Form	0,930	396	708,830	122,977
Bedeutung	0,944	404	712,455	129,389

Hinweis: Bei den Konditionen wurden nur die Reaktionszeiten der richtigen Antworten für die Berechnung der Reaktionszeit in ms verwendet.

2 wird auch hier ein lineares Regressionsmodell mit einer Random-Effects-Struktur des Teilnehmereffekts, bei dem Clusterbildungen in der Kovarianzmatrix (teilnehmerspezifische Varianzen der Reaktionszeiten) unterstellt werden, geschätzt (Tabelle 4.7). Die sich durch die Targethäufigkeit systematisch unterscheidenden Trials der Kondition 2 („keine Relation“) wurden bei der Modellierung dahingehend berücksichtigt, dass ein Interaktionseffekt für die Kondition „keine Relation“ in Gruppe A (hochfrequente Targets) spezifiziert wurde. Gemessen am z-Wert ist dieser Effekt der stärkste Effekt im gesamten Modell, so dass eine höhere Vorkommenshäufigkeit der Targets tatsächlich zu schnelleren durchschnittlichen Reaktionszeiten zu führen scheint. Durch den Interaktionseffekt sind die einzelnen Konditionen jetzt im Hinblick auf die Referenzkategorie „keine Relation“ in Gruppe B und Pseudo-Targets zu interpretieren.

Tabelle 4.7: Einflussfaktoren auf die Reaktionszeit – Random-Effects-Modell

	Reaktionszeit	
Target/Pseudo-Target	-26,42*	(-2,14)
Trialnummer	-0,328	(-1,24)
Trialnummer (quad.)	0,000976	(0,89)
keine Relation (a)	-129,1***	(-9,85)
Phonästhem	-29,17**	(-2,96)
Bedeutung	-27,80*	(-2,23)
Form	-30,84**	(-3,17)
Konstante	787,7***	(44,83)
Beobachtungen	2866	

z-Werte in Klammern

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Relationseffekt geschätzt. Die Anpassungsgüte R^2 kann bei Paneldatenmodellen nicht sinnvoll interpretiert werden und wird deshalb nicht berichtet.

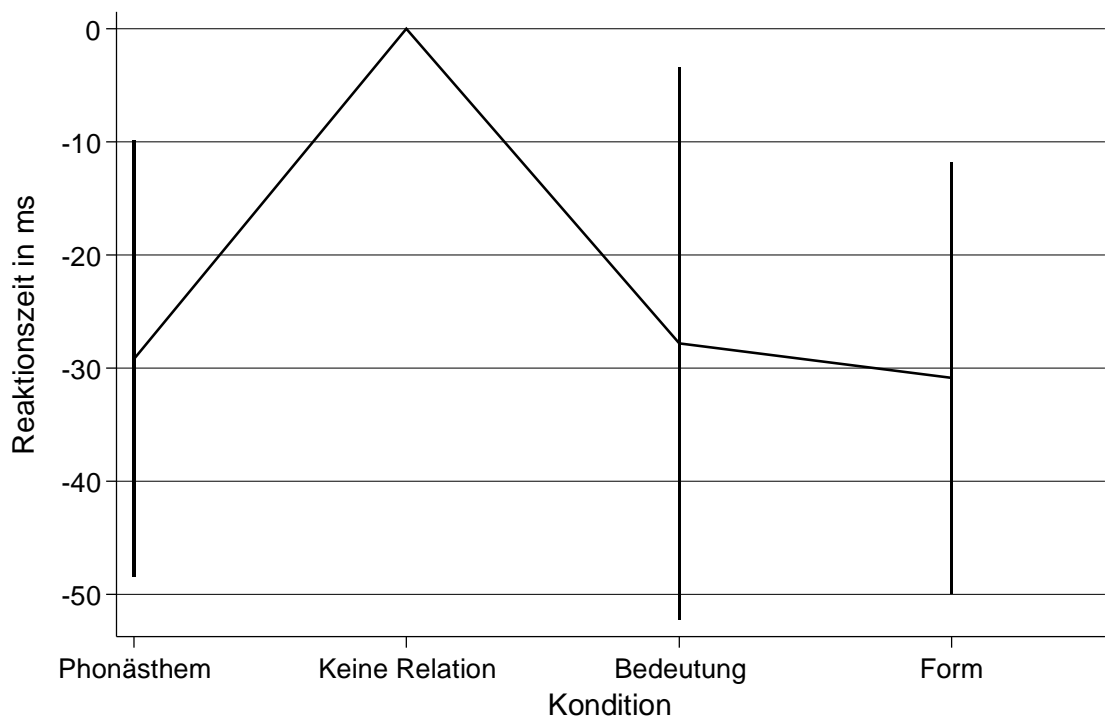
Bei Spezifikation des Interaktionseffekts für die Kondition 2A sind die Konditionseffekte alle signifikant größer als in der Referenzkategorie 2B. Darüber hinaus lassen sich keine Unterschiede zwischen den Konditionen nachweisen (Abbildung 4.8). Obwohl der lineare und der quadrierte Effekt der Trialnummer hier nicht signifikant sind und somit keinen Hinweis auf Lern- oder Ermüdungseffekte liefern, wurde auf die Aufnahme dieser Effekte in das Modell nicht verzichtet, da bereits im vorangegangenen Experiment gezeigt werden konnte, dass es sich hierbei um durchaus relevante Störfaktoren handelt, deren Vernachlässigung selbst bei nicht-signifikanten Effekten zu einer Verzerrung der übrigen Schätzkoeffizienten führen kann.

4.3.4 Diskussion

Obgleich die Effekte für die relationsbehafteten Konditionen bei Spezifikation des Interaktionseffekts für die Kondition 2A alle signifikant größer sind als für die Trials in der Referenzkategorie (2B), muss man dabei beachten, dass bei einer umgekehrten Spezifikation des Interaktionseffekt für Kondition 2B (niedrigfrequente Targets) mit einem gegenläufigen Konditionseffekt hinsichtlich der Referenzkategorie 2A zu rechnen wäre³⁵. Zentral für diesen

³⁵Das heißt, dass der Konditionseffekt von der Vergleichsgruppe abhängt: Bewertet man diesen Effekt im Hinblick auf die relationslose Kondition 2B, fällt er groß aus, da die Targets in 2B besonders niedrigfrequent sind; bewertet man ihn dagegen im Hinblick auf Gruppe 2A, so fällt er klein aus, da die Targets in dieser

Abbildung 4.8: Marginsplot mit 95 % Konfidenzintervall für die Effekte der Konditionen



dritten, modifizierten Replikationsversuch ist deswegen das Fehlen von signifikanten Unterschieden zwischen den relationsbehafteten Konditionen.

Das Ausbleiben eines signifikanten Ermüdungseffekts trotz verfünffacher Trialzahl scheint von den Teilnehmern durch die Pausen effizienter kompensiert worden zu sein als es bei den englischsprachigen Teilnehmern der Fall war³⁶.

Der aus Bergen (2004) hervorgehende signifikante Effekt der Phonästhem-Kategorie konnte auch bei einer Verfünffachung der Fallzahl unter Berücksichtigung von Lern- und Ermüdungseffekten nicht repliziert werden. Stattdessen kann gezeigt werden, dass es sich bei der Target-Token-Frequenz um eine zentrale Drittvariable handelt, von der angenommen werden kann, dass sie nicht nur für die vorliegende Experimentalreihe, sondern für alle wortbasierten Primingstudien von besonderer Bedeutung ist.

An dieser Stelle könnte man einwenden, dass der Begriff Target-Token-Frequenz die in Abschnitt 3.2.3 eingeführte Differenzierung umgeht. Token-Frequenz kann nämlich einerseits als Verarbeitungshäufigkeit aufgefasst werden, andererseits aber auch als Vorkommenshäufigkeit. Dabei ist erstere der Mikroebene und letztere der Makroebene zuzuordnen.

Wenn man nun behaupten möchte, dass sich die aus dem COSMAS II-Korpus erhobenen Frequenzwerte der Vorkommenshäufigkeit der Targets tatsächlich auf die Gebrauchshäufigkeit dieser Targets bei den einzelnen Teilnehmern auswirkt, so muss erst in einem separaten Argumentationsschritt untersucht werden, ob eine solche Übertragbarkeit überhaupt gegeben ist. Intuitiv mag es abwegig erscheinen anzunehmen, dass Sprecher die relative Vorkommenshäufigkeit von Wörtern hinreichend genau schätzen können, andererseits wäre es falsch eine Unabhängigkeit von Mikro- und Makrophänomenen zu unterstellen. Um dieses Problem zu lösen, ist im Folgenden ein Rankingverfahren durchgeführt worden.

4.4 Ranking lexikalischer Zielreize

4.4.1 Teilnehmer

Die Testpersonen setzen sich aus 38 monolingualen Studenten mit der Muttersprache Deutsch zusammen, die über Aushänge auf dem Campus der Ludwig-Maximilians-Universität Mün-

Gruppe hochfrequent sind.

³⁶An dieser Stelle ließe sich spekulieren, dass es den größtenteils universitätsexternen Teilnehmern der ersten beiden Replikationsversuche nicht so sehr um die Sache gegangen ist als um die Gelegenheit in nur wenigen Minuten und mit geringem Aufwand 10 Euro zu verdienen. Damit sollen externe Teilnehmer nicht diskriminiert werden, denn solche Beweggründe können auch für vereinzelte studentische Probanden nicht ausgeschlossen werden; es geht schlichtweg darum Vermutungen anzustellen, mit deren Hilfe das Auftreten bzw. Ausbleiben dieser Effekte erklärt werden könnte.

chen angeworben worden sind. Für ihre Teilnahme erhielten sie jeweils 10 Euro Aufwandsentschädigung.

4.4.2 Experimentaldesign, Aufgabenstellung und Ablauf

Die Idee zu diesem für die Psycholinguistik neuen Verfahren stammt aus einem Klassiker der empirischen Sozialforschung von Andreas Diekmann (2001):

„Das Ranking-Verfahren sieht[...] vor, daß die Befragten die einzelnen Themen nach ihrer Wichtigkeit in eine Reihenfolge bringen. So werden die Themen auf Kärtchen geschrieben, die der Befragte dann nach der Wichtigkeit sortieren soll.“
(Diekmann 2001:393)

Dass es sich beim Ranking um ein in der Linguistik unerprobtes Verfahren handelt, mag zum einen daran liegen, dass die Lösung linguistischer Fragestellungen mit den Methoden der experimentellen Psychologie einen Aufwand darstellt, den man nicht auch noch durch die Hinzunahme weiterer fachfremder Methoden erhöhen möchte; zum anderen liegt die Vermutung angesichts der Ausführungen in Abschnitt 3.1.2 nahe, dass die Schlichtheit eines solchen Experiments von dessen Aussagepotential ablenken könnte. Nichtsdestotrotz erschien es für die im vorhergehenden Abschnitt beschriebene Problematik gerade im Hinblick auf die fehlenden Erfahrungswerte für dieses Verfahren innerhalb der Linguistik, aber nicht zuletzt auch angesichts der relativ unaufwendigen Durchführbarkeit als Mittel der Wahl.

Die Teilnehmer wurden diesmal, anders als in den drei vorangegangenen Replikationsversuchen, gemeinsam in einen Raum gebeten, wo unter Anwesenheit der Versuchsleitung, die sich zu den Teilnehmern setzte, das Experiment durchgeführt wurde. Die Tische waren dabei in einem Hufeisen angeordnet, wobei die Versuchsleitung für alle sichtbar zwischen den Teilnehmern Platz nahm. Jeden Teilnehmer erwarteten an seinem Platz dieselben Utensilien: die Versuchsanleitung, Karteikärtchen, Gummibänder und drei Listen.

Zuvor sind (unter Abwesenheit der Teilnehmer) alle lexikalisierten Targets des dritten Replikationsversuchs einzeln in Blockschrift auf Karteikärtchen notiert worden. Ursprünglich war es gedacht, 19 Personen je 125 Karten³⁷ ranken zu lassen, was dann intuitiv als zu weitläufig empfunden wurde, so dass die Karten halbiert und die Teilnehmer dafür verdoppelt wurden. Letztendlich hatten 38 Teilnehmer etwa 60 bis 65 Karten zu sortieren, was als überschaubar eingestuft wurde.

³⁷Diese Zahl setzt sich aus 25 lexikalisierten Targets der 4 Konditionen (Phonästhem, Form, Bedeutung, keine Relation) aus dem dritten (deutschsprachigen) Replikationsversuch zusammen zuzüglich der niedrigfrequenten Targets aus der Kondition 2B.

Die Versuchsanleitung liest sich wie folgt:

„Du siehst ein Päckchen Karten vor dir und eine Liste mit drei Seiten. (Die Liste ist vorerst nicht wichtig, du wirst sie aber später brauchen). Auf jeder einzelnen Karte dieses Päckchens steht ein Wort (genauer gesagt ein Verb). Um das Päckchen sind drei Gummis gewickelt, die du zunächst entfernst. Du wirst sie aber noch benötigen.

1. Als erstes teilst du bitte die Wörter in DREI Stapel auf³⁸:

Stapel 1: Wörter, denen du häufig begegnest (häufige Wörter)

Stapel 2: Wörter, denen du gelegentlich begegnest (nicht so häufige Wörter)

Stapel 3: Wörter, denen du selten begegnest (seltene Wörter)

Die Stapel müssen NICHT gleichmäßig groß sein.

Teile die Karten intuitiv und möglichst zügig, also ohne lange zu überlegen, diesen drei Kategorien zu.

2. Nachdem du alle Karten nach diesen drei Kategorien sortiert hast, nimmst du nun deinen Stapel 1 mit den HÄUFIGEN Wörtern und legst alle Karten einzeln vor dich hin, so dass du einen Überblick über alle Karten dieses Stapels hast. Das sieht in etwa so aus (die Karten können auch etwas überlappen; wichtig ist der Überblick)³⁹

Nun sortierst du diese Karten wieder nach der Häufigkeit der einzelnen Wörter, beginnend mit dem HÄUFIGSTEN Wort, endend mit dem am wenigsten häufigen Wort dieses Stapels 1.

Du hast schlussendlich den sortierten Stapel 1 in der Hand, wobei das häufigste Wort OBEN liegt und das am wenigsten häufige Wort GANZ UNTEN im Stapel 1.

3. Notiere bitte das oberste d.h. häufigste Wort (oberste/erste Karte) als Nummer 1 auf der 1. Seite der Liste, die du zum Wortstapel erhalten hast. Deine Liste hat drei Seiten: häufige Wörter (Seite 1), weniger häufiger Wörter (Seite 2) und seltene Wörter (Seite 3). Das häufigste Wort des Stapels 1 „Häufige Wörter“ notierst du demnach auf Seite 1 als Nummer 1.

Bitte notiere auch das unterste/am wenigsten häufige Wort des Stapels 1 auf Seite 1 in der Position ganz zum Schluss (einfach als letztes Wort unter Nummer 30).

Die Worte, die zwischen dem häufigsten und am wenigsten häufigen Wort dieses Stapels 1 liegen, werden von MIR im Nachhinein in die Liste eingetragen (so hast

³⁸Nach dieser Aufzählung war in der Instruktion eine simple Graphik dieser Dreiteilung abgebildet.

³⁹Nachfolgend war in der Instruktion eine simple Graphik dieser Anordnung abgebildet.

du weniger Arbeit).

4. Jetzt binde bitte einen Gummi um deinen Stapel 1 (damit die Sortierung nicht durcheinander kommt).

5. Bitte verfare nach gleichem Muster auch für die anderen beiden Stapel, nämlich Stapel 2 "weniger häufige Wörter" und Stapel 3 „seltene Wörter“. Das häufigste Wort des 2. Stapels trägst du bitte nach demselben Schema wie oben beschrieben auf Seite 2 als Nummer 1 ein, und das letzte Wort dieses Stapels 2 unter Nummer 30 auf Seite 2. Entsprechend verfahrst du mit dem häufigsten und am wenigsten häufigen Wort des Stapels 3 (dann auf Seite 3).

6. Bitte vergiss nicht, um jeden deiner drei sortierten Stapel zum Schluss einen Gummi zu binden.

7. Nachdem du die häufigsten und am wenigsten häufigen Wörter deiner drei Stapel in die Liste eingetragen hast, gib bitte die Liste und deine drei Stapel mit den Gummis drum herum bei mir ab. Dann erhältst du deine Vergütung.

Danke für deine Teilnahme!“

Es wird deutlich, dass es sich beim Ranking, ähnlich wie beim Priming, um ein Verfahren handelt, das zwar unaufwendig in der Umsetzung, nicht jedoch in der Vorbereitung ist.

Die so von den Teilnehmern gerankten Karteikärtchenstapel wurden nach der Verabschiedung von der Versuchsleitung entpackt und die Worte der Reihenfolge nach in die Liste des jeweiligen Teilnehmers übertragen. Wichtig erscheint dabei in der Instruktion die Formulierung „Worte, denen du [...]begegnest“. Damit sollte ein Bezug zu erfahrenen, d.h. gelesenen und gehörten, Worten geschaffen werden (d.h. zum interessierenden Faktor „Vorkommenshäufigkeit“). Bewusst wurde von der Formulierung „Worte, die du [...] gebrauchst“ abgesehen, da damit ein Bezug zur Verarbeitungshäufigkeit von Wörtern geschaffen wird, den es hier bestmöglich zu umgehen galt. Auf diese Differenzierung wurde jedoch nicht explizit hingewiesen, um eine unnötige Überforderung der Teilnehmer zu vermeiden.

Abschließend wurde in der Nachbearbeitung jedem Target die entsprechenden Ränge aus den Teilnehmerlisten zugeordnet und diese dann gemittelt. Für alle (lexikalisierten) Targets wurden zu Vergleichszwecken Häufigkeitsränge aus dem COSMAS II-Korpus berechnet.

4.4.3 Diskussion der Ergebnisse

Wie sich bei der Auswertung des dritten Experiments gezeigt hat, ist die Vorkommenshäufigkeit lexikalischer Targets, die sich z. B. dem COSMAS II-Korpus entnehmen lässt, ein wesentlicher Störfaktor, wenn es darum geht, Unterschiede zwischen den interessierenden Kon-

ditionen und Konditionseffekte festzustellen. Streng genommen muss dieser Faktor jedoch nur dann berücksichtigt werden, wenn unterstellt wird, dass Testpersonen die Varianzen bei der Vorkommenshäufigkeit von Wörtern sehr genau einschätzen können.

Abgegrenzt werden sollte dieser Faktor von der potentiellen Drittvariablen der lexikalischen Verarbeitungshäufigkeit. Es kann nämlich nicht vorausgesetzt werden, dass Worte, die als häufig beurteilt werden können, von allen Testpersonen im Alltag auch häufig gebraucht werden. Dies hängt allgemein von dem Sprachstil bzw. dem Idiolekt einzelner Personen ab, der wiederum stark von dem jeweiligen sozialen Umfeld beeinflusst wird⁴⁰.

Intuitiv wäre aus folgenden Gründen anzunehmen, dass Sprecher Vorkommenshäufigkeiten (unter Abgrenzung von den eigenen Verarbeitungshäufigkeiten) nur mit geringer Genauigkeit würden schätzen können:

- Schriftsprache (aus denen sich der Großteil der Korpora zusammensetzt) kann nicht mit gesprochener Sprache bzw. Alltagssprache gleichgesetzt werden. Warum sollten Sprecher Wörter nicht ausschließlich anhand der eigenen Verarbeitungshäufigkeit beurteilen?
- Schriftsteller und Journalisten, deren Texte Bestandteile der Korpora sind, verfügen möglicherweise über einen völlig anderen Wortschatz als studentische Versuchsteilnehmer.
- Es kann nicht davon ausgegangen werden, dass der Bruchteil an schriftsprachlichen Quellen, mit dem einzelne Sprecher konfrontiert werden, Korpusverteilungen abbildet.
- Korpushäufigkeiten bilden nicht unbedingt die gegenwärtige Häufigkeit von Worten wieder, sondern setzen sich zum Teil auch aus veralteten Texten zusammen.

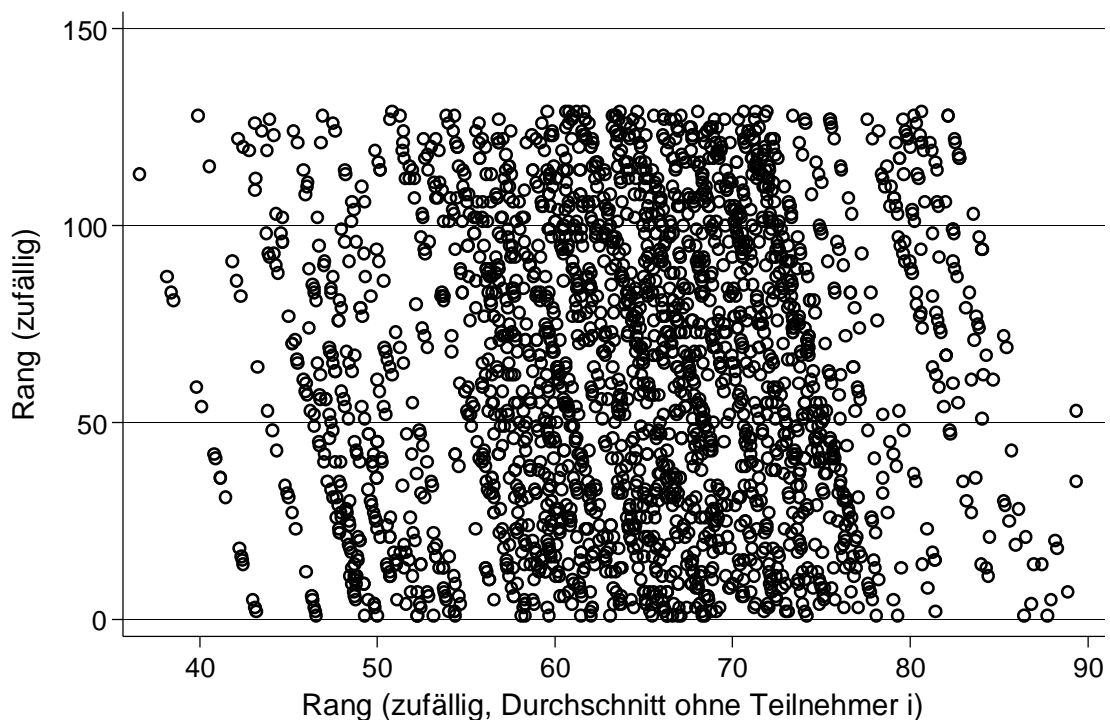
Intuitiv erscheint es abwegig, dass es einzelnen Versuchsteilnehmern gelingen könnte 60-65 Wörter innerhalb einer zeitlich begrenzten Experimentalsituation weitgehend übereinstimmend nach der Vorkommenshäufigkeit zu sortieren. Dieses Problem lässt sich jedoch möglicherweise dadurch lösen, dass durchschnittliche Ränge verwendet werden, die aus den Rankings *mehrerer* unabhängiger Versuchsteilnehmer berechnet werden. Dass so eine Herangehensweise erfolgreich sein kann, zeigen z. B. Fernsehsendungen wie „Wer wird Millionär?“, wo der Kandidat gegebenenfalls das Publikum nach der richtigen Antwort fragen kann. Auch hier wird die Einschätzung des Publikums gemittelt, so dass z.B. 10% Antwort A und 90%

⁴⁰Man kann es riskieren diese Drittvariable unter der Annahme zu vernachlässigen, dass die Versuchspersonen (als einheimische Studenten) ähnlichen sozialen Randbedingungen unterworfen sind, was nicht bedeutet, dass es sich nicht lohnen könnte diese mit separaten Experimenten zu untersuchen.

Antwort B als richtig erachten, wobei der Kandidat meistens richtig antwortet, wenn er wie auch James Surowiecki (2005) in seinem Bestseller der „Weisheit der Vielen“ folgt.

Ein erster Schritt ist hierbei die Beurteilung der internen Validität der im Experiment erfassten Rankings. Hierfür gilt es herauszufinden, wie stark das Ranking einzelner Versuchsteilnehmer mit dem *durchschnittlichen* Ranking der übrigen Versuchsteilnehmer übereinstimmt. Eine Referenzsituation stellt hierbei die Vergabe zufälliger Rankings der Teilnehmer dar (Abbildung 4.9). Deutlich wird hier, dass sich bei einer zufälligen Vergabe von Rankings, also z.B.

Abbildung 4.9: Scatterplot der Rankings einzelner Teilnehmer gegen den Durchschnitt der übrigen Teilnehmer (Simulation)



Hinweis: Ein Datenpunkt ergibt sich jeweils aus dem Ranking des Wortes j durch Teilnehmer i (y-Koordinate) und dem durchschnittlichen Ranking des Wortes j durch alle übrigen Teilnehmer ohne i (x-Koordinate).

dann, wenn die Teilnehmer die Karteikarten mit den Wörtern einfach nur beliebig ordnen würden, kein ersichtliches Muster ergibt. Darüber hinaus kann diese zufällige Referenzsituation aber auch dazu genutzt werden, um kritische Werte für Signifikanztests der Korrelationen von Rankings einzelner Teilnehmer mit dem Durchschnitt der übrigen Teilnehmer zu bestimmen. Hierbei gilt es einerseits zu beachten, dass die Rankings unter der Nullhypothese (H_0 : Es existiert kein Zusammenhang zwischen dem Ranking einzelner Teilnehmer mit dem Durchschnitt der übrigen Teilnehmer) einer Gleichverteilung folgen, wodurch standardmäßige Signifikanz-

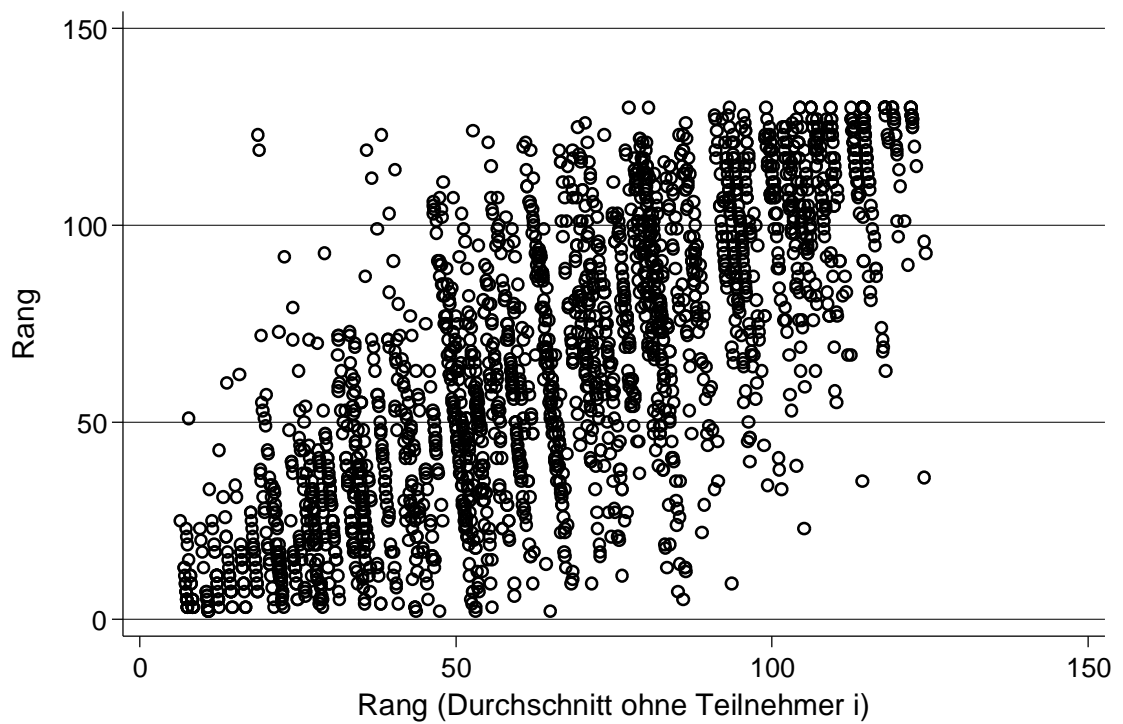
tests für Korrelationen, die eine Normalverteilung der beiden Variablen voraussetzen, nicht mehr gültig sind. Andererseits ist es jedoch keineswegs notwendig, Rangkorrelationskoeffizienten zu berechnen, die die Nichtlinearität der Zusammenhänge zwischen ordinalen Variablen berücksichtigen, da die in dieser Situation zu erwartenden Zusammenhänge linear sein sollten. Somit können kritische Werte für signifikante Korrelationen nur durch Simulationen bestimmt werden, wobei Korrelationen in einer Größenordnung von $\rho = 0,25$ bei gegebener Fallzahl bereits als signifikant anzusehen sind.

Interessant ist nun, dass im Experiment nur ein einziger Teilnehmer diesen kritischen Wert unterschreitet ($\rho = 0,12$), während für die übrigen Teilnehmer ausgesprochen hohe Korrelationen (ρ zwischen 0,55 und 0,97) charakteristisch sind. Anders als bei „Wer wird Millionär?“, wo es durchaus der Fall sein könnte, dass nur ein einziger Publikumsgast, entgegen der „Weisheit der Vielen“ die richtige Antwort kennt, gibt es hier keine richtige Antwort – die Vorkommenshäufigkeit von Wörtern ist das, was von der Mehrheit der Teilnehmer des Experiments als vorkommenshäufig erachtet wird. Deshalb wird im Folgenden bei der Berechnung der durchschnittlichen Rankings auf die Rankings des abweichenden Teilnehmers verzichtet, wobei diese „Korrektur“ letztendlich nur zu minimalen numerischen Abweichungen der auf diesen durchschnittlichen Rankings basierenden Schätzungen führt. Dies zeigt sich auch bei der graphischen Darstellung (Abbildung 4.10), die jetzt entgegen der zufälligen Situation (Abbildung 4.9) ein deutliches Muster erkennen lässt. Diesem replikationsfähigen Ergebnis zufolge können lexikalische Vorkommenshäufigkeiten von unabhängigen Versuchsteilnehmern weitgehend übereinstimmend eingeschätzt werden. Hierbei stellt sich nun die Frage, ob und inwiefern diese Häufigkeiten mit den Korpusfrequenzen gleichzusetzen sind. Wird die Korrelation der durchschnittlichen Vorkommenshäufigkeit mit den Rängen der gleichen Wörter aus dem COSMAS II-Korpus berechnet, so spricht ein Korrelationskoeffizient ρ von 0,68 für einen ausgesprochen starken und auch signifikanten Zusammenhang zwischen diesen beiden Häufigkeitsmessungen, wobei die verbleibenden Abweichungen möglicherweise auf die weiter oben genannten Einwände gegen eine Schätzbarkeit von Korpushäufigkeiten zurückgeführt werden können. Auch hier kann das Ergebnis graphisch mit einem Scatterplot dargestellt werden (Abbildung 4.11).

4.4.4 Schlussfolgerungen für den dritten Replikationsversuch

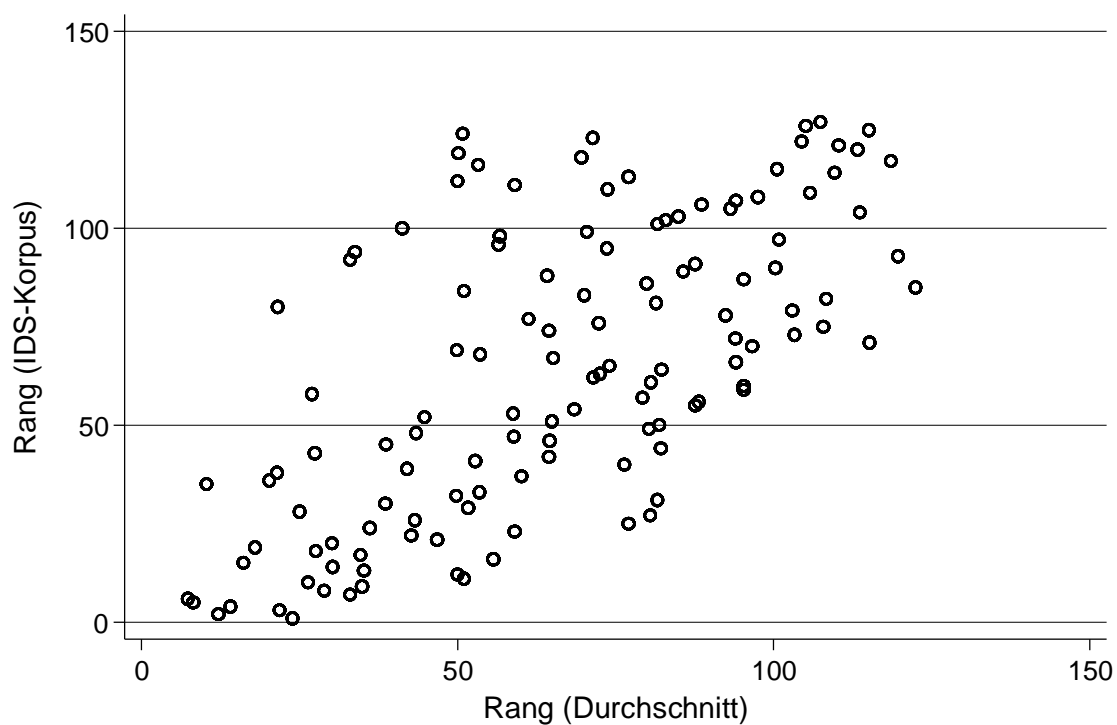
Nun, da sich herausgestellt hat, dass Sprecher über die Fähigkeit verfügen, Korpusfrequenzen mit bemerkenswerter Genauigkeit zu approximieren, stellt sich die Frage, ob und inwiefern diese Vorkommenshäufigkeiten tatsächlich mit den Ergebnissen von Experiment 3 in Zusammenhang stehen. Bisher wurde durch die gezielte Manipulation der relationslosen

Abbildung 4.10: Scatterplot der Rankings einzelner Teilnehmer gegen den Durchschnitt der übrigen Teilnehmer



Hinweis: Ein Datenpunkt ergibt sich jeweils aus dem Ranking des Wortes j durch Teilnehmer i (y -Koordinate) und dem durchschnittlichen Ranking des Wortes j durch alle übrigen Teilnehmer ohne i (x -Koordinate). Auf Teilnehmer 19 wurde wegen einer niedrigen Korrelation der Rankings mit den übrigen Teilnehmern verzichtet.

Abbildung 4.11: Scatterplot der IDS-Ränge gegen die durchschnittlichen Rankings der Teilnehmer



Hinweis: Ein Datenpunkt ergibt sich jeweils aus dem durchschnittlichen Ranking des Wortes j durch die Teilnehmer (x -Koordinate) und dem Ranking des Wortes j im IDS-Korpus (y -Koordinate). Auf Teilnehmer 19 wurde wegen einer niedrigen Korrelation der Rankings mit den übrigen Teilnehmern verzichtet.

Kondition gezeigt, dass die Vorkommenshäufigkeit einen potentiellen Störfaktor darstellt. Allerdings wurde für die Teilgruppen 2A und 2B eine starke Polarität im Hinblick auf ihre Token-Häufigkeiten erzeugt, so dass die Effektstärke dieser Drittvariablen allein damit noch nicht beurteilt werden kann⁴¹. Um das herauszufinden, werden sowohl die selbst erhobenen durchschnittlichen Vorkommenshäufigkeiten als auch die KOSMAS II-Ränge den Daten aus dem dritten Experiment zugespielt, so dass nun für jedes lexikalisierte Target im Experiment zwei Variablen vorliegen, nämlich a) die Vorkommenshäufigkeit aus dem Ranking und b) den COSMAS II-Rang des jeweiligen Targets. Dabei ist anzunehmen, dass die Reaktionszeit mit der Ranghöhe positiv korreliert ist, d.h. dass mit steigendem Target-Rang (sowohl aus dem Ranking als auch aus dem COSMAS II-Korpus) auch die Reaktionszeit steigt.

Deutlich wird in Tabelle 4.8, dass die Konditionen mit den langsamsten bzw. schnellsten Reaktionen die höchsten bzw. niedrigsten Rankings aufweisen. Auf Bergens (2004) Ergebnisse übertragen, entsteht der Eindruck, als wären der von ihm berichtete signifikante Primingeffekt der Phonästhem-Kondition maßgeblich von den Token-Frequenzen des dazugehörigen Targetsets erzeugt worden. Zumindest seine eigenen Frequenzangaben aus dem Brown-Korpus (Kucera & Francis 1967) und aus dem Thorndike-Lorge-Korpus (1944) zeigen die insgesamt höchsten Frequenzwerte für die Phonästhem-Kondition.

Tabelle 4.8: Durchschnittliche Reaktionszeiten, Rankings und Ränge nach Konditionen

Kondition	MW Reaktionszeit	MW Rang Ranking	MW Rang COSMAS II
Keine Relation	665,725	59,508	45,451
Keine Relation (A)	606,685	33,051	14,477
Keine Relation (B)	745,037	94,674	86,621
Phonästhem	711,474	71,890	81,388
Form	708,830	57,434	65,943
Bedeutung	712,455	62,998	65,477

Hinweis: Bei den Konditionen wurden nur die Reaktionszeiten der richtigen Antworten bei lexikalisierten Worten für die Berechnung der Reaktionszeit in ms verwendet. Dass niedrige bzw. hohe Rankings bei den COSMAS II-Daten niedriger bzw. höher ausfallen, ist auf die Durchschnittsbildung der Ränge im Experiment zurückzuführen, die zwangsläufig mit einer Verringerung der Varianz bzw. Verkleinerung des Wertebereichs einhergeht.

Im Übrigen ist für die Ergebnisse des dritten Replikationsversuchs interessant, dass sich die Reaktionszeiten für die Form-Gruppe – vermutlich aufgrund der systematisch höherfre-

⁴¹Es stellt sich also die Frage, ob dieser Störeffekt so stark ist, dass er die feinen (insignifikanten) Unterschiede zwischen den Konditionen erzeugt haben könnte. Falls ja, so wäre nicht anzunehmen, dass mit einer Samplevergrößerung die interessierenden Relationseffekte gemessen werden können.

quenten Primes – weniger stark von den anderen Konditionen unterscheiden als in den ersten beiden Experimenten. Dies kann als Hinweis darauf interpretiert werden, dass die relative Prime-Target-Häufigkeiten tatsächlich als Störfaktor für phonologisches Priming berücksichtigt werden sollte (vgl. Grainger 1990).

Abschließend gilt es noch die Rankings als Drittvariable in das Regressionsmodell zur Auswertung des dritten Experiments (Tabelle 4.7) aufzunehmen, da so auch gezeigt werden kann, wie groß der Erklärungsbeitrag von Frequenzeffekten ist. Nur so lässt sich die Möglichkeit ausschließen, dass Bergens (2004) signifikante Konditionseffekte ein Artefakt unterschiedlicher Worthäufigkeiten sein könnten. Problematisch ist hierbei jedoch, dass der Effekt des Rankings nicht zwangsläufig linear verlaufen muss, so dass z.B. die Reaktionszeiten der Ränge 1 bis 50 relativ ähnlich sein könnten, um dann bei seltenen Wörtern exponentiell anzusteigen. Um hier die richtige funktionale Spezifikation auszuwählen, wurde eine Regressionsdiagnostik mit einem Component-plus-Residual-Plot (vgl. Hamilton 2009) durchgeführt, anhand dessen sich die Linearität von Effekten in Regressionsmodellen überprüfen lässt (Abbildung 4.12). Hier zeigt sich, dass sowohl der Effekt der empirisch bestimmten Worthäufigkeiten als auch der Effekt der COSMAS II-Ränge (keine Abbildung) die Reaktionszeiten linear beeinflussen, so dass keine weiteren Anpassungen am Modell mehr vorgenommen werden müssen.

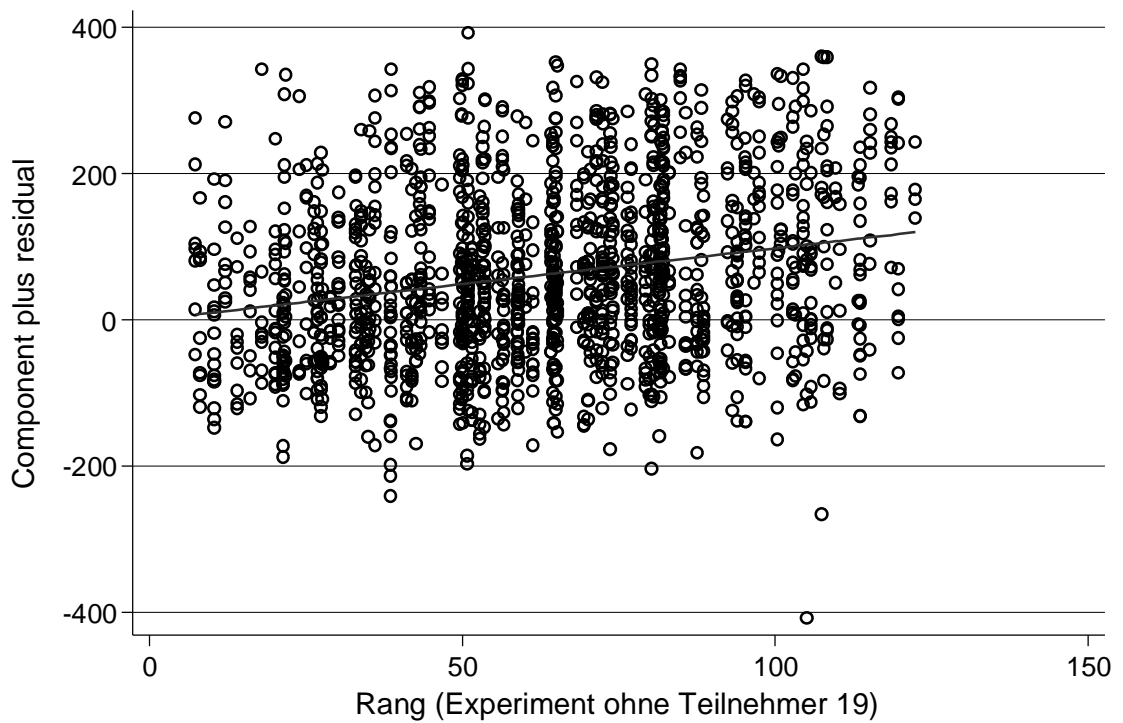
Wie bereits bei den vorangegangenen Experimenten wird auch hier ein lineares Regressionsmodell mit einer Random-Effects-Struktur des Teilnehmereffekts geschätzt, bei dem Clusterbildungen in der Kovarianzmatrix (teilnehmerspezifische Varianzen der Reaktionszeiten) unterstellt werden (Tabelle 4.9). Deutlich wird nun, dass Frequenzeffekte eine zentrale Rolle spielen, wenn es darum geht, Reaktionszeiten zu erklären, wobei die empirisch bestimmten Verarbeitungshäufigkeiten aufgrund des größeren z-Werts tatsächlich besser geeignet scheinen, um Vorkommenshäufigkeiten abzubilden als die COSMAS II-Ränge.

Da die Aufnahme der Frequenzeffekte zu einer Verringerung der z-Werte der Konditionseffekte führt, kann nicht ausgeschlossen werden, dass es sich bei den Konditionseffekten in Bergens (2004) Experiment um Scheinkorrelationen handelt, die maßgeblich von Worthäufigkeiten verursacht werden.

In der eigenen Empirie weist die Phonästhem-Relation bei einer Kontrolle von Frequenzeffekten die niedrigsten Reaktionszeiten auf. Dass dieses Phänomen jedoch keinesfalls überbewertet werden sollte, wird aus dem Marginsplot (Abbildung 4.13) deutlich, wo die Konfidenzintervalle der Effekte alle die Null mit einschließen.

Mit diesem Ergebnis kann die Validität sämtlicher Experimente, die Frequenzeffekte bei der Auswertung nicht als Drittvariable berücksichtigen (u.a. Bergen 2004, Feldman 2000), in Frage gestellt werden. Da es in der Realität ausgesprochen unwahrscheinlich ist, dass zwei

Abbildung 4.12: Component-plus-Residual-Plot für den Effekt der empirisch bestimmten Worthäufigkeiten (Ranking)



Hinweis: Auf Teilnehmer 19 wurde wegen einer zu niedrigen Korrelation mit den durchschnittlichen Rängen der übrigen Teilnehmer verzichtet. Wird dieser Ausschluss nicht vorgenommen, verläuft der Effekt nahezu identisch.

Tabelle 4.9: Einflussfaktoren auf die Reaktionszeit – Random-Effects-Modell

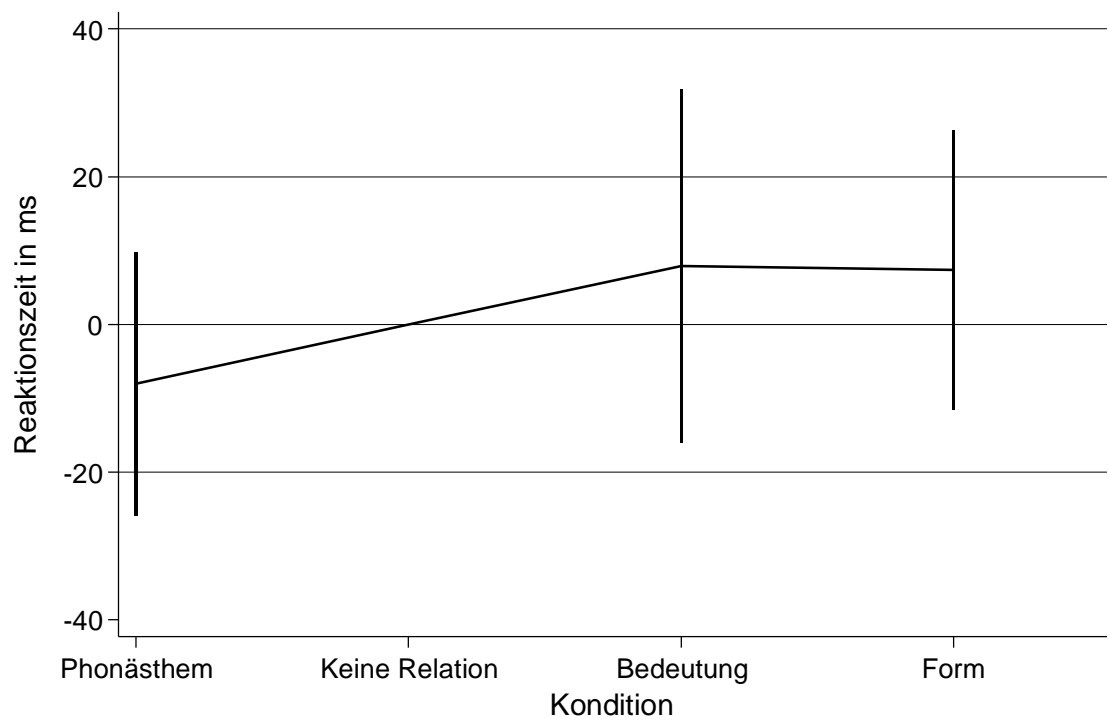
	Reaktionszeit	
Trialnummer	-0,415	(-1,37)
Trialnummer (quad.)	0,00141	(1,08)
keine Relation (A)	-54,64***	(-3,37)
Phonästhem	-8,080	(-0,89)
Bedeutung	7,898	(0,65)
Form	7,363	(0,76)
Rang (COSMAS II)	0,239*	(2,37)
Rang (Ranking)	0,982***	(7,42)
Konstante	652,2***	(39,98)
Beobachtungen	1517	

z-Werte in Klammern

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht eine Referenzkategorie definiert werden, für die dann kein Relationseffekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Konditions-Effekt geschätzt. Die Anpassungsgüte R^2 kann bei Paneldatenmodellen nicht sinnvoll interpretiert werden und wird deshalb nicht berichtet. Da für Pseudo-Targets keine Verarbeitungshäufigkeiten berechnet werden können, reduziert sich die Fallzahl gegenüber dem dritten Experiment auf etwa die Hälfte. Auf Teilnehmer 19 wurde bei der Berechnung der Verarbeitungshäufigkeiten wegen einer zu niedrigen Korrelation mit den durchschnittlichen Rängen der übrigen Teilnehmer verzichtet. Wird dieser Ausschluss nicht vorgenommen, verläuft der Effekt nahezu identisch und zeigt sogar einen etwas höheren z-Wert.

Abbildung 4.13: Marginsplot mit 95% Konfidenzintervall für die Effekte der Konditionen



unterschiedliche Worte eine identische Vorkommenshäufigkeit aufweisen, kann es angesichts der Stärke dieser Störvariablen nicht genügen, wie bei Bergen (2004) eine Konstanzhaltung der Frequenz durch eine Auswahl ähnlich häufiger Targets zu erwirken.

Erschwerend kommt hinzu, dass die Verarbeitungshäufigkeit von Wörtern zwischen den Teilnehmern variieren kann und nicht unbedingt mit der Vorkommenshäufigkeit dieser Wörter übereinstimmt, weshalb diese Faktoren bereits im Experimentaldesign berücksichtigt sein sollten.

Um möglicherweise doch noch Konditionseffekte messen zu können, wird im Folgenden ein erster Replikationsversuch von Bergen (2010) dokumentiert. Dabei handelt es sich um den Nachfolgeversuch des Urexperiments von Bergen (2004), das von Bergen (2010) zum Zweck der Eliminierung des Einflusses von Target-Häufigkeiten modifiziert worden ist⁴².

4.5 Replikationsversuch von Bergen (2010)

4.5.1 Teilnehmer

Die Testpersonen setzen sich aus 16 monolingualen Studenten mit der Muttersprache Deutsch zusammen, die über Aushänge auf dem Campus der Ludwig-Maximilians-Universität München angeworben worden sind. Für ihre Teilnahme erhielten sie jeweils 10 Euro Aufwandsentschädigung.

4.5.2 Experimentaldesign, Aufgabenstellung und Ablauf

Sechs Jahre nach Veröffentlichung von *The psychological reality of phonaesthemes* in der *Language* berichtete Benjamin Bergen auf der Sound Symbolism-Konferenz an der Emory Universität in Atlanta im US-Staat Georgia von einem auf dieser Publikation aufbauenden Versuch, mit dem es ihm gelungen ist den Frequenzfaktor bereits im Vorfeld der Versuchswiederholung zu eliminieren. Leider ist dieser Bericht bis heute in keiner Fachzeitschrift veröffentlicht worden, sondern kursiert in Form einer Folienpräsentation im Internet. Obgleich die Idee zu dem darin beschriebenen Experimentaldesign positiv hervorgehoben werden kann, wird dieser gute erste Eindruck durch fehlende bzw. unvollständige Beschriftung der Graphiken, die die berichteten Effekte veranschaulichen sollen, stark gemindert. Alles in allem gelingt es Bergen (2010) sein eigenes Resultat hinsichtlich des signifikanten Phonästhem-Effekts

⁴²Damit kann Bergen (2010) als erster, stark modifizierter Replikationsversuch von Bergen (2004) aufgefasst werden, den der Autor selbst durchgeführt hat.

in Bergen (2004) zu replizieren, die dazugehörigen numerische Daten sucht man in der Präsentation jedoch vergebens. Dass Bergens Auseinandersetzung⁴³ mit der Thematik unabhängig von der eigenen Versuchsreihe (abgesehen vom gemeinsamen Ausgangspunkt) letztendlich zu derselben Schlussfolgerung hinsichtlich der Tokenfrequenzkontrolle⁴⁴ geführt hat, bestätigt die Notwendigkeit einer Fortführung der Experimentalreihe.

Da Bergen (2010) seine eigenen Resultate auch unter Kontrolle von Frequenzeffekten replizieren kann, führt ihn natürlich zu dem gegenläufigen Schluss, nämlich dass Token-Frequenz für phonästhemische Primingeffekte keine Rolle spielt. Damit lag es nahe auch diese Publikation einem Replikationsversuch zu unterziehen. Nach Vorbild des in dem Paper von 2010 skizzierten Designs wurden allen Teilnehmern unter denselben Bedingungen wie bei den ersten beiden Replikationsversuchen von Bergen (2004) insgesamt jeweils 120 Trials präsentiert, von denen die Hälfte lexikalisierte Targets der vier genannten Konditionen enthielten, d.h. 15 deutsche Wortpaare pro Kondition⁴⁵. Wie zuvor auch war die Abfolge der Trials bei allen Teilnehmern identisch und erfolgte in zufälliger Reihenfolge. Der Einfluss der Target-Häufigkeit wurde umgangen, indem das Target-Set für alle vier Konditionen konstant gehalten wurde, d.h. jede Kondition enthielt dasselbe phonästhemische⁴⁶ Target-Set.

Bei der Zusammenstellung der Phonästhem-Trials wurden die Phonästhem SCHM- (z.B. *SCHMollen* – *SCHMunzeln*), SCHN- (z.B. *SCHNupfern* – *SCHNiefen*), BL- (z.B. *BLinzeln* – *BLicken*), GL- (z.B. *GLimmen* – *GLotzen*) und FL- (z.B. *FLattern* – *FLimmern*) eingebaut, deren Kognaten auch in der Skizzierung von Bergens (2010) eigener Replikation genannt werden (mit Ausnahme von SCHM- respektive SM-). Auch hier sollten die Probanden über eine Ja- und eine Nein-Taste entscheiden, ob ein auf dem Computerbildschirm erscheinender Zielreiz ein gültiges deutsches Wort darstellt oder nicht.

4.5.3 Ergebnisse

Deutlich wird bei Tabelle 4.10, dass bei der relationslosen Kondition („keine Relation“) die meisten Fehler gemacht werden. Anders als in den vorangegangenen Experimenten sind jetzt

⁴³Der Schluss liegt nahe, dass diese erneute Auseinandersetzung mit der Veröffentlichung von 2004 bei Bergen nicht zuletzt von dem Wissen um die unabhängigen Replikationsversuche, wie sie in dieser Arbeit dokumentiert sind, angeregt worden ist.

⁴⁴Gemeint ist die Schlussfolgerung, dass eine ideale Tokenfrequenzkontrolle darin besteht, diese Drittvariable durch eine Überarbeitung des Versuchsdesigns zu eliminieren.

⁴⁵Insgesamt erhielt jeder Teilnehmer 120 Wortpaare, bestehend aus 60 Pseudowörtern und 60 Wörtern, davon 15 Phonästhem-, 15 Form-, 15 Bedeutungs- und 15 relationslose Trials

⁴⁶Dies hätte anders auch nicht gelöst werden können, da ein phonästhemisches Prime-Set zwingend nach einem phonästhemisches Targetset verlangt und die Konstanthaltung des Targetsets über die Konditionen hinweg vorausgesetzt wurde.

jedoch die Reaktionen der Teilnehmer in der Phonästhem-Kondition am schnellsten.

Tabelle 4.10: Fehlerraten, Fallzahlen, durchschnittliche Reaktionszeiten und Standardabweichungen nach Konditionen

Kondition	Fehlerrate	N (richtig)	MW Reaktionszeit	Standardabw.
Pseudo-Target	0,970	861	695,587	122,440
Keine Relation	0,957	202	644,262	123,870
Phonästhem	0,981	203	639,295	138,738
Form	0,995	214	647,264	130,279
Bedeutung	0,981	209	644,805	133,657

Hinweis: Bei den Konditionen wurden nur die Reaktionszeiten der richtigen Antworten für die Berechnung der Reaktionszeit in ms verwendet.

Wie bei den vorangegangenen Experimenten wird auch hier ein lineares Regressionsmodell mit einer Random-Effects-Struktur des Teilnehmereffekts, bei dem Clusterbildungen in der Kovarianzmatrix (teilnehmerspezifische Varianzen der Reaktionszeiten) unterstellt werden, geschätzt. Designbedingt gilt es hier jedoch einige Besonderheiten zu beachten. Einerseits ist es nämlich so, dass durch die Wiederholung der Targets besonders starke Lerneffekte plausibel erscheinen, was dazu führen könnte, dass die Reaktionszeiten sich mit den Wiederholungen verkürzen. Dieser Effekt kann durch Dummyvariablen für die zweite, dritte und vierte Wiederholung eines Targets im Modell berücksichtigt werden. Andererseits bietet die Wiederholung der Targets nun auch die Möglichkeit, unabhängig vom Effekt der Konditionen Effekte einzelner Worte zu berücksichtigen. Dass eine derartige Herangehensweise sinnvoll ist, zeigt Abbildung 4.14. Tatsächlich ist es nämlich so, dass – z.B. aufgrund unterschiedlicher Verarbeitungshäufigkeiten – Worte mit relativ kurzen (z.B. *flüstern*) und relativ langen Reaktionszeiten (z.B. *flimmern*) beobachtet werden, die darüber hinaus auch wortspezifische Varianzen der Reaktionszeiten aufweisen könnten. Obwohl derartige Worteffekte in diesem Experiment aufgrund des Designs nicht zu einer Verzerrung der Konditionseffekte führen sollten, erscheint es sinnvoll, die Wort-Effekte, wie bereits zuvor die Teilnehmereffekte, als Random-Effects zu modellieren und bei der Struktur der Kovarianzmatrix zu berücksichtigen. Vorausgesetzt, die Random-Effects-Annahme ist korrekt, was im Rahmen eines Hausman-Tests (vgl. Wooldridge 2003) bestätigt werden konnte, so setzt dieses verfeinerte Modell etwas weniger restriktive Annahmen⁴⁷ für Signifikanztests voraus. Umgesetzt wurde diese Herangehensweise in Modell 2, während Modell 1 eine Schätzung des ursprünglichen Random-Effects-Modells darstellt (Tabelle 4.11). Letztendlich ergeben sich durch die beiden Modellierungsvarianten

⁴⁷Die Varianzen der Reaktionszeiten sind nicht für alle Worte eines Teilnehmers konstant.

Abbildung 4.14: Unterschiedliche Reaktionszeiten bei unterschiedlichen Targets

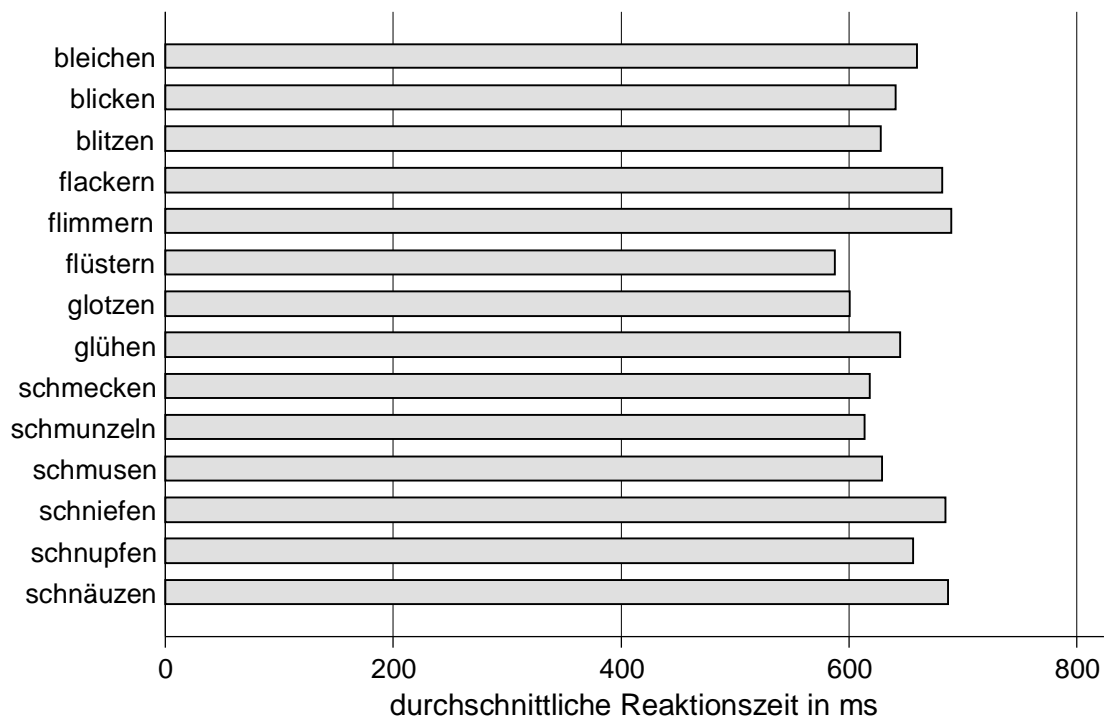


Tabelle 4.11: Einflussfaktoren auf die Reaktionszeit – Random-Effects-Modelle

	(1) Reaktionszeit	(2) Reaktionszeit
Target/Pseudo-Target	–32,17*** (–3,90)	–22,14 (–1,65)
Trialnummer	–1,422** (–3,18)	–1,518** (–3,04)
Trialnummer (quad.)	0,00558 (1,75)	0,00855* (2,16)
Phonästhem	–29,00** (–3,27)	–26,41** (–2,63)
Bedeutung	–23,24** (–2,66)	–15,70 (–1,46)
Form	–2,856 (–0,45)	4,112 (0,41)
Wiederholung_2	22,02* (2,20)	3,592 (0,30)
Wiederholung_3	–28,58* (–2,10)	–46,08*** (–3,47)
Wiederholung_4	–13,86 (–1,02)	–38,51** (–2,82)
Konstante	759,0*** (33,15)	751,2*** (48,80)
Beobachtungen	1689	1689

z-Werte in Klammern

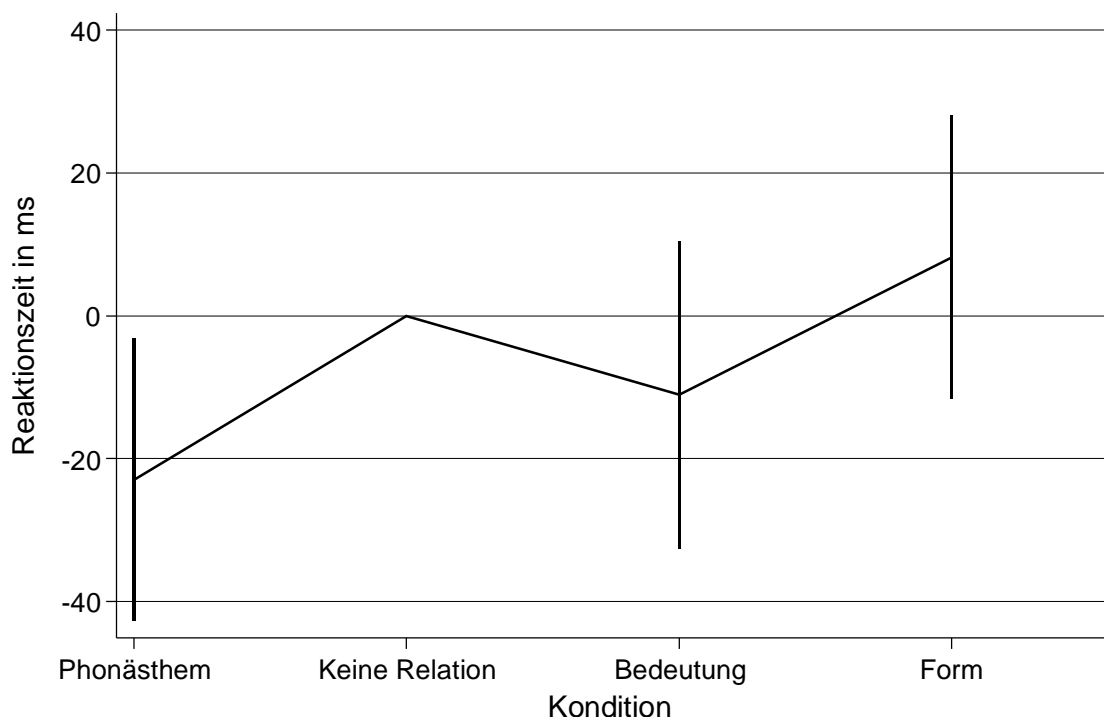
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht, eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Relationseffekt geschätzt. Die Anpassungsgüte R^2 kann bei Paneldatenmodellen nicht sinnvoll interpretiert werden und wird deshalb nicht berichtet.

keine größeren Unterschiede, obwohl insbesondere der Wiederholungseffekt, der im zweiten Modell eindeutig als Lerneffekt identifiziert werden kann, gegen das erste Modell spricht. Die fehlende Signifikanz des Wort-Effekts im zweiten Modell erklärt sich dadurch, dass hier jedem lexikalisierte Target bzw. nicht-lexikalisierten Pseudo-Target eine spezifische Varianz der Reaktionszeit unterstellt wird, so dass Unterschiede zwischen Gruppierungen von Worten nun möglicherweise etwas schwerer nachzuweisen sind. Dies könnte einerseits auch die Ursache für die nun fehlende Signifikanz der Bedeutungskondition sein (im ersten Modell sind Bedeutungskondition und Phonästhem-Kondition signifikant). Andererseits spricht der nach wie vor signifikante Effekt der Phonästhem-Kondition dafür, dass schnelle Reaktionszeiten bei Phonästhemen möglicherweise doch nicht auf den Bedeutungszusammenhang zwischen Prime und Target reduziert werden können.

Wie der Marginsplot (Abbildung 4.15) zeigt, unterscheiden sich die Effekte für Phonästhem und Bedeutung jedoch nicht signifikant voneinander, da die Konfidenzintervalle jeweils den Mittelwert der anderen Kategorie mit einschließen.

Abbildung 4.15: Marginsplot mit 95 % Konfidenzintervall für die Effekte der Konditionen



4.5.4 Diskussion

Hinsichtlich des überraschenden Ergebnisses des Replikationsversuchs von Bergen (2010), bei der die Target-Token-Frequenz dadurch homogenisiert werden konnte, dass jeder Kondition dasselbe Target-Set zugespielt worden ist, ergeben sich neue Schlussfolgerungen für eine Theorie lexikalischer Wissensrepräsentation. Anders als in Bergen (2004) und (2010) beschleunigen *sowohl* die Phonästhem-Kondition *als auch* die Bedeutungskondition die durchschnittlichen Reaktionszeiten der Teilnehmer in signifikanter Weise. Zentral ist dabei jedoch, dass kein signifikanter Unterschied zwischen diesen beide Konditionen zu verzeichnen ist. Ähnlich wie bei den morphologischen Primingeffekten in Rastle et al. (2000), wo Primingstärke eine Funktion der semantischen Beziehung zwischen Prime und Target darstellt⁴⁸, scheint es so, als wären *die Effekte für die Phonästhemie vollständig auf ihre Bedeutungskomponente zurückzuführen*.

Hervorzuheben wäre auch der Umstand, dass die Stimulus-Onset-Asynchrony, die u.a. von Feldman (2000) als wichtiger Parameter für die Messung von Konditionseffekten berichtet wird, über alle Replikationsversuche hinweg konstant gehalten wurde. Angesicht der Berichte von Feldman (2000) und Rastle et al. (2000) wären bei einer Primedauer von 150 ms lediglich messbare Bedeutungseffekte zu erwarten. Auch Frost et al. 1997 können bei kurzen SOA keine Bedeutungseffekte verzeichnen, dafür aber Perea & Gotor (1997). Überbewertet werden sollten diese Vorgaben also nicht, da z.B. Diependaele et al. (2009) morphologische Effekte bereits bei 67 ms Primedauer beobachten, in einem Bereich also wo Feldman (2000) nur Formeffekte (die Rastle et al. 2000 weder bei kurzen noch bei langen SOA messen) misst und Kiear et al. (2008) nur semantische Effekte berichten, jedoch keine Formeffekte.

Nicht zuletzt ist eine frappierende Übereinstimmung der eigenen Resultate mit der Studie von Rastle et al. (2000) hervorzuheben, die für die vorliegende Arbeit gerade deswegen von besonderer Bedeutung ist, da es den Autoren überhaupt nicht um die Messung von Phonästhemeffekten geht. Die Effekte der phonästhemischen Kondition, die dem von ihnen verwendeten Term (-M+S+O) zugrundeliegt und lediglich als Kontrollgruppe für Morphemeffekte dient, lassen sich bei Rastle et al. (2000) ebensowenig von der semantischen Kondition unterscheiden wie in der eigenen Studie: „[P]riming effects for -M+S+O pairs did not differ significantly from priming effects for -M+S-O pairs“ (Rastle et al. 2000:525). Berücksichtigt man die SOA allen Einwänden zum Trotz, so darf nicht unerwähnt bleiben, dass diese Effekte in Rastle et al. (2000) bei einer SOA von 230 ms, die eine ähnliche Primedauer impliziert wie in vorliegender

⁴⁸„The amount of priming in the two morphologically related conditions varied as a function of the semantic relationship between prime and target, such that transparent (+M+S+O) pairs produced more priming than opaque (+M+S-O) pairs.“ (Rastle et al. 2000:516)

Arbeit, auch signifikant werden.

Obwohl nun vieles dafür spricht, die Versuchsreihe an dieser Stelle zu schließen, sollte im Folgenden dem eigenen Anspruch gerecht werden, die Validität von Versuchsergebnissen nicht mit einem Einzelversuch zu begründen. Insbesondere ein Experiment, aus dem bedeutende theoretische Schlussfolgerungen abgeleitet werden, sollte replizierbar sein und auch repliziert werden, so dass der Replikationsversuch der eigenen Ergebnisse als abschließendes Kontrollexperiment aufgefasst werden kann.

4.6 Replikationsversuch der eigenen Ergebnisse mit auditivem Priming

4.6.1 Teilnehmer

Die Testpersonen setzen sich aus 30 monolingualen Studenten mit der Muttersprache Deutsch zusammen, die über Aushänge auf dem Campus der Ludwig-Maximilians-Universität München angeworben worden sind⁴⁹. Für ihre Teilnahme erhielten sie jeweils 10 Euro Aufwandsentschädigung.

4.6.2 Experimentaldesign, Aufgabenstellung und Ablauf

Versuchsbedingungen und Ablauf waren dieselben wie im Replikationsversuch von Bergen (2010) mit dem Unterschied, dass dieses Mal statt eines visuellen Primings ein auditives Priming durchgeführt wurde⁵⁰. Das heißt, dass das Stimulimaterial aus dem vorangegangenen Versuch nun in Form von Audiofiles vorlag. Die Teilnehmer erhielten die Instruktion wie die Probanden in den Primingversuchen zuvor in schriftlicher Form am Computermonitor, der während des Ablaufs des Experiments bis auf die Pausenankündigungen schwarz blieb.

⁴⁹In der Datentabelle, die dieser Arbeit beiliegt, beginnt die Zählung der Teilnehmer mit der Nummer 2 (endend bei 31), da der Testlauf als Teilnehmer 1 gespeichert worden war, und aus der Analyse ausgeschlossen werden musste.

⁵⁰Eine weitere Abweichung bestand in der Trialzahl. Anstelle von 120 Stimulipaaren kamen nur 116 Paare pro Teilnehmer zum Einsatz, da Trials 10,12,23 und 30 aufgrund der unbefriedigenden Tonqualität der Audiofiles vor der Versuchsdurchführung für alle Teilnehmer gestrichen worden sind. Konkret sind das die Paare *striegeln–gleißend*, *glänzen–gleißend*, *glücken–gleißend*, *strahlen–gleißend* betroffen, bei denen es nicht gelungen ist, den stimmhaften alveolaren Plosiv deutlich genug zu artikulieren, ohne künstlich zu klingen, so dass *gleißend* Gefahr lief als **gleißen* und somit als Pseudowort aufgefasst zu werden. Die übrigen 116 Trials bestanden aus 60 Pseudowort- und 56 Wortpaaren, davon 14 Form-, 14 Phonästhem-, 14 Bedeutungs- und 14 relationslose Paare

Alle lexikalischen Stimuli wurden von der Versuchsleitung standarddeutsch auf Band gesprochen. Die Implementierung dieser Files gestaltete sich mit PXLab ausgesprochen mühsam, da diese Anwendung lediglich Wav-Dateien mit der Bit-Rate 88 kBits/s, dem Audioformat PCM sowie der Abtastrate 11 kHz abspielt (was im Manual leider nicht vermerkt war⁵¹). Ein Vorgehen nach dem Trial-and-Error-Prinzip hat letztendlich zum Ziel geführt, wobei die Audiofiles mit Hilfe von Audacity, einem freien Audioeditor- und -recorder, aufgenommen worden sind. Eine Veränderung der Modalität brachte zusätzliche Veränderungen im Design mit sich: so konnte die Primedauer nicht mehr konstant gehalten werden, weil die Tonqualität darunter gelitten hätte. Auch wurden nun nicht mehr Ja- und Nein-Reaktionen der Testpersonen abgefragt, sondern lediglich Ja-Reaktionen, eine Vereinfachung, die für den erhöhten Konzentrationsaufwand bei der klanglichen Unterscheidung von Prime und Target kompensieren sollte. Die Unterscheidbarkeit von Prime und Target wurde dadurch verbessert, dass die Primes (anders als z.B. bei Kouider et al. 2009 und Kotz et al. 2002) etwas leiser und in einer niedrigeren Frequenz dargeboten wurden. Bei dem eingesetzten Stereo-Kopfhörer handelt es sich um den Porta Pro von KOSS.

4.6.3 Ergebnisse

Deutlich wird in Tabelle 4.12, dass bei den Pseudo-Targets die meisten Fehler gemacht werden. Aussagen über die Reaktionszeit sind hier jedoch nicht möglich, da als richtige Reaktion auf nicht-lexikalisierte Targets designbedingt das Ausbleiben einer Reaktion gewertet wird und somit nicht zwischen richtigen und zu langsamen Antworten unterschieden werden kann. Aufgrund der relativ großen Fallzahlen wäre es bereits an dieser Stelle möglich, signifikante Unterschiede der durchschnittlichen Reaktionszeiten zwischen den Konditionen mit einer einfaktoriellen Varianzanalyse (vgl. Abschnitt 4.1.3) zu belegen. Dies erübrigt sich jedoch aufgrund der zahlreichen potentiellen Störvariablen, die sowohl zu einer Verzerrung der Ergebnisse als auch zu einer Verletzung der Annahmen für die Varianzanalyse bzw. das Regressionsmodell führen könnten. Zusammengefasst handelt es sich dabei im Einzelnen um folgende Effekte:

- Teilnehmereffekt
- Lerneffekt
- Ermüdungseffekt

⁵¹Leider ist der Entwickler von PXLab, Dr. Irtel, inzwischen verstorben. Eine Nachfrage beim Institut für Deutsche Sprache (IDS) in Mannheim hat aufgrund von fehlender Resonanz zu dem Schluss geführt, dass wohl zur Zeit nicht genügend Kapazitäten vorhanden sind, um seine Arbeit an PXLab fortzuführen.

Tabelle 4.12: Fehlerraten, Fallzahlen, durchschnittliche Reaktionszeiten und Standardabweichungen nach Konditionen

Kondition	Fehlerrate	N (richtig)	MW Reaktionszeit	Standardabw.
Pseudo-Target	0,947	1705	—	—
Phonästhem	0,990	416	1009,586	205,364
Keine Relation	0,983	413	1032,902	204,392
Bedeutung	0,993	417	979,007	213,341
Form	0,981	412	1023,146	225,711

Hinweis: Designbedingt können für nicht-lexikalisierte Targets (Pseudo-Targets) keine Reaktionszeiten berechnet werden. Bei den Konditionen wurden nur die Reaktionszeiten der richtigen Antworten für die Berechnung der Reaktionszeit in ms verwendet.

- Frequenzeffekt
- Wort-Effekt
- Wiederholungseffekt
- Effekt der Wortlänge

Ausführlich diskutiert und empirisch belegt wurde der Einfluss der entsprechenden Variablen bereits in den Abschnitten 4.1.3 (Teilnehmereffekt), 4.2.3 (Lerneffekt, Ermüdungseffekt), 4.3.3 (Frequenzeffekt) und 4.5.3 (Wort-Effekt, Wiederholungseffekt). Neu hinzugekommen ist nur der „Effekt der Wortlänge“, bei dem es sich um ein designspezifisches Artefakt des auditiven Primings handelt. Ursache hierfür ist die Tatsache, dass bereits während des Abspielens der akustischen Targets Reaktionen möglich sein müssen⁵², da sehr schnelle Reaktionen ansonsten nicht erfasst werden könnten. Dadurch kommt es bei längeren Audiofiles bzw. bei längeren Wörtern zu einer systematischen Erhöhung der Reaktionszeiten. Eine andere Erklärung für diesen Effekt könnten aber auch geringfügige Abweichungen bei der Qualität der Audiofiles sein. Besonders deutlich fällt dies für die Targets „bleichen“, „flüstern“ und „glotzen“ aus, für die jeweils sehr langsame bzw. sehr schnelle Reaktionen charakteristisch sind (Abbildung 4.16). Wie bereits bei den vorangegangenen Experimenten bietet es sich auch hier an, Regressionsmodelle für Paneldaten (Abschnitt 4.1.3) zu schätzen, bei dem die genannten Effekte als Kontrollvariablen in die Modelle mitaufgenommen werden (vgl. Tabelle 4.13). Ebenso wie im Replikationsversuch von Bergen (2010) beschleunigen sowohl die Bedeutungs- als auch die Phonästhem-Beziehung die durchschnittliche Reaktionszeit der Teilnehmer signifikant. Die

⁵²Auch die Umsetzung dieser Bedingung stellte sich bei der Arbeit mit PXLab als recht zeitintensiv heraus.

Abbildung 4.16: Unterschiedliche Reaktionszeiten bei unterschiedlichen Targets

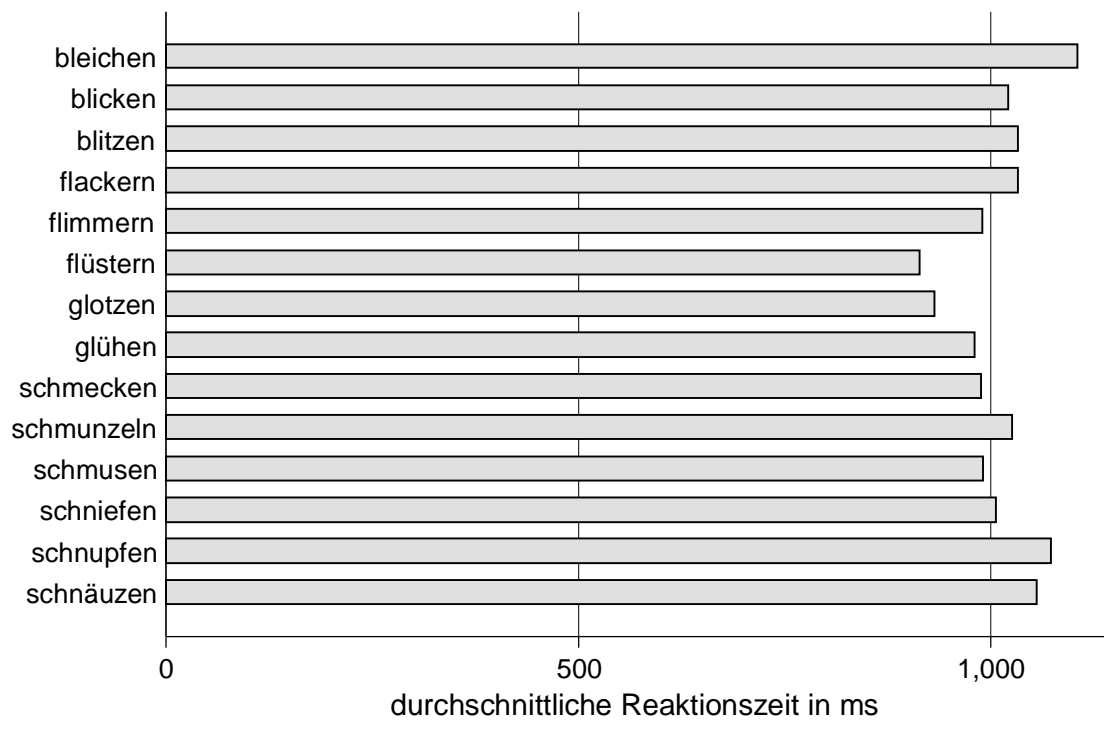


Tabelle 4.13: Einflussfaktoren auf die Reaktionszeit – Random-Effects-Modelle

	(1) Reaktionszeit	(2) Reaktionszeit
Target/Pseudo-Target	1779,4*** (-82,47)	-1802,1*** (-69,24)
Trialnummer	-0,454 (-1,96)	-0,746 (-1,09)
Trialnummer (quad.)	0,00501** (2,73)	0,00567 (1,07)
Phonästhem	-23,15** (-2,65)	-25,02* (-2,30)
Bedeutung	-41,73** (-3,24)	-38,88*** (-3,31)
Form	-2,427 (-0,30)	0,771 (0,07)
Wiederholung_2	-34,79*** (-3,50)	-41,85** (-3,14)
Wiederholung_3	-76,62*** (-7,60)	-74,00*** (-4,07)
Wiederholung_4	-71,65*** (-5,69)	-62,28* (-2,12)
Konstante	2857,0*** (521,18)	2892,4*** (92,73)
Beobachtungen	3363	3363

z-Werte in Klammern

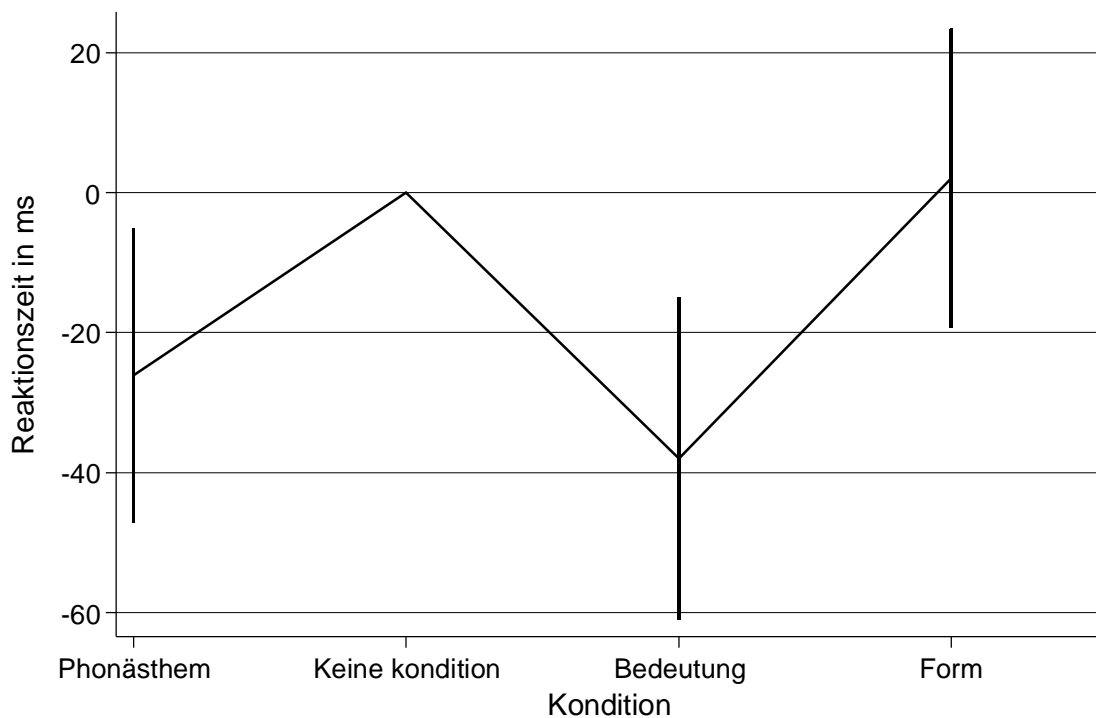
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hinweis: Wird der Relationseffekt in das Modell aufgenommen, muss für die 4 Konditionen, bei denen eine Beziehung zwischen Prime und Target besteht eine Referenzkategorie definiert werden, für die dann kein Effekt berechnet werden kann. Um dieses Problem zu umgehen, wurde ein Modell ohne Relationseffekt geschätzt. Die Anpassungsgüte R^2 kann bei Paneldatenmodellen nicht sinnvoll interpretiert werden und wird deshalb nicht berichtet.

Form-Beziehung weist dagegen keine signifikanten Unterschiede zur relationslosen Referenzkategorie auf. Interessant ist nun die Frage, ob sich die durchschnittliche Reaktionszeit in der Bedeutungsrelation signifikant von der phonästhemischen Relation unterscheidet.

Der Marginsplot (Abbildung 4.17) zeigt, dass sich die beiden Beziehungen nicht signifikant voneinander unterscheiden. Aufgrund des etwas stärkeren Effekts der Bedeutungsrelation kann ein Beleg für die besondere Qualität der Phonästhem-Beziehung nicht geleistet werden kann.

Abbildung 4.17: Marginsplot mit 95 % Konfidenzintervall für die Effekte der Konditionen



5 Zusammenfassung

Nachdem mit jedem Replikationsversuch Schritt für Schritt Störfaktoren detektiert und in den statistischen Modellen als Drittvariablen berücksichtigt worden sind (u.a. Randomisierungseffekte, Frequenz-Effekte, Lern- und Ermüdungseffekte, Fallzahlen), kann letztendlich der interessierende Effekt der phonästhemischen Kondition beurteilt werden. Als zentrales Ergebnis kann dabei der Einfluss der semantischen und phonästhemischen Relationen auf die Reaktionszeiten angesehen werden, der trotz eines Wechsels von der visuellen zur auditorischen Modalität bestätigt werden konnte.

Dass die Effekte der semantischen und phonästhemischen Konditionen sich auch bei einer effizienten Modellierung des Teilnehmereffekts und des Effekts der Wortlänge durch Random-Effects nicht signifikant voneinander unterscheiden, spricht dafür, dass Phonästhemie über den Bedeutungszusammenhang ihrer Lexeme hinaus *keine* kognitiv funktionale Qualität aufweisen.

Aus statistischer Sicht handelt es sich bei den Ergebnissen sowohl im Hinblick auf die Effekte der Bedeutungs- und Phonästhem-Konditionen als auch auf die der potentiellen Störvariablen um starke Effekte, die einen deutlichen Beitrag zur Erklärung der Reaktionszeiten liefern. Interessant ist zudem, dass bei Berücksichtigung potentieller Störvariablen ein relativ großer Anteil der Varianz der Reaktionszeiten erklärt werden kann. Dies wiederum spricht dafür, dass der Einfluss weiterer potentieller Störfaktoren, die nicht erfasst werden konnten, relativ gering sein dürfte.

Für die Theorie des Phonästhem ergeben sich darauf aufbauend bedeutsame Schlussfolgerungen. Nicht nur ist die unabhängige und von Bergen (2004, 2010) unbeachtete Studie von Rastle et al. (2000)¹ zu vergleichbaren Ergebnissen hinsichtlich der Bedeutungs- und Phonästhem-

¹Der Umstand, dass es in Rastle et al. (2000) überhaupt nicht um die Überprüfung kognitiver phonästhemischer Funktionalität gegangen ist, kann bereits als Hinweis auf die Validität dieser Teilergebnisse aufgefasst werden. Nichtsdestotrotz bietet es sich an auch diese Studie einem akribischen Replikationsversuch zu unterziehen.

Tabelle 5.1: Synopsis sämtlicher Replikationsversuche inklusive des Urexperiment von Bergen (2004)

CR	DV	FK	TN	SG	KD	RZ	FR	SA	FZ	SF
U	(FE)	(✓)	E	20	PÄ	606,70	–	–	–	✓
					PPÄ	658,70	–	–	–	x
					BD	642,70	–	–	–	x
					FM	668,20	–	–	–	x
					K	665,30	–	–	–	x
RV1_U		x	E	20	PÄ	688,82	1,00	123,57	82	x
					PPÄ	641,98	0,99	166,31	82	x
					BD	640,02	1,00	109,47	86	✓
					FM	645,02	0,98	98,47	84	✓
					K	683,50	0,96	127,14	79	✓
RV2_U	DE LE EE	x	E	21	PÄ	627,60	0,99	141,23	103	x
					PPÄ	647,44	0,79	120,10	78	x
					BD	645,11	0,98	131,48	98	x
					FM	632,65	0,99	131,05	103	x
					K	646,58	0,98	112,96	100	x
RV3_U	FE DE LE EE FZ	✓	D	18	PÄ	711,47	0,87	130,38	380	✓
					BD	712,46	0,94	129,39	404	✓
					FM	708,83	0,93	122,98	396	✓
					Kf	606,69	0,97	107,42	223	✓
					Ki	745,04	0,84	119,06	166	✓
	Kg	665,73	0,91	131,62	389	✓				
RV1_M	FE DE LE EE	✓✓	D	16	PÄ	639,30	0,98	138,74	203	✓
					BD	644,81	0,98	133,66	209	✓
					FM	647,26	0,99	130,28	214	x
					K	644,26	0,96	123,87	202	✓
RV1_E	FE DE LE EE FZ	✓✓	D	30	PÄ	1009,6	0,99	205,36	416	✓
					BD	979,0	0,99	213,34	417	✓
					FM	1023,2	0,98	225,71	412	x
					K	1032,9	0,98	204,39	413	✓

Erläuterung: **CR** = Code des Replikationsversuchs (RV = Replikationsversuch, U = Urexperiment Bergen (2004), M = Modifiziertes Urexperiment Bergen (2010), E = eigene Ergebnisse), **DV** = berücksichtigte Drittvariable (DE = Designeffekt, LE = Lerneffekt, EE = Ermüdungseffekt, FK = Frequenzeffekt), **FK** = Frequenzeffekt-Kontrolle (✓ = kontrolliert, ✓✓ = homogenisiert), **TN** = Teilnehmer (D = Muttersprache Deutsch, E = Muttersprache Englisch), **SG** = Samplegröße, **KD** = Kondition (PÄ = Phonästhem, PPÄ = Pseudo-Phonästhem, BD = Bedeutung, FM = Form, K = keine Relation; f = frequent, i = infrequent, g = gesamt), **RZ** = Mittelwert der Reaktionszeiten, **FR** = Fehlerrate, **SA** = Standardabweichung, **FZ** = Fallzahl, **SF** = Signifikanz (✓ = ja, x = nein)

Effekte gelangt wie die vorliegende Arbeit, was als Hinweis darauf gewertet werden kann, dass es sich bei den berichteten Effekten um robuste Effekte handelt. Auch lassen sich auf Grundlage der hier vorgestellten eigenen Ergebnisse die in Abschnitt 2.2 diskutierten kognitiven Modelle lexikalischer Wissensrepräsentation von Bybee (2001) und Johnson (1987, 2007) auf ihre Erklärungsadäquatheit hin beurteilen.

„*The semantic connections are the strongest and the most important in determining the closeness of the relations among words. [...] The strongest sort of relation between forms is this [sic] morphological relation which consists of parallel semantic and phonological connections. Morphological relations can be closer or more distant, depending upon three factors: 1) the degree of semantic relatedness, which is determined by the number of shared features, and the nature of shared features. 2) The extent of phonological similarity between the items (e.g. *sing* and *sang* are more closely related than *bring* and *brought*). 3) Word frequency: high-frequency words form more distant connections than low-frequency words. In the case of morphologically complex words, this means that high-frequency words undergo less analysis, and are less dependent on their related base words than low-frequency words.*“ (Bybee 1985:118, eigene Kursivierung).

Das in Abschnitt 2.2 entworfene Netzwerk-Modell (am Beispiel des Phonästhem -[V]RR-), das aus Bybees (1985, 2004) vom Konnektionismus beeinflussten Theorie des mentalen Lexikons folgt, wäre mit den Ergebnissen Bergens von Anfang an nicht vollständig vereinbar gewesen. Zwar kann Bergens Resultaten ein zwei-faktorieller Suchalgorithmus zugrunde gelegt werden, so dass sie im Hinblick auf den „number of shared features“-Aspekt in Bybees Modell zu passen scheinen², jedoch ist in ihrer Theorie der semantische Aspekt dem morphologischen übergeordnet. Für die empirische Arbeit bedeutet dies, dass überhaupt nicht

²Der Verarbeitungsvorteil gegenüber der einfaktoriellen Suchfunktion beruht dabei auf einer Reduktion der zu vergleichenden Elemente: Während innerhalb eines einfaktoriellen Suchalgorithmus jeder neu abgerufene Lexikoneintrag mit allen anderen gespeicherten Einträgen abgeglichen werden muss, wird dieser Abgleich innerhalb einer zweifaktoriellen Suche auf der Basis einer Vorauswahl durch den zweiten Faktor beschleunigt. Dieses Prinzip kann mit einer Spielkisten-Analogie veranschaulicht werden: es ist anzunehmen, dass bei der Suche mit geschlossenen Augen in einer Spielkiste, die mit diversen geometrischen Objekten wie Pyramiden, Zylindern, Quadern, Würfeln und Kugeln gefüllt ist, das manuelle Aussortieren aller Würfel (Faktor 1) länger dauern wird als wenn noch der Umstand hinzukommt, dass alle Würfel klebrig sind (Faktor 2). Der von Bergen (2004, 2010) berichtete signifikante Effekt der Phonästhem gegenüber allen anderen Relationen würde sich vor diesem Hintergrund aus einem kombinierten Verarbeitungsvorteil formaler *und* inhaltlicher Beziehungen gegenüber den einfaktoriellen Relationen der Vergleichsgruppen (für die z.B. nur ein inhaltlicher Zusammenhang charakteristisch ist) ergeben. Da weder Feldman (2000), auf deren morphologisches Priming dieser Schluss übertragen werden kann, noch Bergen alternative Theorien, die ihre Experimente begründen könnten, beisteuern, scheint zunächst nichts gegen eine Erklärung im Sinne der Spielkisten-Analogie zu sprechen.

davon auszugehen ist, in einem Priming-Experiment ausschließlich phonästhemische Effekte zu messen. Nicht nur im Hinblick auf die theoretischen Vorgaben, sondern auch angesichts der relativ starken und robusten Effekte semantischer Beziehungen in verwandten Primingstudien (u.a. Neely 1991, Marslen-Wilson et al. 1994b, Anderson & Holcomb 1995, Kotz et al. 2002), hätte Bergen (2004, 2010), der lediglich signifikante Effekte für die für ihn zentrale Phonästhem-Kondition berichtet, das Fehlen eines signifikanten Bedeutungseffekts zu denken geben müssen.

Obgleich Bergens Resultate sich innerhalb der eigenen Versuchreihe nicht replizieren ließen, spricht nichts dagegen, zumindest für das Morphempriming von einem kognitiven Verarbeitungsvorteil auf Grundlage eines zweifaktoriellen Algorithmus auszugehen. Zumindest solange keine plausiblere Theorie dafür vorgebracht werden kann, wie man sich ein kognitives Korrelat für Kategorialität überhaupt vorzustellen hat. Ein derartiger Ansatz hätte zudem den Vorteil, ökonomisch zu sein, wobei Ökonomie nicht mit Trivialität verwechselt werden sollte³. Einfache Erklärungen sind komplexen Theorien aus wissenschaftstheoretischer Sicht schon deswegen vorzuziehen, da sie leichter zu überprüfen sind (Popper 2002).

Die eigenen Ergebnisse dagegen sprechen für den Ansatz von Bybee (1985, 2004), die in ihrem Netzwerkmodell von einem übergeordneten Bedeutungseffekt ausgeht. Unter Berücksichtigung der für die Bedeutungsseite von Phonästhemem spezifizierten Image Schemas (vgl. Abschnitt 2.2) in Anlehnung an Johnson (1987, 2007) ergibt sich für die Theorie folgende Schlussfolgerung:

Die signifikanten Effekte der sich nicht signifikant voneinander unterscheidenden Konditionen der semantischen Beziehung und der phonästhemischen Beziehung deuten darauf hin, dass die Lexeme der phonästhemischen Targets zu ein und demselben Image Schema gehören wie die Targetlexeme der Bedingungskondition. Ein Beleg für eine besondere Qualität der Phonästhemie, die über den Bedeutungszusammenhang zwischen ihren Lexemen hinausgeht, kann auf Grundlage dieser Versuchsreihe nicht geleistet werden.

Obgleich nichts dagegen spricht Phonästhemie im Rahmen der konnektionistischen Schema-Theorie im Sinne von Generalisierungen, die als spontane Ordnungen aus Beziehungssystemen hervorgehen (vgl. Hayek 1952) als psychologisch real zu betrachten, geben die Resultate dieser umfassenden Empirie keinen Hinweis auf eine kognitive Funktionalität von Phonäs-

³Zur Bedeutung der Ökonomie-Maxime für die Sprachwandeltheorie vgl. Keller (2003), für die Rechtswissenschaften Friedmann (1999), für die Volkswirtschaftslehre Ramb et al. (1993), für die Biologie Kokko et al. (2002) und für die Psychologie McGonigle & Chalmers (1998, 2002) sowie die Ausführungen in Abschnitt 3.2.3.

themen – so wie es von kompositionellen Theorien der Morphologie vorausgesagt worden wäre.

Zu Beginn dieser Arbeit stand die Frage, ob „die ‚linguistische Evidenz‘ zur Begründung ‚psychologischer Realität‘“ (Chomsky 1981:111f.), in diesem Fall der psychologischen Realität des Phonästhem, ausreichen kann. Auch ohne den hier dargelegten theoretischen und empirischen Hintergrund lässt sich sagen, dass auf der Grundlage rein sprachlicher Argumente keine umfassenden Erkenntnisse über die kognitive Repräsentation von Phonästhem zu gewinnen sind. Angesichts der methodischen und methodologischen Unzulänglichkeiten einschlägiger psycholinguistischer Primingsstudien wird die psychologische Evidenz diesem Anspruch jedoch ebensowenig gerecht.

Mit der hier vorgestellten Methodologie soll ein Beitrag zur Überwindung dieses Dilemmas geleistet werden.

Literatur

1. ABELIN, A. 1999. *Studies in sound symbolism*. Göteborg University dissertation, Göteborg.
2. ABELIN, A. 2012. Relative frequency and semantic relations as organizing principles for the psychological reality of phonaesthemes. In: *Selected articles from UK-CLA meetings*, 1: 128–145.
3. ADAMS, V. 1973. *Introduction to modern English word-formation*. Longman, London.
4. ADAMS, V. 2001. *Complex Words in English*. Longman, London.
5. ALBERT, R. & KOSTER, C.J. 2002. *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch*. Narr, Tübingen.
6. ALEGRE, M. & GORDON, P. 1999. Rule-based versus associative processes in derivational morphology. In: *Brain and Language*, 68: 347-354.
7. ALLEN, P. A., MCNEAL, M. & KVAK, D. 1992. Perhaps the lexicon is coded as a function of word frequency. In: *Journal of Memory and Language*, 31: 826–844.
8. ÁLVAREZ, C. J., CARREIRAS, M. & TAFT, M. 2001. Syllables and morphemes: Contrasting frequency effects in Spanish. In: *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27: 545-555.
9. AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE (AAAS). 2008. Science update: Brain capacity, September 16th 2008.<http://www.scienceupdate.com/2008/09/brain-capacity/> (aufgerufen am 01.04.2012)
10. ANDERSON, J. R. 1976. *Language, memory, and thought*. Lawrence Erlbaum Associates, Hillsdale.
11. ANDERSON, J. E. & HOLCOMB, J. P. 1995. Auditory and visual semantic priming using different stimulus onset asynchronies: An event-related brain potential study. In: *Psychophysiology*, 32: 177–190.
12. ANDREWS, S. 1992. Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18: 234–254.

13. ANDREWS, S. 1996. Lexical retrieval and selection processes: Effects of transposed-letter confusability. In: *Journal of Memory and Language*, 35: 775-800.
14. ARGOUD, L. 2008. Les ‚mots en bl-‘ du lexique anglais : étude de la structuration des données dans une optique lexico cognitive. In: *Lexis*, 2: 43-76.
15. ARONOFF, M. 1981. *Word formation in Generative Grammar*. 2. Aufl. The MIT Press, Cambridge.
16. ASHER, J. E., LAMB, J. A., BROCKLEBANK, D., CAZIER, J. B., MAESTRINI, E., ADDIS, L., SEN, M., BARON-COHEN, S. & MONACO, A. P. 2009. A whole-genome scan and fine-mapping linkage study of auditory-visual synesthesia reveals evidence of linkage to chromosomes 2q24, 5q33, 6p12, and 12p12. In: *The American Journal of Human Genetics*, 84, 2: 1-7.
17. ATKINSON, R. C. & SHIFFRIN, R. M. 1968. Human memory: A proposed system and its control processes. In: SPENCE, K. W. & SPENCE, J. T. (Hrsg.), *The psychology of learning and motivation: Advances in research and theory*, vol. 2, 742-775. Academic Press, New York.
18. AUSTERLITZ, R. 1994. Finnish and Gilyak sound symbolism – an interplay between system and history. In: HINTON, L., NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound symbolism*, 249-260. Cambridge University Press, Cambridge.
19. BAAYEN, R. H. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge University Press, Cambridge.
20. BAAYEN, R. H., DIJKSTRA, T. & SCHREUDER, R. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. In: *Journal of Memory and Language*, 37: 94-117.
21. BAAYEN, R. H., WORM, L. H. & AYCOCK, J. 2007. Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. In: *The Mental Lexicon*, 2, 3: 419-463.
22. BADECKER, W. & CARAMAZZA, A. 1991. Morphological composition in the lexical output system. In: *Cognitive Neuropsychology*, 8: 335-367.
23. BALD, W.-D. 1977. *Testmethoden und Linguistische Theorie. Beispiele und Übungen zur Linguistischen Methodologie*. Niemeyer, Tübingen.
24. BALLMER, T. T. 1976. Inwiefern ist Linguistik empirisch? In: Wunderlich, D. (Hrsg.), *Wissenschaftstheorie der Linguistik*, 6-53. Athenäum, Kronberg.
25. BALLMER, T. T. 1997. *Deutsche Verben. Eine sprachanalytische Untersuchung des Deutschen Verbwortschatzes*. Gunter Narr, Tübingen.

26. BALOTA, D. A. 1983. Automatic semantic activation and episodic memory encoding. In: *Journal of Verbal Learning and Verbal Behavior*, 22: 88-104.
27. BALOTA, D. A., CORTESE, M., HUTCHISON, K., NEELY, J., NELSON, D., SIMPSON, G. & TREIMAN, R. 2007. *The English lexicon project: A webbased repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Washington University, Washington. <http://elexicon.wustl.edu/> (aufgerufen am 23.10.2014)
28. BARTLETT, F. C. 1932. *Remembering: A study in experimental and social psychology*. Cambridge University Press, Cambridge.
29. BERGEN, B. K. 2004. The psychological reality of phonaesthemes. In: *Language*, 80, 2: 291-311.
30. BERGEN, B. K. 2010. Phonaesthemes: Frequency and psychological reality. In: *Emory Sound Symbolism Workshop „Challenging the Arbitrariness of Language“*, Atlanta. <http://psychology.emory.edu/soundsymbolismworkshop2010/> (aufgerufen am 07.06.2012)
31. BERLIN, B. 1994. Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In: HINTON, L., NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound Symbolism*, 76-93. Cambridge University Press, Cambridge.
32. BERMAN J. M. 1961. Contribution on blending. In: *Zeitschrift für Anglistik und Amerikanistik*, 9: 278-281.
33. BERTRAM, H. 1994. Die Stadt, das Individuum und das Verschwinden der Familie. In: *Aus Politik und Zeitgeschichte*, 44, B 29/30: 15-35. <http://edoc.hu-berlin.de/humboldt-vl/bertram-hans/PDF/Bertram.pdf> (aufgerufen am 23.10.2014)
34. BERTRAM, R., BAAYEN, R. H. & SCHREUDER, R. 2000. Effects of family size for complex words. In: *Journal of Memory and Language*, 42: 390-405.
35. BIBLIOGRAPHISCHES INSTITUT GMBH. 2000. *Duden. Die deutsche Rechtschreibung*. 22. Aufl. Dudenverlag, Mannheim.
36. BIBLIOGRAPHISCHES INSTITUT GMBH. 2012. Duden Online. <http://www.duden.de/woerterbuch> (aufgerufen am 25.10.2014)
37. BLEVINS, J. P. & BLEVINS, J. 2009. *Analogy in grammar: Form and acquisition*. Oxford University Press, Oxford.
38. BLOOMFIELD, L. 1933. *Language*. Henry Holt, New York.
39. BLUST, R. A. 2003. The phonestheme ŋ in Austronesian languages. In: *Oceanic Linguistics*, 42, 1: 187-212.
40. BOERSMA, P. & HAMANN, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. In: *Phonology*, 25, 2: 217-270.

41. BOLINGER, D. 1950. Rime, assonance, and morpheme analysis. In: *Word*, 6: 117-36.
42. BOOIJ, G. 2004. Usage-based phonology. Review of J. L. Bybee, *Phonology and language use*. In: *Studies in Language*, 28: 225-237.
43. BOOIJ, G. 2005. *The grammar of words. An introduction to linguistic morphology*. Oxford University Press, Oxford.
44. BOURDIEU, P. 1987. *Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft*. Suhrkamp, Frankfurt a. M.
45. BOWERS, J. S. & MICHITA, Y. 1998. An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script Kanji–Hiragana priming. In: *Psychonomic Bulletin and Review*, 5: 259-264.
46. BOWERS, J. S. 2000. In defense of abstractionist theories of word identification and repetition priming. In: *Psychonomic Bulletin and Review*, 7: 83–99.
47. BOYLAND, J. T. 1996. *Morphosyntactic change in progress: a psycholinguistic approach*. University of California dissertation, Berkeley.
48. BRANG, D., ROUW, R., RAMACHANDRAN, V. S., & COULSON, S. 2011. Similarly shaped letters evoke similar colors in grapheme-color synesthesia. In: *Neuropsychologia*, 49: 1355-1358.
49. BRAUN, N. & GAUTSCHI, T. 2011. *Rational Choice Theory*. Juventa, München.
50. BRITISH NATIONAL CORPUS. 2007. Version 3 (BNC XML edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. <http://www.natcorp.ox.ac.uk/> (aufgerufen am 04.06.2012)
51. BROADBENT, D. E. & GREGORY, M. 1968. Visual perception of words differing in letter digram frequency. In: *Journal of Verbal Learning and Verbal Behavior*, 7: 569-571.
52. BROSCART, J. 2007. *Über die Sprachwissenschaft hinaus. Sprache und Linguistik aus transdisziplinärer Sicht*. LIT Verlag, Berlin.
53. BROWN, A. S. 1991. A review of the tip-of-the-tongue experience. In: *Psychological Bulletin*, 109, 2: 204-223.
54. BROWN, R. W., BLACK, A. H. & HOROWITZ A. E. 1955. Phonetic symbolism in natural languages. In: *Journal of Abnormal and Social Psychology*, 50: 388-393.
55. BROWN, R. W. & MCNEILL, D. 1966. The „tip of the tongue“ phenomenon. In: *Journal of Verbal Learning and Verbal Behavior*, 5, 4: 325–337.
56. BROWN R. W. & NUTTALL, R. 1959. Method in phonetic symbolism experiments. In: *Journal of Abnormal and Social Psychology*, 59: 441-445.

57. BUSSE, D. 2007. Sprache – Kognition – Kultur. Der Beitrag einer linguistischen Epistemologie zur Kognitions- und Kulturwissenschaft. In: *Jahrbuch der Heinrich-Heine-Universität Düsseldorf 2006/2007*: 267-279.
58. BUSSE, D. 2012. *Frame-Semantik - Ein Kompendium*. De Gruyter, Berlin.
59. BUTTERWORTH, B. 1983. Lexical representation. In: BUTTERWORTH, B. (Hrsg.), *Language Production*, vol. 2, 257-294. Academic Press, London.
60. BYBEE, J. 1985. *Morphology*. John Benjamins, Amsterdam.
61. BYBEE, J. 1988. Morphology as lexical organization. In: HAMMOND, M. & NOONAN, M. (Hrsg.), *Theoretical approaches to morphology*, 119-141. Academic Press, San Diego.
62. BYBEE, J. 1995. Regular morphology and the lexicon. In: *Language and Cognitive Processes*, 10: 425-455.
63. BYBEE, J. 2004. *Phonology and language use*. Cambridge University Press, Cambridge.
64. BYBEE, J. 2007. *Frequency of use and the organization of language*. Oxford University Press, Oxford.
65. BYBEE, J. & HOPPER, P. (Hrsg.) 2001. *Frequency and the emergence of linguistic structure*. John Benjamins, Amsterdam.
66. BYBEE, J. & MODER, C. L. 1983. Morphological classes as natural categories. In: *Language*, 59: 251-270.
67. CALLEJAS, A., ACOSTA, A., & LUPIANEZ, J. 2007. Green love is ugly: Emotions elicited by synaesthetic grapheme-color perceptions. In: *Brain Research*, 1127: 99-107.
68. CAMERER, C. F. & FEHR, E. 2004. Measuring social norms and preferences using experimental games: a guide for social scientists. In: HENRICH, J., BOYD, R., BOWLES, S., CAMERER, C., FEHR, E. & GINTIS, H. (Hrsg.), *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies*, 55-95. Oxford University Press, Oxford.
69. CAMPBELL, D. T. & STANLEY, J. C. 1973. *Experimental and quasi-experimental designs for research*. McNally, Chicago.
70. CARAMAZZA, A., LAUDANNA, A. & ROMANI, C. 1988. Lexical access and inflectional morphology. In: *Cognition*, 28: 297-332.
71. CARREIRAS, M., ÁLVAREZ, C. J. & VEGA, M. de 1993. Frequency and visual word recognition in Spanish. In: *Journal of Memory and Language*, 32: 766-780.

72. CARREIRAS, M., PEREA, M. & GRAINGER, J. 1997. Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. In: *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23: 857–871.
73. CARROLL, L. 2003. *Alice's adventures in wonderland and through the looking glass*. Illustrations by John Tenniel. Penguin Classics, London.
74. CHAMBERS, S. M. 1979. Letter and order information in lexical access. In: *Journal of Verbal Learning and Verbal Behavior*, 18: 225-241.
75. CHEESMAN, J. & MERIKLE, P. M. 1984. Priming with and without awareness. In: *Perception and Psychophysics*, 36: 387-395.
76. CHÉREAU, C., GASKELL, M. G. & DUMAY, N. 2007. Reading spoken words: Orthographic effects in auditory priming. In: *Cognition*, 102: 341–360.
77. CHIALANT, D. & CARAMAZZA, A. 1995. Where is morphology and how is it processed? The case of written word recognition. In: FELDMAN, L. B. (Hrsg.), *Morphological aspects of language processing*, 55-76. Lawrence Erlbaum Associates, Hillsdale.
78. CHOLIN, J., LEVELT, W. J. M., & SCHILLER, N. O. 2006. Effects of syllable frequency in speech production. In: *Cognition*, 99: 205-235.
79. CHOMSKY, N. 1981. *Regeln und Repräsentationen*. Suhrkamp, Frankfurt a. M.
80. CHRISTIANSEN, M. H. 2010. A possible division of labor between arbitrary and aystematic sound-meaning mappings in language. In: *Emory Sound Symbolism Workshop „Challenging the Arbitrariness of Language“*, Atlanta. <http://psychology.emory.edu/soundsymbolismworkshop2010/> (aufgerufen am 07.06.2012)
81. CHUNG, I. 2009. Suppletive verbal morphology in Korean and the mechanism of vocabulary insertion. In: *Journal of Linguistics*, 45: 533–567.
82. CIENKI, A. 1998. STRAIGHT: An image schema and its metaphorical extensions. In: *Cognitive Linguistics*, 9: 107-149.
83. CLAHSSEN, H., ROTHWEILER, M., WOEST, A. & MARCUS, G. 1992. Regular and irregular inflection in the acquisition of German noun plurals. In: *Cognition*, 45: 225-255.
84. CLARK, H. H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. In: *Journal of Verbal Learning and Verbal Behavior*, 12: 335–359.
85. CLAUSNER, T. & CROFT, W. 1999. Domains and image schemas. In: *Cognitive Linguistics*, 10: 1-31.
86. COLEMAN, J. S. 1991. *Grundlagen der Sozialtheorie, Band 1: Handlungen und Handlungssysteme*. Oldenbourg, München.

87. COLTHEART, M., DAVELAAR, E., JONASSON, J.F. & BESNER, D. 1977. Access to the internal lexicon. In: DORNIC, S. (Hrsg.), *Attention and performance VI*, 535-555. Lawrence Erlbaum Associates, Hillsdale.
88. COLTHEART, M., RASTLE, K., PERRY, C., LANGDON, R. & ZIEGLER, J. 2001. The DRC model: A model of visual word recognition and reading aloud. In: *Psychological Review*, 108: 204–256.
89. COSMAS II. CORPUS SEARCH, MANAGEMENT AND ANALYSIS SYSTEM. 1991-2010. Institut für Deutsche Sprache, Mannheim. <http://www.ids-mannheim.de/cosmas2/> (aufgerufen am 10.06.2012)
90. CROFT, W. 2000. *Explaining language change: An evolutionary approach*. Longman, London.
91. CROFT, W. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press, Oxford.
92. CYTOWIC, R. E. 1995. Synesthesia, phenomenology & neuropsychology: a review of current knowledge. In: *Psyche*, 2, 10. <http://www.theassc.org/files/assc/2346.pdf> (aufgerufen am 23.10.2014)
93. DAMIAN, M.F. & MARTIN, R. C. 1999. Semantic and phonological codes interact in single word production. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25: 345–361.
94. DAUGHERTY, K. & SEIDENBERG, M. S. 1992. Rules or connections? The past tense revised. In: *Fourteenth Annual Conference of the Cognitive Science Society*: 259-64. Lawrence Erlbaum Associates, Hillsdale.
95. DAUGHERTY, K. & SEIDENBERG, M. S. 1994. Beyond rules and exceptions: A connectionist approach to inflectional morphology. In: LIMA, S.D., CORRIGAN, R.L. and IVERSON, G.K. (Hrsg.), *The reality of linguistic rules*, 353-388. John Benjamins, Amsterdam.
96. DAVIES, M. 2008-. *The Corpus of Contemporary American English: 425 million words, 1990-present*. <http://corpus.byu.edu/coca/> (aufgerufen am 04.06.2012)
97. DAVIS, C. 2005. N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. In: *Behavior Research Methods*, 37: 65-70.
98. DAVIS, C., CASTLES, A. & IAKOVIDIS, E. 1998. Masked homophone and pseudohomophone priming in children and adults. In: *Language and Cognitive Processes*, 13: 625-651.
99. DAVIS C. & LUPKER S.J. 2006. Masked inhibitory priming in English: Evidence for lexical inhibition. In: *Journal of Experimental Psychology: Human Perception and Performance*, 32: 668-687.

100. DAVIS, R. 1961. The fitness of names to drawings. A cross-cultural study in Tanganyika. In: *British Journal of Psychology*, 52: 259-268.
101. DER DIGITALE GRIMM. 2004. Elektronische Ausgabe der Erstbearbeitung. Digitalisiert von Hans-Werner Bartz, Thomas Burch, Ruth Christmann, Kurt Gärtner, Vera Hildenbrandt, Thomas Schares und Klaudia Wegge. Hg. vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften. Frankfurt am Main.
102. DETGES, U. 2001. *Grammatikalisierung. Eine kognitiv-pragmatische Theorie, dargestellt am Beispiel romanischer und anderer Sprachen*. Habilitationsschrift. Universität Tübingen, Tübingen.
103. DEUTSCH-PLATTDEUTSCH-WÖRTERBUCH. 2012. <http://www.deutsch-plattdeutsch.de/> (aufgerufen am 21.08.2012)
104. DHAR, R. 1997. Context and task effects on choice deferral. In: *Marketing Letters*, 8, 1: 119–130.
105. DIEKMANN, A. 2001. *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Rowohlt, Hamburg.
106. DIEKMANN, A. 2008. Soziologie und Ökonomie: Der Beitrag experimenteller Wirtschaftsforschung zur Sozialtheorie. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 60: 528-550.
107. DIEPENDAELE, K., SANDRA, D. & GRAINGER, J. 2009. Semantic Transparency and Masked Morphological Priming: The Case of Prefixed Words. In: *Memory and Cognition*, 37: 895-908.
108. DIFFLOTH, G. 1994. i: big, a: small. In: HINTON, NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound symbolism*, 107-114. Cambridge University Press, Cambridge.
109. DIMOCK, E. 1957. Symbolic forms in Bengali. In: *Bulletin of the Deccan College Research Institute*, 18: 22-29.
110. DITTMAR, N. 2005. Möglichkeiten und Grenzen einer soziolinguistischen Theorie. In: AMMON, U., DITTMAR, N., MATTHEIER, K. J. & TRUDGILL, P. (Hrsg.), *Soziolinguistik. Ein internationales Handbuch*, Bd. 2, 930-945. De Gruyter, Berlin.
111. DRELLISHAK, S. 2006. Statistical techniques for detecting and validating phonesthemes. <http://depts.washington.edu/uwcl/matrix/sfd/Drellishak%20-%20Phonesthemes.pdf> (aufgerufen am 23.10.2014)
112. DREWS, E. & ZWITSERLOOD, P. 1995. Orthographic and morphological similarity in visual word recognition. In: *Journal of Experimental Psychology: Human Perception and Performance*, 21, 5: 1098-1116.

113. DÜRSCHIED, C. 2005. E-Mail – verändert sie das Schreiben? In: SIEVER, T., SCHLOBINSKI, P., RUNKEHL, J. (Hrsg.), *Websprache.net. Sprache und Kommunikation im Internet*, 85-97. De Gruyter, Berlin.
114. DUMAY, N., BENRAISS, A., BARRIOL, B., COLIN, C., RADEAU, M., BESSON, M. 2001. Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. In: *Journal of Cognitive Neuroscience*, 13, 1: 121-143.
115. DURAND, M. 1961. Les impressifs en Vietnamien, étude préliminaire. In: *Bulletin de la Societe des Etudes Indo Chinoises*, 36, 1: 5-51.
116. EDDINGTON, D. 2000. Analogy and the dual-route model of morphology. In: *Lingua*, 110: 281–298.
117. ELLIS H. & ELLIS, A. 1998. Why we study... repetition priming. In: *The Psychologist*. http://www.thepsychologist.org.uk/archive/archive_home.cfm/volumeID_11-editionID_68-ArticleID_284-getfile_getPDF/thepsychologist%5Crepitition.pdf (aufgerufen am 14.11.2013)
118. ELMAN, J. L., BATES, E. A., JOHNSON, M. H., KARMILOFF-SMITH, A., PARISI, D. & PLUNKETT, K. 1996. *Rethinking innateness: A connectionist perspective on development*. The MIT Press, Cambridge.
119. EVANS, V. & GREENE, M. 2006. *Cognitive Linguistics: An Introduction*. Lawrence Erlbaum Associates, Hillsdale.
120. EVERS, M. & GROLLE, J. 2006. Es war der Höhenrausch. Interview mit Andreas Diekmann. In: *Der Spiegel*, 2. <http://www.spiegel.de/spiegel/print/d-45280087.html> (aufgerufen am 02.06.2012)
121. FAHRMEIR, L., KÜNSTLER, R., PIGEOT, I. & TUTZ, G. 2004. *Statistik – Der Weg zur Datenanalyse*. Springer, Berlin.
122. FEATHERSTONE, S. & STERNEFELD, W. 2007. *Roots. Linguistics in search of ist Evidential Base*. De Gruyter, Berlin.
123. FELDMAN, L. B. (Hrsg.) 1995. *Morphological aspects of language processing*. Lawrence Erlbaum Associates, Hillsdale.
124. FELDMAN, L. B. 2000. Are morphological effects distinguishable from the effects of shared meaning and shared form? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26: 1431-1444.
125. FELDMAN, L. B. & PROSTKO, B. 2002. Graded aspects of morphological processing: Task and processing time. In: *Brain and Language*, 81: 12-27.
126. FELDMAN, L. B. & SOLTANO, E. G. 1999. Morphological priming: The role of prime duration, semantic transparency and affix position. In: *Brain and Language*, 60: 33-39.

127. FELDMAN, L. B., SOLTANO, E. G., PASTIZZO, M. J. & FRANCIS, S. E. 2004. What do graded effects of semantic transparency reveal about morphological processing? In: *Brain and Language*, 90: 17-30.
128. FIELD, J. 2004. An insight into listeners' problems: too much bottom-up or too much top-down? In: *System* 32: 363– 377.
129. FILLMORE, C. J. 1977. Scenes-and-frames semantics. In: ZAMBOLLI, A. (Hrsg.), *Linguistic Structure Processing*, 55–82. North Holland Publishing Company, Amsterdam.
130. FILLMORE, C. J. 1985. Frames and the semantics of understanding. In: *Quaderni di Semantica*, 6, 2: 222-254.
131. FILLMORE, C. J. & ATKINS, B. T. S. 1992. Towards a frame-based organization of the lexicon: the semantics of RISK and its neighbors. In: LEHRER, A. & KITTAY, E. (Hrsg.), *Frames, fields, and contrasts: new essays in lexical organization*, 75-102. Lawrence Erlbaum Associates, Hillsdale.
132. FISCHER, K. & STEFANOWITSCH, A. (Hrsg.) 2006. *Konstruktionsgrammatik: Von der Anwendung zur Theorie*. Stauffenberg, Tübingen.
133. FIRTH, J. R. 1930. *Speech*. Oxford University Press, London.
134. FITCH, W. T. 2000. The evolution of speech: a comparative review. In: *Trends in Cognitive Sciences*, 4: 258-267.
135. FORSTER, K. I., DAVIS, C., SCHOKNECHT, C. & CARTER, R. 1987. Masked priming with graphemically related forms: Repetition or partial activation? In: *Quarterly Journal of Experimental Psychology*, 39: 211-251.
136. FORSTER, K. I. & SHEN, D. 1996. No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 696–713.
137. FOX, C. 1935. An experimental study of naming. In: *The American Journal of Psychology*, 47, 4: 545-579.
138. FRAUENFELDER, U. H. & PETERS, G. 1998. Simulating the time course of word recognition: An analysis of lexical competition in TRACE. In: GRAINGER, J. & JACOBS, A. M. (Hrsg.), *Symbolic/Connectionist approaches to human cognition*, 101-146. Lawrence Erlbaum Associates, Mahwah.
139. FRIEDERICI, A. D., PFEIFER, E. & HAHNE, A. 1993. Event-related brain potential during natural speech processing: Effects of semantic, morphological and syntactic violations. In: *Cognitive Brain Research*, 1,3: 183–192.
140. FRIEDERICI, A. D. & WESSELS, J. M. 1993. Phonotactic knowledge and its use in infant speech perception. In: *Perception and Psychophysics*, 54: 287–295.

141. FRIEDMANN, D. D. 1999. *Hidden order: the economics of everyday life*. Harper Collins, New York.
142. FROMKIN, V., RODMAN, R. & HYAMS, N. 2003. *An Introduction to Language*. 7. Aufl. Heinle, Boston.
143. FROST, R., DEUTSCH, A., GILBOA, O., TANNENBAUM, M. & MARSLEN-WILSON, W. D. 2000. Morphological priming: Dissociation of phonological, semantic, and morphological factors. In: *Memory and Cognition*, 28: 1277-1288.
144. FROST, R., FORSTER, K. I. & DEUTSCH, A. 1997. What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23: 829-856.
145. FUKUDA, H. 2003. *Jazz Up Your Japanese with Onomatopoeia*. Kodansha International, Tokyo.
146. GADENNE, V. 1976. *Die Gültigkeit psychologischer Untersuchungen*. Kohlhammer, Stuttgart.
147. GÄRTNER, H.-M. 2004. On the OT-status of „Unambiguous Encoding“. In: BLUTNER, R., ZEEVAT, H. (Hrsg.), *Optimality Theory and pragmatics*, 154-172. Palgrave Macmillan, Houndmills.
148. GARCIA-ALBEA, J. E., SANCHEZ-CASAS, R. & IGOA, J. M. 1998. The contribution of word form and meaning to language processing in Spanish: Some evidence from monolingual and bilingual studies. In: HILLERT, D. (Hrsg.), *Sentence processing: A cross-linguistic perspective*, 183–209. Academic Press, New York.
149. GEBELS, G. 1969. An investigation of phonetic symbolism in different cultures. In: *Journal of Verbal Learning*, 8: 310-312.
150. GERNERT, D. 2007. Ockham's razor and its improper use. In: *Journal of Scientific Exploration*, 21,1: 135–140.
151. GIBBS, R. W. Jr., BEITEL, D. A., HARRINGTON, M. & SANDERS, P. E. 1994. Taking a stand on the meanings of stand: Bodily experience as motivation for polysemy. In: *Journal of Semantic*, 11: 231-251.
152. GIRAUDO, H. & GRAINGER, J. 2000. Effects of prime word frequency and cumulative root frequency in masked morphological priming. In: *Language and Cognitive Processes*, 15: 421-444.
153. GOLDBERG, A. E. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, Chicago.
154. GOLDBERGER, A. 1991. *A course in Econometrics*. Harvard University Press, Harvard.

155. GOLDINGER, S.D. 1998. Echoes of echoes? An episodic theory of lexical access. In: *Psychological Review*, 105: 251-279.
156. GOLDINGER, S.D. & AZUMA, T. 2004. Episodic memory reflected in printed word naming. In: *Psychonomic Bulletin and Review*, 11: 716-722.
157. GONNERMAN, L. M., SEIDENBERG, M. S. & ANDERSEN, E. S. 2007. Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. In: *Journal of Experimental Psychology: General*, 136: 323-345.
158. GORR, W. L., NAGIN, D. & SZCZYPULA, J. 1994. Comparative study of artificial neural network and statistical models for predicting student grade point averages. In: *International Journal of Forecasting*, 10: 17-34.
159. GRAINGER, J. 1990. Word frequency and neighborhood frequency effects in lexical decision and naming. In: *Journal of Memory and Language*, 29: 228-244.
160. GRAINGER, J., COLE, P. & SEGUI, J. 1991. Masked morphological priming in visual word recognition. In: *Journal of Memory and Language*, 30: 370-384.
161. GRAINGER, J. & FERRAND, L. 1994. Phonology and orthography in visual word recognition: Effects of masked homophone primes. In: *Journal of Memory and Language*, 33: 218-233.
162. GRAINGER, J. & JACOBS, A. M. 1996. Orthographic processing in visual word recognition: A multiple read-out model. In: *Psychological Review*, 103: 518-565.
163. GRAINGER, J., MUNEAX, M., FARIOLI, F. & ZIEGLER, J. C. 2005. Effects of phonological and orthographic neighbourhood density interact in visual word recognition. In: *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58A, 6: 981-998.
164. GRAINGER, J., O'REGAN, K., JACOBS, A. & SEGUI, J. 1989. On the role of competing word units in visual word recognition: The neighborhood frequency effect. In: *Perception and Psychophysics*, 45: 189-195.
165. GRIES, S. T. 2008. *Statistik für Sprachwissenschaftler*. Vandenhoeck & Ruprecht, Göttingen.
166. GRIMM, H. & ENGELKAMP, J. 1981. *Sprachpsychologie. Handbuch und Lexikon der Psycholinguistik*. Erich Schmidt, Berlin.
167. GRIMM, J. & GRIMM, W. 1854-1960. *Deutsches Wörterbuch*. S. Hirzel, Leipzig.
168. GROSS, M., SAY, T., KLEINGERS, M., CLAHSSEN, H. & MUENTE, T. F. 1998. Human brain potentials to violations in morphologically complex Italian words. In: *Neuroscience Letters*, 241: 83-86.

169. HACKL, P. 2005. *Einführung in die Ökonometrie*. Pearson, München.
170. HAHN, U. & NAKISA, R. C. 2000. German inflection: Single route or dual route? In: *Cognitive Psychology*, 41: 313-360.
171. HAMANO, S. 1998. *The sound-symbolic system of Japanese*. Cambridge University Press, Cambridge.
172. HAMILTON, L. 2009. *Statistics with Stata*. 10th edition. Cengage Learning, Belmont.
173. HAMPE, B. (Hrsg.) 2005. *From perception to meaning: image schemas in Cognitive Linguistics*. De Gruyter, Berlin.
174. HANN, T. H. & STEURER, E. 1996. Much ado about nothing? Exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data. In: *Neurocomputing*, 10: 323–339.
175. HARM, M. W. & SEIDENBERG, M. S. 2004. Computing the meanings of words in reading: Division of labor between visual and phonological processes. In: *Psychological Review*, 111: 662-720.
176. HARPER, D. 2001-12. *Online Etymology Dictionary*. <http://www.etymonline.com/index.php> (aufgerufen am 09.06.2012)
177. HARRISON, K. D. 2004. South Siberian sound symbolism. In: VAJDA, E. J. (Hrsg.), *Languages and Prehistory of Central Siberia*, 199-214. John Benjamins, Amsterdam.
178. HASPELMATH, M. 2002. *Understanding Morphology*. Oxford University Press, Oxford.
179. HAY, J. 2001. Lexical Frequency in Morphology: Is Everything Relative? In: *Linguistics*, 39: 1041-1070.
180. HAYEK, F. A. von 1952. *The sensory order: An inquiry into the foundations of theoretical psychology*. Routledge & Kegan Paul PLC, London.
181. HAYS, T. E. 1994. Sound symbolism, onomatopoeia, and New Guinea frog names. In: *Journal of Linguistic Anthropology*, 4: 153-174.
182. HEBB, D. O. 1949. *The organization of behavior*. Wiley & Sons, New York.
183. HELLWIG, E., KLIMEK, J. & ATTIN, T. 2009. *Einführung in die Zahnerhaltungskunde*. 5. Auflage. Deutscher Zahnärzterverlag, Köln.
184. HERBERG, D., KINNE, M. & STEFFENS, D. 2004. *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen*. Unter Mitarbeit von Elke Tellenbach und Doris al-Wadi. De Gruyter, Berlin. <http://www.owid.de/wb/neo/start.html> (aufgerufen am 09.06.2012)
185. HOCKETT, C. 1954. Two models of grammatical description. In: *Word*, 10: 210–234.

186. HOFFMANN, K. 1952. Wiederholende Onomatopoetika im Altindischen. In: *Indogermanische Forschungen*, 60: 254-264.
187. HOHENHAUS, P. 1996. *Ad-hoc-Wortbildung. Terminologie, Typologie und Theorie kreativer Wortbildung im Englischen*. Peter Lang, Frankfurt a. M.
188. HOLTHAUSEN, K. & ZICHE, P. 2007. Neuronale Netze für die Geisteswissenschaften. In: *Akademie Aktuell – Zeitschrift der Bayerischen Akademie der Wissenschaften*, 1: 32-35.
189. HONDRICH, K. O. 2001. Mehrheitsmoral und Elitenmoral. In: *Merkur*, 55, 7: 572-585.
190. HORN, L. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: SCHIFFRIN, D. (Hrsg.), *Meaning, form and use in context*, 11-42. Georgetown University Press, Washington.
191. HOUSEHOLDER, F. W. 1946. On the problem of sound and meaning, an English phonestheme. In: *Word*, 2: 83-84.
192. HOUSEHOLDER, F. W. 1987. Old bottles or the limits of arbitrariness. In: *Studia Linguistica*, 32: 97-106.
193. HUMPHREYS, G., EVETT, L. & QUINLAN, P. 1990. Orthographic processing in visual word identification. In: *Cognitive Psychology*, 22: 517-560.
194. HUTCHINS, S. S. 1998. *The psychological reality, variability, and compositionality of English phonesthemes*. Emory University dissertation, Atlanta.
195. IMAI, M., KITA, S. & KANTARTZIS, K. 2010. Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. In: SMITH, A. D. M., SCHOUWSTRA, M., BOER, B. de & SMITH, K. (Hrsg.), *Proceedings of the 8th international conference on the evolution of language (EVOLANG8)*, 206-213. World Scientific, Singapore.
196. INTERNATIONAL COMPUTER SCIENCE INSTITUTE (ICSI). 2012. *The FrameNet project*. <https://framenet.icsi.berkeley.edu/fndrupal/home> (aufgerufen am 23.06.2012)
197. IRTEL, H. 2007. *PXLab: The Psychological Experiments Laboratory*. Version 2.1.11. <http://www.pxlab.de/> (aufgerufen am 09.11.2013)
198. JACOBY, L. L. 1983. Remembering the data: Analyzing interactive processes in reading. In: *Journal of Verbal Learning and Verbal Behavior*, 22: 485-508.
199. JAEGER, J. J., LOCKWOOD, A. H., KEMMERER, D. L., VALIN Jr., R. D. van, MURPHY, B. W. & KHALAK, H. G. 1996. A positron emission tomographic study of regular and irregular verb morphology in English. In: *Language*, 72: 451-497.

200. JESPERSEN, O. 1928. *Language. Its nature, development, and origin*. George Allen & Unwin, London. http://www.archive.org/stream/languageitsnatur00jespuoft/languageitsnatur00jespuoft_djvu.txt (aufgerufen am 12.07.2012)
201. JOANISSE, M. F. & SEIDENBERG, M. S. 1999. Impairments in verb morphology after brain injury: A connectionist model. In: *Proceedings of the National Academy of Sciences (PNAS)*, 96: 7592–7597.
202. JOHNSON, K. 2008. *Quantitative methods in linguistics*. Wiley-Blackwell, Oxford.
203. JOHNSON, M. 1987. *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press, Chicago.
204. JOHNSON, M. 2005. The philosophical significance of image schemas. In: HAMPE, B. (Hrsg.), *From perception to meaning: image schemas in Cognitive Linguistics*, 15–33. De Gruyter, Berlin.
205. JOHNSON, M. 2007. *The meaning of the body*. University Of Chicago Press, Chicago.
206. JONG, N. H. de, SCHREUDER, R. & BAAYEN, R. H. 2000. The morphological family size effect and morphology. In: *Language and Cognitive Processes*, 15: 329–365.
207. JOSEPH, B. D. 1987. On the use of iconic elements in etymological investigation: Some case studies from Greek. In: *Diachronica*, 4: 1–26.
208. JOSEPH, B. D. 1991. A Greek Perspective on the Question of Arbitrariness of Linguistic Signs. In: *Modern Greek Studies year book*, 7: 335-352.
209. JOSEPH, B. D. 1994. Modern Greek ts: beyond sound symbolism. In: HINTON, L., NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound Symbolism*, 222-236. – Cambridge University Press, Cambridge.
210. KAKEHI, H., TAMORI, I. & SCHOURUP, L. 1996. *Dictionary of Iconic Expressions in Japanese*. De Gruyter, Berlin.
211. KAMANN, W. K. 2001. *Die Goldhämmerfüllung – alte Kamelle aber topaktuell*. In: *Zahnärztliche Mitteilungen (ZM)*, 91: 26-31.
212. KAMANN, W. K. & GÄGLER, P. 1999. Zur Funktionszeit von Amalgam, Composite und Goldhämmerfüllungen. In: *ZWR - Das Deutsche Zahnärzteblatt*, 108: 270-273.
213. KANT, I. 1990. *Kritik der reinen Vernunft*. Nach der 1. und 2. Orig.-Ausgabe hrsg. von Raymund Schmidt. 3. Aufl. Meiner, Hamburg.
214. KARUNANITHI, N., GRENNEY, W. J., WHITLEY, D. & BOVEE, K. 1994. Neural networks for river flow prediction. In: *Journal of Computing in Civil Engineering*, 8, 2: 201–220.
215. KATAMBA, F. 1993. *Morphology*. Palgrave, London.

216. KAY, P. 1997. *Words and the grammar of context*. Stanford University Press, Stanford.
217. KAY, P. & FILLMORE, C. J. 1999. Grammatical constructions and linguistic generalizations: the What's X Doing Y? construction. In: *Language*, 75, 1: 1-33.
218. KELLER, R. 2003. *Sprachwandel: Von der unsichtbaren Hand in der Sprache*. UTB, Stuttgart.
219. KELLER, R. 2009. Konventionen, Regeln, Normen. Zum ontologischen Status natürlicher Sprachen. In: KONOPKA, M. & STRECKER, B. (Hrsg.), *Deutsche Grammatik - Regeln, Normen, Sprachgebrauch*, 9-22. De Gruyter, Berlin.
220. KEMPLEY, S. T. & MORTON, J. 1982. The effects of priming with regularly and irregularly related words in auditory word recognition. In: *British Journal of Psychology*, 73: 441-454.
221. KEPSEK, S. & REIS, M. (Hrsg.) 2005. *Linguistic evidence. Empirical, theoretical and computational perspectives*. – De Gruyter, Berlin.
222. KERR, N. L. 1998. HARKing: Hypothesizing after the results are known. In: *Personality and Social Psychology*, 2: 196-217.
223. KIELAR, A. & JOANISSE, M. F. 2011. The role of semantic and phonological factors in word recognition: An ERP cross-modal priming study of derivational morphology. In: *Neuropsychologia*, 49, 2: 161-177.
224. KIELAR, A. & JOANISSE, M. F. & HARE, M. L. 2008. Priming English past tense verbs: Rules or statistics? In: *Journal of Memory and Language*, 58: 327-346.
225. KIM, K.-O. 1977. Sound symbolism in Korean. In: *Journal of Linguistics*, 13: 67-75.
226. KIRCHGÄSSNER, G. 2008. *Homo oeconomicus: Das ökonomische Modell individuellen Verhaltens und seine Anwendung in den Wirtschafts- und Sozialwissenschaften*. Mohr-Siebeck, Tübingen.
227. KLINK, R. R. 2000. Creating brand names with meaning. The use of sound symbolism. In: *Marketing Letters*, 11: 5-20.
228. KLINK, R. R. 2003. Creating meaningful brands: The relationship between brand name and brand mark. In: *Marketing Letters*, 14, 3: 143-157.
229. KNOWLES, G. 1987. *Patterns of spoken English: An introduction to English phonetics*. Longman, London.
230. KNUTH, D. E. 1973. *The art of computer programming, vol. 3: Sorting and searching*. Addison-Wesley, Reading.
231. KÖHLER, W. 1929. *Gestalt psychology*. Liveright, New York.

232. KÖHLER, W. 1947. *Gestalt psychology: an introduction to new concepts in modern psychology*. A revised edition of the 1929 book. Liveright, New York.
233. KOHN, E. O. 2005. Runa realism: Upper Amazonian attitudes to nature knowing. In: *Ethnos*, 70, 2: 179–196.
234. KOKKO, H., JOHNSTONE, R. A. & WRIGHT, J. 2002. The evolution of parental and alloparental effort in cooperatively breeding groups: when should helpers pay to stay? In: *Behavioral Ecology*, 13, 3: 291-300.
235. KORIAT, A. & LEVY, I. 1977. The symbolic implications of vowels and of their orthographic representations in two natural languages. In: *Journal of Psycholinguistic Research*, 6, 2: 93-103.
236. KOTZ, S. A. 1998. Comparing the auditory and visual sequential priming paradigm: An event-related potential study. In: *5th Meeting of the Cognitive Neuroscience Society*. http://cognet.mit.edu/library/conferences/view?conference_id=1 (aufgerufen am 04.06.2012).
237. KOTZ, S. A., CAPPA, S. F., CRAMON, D. Y. von & FRIEDERICI, A. D. 2002. Modulation of the lexical-semantic network by auditory semantic priming: An event-related functional MRI study. In: *NeuroImage*, 17, 4: 1761-1772.
238. KOUIDER, S. & DUPOUX, E. 2009. Episodic accessibility and morphological processing: Evidence from long-term auditory priming. In: *Acta Psychologica*, 130, 1: 38–47.
239. KUCERA, H. & FRANCIS, W. N. 1967. *Computational Analysis of Present-day American English*. Brown University Press, Providence.
240. KUHN, T. 1996. *The structure of scientific revolutions*. 3rd edition. University of Chicago Press, Chicago.
241. KUIKEN, F. & VEDDER, I. 2007. Task complexity and measures of linguistic performance in L2 writing. In: *International Review of Applied Linguistics in Language Teaching*, 45: 261–284.
242. KUPERMAN, V., BERTRAM, R. & BAAYEN, R. H. 2008. Morphological dynamics in compound processing. In: *Language and Cognitive Processes*, 23: 1089-1132.
243. LABOV, W. 1963. The social motivation of a sound change. In: *Word*, 18: 1-42.
244. LABOV, W. 2006. *The Social Stratification of English in New York City*. Second edition. Cambridge University Press, Cambridge.
245. LAKOFF, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago University Press, Chicago.
246. LAKOFF, G. & TURNER, M. 1989. *More than cool reason: A field guide to poetic metaphor*. University of Chicago Press, Chicago.

247. LAKOFF, G. & JOHNSON, M. 1999. *Philosophy in the flesh*. Basic Books, New York.
248. LANGACKER, R. W. 1987. *Foundations of Cognitive Grammar, vol. I*. Stanford University Press, Stanford.
249. LANGACKER, R. W. 1991. *Foundations of Cognitive Grammar, vol. II*. Stanford University Press, Stanford.
250. LANGACKER, R. W. 2002. *Concept, image, and symbol*. De Gruyter, Berlin.
251. LATHA, P., GANESAN, L. & ANNADURAI, S. 2009. Face recognition using neural networks. In: *Signal Processing: An International Journal (SPIJ)*, 3, 5: 153-160.
252. LAWLER, J. 1990. Women, men, and bristly things: The phonosemantics of the BR-
assonance in English. In: *Michigan Working Papers in Linguistics*, 1, 1: 27-43.
253. LEHTONEN, M., NISKA, H., WANDE, E., NIEMI, J. & LAINE, M. 2006. Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. In: *Journal of Psycholinguistic Research*, 35, 2: 121-146.
254. LEO DICTIONARY TEAM 2006-12. *LEO - Deutsch-Englisches Wörterbuch*. <http://dict.leo.org/> (aufgerufen am 09.06.2012)
255. LEVELT, W. J. M., ROELOFS, A., & MEYER, A. S. 1999. A theory of lexical access in speech production. In: *Behavioral and Brain Sciences*, 22: 1-75.
256. LEVELT, W. J. M., & WHEELDON, L. R. 1994. Do speakers have access to a mental syllabary? In: *Cognition*, 50: 239-269.
257. LIBBEN, G. & JAREMA, G. 2002. Mental lexicon research in the new millennium. In: *Brain and Language*, 81: 2-11.
258. LIBERMAN, A. 1990. Etymological studies III: Some Germanic words beginning with FL-. Language at play. In: *General Linguistics*, 30, 2: 81-107.
259. LIEBER, R. 1992. *Deconstructing Morphology*. University of Chicago Press, Chicago.
260. LINDBLOM, B. 1990. Explaining phonetic variation: a sketch of the H& H theory. In: HARDCASTLE, W. J., MARCHAL, A. (Hrsg.), *Speech production and speech modeling*, 403-439. Kluwer, Dordrecht.
261. LOGAN, G. D. 1990. Repetition priming and automaticity: Common underlying mechanisms? In: *Cognitive Psychology*, 22: 1-35.
262. LOWREY, T. M. & SHRUM, L. J. 2007. Phonetic symbolism and brand name preference. In: *Journal of Consumer Research*, 34: 406-414.
263. LUCE, P. A. & PISONI, D. B. 1998. Recognizing spoken words: The neighborhood activation model. In: *Ear and Hearing*, 19: 1-36.

264. LÜDELING, A. & JONG, N. H. de 2002. German particle verbs and word-formation. In: DEHE, N., JACKENDOFF, R., MCINTYRE, A. & URBAN, S. (Hrsg.), *Verb–particle explorations*, 315–333. De Gruyter, Berlin.
265. MACKAY, D. G. 1982. The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. In: *Psychological Review*, 89: 60-94.
266. MAGNUS, M. 2000. *What's in a word? Evidence for phonosemantics*. University of Trondheim Dissertation, Trondheim.
267. MAGNUSON, J. S., DIXON, J. A., TANENHAUS, M. K. & ASLIN, R. N. 2007. The dynamics of lexical competition during spoken word recognition. In: *Cognitive Science*, 31: 1–24.
268. MALTZMAN, I. MORRISETT, L. Jr. & BROOKS, L. O. 1956. An investigation of phonetic symbolism. In: *Journal of Abnormal and Social Psychology*, 53: 245-251.
269. MANSKI, C. F. 1988. *Analog estimation methods in econometrics*. Chapman and Hall, New York.
270. MARCHMAN, V. A. 1997. Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. In: *Cognitive Science*, 21: 283-304.
271. MARCUS, G. F., BRINKMANN, U., CLAHSSEN, H., WIESE, R., WOEST, A. & PINKER, S. 1995. German inflection: The exception that proves the rule. In: *Cognitive Psychology*, 29: 189-256.
272. MARIL, A., WAGNER, A. D. & SCHACTER, D. L. 2001. On the tip of the tongue: an event-related fMRI study of semantic retrieval failure and cognitive conflict. In: *Neuron*, 31: 653–660.
273. MARSLLEN-WILSON, W. D. 1987. Functional parallelism in spoken word-recognition. In: *Cognition*, 25: 71-102.
274. MARSLLEN-WILSON, W. D. & TYLER, L. K. 1997. Dissociating types of mental computation. In: *Nature*, 387: 592–594.
275. MARSLLEN-WILSON, W. D. & TYLER, L. K. 1998. Rules, representations and the English past tense. In: *Trends in Cognitive Sciences*, 2, 11: 428–435.
276. MARSLLEN-WILSON, W. D. & TYLER, L. K. 2007. Morphology, language and the brain: The decompositional substrate for language comprehension. In: *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362: 823–836.
277. MARSLLEN-WILSON, W. D., TYLER, L. K., WAKSLER, R. & OLDER, L. 1994b. Morphology and meaning in the English mental lexicon. In: *Psychological Review*, 101: 3–33.

278. MARSLEN-WILSON, W. D. & WARREN, P. 1994a. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. In: *Psychological Review*, 101: 653-675.
279. MARSLEN-WILSON, W. D. & ZWITSERLOOD, P. 1989. Accessing spoken words: The importance of word onsets. In: *Journal of Experimental Psychology: Human Perception and Performance*, 15: 576-585.
280. MARSOLEK, C. J. & BURGUND, E. D. 1997. Computational analyses and hemispheric asymmetries in visual-form recognition. In: CHRISTMAN, S. (Hrsg.), *Cerebral asymmetries in sensory and perceptual processing*, 125-158. Elsevier, Amsterdam.
281. MARTIN, C. D., THIERRY, G. 2008. Interplay of orthography and semantics in reading: An event-related potential study. In: *Neuroreport*, 19, 15: 1501-1505.
282. MARTIN, S. 1962. Phonetic symbolism in Korean. In: POPPE, N. (Hrsg.), *American studies in Altaic linguistics. Uralic and Altaic series vol. 13*, 177-189. Indiana University Press, Bloomington.
283. MARTINET, A. 1962. *A functional view of language*. Oxford University Press, Oxford.
284. MCCLELLAND, J. L. & PATTERSON, K. 2002a. Rules or connections in past tense inflections: What does the evidence rule out? In: *Trends in Cognitive Neurosciences*, 6, 11: 465-472.
285. MCCLELLAND, J. L. & PATTERSON, K. 2002b. 'Words or Rules' cannot exploit the regularity in exceptions. In: *Trends in Cognitive Neurosciences*, 6, 11: 464-465.
286. MCCLELLAND, J. L. & PATTERSON, K. 2003. Differentiation and integration in human language. In: *Trends in Cognitive Sciences*, 7, 2: 63-64.
287. MCCLELLAND, J. L. & RUMELHART, D. E. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. In: *Psychological Review*, 88: 375-407.
288. MCCORMICK, S. F., BRYLSBAERT, M. & RASTLE, K. 2009. Is morphological decomposition limited to low-frequency words? In: *The Quarterly Journal of Experimental Psychology*, 62, 9: 1706-1715.
289. MCGONIGLE, B. O. & CHALMERS, M. 1998. Rationality as optimised self-regulation. In: OAKSFORD, M., CHATER, N. (Hrsg.), *Rational models of cognition*, 501-534. Oxford University Press, Oxford.
290. MCGONIGLE, B. O. & CHALMERS, M. 2002. The growth of cognitive structure in monkeys and men. In: FOUNTAIN, S. B., BUNSEY, M. D., DANKS, J. H. & MCBEATH, M. K. (Hrsg.), *Animal cognition and sequential behaviour: Behavioral, biological and computational perspectives*, 269-314. Kluwer Academic, Boston.

291. MEUNIER, F. & SEGUI, J. 1999. Frequency effects in auditory word recognition: The case of suffixed words. In: *Journal of Memory and Language*, 41: 327–344.
292. MERRIAM-WEBSTER.COM 2012. <http://www.merriam-webster.com/>(aufgerufen am 31.07.2012)
293. MIOZZO, M. 2003. On the processing of regular and irregular forms of verbs and nouns: Evidence from neuropsychology. In: *Cognition*, 87: 101–127.
294. MONAGHAN, P., CHATER, N. & CHRISTIANSEN, M. H. 2005. The differential contribution of phonological and distributional cues in grammatical categorization. In: *Cognition*, 96: 143–182.
295. MONAGHAN, P., CHRISTIANSEN, M. H. & CHATER, N. 2007. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. In: *Cognitive Psychology*, 55: 259–305.
296. MONSELL, S., & HIRSH, K. W. 1998. Competitor priming in spoken word recognition. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24: 1495–1520.
297. MORTON, E. S. 1994. Sound symbolism and its role in non-human vertebrate communication. In: HINTON, L., NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound symbolism*, 348–365. Cambridge University Press, Cambridge.
298. MORTON, J. 1970. A functional model for memory. In: NORMAN, D. A. (Hrsg.), *Models of human memory*, 203–254. Academic Press, New York.
299. MORTON, J. 1979. Word recognition. In: MORTON, J. & MARSHALL, J. C. (Hrsg.), *Structures and processes*, 109–156. The MIT Press, Cambridge.
300. MOSS, H. E., OSTRIN, R. K., TYLER, L. K. & MARSLER-WILSON, W. D. 1995. Accessing different types of lexical semantic information: Evidence from priming. In: *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21: 863–883.
301. MÜNTE, T. F., SAY, T., CLAHSSEN, H., SCHILTZ, K. & KUTAS, M. 1998. Decomposition of morphologically complex words in English: Evidence from event-related brain potentials. In: *Cognitive Brain Research*, 7: 241–253.
302. MUSGRAVE, A. 1980. Stützung durch Daten, Falsifikation, Heuristik und Anarchismus. In: RADNITZKY, G. & ANDERSSON, G. (Hrsg.), *Fortschritt und Rationalität der Wissenschaft*, 199–220. Mohr Siebeck, Tübingen.
303. NAKISA, R. C., PLUNKETT, K. & HAHN, U. 1998. A cross-linguistic comparison of single and dual-route models of inflectional morphology. In: BROEDER, P. & MURRE, J. (Hrsg.), *Cognitive models of language acquisition*. The MIT Press, Cambridge.

304. NAPPS, S. E. 1989. Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships? In: *Memory and Cognition*, 17: 729-739.
305. NEELY, J. H. 1991. Semantic priming effects in visual word recognition: A selective review of current findings and theory. In: BESNER, D. & HUMPHREYS, G. W. (Hrsg.), *Basic processes in reading: Visual word recognition*, 264-336. Lawrence Erlbaum Associates, Hillsdale.
306. NEWMAN S. S. 1933. Further experiments in phonetic symbolism. *American Journal of Psychology*, 45: 53-75.
307. NEYMAN, J. & PEARSON, E. 1933. On the problem of the most efficient tests of statistical hypotheses. In: *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, 231: 289–337.
308. NGO, M. K., MISHRA, R. & SPENCE, C. 2011. Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. In: *Food Quality and Preference*, 6: 567-572.
309. NIELSEN, A., & RENDALL, D. 2011. The sound of round: Evaluating the role of consonants in the classic Takete-Maluma phenomenon. In: *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65: 115–124.
310. NORRIS, D. 1994. Shortlist: A connectionist model of continuous speech recognition. In: *Cognition*, 52: 189-234.
311. NORRIS, D., MCQUEEN, J. M. & CUTLER, A. 2002. Bias effects in facilitatory phonological priming. In: *Memory and Cognition*, 30: 399–411.
312. NUCKOLLS, J. B. 1996. *Sounds like life: Sound symbolic grammar, performance and cognition in Pastaza Quechua*. Oxford University Press, New York.
313. OBER, B. & SHENAUT, G. K. 1988. Lexical decision and priming in Alzheimer's disease. In: *Neuropsychologia*, 26: 273-286.
314. ODELL, K., MCNEIL, M., ROSENBEK, J. C. & HUNTER, L. 1990. Perceptual characteristics of consonant production by apraxic speakers. In: *Journal of Speech and Hearing Disorders*, 55: 345-359.
315. O'GRADY, W. 2001. *Contemporary Linguistics*. 4. Aufl. Pearson ESL, London.
316. OHALA, J. J. 1994. The frequency codes underlies the sound symbolic use of voice pitch. In: HINTON, L., NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound symbolism*, 325-347. Cambridge University Press, Cambridge.
317. ORNE, M. T. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. In: *American Psychologist*, 17: 776-783.

318. OSBORNE, J. W. & OVERBAY, A. 2004. The power of outliers (and why researchers should always check for them). In: *Practical Assessment, Research and Evaluation*, 9, 6: 1-12.
319. OTIS, K. & SAGI, E. 2008. Phonaesthemes: A corpus-based analysis. In: B. C., MCRAE, K. & SLOUTSKY, V. M. (Hrsg.), *Proceedings of the 30th annual meeting of the cognitive science society*. <http://csjarchive.cogsci.rpi.edu/proceedings/2008/pdfs/p65.pdf> (aufgerufen am 23.10.2014)
320. PASTIZZO, M. J. & FELDMAN, L. B. 2002. Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28: 244-249.
321. PASTIZZO, M. J. & FELDMAN, L. B. 2009. Multiple dimensions of relatedness among words: Conjoint effects of form and meaning in word recognition. In: *The Mental Lexicon*, 4, 1: 1-25.
322. PAWLEY, A. & SYDER, F. H. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In: RICHARDS, J. C. & SCHMIDT, R. W., *Language and communication*, 191-225. Longman, London.
323. PEEREMAN, R. & CONTENT, A. 1995. Neighborhood size effect in naming: Lexical activation or sublexical correspondences? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 2: 409-421.
324. PEIRCE, C. S. 1998. What is a sign? In: THE PEIRCE EDITION PROJECT (Hrsg.), *The essential Peirce, Volume 2: Selected philosophical writings (1893-1913)*, 4-10. Indiana University Press, Indiana.
325. PEREA, M. & CARREIRAS, M. 1998. Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. In: *Journal of Experimental Psychology: Human Perception and Performance*, 24: 134-144.
326. PEREA, M. & GOTOR, A. 1997. Associative and semantic priming effects occur at very short SOAs in lexical decision and naming. In: *Cognition*, 62: 223-240.
327. PERRE, L., MIDGLEY, K., ZIEGLER, J. C. 2009. When beef primes reef more than leaf: Orthographic information affects phonological priming in spoken word recognition. In: *Psychophysiology*, 46: 739-746.
328. PEREA, M. & ROSA, E. 2002. The effects of associative and semantic priming in the lexical decision task. In: *Psychological Research*, 66: 180-194.
329. PERKELL, J. S., MAJID, Z., MATTHIES, M. L. & LANE, H. 2002. Economy of effort in different speaking conditions. A preliminary study of intersubject differences and modeling issues. In: *The Journal of the Acoustical Society of America*, 112, 4: 1627-1641.

330. PETERMANN, S. 2001. *Soziale Vernetzung städtischer und ländlicher Bevölkerungen am Beispiel der Stadt Halle. Abschlussbericht und Codebuch*. Der Hallesche Graureiher 2001-2. Forschungsberichte des Instituts für Soziologie. Martin-Luther-Universität Halle-Wittenberg, Halle.
331. PHILPS, D. 2012. Submorphemes: backtracking from English ‚kn- words‘ to the emergence of the linguistic sign. In: *Miranda (Online)*, 7. http://www.miranda-ejournal.eu/sdx2/miranda/article.xsp?numero=7&id_article=Article_11-905 (aufgerufen am 28.01.13)
332. PIERREHUMBERT, J., BENT, T., MUNSON, B., BRADLOW, A. & BAILEY, J.M. 2004. The influence of sexual orientation on vowel production. In: *Journal of the Acoustic Society of America*, 116, 4: 1905-1908.
333. PINKER, S. 1991. Rules of language. In: *Science*, 253: 530-534.
334. PINKER, S. 1997. Words and rules in the human brain. In: *Nature*, 387: 547-548.
335. PINKER, S. & PRINCE, A. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In: *Cognition*, 28: 73-193.
336. PINKER, S. 1994 & PRINCE, A. Regular and irregular morphology and the psychological status of rules of grammar. In: LIMA, S. D., CORRIGAN, R. L. & IVERSON, G. K. (Hrsg.), *The reality of linguistic rules*, 321-351. John Benjamins, Amsterdam.
337. PLAUT, D. C. & GONNERMAN, L. M. 2000. Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? In: *Language and Cognitive Processes*, 15, 4/5: 445-485.
338. PLUNKETT, K. & MARCHMAN, V. 1991. U-Shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. In: *Cognition*, 38: 43-102.
339. PLUNKETT, K. & MARCHMAN, V. 1993. From rote learning to system building: acquiring verb morphology in children and connectionist nets. In: *Cognition*, 48: 21-69.
340. POKORNY, J. 1959-1969. *Indogermanisches etymologisches Wörterbuch*. Francke, München.
341. POPPER, K. R. 2002. *The logic of scientific discovery*. Routledge, London.
342. PORTIN, M., LEHTONEN, M., HARRER, G., WANDE, E., NIEMI, J. & LAINE, M. 2008. L1 effects on the processing of inflected nouns in L2. In: *Acta Psychologica*, 128: 452-465.
343. PRASADA, S. & PINKER, S. 1993. Generalisation of regular and irregular morphological patterns. In: *Language and Cognitive Processes*, 8: 1-56.

344. PYLKKÄNEN, L., FEINTUCH, S., HOPKINS, E. & MARANTZ, A. 2004. Neural correlates of the effects of morphological family frequency and family size: An MEG study. In: *Cognition*, 91: B35-B45.
345. RACHEV, S., MITTNIK, S., ERBEN, R. F. & ROMEIKE, F. 2006. Abschied von der Glockenkurve. Neue Ansätze zur Portfolio-Optimierung. In: *Risiko Manager*, 11: 20-22.
346. RADEAU, M., BESSON, M., FONTENEAU, E., & CASTRO, S. L. 1998. Semantic, repetition, and rime priming between spoken words: Behavioral and electrophysiological evidence. In: *Biological Psychology*, 48: 183–204.
347. RADEAU, M., MORAIS, J., & SEGUI, J. 1995. Phonological priming between monosyllabic spoken words. In: *Journal of Experimental Psychology: Human Perception and Performance*, 21: 1297–1311.
348. RAITH, A. 2012. Der Mundarttod tritt 2050 ein. Interview mit Bernhard Stör. In: *IDO-WA*, 5. <http://www.idowa.de/home/archiv/artikel/2012/05/09/der-mundarttod-tritt-2050-ein.html> (aufgerufen am 02.07.2012)
349. RAJ, A. & CHEN Y. H. 2011. The Wiring Economy Principle: Connectivity determines anatomy in the human brain. In: *PLoS ONE*, 6, 9. <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0014832>(aufgerufen am 03.06.2012)
350. RAMACHANDRAN, V. S. & HUBBARD, E. M. 2001. Synesthesia – a window into perception, thought, and language. In: *Journal of Consciousness Studies*, 8: 3-34.
351. RAMB, B.-T., TIETZEL, M. & HARTWIG, K.-H. 1993. *Ökonomische Verhaltenstheorie*. Vahlen, München.
352. RASTLE, K., DAVIS, M. H., MARSLEN-WILSON, W. D. & TYLER, L. K. 2000. Morphological and semantic effects in visual word recognition: A time course study. In: *Language and Cognitive Processes*, 15: 507-538.
353. RAVEH, M. 2002. The contribution of frequency and semantic similarity to morphological processing. In: *Brain and Language*, 81: 312-325.
354. RHEINISCHES MITMACHWÖRTERBUCH. 2007-. Landschaftsverband Rheinland (LVR), Köln. <http://www.mitmachwoerterbuch.lvr.de/> (aufgerufen am 10.06.2012)
355. RHEINISCHES WÖRTERBUCH. 1928-71. Fritz Klopp, Bonn. <http://woerterbuchnetz.de/RhWB/?lemid=RA00001> (aufgerufen am 28.06.2012)
356. RENDALL, D. 2003. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. In: *Journal of the Acoustical Society of America*, 113: 3390-3402.
357. RENDALL, D. 2009. The evolution of language and the (ir)relevance of primate communication. In: *Evolutionary Anthropology*, 18: 230-233.

358. RICE, G. A. & ROBINSON, D. O. 1975. The role of bigram frequency in the perception of words and nonwords. In: *Memory and Cognition*, 3: 513-518.
359. RICKHEIT, G., SICHELSCHEIDT, L. & STROHNER, H. 2002. *Psycholinguistik*. Stauffenburg Verlag, Tübingen.
360. RIETVELD, T. & HOUT, R. van 2005. *Statistics in Language Research: Analysis of Variance*. De Gruyter, Berlin.
361. ROGERS, S. K., & ROSS, A. S. 1975. A cross-cultural test of the Maluma-Takete phenomenon. In: *Perception*, 4, 1: 105–106.
362. ROMANI, C., & GALLUZZI, C. 2005. Effects of syllabic complexity in predicting accuracy of repetition and direction of errors in patients with articulatory and phonological difficulties. In: *Cognitive Neuropsychology*, 22: 817-850.
363. RONNEBERGER-SIBOLD, E. 1997. Sprachökonomie und Wortschöpfung. In: BIRKMANN, T., KLINGENBERG, H., NÜBLING, D. & RONNEBERGER-SIBOLD, E. (Hrsg.), *Vergleichende germanische Philologie und Skandinavistik. Festschrift für Otmar Werner zum 65. Geburtstag*, 249-262. Niemeyer, Tübingen.
364. RUECKL, J. G., MIKOLINSKI, M., RAVEH, M., MINER, C. S. & MARS, F. 1997. Morphological priming, fragment completion and connectionist networks. In: *Journal of Memory and Language*, 36: 382–405.
365. RUMELHART, D. E. & MCCLELLAND, J. L. 1986. On learning the past tense of English verbs. In: MCCLELLAND, J. L. & RUMELHART, D. E. (Hrsg.), *Parallel distributed processing, vol. 2*, 216-271. The MIT Press, Cambridge.
366. SAINZ, J. S. & LAZARO, M. 2009. Early and late effects of morphological decomposition: Brain correlates of family size effects on complex words and pseudowords. In: TAATGEN, N. A. & RIJN, H. van (Hrsg.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society, 2992-2997*. Cognitive Science Society, Austin.
367. SAPIR, E. 1929. A study in phonetic symbolism. In: *Journal of Experimental Psychology*, 12: 225–239.
368. SAUSSURE, F. de 2001. *Grundfragen der allgemeinen Sprachwissenschaft*. 3. Aufl. Hrsg. v. C. Bally und A. Sechehaye. De Gruyter, Berlin.
369. SCARBOROUGH, D. L., CORTESE, C. & SCARBOROUGH, H. S. 1977. Frequency and repetition effects in lexical memory. In: *Journal of Experimental Psychology: Human Perception and Performance*, 3: 1–17.
370. SCHACTER, D. L., CHIU, C. Y. P. & OCHSNER, K. N. 1993. Implicit memory: a selective review. In: *Annual Review of Neuroscience*, 16: 159–182.

371. SCHILLER, K. & LÜBBEN, A. 2010. *Mittelniederdeutsches Wörterbuch*. Digitalisiert von Gerhard Köbler. <http://www.koeblergerhard.de/Mittelniederdeutsch-HP/Einfuehrung-Mnd-eD-HP.htm> (aufgerufen am 21.08.2012)
372. SCHMID, H.-J. 2007. Entrenchment, salience and basic levels. In: GEERAERTS, D. & CUYCKENS, H. (Hrsg.), *The Oxford Handbook of Cognitive Linguistics*, 117-138. Oxford University Press, Oxford.
373. SCHMID, H.-J. 2010. Does frequency in text really instantiate entrenchment in the cognitive system? In: GLYNN, D. & FISCHER, K. (Hrsg.), *Usage-based cognitive semantics. Corpus-driven methods for the study of meaning*, 101-133. De Gruyter, Berlin.
374. SCHMIDT, P. 1977. Sozialtechnologie und Handlungsforschung als experimentelle Handlungs- und Forschungsstrategien. In HAFT, H. & HAMEYER, U. (Hrsg.), *Handlungsorientierte Schulforschungsprojekte: Erfahrungsberichte, Analysen, Kritik*, 223-253. Beltz, Weinheim.
375. SCHREUDER, R. & BAAYEN, R. 1995. Modeling morphological processing. In: FELDMAN, L. B. (Hrsg.), *Morphological aspects of language processing*, 131–154. Lawrence Erlbaum Associates, Hillsdale.
376. SCHREUDER, R. & BAAYEN, R. H. 1997. How complex simplex words can be. In: *Journal of Memory and Language*, 37: 118–139.
377. SEARS, C. R., HINO, Y. & LUPKER, S. J. 1995. Neighborhood frequency and neighborhood size effects in visual word recognition. In: *Journal of Experimental Psychology: Human Perception and Performance*, 21: 876–900.
378. SEGUI, J. & GRAINGER, J. 1990. Priming word recognition with orthographic neighbours: Effects of relative prime-target frequency. In: *Journal of Experimental Psychology: Human Perception and Performance*, 16: 65–76.
379. SEIDEL, F. VOLPE, M. & PSOTTA, K. 2012. Bayerns Torgasmus. *Bild Digital 2011*. <http://www.bild.de/sport/fussball/bayern-muenchen/torgasmus-gegen-hoffenheim-auch-gegen-basel-23094098.bild.html> (aufgerufen am 09.06.2012)
380. SEIDENBERG, M. S. 1992. Connectionism without tears. In: DAVIS, S. (Hrsg.), *Connectionism: Theory and practice*, 84-122. Oxford University Press, New York.
381. SEIDENBERG, M. S. & GONNERMAN, L. M. 2000. Explaining derivational morphology as the convergence of codes. In: *Trends in Cognitive Sciences*, 4: 353-361.
382. SEIFART, F. 2000. *Grundfragen bei der Dokumentation bedrohter Sprachen*. http://www.eva.mpg.de/lingua/staff/seifart/pdf/papers/Seifart2000_ArbpapKoeln.pdf (aufgerufen am 02.07.2012)
383. SERENO, J. A. 1994. Phonosyntactics. In: HINTON, L., NICHOLS, J. & OHALA, J. J. (Hrsg.), *Sound symbolism*, 263-275. Cambridge University Press, Cambridge.

384. SERENO, J. A. & JONGMAN, A. 1997. Processing of English inflectional morphology. In: *Memory and Cognition*, 25: 425-37.
385. SHADISH, W. R., COOK, T. D. & CAMPBELL, D. T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company, Boston.
386. SHARLIN, M. 2009. Sounds Like...: Understanding Japanese Sound Symbolism. <http://www.swarthmore.edu/SocSci/Linguistics/2010theses/NaomiSharlinBMC.pdf> (aufgerufen am 25.10.2014)
387. SHELTON, J. R. & MARTIN, R. C. 1992. How semantic is automatic semantic priming? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18: 1191–1210.
388. SHEN, D. & FORSTER, K. I. 1999. Masked phonological priming in reading Chinese words depends on the task. In: *Language and Cognitive Processes*, 14: 429-459.
389. SHIN, S. 2005. *Grammaticalization of politeness: A contrastive study of German, English and Korean*. Doctoral thesis. University of California, Berkeley.
390. SLOWIACZEK, L. M., MCQUEEN, J., SOLTANO, E. G. & LYNCH, M. 2000. Phonological representations in prelexical speech processing: Evidence from form-based priming. In: *Journal of Memory and Language*, 43: 530–560.
391. SMOLENSKY, P. 1995. Reply: Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture. In: MACDONALD, C. & MACDONALD, G. (Hrsg.), *Connectionism: Debates on psychological explanation*, 223-290. Basil Blackwell, Oxford.
392. SPINELLI, E., SEGUI, J., & RADEAU, M. (2001). Phonological priming in spoken word recognition with bisyllabic targets. In: *Language and Cognitive Processes*, 16: 367–392.
393. SOLTE-GRESSER, C., STRUVE, K. & UECKMANN, N. (Hrsg.). 2005. *Von der Wirklichkeit zur Wissenschaft: Aktuelle Forschungsmethoden in den Sprach-, Literatur- und Kulturwissenschaften*. LIT Verlag, Hamburg.
394. SONNEMANN, I. 2004. *Kommunikation im Wandel: Auswirkungen des Mobilfunks auf die Gesellschaft*. http://publikationen.ub.uni-frankfurt.de/files/5005/I_Sonnemann.pdf (aufgerufen am 02.07.2012)
395. SONNENSTUHL, I., EISENBEISS, S. & CLAHSSEN, H. 1999. Morphological priming in the German mental lexicon. In: *Cognition*, 72: 203–236.
396. SPILLING, K. 2006-7. *Wortideen von Kindern*. <http://www.ideesamkeit.de/html/kindeergarten-kinder-sprache-wo.html> (aufgerufen am 08.06.2012)

397. SPLIETH, C., ROSIN, M., STEFFEN, H. & MEYER, G. 2000. Minimal-invasive Restaurationen. In: *ZWR - Das Deutsche Zahnärzteblatt*, 11: 634-637.
398. SPRECHENDER SPRACHATLAS VON BAYERN. 2006-2008. Bayerische Landesbibliothek Online (BLO), München. <http://sprachatlas.bayerische-landesbibliothek-online.de/> (aufgerufen am 12.07.2012)
399. STANNERS, R. F., NEISER, J. J., HERNON, W. P. & HALL, R. 1979. Memory representation for morphologically related words. In: *Journal of Verbal Learning and Verbal Behavior*, 18: 399-412.
400. STEGBAUER, C. 2002. *Reziprozität. Einführung in soziale Formen der Gerechtigkeit*. Westdeutscher Verlag, Wiesbaden.
401. STEINMETZ, S. & KIPFER, B. A. 2006. *The life of language. The fascinating ways words are born, live & die*. Random House, New York.
402. STEMBERGER, J. P. 1994. Rule-less morphology at the phonology-lexicon interface. In: LIMA, S. D., CORRIGAN, R. & IVERSON, G. K. (Hrsg.), *The reality of linguistic rules*, 147-169. John Benjamins, Amsterdam.
403. STEMBERGER, J. P. & MACWHINNEY, B. 1986. Frequency and the lexical storage of regularly inflected forms. In: *Memory and Cognition*, 14: 17-26.
404. STEMBERGER, J. P. & MACWHINNEY, B. 1988. Are inflected forms stored in the lexicon? In: HAMMOND, M. & NOONAN, M. (Hrsg.), *Theoretical approaches to morphology*, 101-116. Academic Press, San Diego.
405. STICKEL, G. (Hrsg.) 1997. *Varietäten des Deutschen. Regional- und Umgangssprachen. Jahrbuch 1996 des Instituts für deutsche Sprache*. 476 S. Mit 24 Abb. und 33 Karten. De Gruyter, Berlin.
406. STOCKALL, L. & MARANTZ, A. 2006. A single route, full decomposition model of morphological complexity. In: *The Mental Lexicon*, 1, 1: 85-123.
407. STOLZ, J. A. & FELDMAN, L. B. 1995. The role of orthographic and semantic transparency of the base morpheme in morphological processing. In: FELDMAN, L. B. (Hrsg.), *Morphological aspects of language processing*, 109-129. Lawrence Erlbaum Associates, Hillsdale.
408. STOLZ, T. & KOLBE, K. (Hrsg.) 2003. *Methodologie in der Linguistik*. Peter Lang, Frankfurt a. M..
409. SUROWIECKI, J. 2005. *The wisdom of crowds - why the many are smarter than the few*. First Anchor Books Edition, New York.
410. TAFT, M. 1979. Recognition of affixed words and the word frequency effect. In: *Memory and Cognition*, 7: 263-272.

411. TAFT, M. 1985. The decoding of words in lexical access: A review of the morphographic approach. In: BESNER, D., WALKER, T. & MACKINNON, G. (Hrsg.), *Reading research: Advances in theory and practice, Vol. 5*, 83-123. Academic Press, New York.
412. TAFT, M. 1994. Interactive-activation as a framework for understanding morphological processing. In: *Language and Cognitive Processes*, 9: 271–294.
413. TAFT, M. 2004. Morphological decomposition and the reverse base frequency effect. In: *The Quarterly Journal of Experimental Psychology*, 57A, 4: 745–765.
414. TAFT, M. & FORSTER, K. I. 1975. Lexical storage and retrieval of prefixed words. In: *Journal of Verbal Learning and Verbal Behavior*, 14: 638-647.
415. TAFT, M. & GRAAN, F. van 1998. Lack of phonological mediation in a semantic categorization task. In: *Journal of Memory and Language*, 38: 203-224.
416. TAYLOR, J. 2002. *Cognitive Grammar*. Oxford University Press, Oxford.
417. TENPENNY, P. L. 1995. Abstractionist versus episodic theories of repetition priming and word identification. In: *Psychonomic Bulletin and Review*, 2: 339-363.
418. THE BRITISH NATIONAL CORPUS.2007. Version 3 (BNC XML Edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. <http://www.natcorp.ox.ac.uk/> (aufgerufen am 09.11.2013)
419. THORNDIKE, E. L. & LORGE, I. 1944. *The Teacher's Word Book of 30,000 Words*. Bureau of Publications, Teachers College, Columbia University, New York.
420. TOMASELLO, M. 2003. *Constructing a language. A usage-based theory of language acquisition*. Harvard University Press, Cambridge.
421. TROST, B. M. 1995. Atom economy. A challenge for organic synthesis. In: *Angewandte Chemie - International Edition*, 34, 3: 259–281.
422. TSUR, R. 2006. Size–sound symbolism revisited. In: *Journal of Pragmatics*, 38: 905–924.
423. TSURU, S. & FRIES, H. S. 1933. A problem in meaning. In: *The Journal of General Psychology*, 8, 1: 281-284.
424. TULDAVA, J. 1995. *Methods in Quantitative Linguistics*. Wissenschaftlicher Verlag Trier, Trier.
425. TYLER, L. K., MORNAY-DAVIES, P. de, ANOKHINA, R., LONGWORTH, C., RANDALL, B. & MARSLEN-WILSON, W. D. 2002a. Dissociation in processing past tense morphology: Neuropathology and behavioral studies. In: *Journal of Cognitive Neuroscience*, 14, 1: 79–94.
426. TYLER, L. K., RANDALL, B. & MARSLEN-WILSON, W. D. 2002b. Phonology and neuropsychology of the English past tense. In: *Neuropsychologia*, 40: 1154–1166.

427. ULLMAN, M. T. 1999. Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. In: *Language and Cognitive Processes*, 14: 47- 67.
428. UZENDOSKI, M. 2012. Beyond orality: Textuality, territoriality, and ontology among Amazonian people. In: *HAU: The Journal of Ethnographic Theory*, 2, 1: 55–80.
429. VAUX, B. & COOPER, J. 1999. *Introduction to Linguistic field methods*. Lincom Europa, München.
430. VENDRYÈS, J. 1925. *Language: A linguistic introduction to history*. Routledge, Trench, Trübner, London.
431. VOGA, M. & GRAINGER, J. 2004. Masked morphological priming with varying degrees of form overlap: Evidence from Greek verbs. In: *Current Psychology Letters*, 13, 2. <http://cpl.revues.org/index422.html> (aufgerufen am 04.06.2012)
432. WEIS, E. 2002. *PONS Kompaktwörterbuch Englisch-Deutsch / Deutsch-Englisch*. Ernst Klett Sachbuch Verlag, Stuttgart.
433. WERNER, O. 1989. Sprachökonomie und Natürlichkeit im Bereich der Morphologie. In: *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 42: 34-47.
434. WESTBURY, C. 2005. Implicit sound symbolism in lexical access: Evidence from an interference task. In: *Brain and Language*, 93, 1: 10-19.
435. WESTBURY, C. & BUCHANAN, L. 2002. The probability of the least likely non length-controlled bigram affects lexical decision reaction times. In: *Brain and Language*, 81, 1-3: 66-78.
436. WHITNEY, W. D. 1877. The Principle of Economy as a phonetic force. In: *Transactions of the American Philological Association*, 8: 123-134.
437. WIKIPEDIA CONTRIBUTORS. 2012a. Mondo film. *Wikipedia: The Free Encyclopedia*. Zuletzt bearbeitet am 21. May 2012 um 02:49. http://en.wikipedia.org/wiki/Mondo_film (aufgerufen am 09.06.2012)
438. WIKIPEDIA CONTRIBUTORS. 2012b. Buildering. *Wikipedia: The Free Encyclopedia*. <http://de.wikipedia.org/wiki/Buildering> (aufgerufen am 10.06.2012)
439. WIKIPEDIA CONTRIBUTORS. 2013. Synonym. *Wikipedia: The Free Encyclopedia*. http://de.wikipedia.org/wiki/Synonym#Partielle_Synonymie_.28Bedeutungs.C3.A4hnlichkeit.29 (aufgerufen am 24.10.2014)
440. WILDER, C., GÄRTNER, H.-M. & BIERWISCH, M. (Hrsg.) 1997. *The role of economy principles in Linguistic theory*. Akademie-Verlag, Berlin.

441. WILSON, M. D. 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. In: *Behavioural Research Methods, Instruments and Computers*, 20, 1: 6-11.
442. WINTER, B. & GRAWUNDER, S. 2010. Acoustic correlates of politeness: prosodic and voice quality measures in polite and informal speech of Korean and German speakers. In: *Speech Prosody 2010*: 1-4. http://www.bodowinter.com/papers/Grawunder_Winter_2010_Korean_German_politeness.pdf (aufgerufen am 18.07.2012)
443. WOOLDRIDGE, J. M. 2003. *Introductory econometrics. A modern approach*. Thomson South-Western, Mason.
444. WORBS, E., MARKOWSKI, A. & MEGERLE, A. 2007. *Polnisch-deutsches Wörterbuch der Neologismen. Neuer polnischer Wortschatz nach 1989*. Harrassowitz, Wiesbaden.
445. WURZEL, W. 1997. Natürlicher grammatischer Wandel, ‚unsichtbare Hand‘ und Sprachökonomie - Wollen wir wirklich so Grundverschiedenes? In: BIRKMANN, T., KLINGENBERG, H., NÜBLING, D. & RONNEBERGER-SIBOLD, E. (Hrsg.), *Vergleichende germanische Philologie und Skandinavistik*, 295-308. Niemeyer, Tübingen.
446. YORKSTON, E. A. & MENON, G. 2004. A sound idea: Phonetic effects of brand names on consumer judgments. In: *Journal of Consumer Research*, 31: 43-51.
447. YOSHIDA, H. & SMITH, L. 2006. Dynamic properties of form and meaning and children's learning of verbs. In: *XVth International Conference of Infant Studies, Kyoto, Japan*.
448. YOUNG, P. 2001. *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press, Princeton.
449. ZHANG, G. B., PATUWO, B. E. & HU, M. Y. 1998. Forecasting with artificial neural networks: The state of the art. In: *International Journal of Forecasting*, 14: 35-62.
450. ZLATEV, J. 2007. Embodiment, language and mimesis. In: ZIEMKE, T., ZLATEV, J. & FRANCK, R. (Hrsg.) *Body, language, mind. Vol 1: Embodiment*, 297-337. De Gruyter, Berlin.
451. ZLATEV, J. (im Druck). Love reign o'er me: The intersubjective grounding of language in bodily mimesis. In: ITKONEN, E. & PAJUNEN, A. (Hrsg), *Language, culture and mind*. Turku University Press, Turku. <http://project2.sol.lu.se/sedsu/publications/Brighton-paper.pdf> (aufgerufen am 06.07.2012)