The syllable as a processing unit in speech production: Evidence from frequency effects on coarticulation

Barbara Samlowski

Dissertation zur Erlangung des akademischen Grades
Doctor philosophiae (Dr. phil.)

vorgelegt an der Fakultät für Linguistik und
Literaturwissenschaft der Universität Bielefeld

Betreuer:
Prof. Dr. Petra Wagner (Universität Bielefeld)
Prof. Dr. Bernd Möbius (Universität des Saarlandes)

Dissertation zur Erlangung des akademischen Grades Doctor philosophiae (Dr. phil.)
vorgelegt an der Fakultät für Linguistik und Literaturwissenschaft der Universität Bielefeld
Eröffnung des Promotionsverfahrens am 07.05.2015

Prüfungskommission:
Prof. Dr. Petra Wagner (Universität Bielefeld, Erstgutachterin)
Prof. Dr. Bernd Möbius (Universität des Saarlandes, Zweitgutachter)
Prof. Dr. Friederike Kern (Universität Bielefeld)
Dr, Joana Cholin (Ruhr-Universität Bochum)

Tag der mündlichen Prüfung: 02.11.2015

Gedruckt auf alterungsbeständigem Papier °° ISO 9706

Table of Contents

List of Figures

List of Tables

Acknowledgements

Many people have given me help and support on the long and difficult path towards my PhD degree. First and foremost, I want thank my supervisors, Petra Wagner and Bernd Möbius, who helped me in innumerable ways. Among many other things, they gave me invaluable input and feedback throughout my writing process, were always willing to answer questions on theoretical concerns, experimental designs and interpretation of results, and managed to combine endless patience with constructive pressure and encouragement. Further, I am truly grateful to Joana Cholin and Friederike Kern for their helpful advice and feedback and for being part of my PhD committee.

A huge "Thank You!" goes out to all my work group colleagues in Bonn (especially to Donata Moers, Denis Arnold, Charlotte Wollermann, Christopher Sappok, Christine de Bond and Mareike Ahrens) and Bielefeld (among them Andreas Windmann, Zofia Malisz, Marcin Włodarczak, Juraj Šimko, Laura de Ruiter, Leona Polyanskaya, Mikhail Ordin, Christiane König, Joanna Skubisz, Said Abir Anbari, Simon Betz, Angelika Höhnemann, Valentina Schettino, and Aleksandra Ćwiek) for many interesting and fruitful discussions, countless shared meals and coffee breaks, lots of encouragement, practical help, and useful advise, as well as frequent fun and memorable group activities. It was a great privilege to be able to work with people who view each other not only as colleagues, but also as friends.

My dissertation would not have been possible without the production studies. Special thanks go to Robert Eickhaus for his expertise in setting up and executing the recordings, his patience and flexibility in the scheduling of appointments, and his valued friendship. I am also indebted to all the people who took part in my experiments for their time and effort.

Numerous other people contributed to my thesis in one way or another. Here are a few specific examples (the list is by no means complete). I would like to express my sincere gratitude to:

- Jörg Mayer and Berthold Crysman for valuable comments.
- Volker Boehlke for providing me with the Leipzig database.
- Martin Hartung for sending me orthographic transcripts of the GF corpus.

- Antje Schweitzer for providing and helping me with the German language text normalization module of the Festival Speech Synthesis System.
- Sebastian Framke, Bastian Fischer, Hendrik Hasenbein, and Kai Mismahl for their patience and tenacity in dealing with various computer problems and system administration issues.
- Natalie Lewandowski for letting me use her Matlab scripts for calculating spectral similarity.
- Matthias Sandmann and Patrick Dunkhorst for supporting me with scripts for corpus processing and prominence analysis.

I also want to thank the German Research Association (DFG) and Bielefeld University for providing me with the financial support necessary to complete my thesis.

Finally, I would like to thank my parents and my sister for providing moral support and understanding throughout the dissertation process and for always being available to act as sounding boards and provide constructive criticism, whether it was a question of thinking through new ideas, putting them into writing, practicing for oral presentations, or dealing with the problems and frustrations that inevitably occur along the way.

Zusammenfassung

Im Deutschen wie auch in vielen anderen Sprachen sind Silbenhäufigkeiten sehr ungleich verteilt, so dass einige wenige Silben sehr häufig auftauchen, während die meisten anderen extrem selten sind (Möbius, 2001). Einige kognitionslinguistische Theorien postulieren, dass sich häufige und seltene Silben in der Art unterscheiden, wie sie mental verarbeitet und produziert werden, und dass diese unterschiedliche Verarbeitung Auswirkungen auf ihre letztendliche Aussprache haben kann (Levelt/Wheeldon, 1994; Whiteside/Varley, 1998a Walsh et al., 2007). Ziel der vorliegenden Dissertation war es, in einem stark kontrollierten Vorlese-Experiment zu untersuchen, inwieweit Silbenhäufigkeiten im Deutschen einen Einfluss auf die Aussprache von häufigen und seltenen Silben haben (siehe Kapitel 7). Hierfür wurde zunächst in einer ausführlichen Analyse unterschiedlicher Korpora eine zuverlässige Basis für für Silbenhäufigkeitsangaben im Deutschen geschaffen (siehe Kapitel 3 und 4). Um die Größe der möglichen Effekte besser einschätzen zu können und festzustellen, welche potenziellen Störfaktoren in dem Hauptexperiment berücksichtigt werden müssen, wurde in zwei Produktionsexperimenten (siehe Kapitel 5 und 6) untersucht, welchen Einfluss Wort- und Satzbetonung, Wort- und Satzgrenzen sowie Wortart und Lemmahäufigkeit auf die Silbenaussprache haben. Die Dissertation beginnt mit einer kurzen Vorstellung des Vorhabens und der einzelnen Kapitel (Kapitel 1) sowie einem Überblick über die bisherige Forschung zu der Rolle von Silben in der Sprachproduktion (Kapitel 2). Eine Diskussion der einzelnen Studienergebnisse im Gesamtzusammenhang der Arbeit sowie eine abschließende Zusammenfassung der wichtigsten Schlussfolgerungen finden sich in den Kapiteln 8 und 9.

Hintergrund

Die "Mental Syllabary"-Theorie besagt, dass zumindest die häufigen Silbenformen als ganze Artikulationsabläufe im Gehirn gespeichert werden (Levelt/Wheeldon, 1994). Eine solche Speicherung würde zu einer kürzeren Verarbeitungszeit führen, da die Artikulation dieser Silben nicht jedes Mal erst aus der Artikulation ihrer Einzellaute zusammengesetzt werden müsste. Tatsächlich belegen mehrere Studien, dass Silbenhäufigkeiten einen Einfluss auf Reaktionszeiten bei der Produktion von Pseudowörtern besitzen (Carreiras/Perea, 2004; Cholin et al., 2006; Brendel et al., 2008; Cholin et al., 2009). Eine weitere Konsequenz eines solchen Speichers wäre, dass die Lautübergänge innerhalb der häufigen, als Routinen abgespeicherten Silben besser aufeinander abgestimmt und daher stärker koartikuliert sein sollten als die innerhalb seltener oder unbekannter Silben, deren Artikulation spontan synthetisiert werden muss. Es konnte experimentell bestätigt werden, dass häufige Silben im Vergleich zu seltenen eine stärkere Koartikulation aufweisen (Croot/Rastle, 2004; Flechsig, 2006; Benner et al., 2007; Herrmann et al., 2008). Darüber hinaus zeigen Untersuchungen, dass Patienten mit Sprachapraxie häufige Silben flüssiger und mit weniger Fehlern aussprechen können als seltene (Aichert/Ziegler, 2004; Staiger/Ziegler, 2008). Während einigen Studien zufolge Silbenhäufigkeit auch einen verkürzenden Effekt auf die Silbendauer hat (Levelt/Wheeldon, 1994; Herrmann et al., 2008), zeigen andere Untersuchungen einen Effekt in der gegensätzlichen Richtung, der allerdings möglicherweise auf nicht berücksichtigte Faktoren wie Segmentanzahl und Wortbetonung beruht (Croot/Rastle, 2004; Flechsig, 2006).

Ergänzend zur "Mental Syllabary"-Theorie wurde die "Dual Route"-Hypothese entwickelt, um zu erklären, warum einige Apraxie-Patienten eine generell stark verringerte Koartikulation aufweisen (Whiteside/Varley, 1998a; 1998b; Varley et al., 2006). Dieser Hypothese zufolge stehen bei der Sprachproduktion prinzipiell zwei konkurrierende Produktionsmöglichkeiten zur Verfügung – eine direkte Route, bei der die Aussprache aus einem mentalen Speicher von häufigen Silben oder Wörtern abgerufen werden kann, sowie eine indirekte Route, bei der die Aussprache aus kleineren Elementen zusammengesetzt wird. Die indirekte Route kann dabei auch für häufige Einheiten genutzt werden, wenn etwa der besseren Verständlichkeit halber eine verringerte Koartikulation gewünscht wird oder die direkte Route in irgendeiner Weise gestört ist. Bei Apraxie-Patienten mit einer generell verringerten Koartikulation wird nach dieser Hypothese angenommen, dass der mentale Speicher oder der Zugang zu ihm durch die Krankheit geschädigt ist und die Patienten auf eine ebenfalls beeinträchtigte indirekte Sprachproduktions-Route ausweichen müssen.

Viele Vorhersagen der "Mental Syllabary"-Theorie können auch durch einen exemplartheoretischen Ansatz erklärt werden. Hierbei wird angenommen, dass das Gehirn für häufige Silben, Wörter und sogar ganze Phrasen eine Vielfalt von möglichen Artikulationsabläufen abspeichert, von denen bei der Produktion je nach Kontext einer ausgewählt werden kann (Pierrehumbert, 2001; Walsh et al., 2007). Anders als die klassische "Mental Syllabary"-Theorie, die die gespeicherten Silben als vergleichsweise abstrakte Allzweckbausteine begreift, die an Silbenübergängen eine eher herabgesetzte Koartikulation aufweisen sollten, geht die Exemplartheorie davon aus, dass häufige Silben im Vergleich zu seltenen auch nach außen hin eine höhere Koartikulation haben und insgesamt stärker an den jeweiligen Kontext angepasst und somit variabler sind. Eine deutsche Korpusstudie zeigt, dass die Dauer von häufigen Silben nicht so stark abhängig von der Durchschnittsdauer ihrer einzelnen Segmente ist wie die Dauer von seltenen Silben (Schweitzer/Möbius, 2004).

Korpusanalyse

Bevor Experimente zu Silbenhäufigkeiten gemacht werden konnten, mussten zunächst einmal Gruppen von häufigen und seltenen Silben definiert werden. Auf Grund der ungleichen Häufigkeitsverteilung treten viele der potentiell möglichen Silben auch in großen Korpora gar nicht oder nur sehr selten auf. Vor allem Korpora gesprochener Sprache sind daher meist zu klein, um zuverlässige Aussagen für diese seltenen Silbenformen zu erlauben (Möbius, 2001; Schweitzer/Möbius, 2004). Korpora, die aus geschriebenen Texten zusammengestellt wurden, entsprechen zwar nicht dem eigentlichen Untersuchungsgegenstand "spontane Alltagssprache", liefern aber große Mengen an Datenmaterial. Für diese Arbeit wurden verschiedene deutschsprachige Korpora geschriebener und gesprochener Sprache analysiert und miteinander verglichen (siehe auch Samlowski et al., 2011). Auf diese Weise ließen sich auch Häufigkeiten von Silben berechnen, die zu selten sind, um in den kleineren Korpora gesprochener Sprache zu erscheinen. Von den insgesamt sechs untersuchten Datenbanken bestehen zwei aus schriftlichen Texten (Baroni/Kilgarriff, 2006; Quasthoff et al., 2006), während die restlichen vier sich aus Abschriften von ursprünglich gesprochener Sprache zusammensetzen (Wahlster, 2000; Koehn, 2005; Brysbaert et al., 2011; Schütte, 2014).

Bei allen analysierten Korpora wurden die gesuchten phonetischen Silben automatisch aus orthographischen Texten gewonnen. Ursprünglich schriftsprachliche Korpora besitzen gar keine akustischen Entsprechungen, die sich transkribieren ließen. Aber auch gesprochensprachliche Korpora stehen oft nur als mehr oder weniger stark annotierte Texte zur Verfügung. Eine manuelle Transkription und Silbenermittlung wäre ohnehin sehr aufwendig und in der gewünschten Größenordnung nicht durchführbar. Außerdem ist von der Aufgabenstellung her eine eher breite, am Standard orientierte Transkription erwünscht, wie sie sich auch von einem Transkriptionsprogramm mit einem guten Lexikon bewerkstelligen lässt. Auf diese Weise lässt sich verhindern, dass Koartikulationsphänomene wie Elision und Assimilation zu unterschiedlichen Silbenformen führen. Für die Umwandlung von orthographischem Text zu phonetischen Silbentranskriptionen waren eine Reihe von Verarbeitungsschritten notwendig. Teilweise konnte dabei auf Module oder Unterprogramme aus dem Bereich der Sprachsynthese zurückgegriffen werden (Portele, 1999; Möhler et al., 2000; Schunk, 2004; Schmid et al., 2007). Durch Fehler in den Originaltexten sowie bei den einzelnen Verarbeitungsschritten kann es im Bereich der häufigeren Silbenformen zu verfälschten Häufigkeitsangaben und im Bereich der seltenen Formen zu einer Vielzahl neuer, phonotaktisch unplausibler Silben kommen. Dennoch sollte die Korpusanalyse auf Grund der Menge und Vielfalt der verwendeten Daten eine solide Grundlage für die Auswahl von häufigen und seltenen Silbenstimuli bieten, insbesondere, wenn bei den seltenen Silben berücksichtigt wird, dass sie möglicherweise nicht real existent sind, sondern auf Verarbeitungsfehlern beruhen.

In einer vergleichenden Analyse der einzelnen Korpora (siehe auch Samlowski et al., 2011) konnten Unterschiede im Silbenvokabular von geschriebener und gesprochener Sprache ermittelt und die Robustheit von Silbenhäufigkeiten über verschiedene Kontexte hinweg überprüft werden. Die gewonnenen Erkenntnisse ließen sich dabei nutzbringend in die Gestaltung des Produktionsexperiments zu Silbenhäufigkeiten einbringen. So konnte gezeigt werden, dass bei häufigen Silben Häufigkeitsrangplätze über Korpora hinweg stabiler sind als relative Häufigkeiten, während bei seltenen Silben das Gegenteil der Fall ist. Außerdem wurde in einer Analyse von Silbenhäufigkeiten in den einzelnen Datenbanken im Verhältnis zu ungewichteten Mittelwerten dieser Häufigkeiten in allen sechs Korpora untersucht, inwieweit bestimmte Silben in einem der Korpora vergleichsweise über- oder unterrepräsentiert sind (Kilgarriff, 2009). Bei dieser Untersuchung wurden zahlreiche individuelle Besonderheiten der einzelnen Korpora deutlich. Allerdings fanden sich nur wenige systematische Unterschiede zwischen Spontansprache auf der einen und geplanter Sprache auf der anderen Seite. Dies legt nahe, dass sich, von einigen Ausnahmen abgesehen, Silbenhäufigkeiten aus schriftsprachlichen Korpora generell als eine grobe Annäherung an entsprechende Häufigkeiten in gesprochener Sprache verwenden lassen. Eine Untersuchung der Verteilung der besonders korpustypischen Silben über die Datenbanken hinweg konnte belegen, dass diese häufig auch dann noch überrepräsentiert waren, wenn die Korpusgröße miteinberechnet wurde (Gries, 2008). Die Überrepräsentiertheit dieser Silben in bestimmten Korpora war allerdings unabhängig von ihrer Typehäufigkeit (Conrad et al., 2008), ließ sich also nicht darauf zurückführen, dass sie vorwiegend in bestimmten Fachwörtern auftauchten. Für die Auswahl von Stimuli zu Untersuchungen von Silbenhäufigkeitseffekten ist es daher sinnvoll, neben der Stabilität von Häufigkeiten über Kontexte hinweg auch ihre Typehäufigkeit zu berücksichtigen.

Produktionsexperimente zu Effekten von Prosodie und Grammatik

In zwei Produktionsexperimenten wurden anhand von segmental identischen Wörtern Einflüsse von Wort- und Satzbetonung sowie von folgender Wort- und Satzgrenze einerseits und Effekte von Wortart und Lemmahäufigkeit andererseits untersucht (siehe auch Samlowski et al., 2012; 2013; Samlowski et al., 2014). Die Stimuli wurden so ausgewählt, dass eine möglichst isolierte Analyse der einzelnen Faktoren vorgenommen werden konnte. Auch die Trägersätze wurden sorgfältig geplant, um einerseits die intendierte Wortbedeutung zu verdeutlichen und andererseits zu verhindern, dass die Ergebnisse durch die unterschiedlichen Satzkontexte verfälscht werden. Für jedes direkt verglichene Silbenpaar wurde die vorangehende und nachfolgende Halbsilbe konstant gehalten. Sätze aus beiden Experimenten wurden miteinander vermischt und nacheinander den insgesamt 30 Versuchspersonen auf einem Computerbildschirm präsentiert, so dass sie als wechselseitige Distraktoren dienten. Um die Aufmerksamkeit der Teilnehmer auf den Inhalt der Sätze zu lenken, die sie vorlesen sollten, wurde jedem Satz eine passende Illustration beigefügt. Die Illustrationen wurden mit Hilfe des Programs WordsEye erstellt, das Bildbeschreibungen graphisch umsetzt (Coyne/Sproat, 2001). Die Zielsilben aus den aufgenommenen Sätzen wurden manuell mit Praat annotiert (Boersma, 2001) und in Hinblick auf Silben- und Vokaldauer sowie auf akustische Prominenz im Vergleich zur vorhergehenden und nachfolgenden Kontextsilbe (Tamburini/Wagner, 2007) analysiert. Außerdem wurden Paare von segmental identischen Silben, die vom gleichen Sprecher produziert wurden, dahingehend untersucht, inwieweit sie sich in ihren spektralen Eigenschaften ähneln (Lewandowski, 2011).

In dem ersten der beiden Experimente wurde der Einfluss von Wortbetonung, aber auch von Satzbetonung und Gebundenheit auf die Aussprache von Verbzusätzen untersucht (siehe auch Samlowski et al., 2012; Samlowski et al., 2014). Hierfür wurden die Verbzusätze "um", "unter", "über" und "durch" mit jeweils zwei Verben verbunden, die prosodische Minimalpaare bilden können (z.B. "UMfahren" vs "umFAHren"), und in je sieben unterschiedlichen Sätzen eingebaut. In den jeweils ersten vier Sätzen wurde dabei durch gezielte Formulierung die intendierte Wort- und Satzbetonung manipuliert. In den letzten drei Sätzen wurden gebundene und abgetrennte Zusätze sowie gleichlautende Präpositionen oder

Konjunktionen gegenübergestellt. In den akustischen Analysen zeigte sich ein verlängernder Effekt von Satzgrenzen und Wortbetonung auf Silben- und Vokaldauer. Gebundene Verbzusätze tendierten zu leicht geringeren Silben- und Vokaldauern als gleichlautende Funktionswörter. In wortbetonten Verbzusätzen schien auch die vom Trägersatz suggerierte Satzbetonung einen leichten Einfluss auf die Dauer von Vokalen zu haben. Diese waren in Sätzen mit einem breiten Fokus minimal länger als in Sätzen, die einen Kontrast enthielten, der die Betonung von dem Verb ablenken sollte. Akustische Prominenzwerte wurden von Satzgrenzen, Wortbetonung und Satzbetonung beeinflusst, wobei höhere Dauerwerte mit höheren Prominenzwerten einhergingen. In betonter Wortstellung waren sich Silben über Kontexte ähnlicher als in unbetonter Wortstellung. Eine hohe spektrale Ähnlichkeit über Satzkontexte hinweg trat ebenfalls zwischen Silben in abgetrennten, satzfinalen Verbzusätzen auf, die außerdem Silben in wortbetonten Präfixen ähnlicher waren als Silben in nicht wortbetonten Präfixen.

In dem zweiten Experiment (siehe auch Samlowski et al., 2013; Samlowski et al., 2014) wurden die Funktionswörter "der", "die", "das", "dem" und "den" untersucht, die als bestimmte Artikel, aber auch als Relativ- oder Demonstrativpronomina auftreten können, wobei sich die Verwendung dieser Wörter als bestimmte Artikel in einer Korpusuntersuchung als die bei weitestem häufigste erwies (Baroni/Kilgarriff, 2006). Für jedes analysierte Wort wurden Satzpaare mit Relativ- und Demonstrativpronomina gebildet, die vom unmittelbaren Kontext her einem bestimmten Artikel in einem der anderen Sätze des ersten oder zweiten Experiments entsprachen. Für die Wörter "das" und "den" wurden jeweils drei, für die anderen Wörter je sechs Paare gebildet. Bei letzteren wurden die jeweils zwei möglichen Numera/Genera unterschieden. In akustischen Analysen der Aufnahmen wurde gezeigt, dass Relativpronomina von ihrer Silben- und Vokaldauer her kürzer als Demonstrativpronomina, aber länger als Artikel waren. Während Demonstrativpronomina dazu tendierten, prominenter zu sein als die anderen beiden Wortklassen, waren Prominenzunterschiede zwischen Relativpronomina und bestimmten Artikeln uneindeutig. Während bei einigen Wörtern Relativpronomina ähnliche oder leicht höhere Prominenzwerte erhielten als gleichlautende bestimmte Artikel, waren bei anderen Wörtern die Relativpronomina eindeutig die am wenigsten prominente Wortart. Eine nähere Betrachtung der einzelnen Satzkontexte ergab, dass Prominenzwerte besonders niedrig bei Relativpronomina waren, die den Relativsatz einleiteten und deren vorangehender Kontext durch finale Längung beeinflusst und prominenter gemacht wird. Wenn hingegen die Untergruppe der Zielwörter betrachtet wird, die einer Präposition folgen, sind Relativpronomina prominenter als bestimmte Artikel. In diesem Fall weisen die einzelnen Wortklassen auch ausgeprägtere Dauerunterschiede auf.

Die Experimente zeigen, dass insbesondere Dauerwerte von vielen unterschiedlichen Faktoren beeinflusst werden können. Daher ist es für die Analyse von subtilen Effekten wie etwa dem möglichen Einfluss von Silbenhäufigkeit äußerst wichtig, Wortbetonung, Satzbetonung, syntaktische Grenzen, Satzkontext, Lemmahäufigkeit und Wortklasse der Zielwörter zu kontrollieren. So lässt sich vermeiden, dass der Einfluss der gesuchten Variable von Störeffekten überlagert wird.

Produktionsexperiment zu Silbenhäufigkeiten

Das Hauptexperiment dieser Dissertation untersucht die Produktion von phonologisch ähnlichen Silben mit hoher und geringer Silbenhäufigkeit in Bezug auf Fehlerraten, Dauer, akustische Prominenz (Tamburini/Wagner, 2007), und Ähnlichkeit über unterschiedliche Kontexte hinweg (Lewandowski, 2011). In Anlehnung an Cholin et al. (2006) wurden hierfür acht Quadrupel von CVC-Silben gebildet, bei denen jeweils zwei Silbenpaare den gleichen Unterschied in Onset oder Coda aufweisen, der aber mit entgegengesetzten Häufigkeitsunterschieden einhergeht. Ein Beispiel sind die Silbenpaare [fɪn] / [lɪn] und [fɪç] / [lɪç], bei denen [fɪn] und [lɪç] häufiger sind als [lɪn] und [fɪç]. Auf diese Weise kann der Einfluss von intrinsischer Segmentdauer und von Bigramm- und Einzellauthäufigkeiten kontrolliert werden. Dabei wurde sichergestellt, dass die häufigen Silben jedes Quadrupels in allen sechs untersuchten Korpora unter den häufigsten 500 Types zu finden waren, während die seltenen in diesen Korpora eine relative Häufigkeit von höchstens 5e-04 aufwiesen. Die ausgewählten Silben wurden als betonte Erstsilbe in zweisilbige Nonsense-Wörter eingebettet (z.B. "Fintel" / "Lintel" / "Fichtel" / "Lichtel"). Diese Wörter als fiktive Nachnamen wurden in zwei Gruppen von vier Trägersätzen präsentiert, die von jeweils 15 Versuchspersonen vorgelesen wurden. Um Lern- oder Müdigkeitseffekte zu vermeiden, wurde für jede Versuchsperson eine eigene Reihenfolge der Sätze erstellt. Außerdem wurden die Testsätze mit Distraktorsätzen vermengt, um die Wahrscheinlichkeit zu verringern, dass Versuchspersonen die wechselnden Nachnamen in den immer gleichen Kontexten als kontrastiv auffassen und daher besonders akzentuieren. Die aufgenommenen Sätze wurden automatisch mit dem Segmentationsprogramm MAUS annotiert (Kisler et al., 2012) und anschließend manuell korrigiert.

Eine akustischen Analyse der Zielsilben sowie der unmittelbar vorangehenden und nachfolgenden Kontextsilben ergab, dass Silben- und Vokaldauer von häufigen Silben leicht kürzer war als die von seltenen. Dieser kleine, aber signifikante Effekt spricht für eine mentale Speicherung von Silben als ganzen Artikulationsabläufen. Eine Analyse der Aussprachefehler bei den Zielsilben wies entsprechend der "Mental Syllabary" Theorie zwar mehr Versprecher bei seltenen als bei häufigen Silben auf, der Unterschied war aber zu gering, um aussagekräftig zu sein. Da sich kein Effekt von Silbenhäufigkeit auf die akustische Prominenz der Zielsilben in Vergleich zu ihrem unmittelbaren Kontext ergab, ist nicht davon auszugehen, dass der Dauereffekt auf eine stärkere Akzentuierung von Wörtern mit seltenen Silben beruht. Allerdings zeigen sich einige Einflüsse von Silbenhäufigkeit auf Dauer sowie Dauer- und Prominenzvariabilität der Kontextsilben. Diese lassen sich nicht eindeutig erklären, könnten aber darauf hindeuten, dass Namen mit seltenen Silben manchmal zögernd ausgesprochen oder von Phrasengrenzen gefolgt werden. Die exemplartheoretische Vorhersage, dass häufige Silben variabler in ihrer Aussprache sein sollten als seltene, wurde nur teilweise bestätigt. In Vergleichen von segmental identischen Silben, die vom gleichen Sprecher in unterschiedlichen Kontexten verwendet wurden, ergaben sich für häufige Silben größere Unterschiede in dem Pitch-Akzent-Teil der Prominenzanalyse sowie geringere Werte bei der Untersuchung der spektralen Ähnlichkeit. Andererseits waren Unterschiede in der Vokaldauer bei häufigen Silben geringer als bei seltenen, und die Unterschiede in spektraler Ähnlichkeit waren fast ausschließlich auf eines der acht Quadrupel zurückzuführen. Schließlich gab es Anzeichen, dass die manuelle Korrektur der automatischen Segmentierung bei häufigen Silben stärker ausfiel als bei seltenen, was möglicherweise darauf hindeutet, dass häufige Silben stärker von Koartikulationseffekten betroffen waren, die eine präzise automatische Segmentierung erschwerten.

1 Introduction and Outline

According to a number of linguistic theories, syllables play an important part in the process of speech production. Some approaches propose that at least the more frequent syllable types are stored in the brain as whole articulatory routines. The existence of such a mental syllabary may result in pronunciation differences between frequent and rare syllables. In the main experiment of this thesis, possible effects of syllable frequency on duration, acoustic prominence, or spectral similarity in German were examined in a highly controlled reading task. To determine suitable frequent and rare target stimuli, reliable syllable frequency estimates were obtained by using speech synthesis technology to extract syllable frequencies from various written language corpora and transcriptions of spoken language databases. Two further production experiments were conducted to investigate effects of word and sentence stress, syntactic structure, and lexical class. This made it possible to estimate the influence of these factors on duration, prominence, and spectral similarity, to discover whether they need to be controlled in an examination of syllable frequency effects, and to gain insight on the strength of these effects in relation to each other and to possible effects of syllable frequency.

Chapter 2 of this thesis gives an overview of the literature on the role of syllables and syllable frequency in speech production. Chapter 3 describes the six corpora which were used to investigate syllable frequencies and provides an outline of the various steps in the automatic analysis that lead from orthographic texts to syllabified phonetic transcriptions. Two written-language corpora as well as orthographic transcriptions of three spoken-language databases were analyzed. To obtain syllable frequencies, corpus annotations were removed, the character encoding was unified, the written texts were processed to disambiguate numbers and abbreviations, the orthographic words were transcribed phonetically, and the resulting transcriptions were segmented into syllables. The chapter investigates the sources of error connected with each of these steps and discusses their impact on the resulting frequency results as well as the general consequences of using automatic methods to compute syllable frequencies. Suggestions are made on how to reduce the amount of incorrect syllable types due to processing errors when choosing items for a linguistic experiment. In Chapter 4, corpus statistics are presented and relative syllable frequencies and frequency ranks in the individual corpora are compared. The databases are also investigated in terms of similarities between corpora and characteristic key syllables for each of the corpus domains. The results provide insight on how strongly syllable frequencies vary across different domains and whether written language corpora can be used to adequately estimate frequencies in spontaneous speech. In order to further investigate the extent to which individual syllable frequencies are sensitive to particular words or subject matters, analyses were performed with respect to how many different words contained a particular syllable and how evenly syllable appearances were dispersed across the six corpora as well as within the largest of the databases. An important question addressed in this chapter is how frequencies can best be compared if they are highly unevenly distributed and the corpora they stem from differ considerably in size.

Chapter 5 presents a reading experiment on the effects of stress and syntactic boundaries in German. As target stimuli, verb prefixes were used which form minimal pairs depending on whether the prefix or the main verb receives lexical stress. Interactions of word and sentence stress were analyzed by placing the verbs in carrier sentences constructed to suggest different verb meanings and imply either a broad focus or a narrow focus which does not fall on the target verb. As in some inflected verb forms the prefix is separated from the main verb and placed at the end of the clause, it was also possible to investigate effects of sentence boundary. Two further sentences were composed to compare lexically unstressed, bound prefixes with segmentally identical free prepositions or conjunctions in similar segmental contexts. The reading experiment described in Chapter 6 is concerned with effects of lexical class and lemma frequency. Here, German function words were investigated which occur as segmentally identical demonstrative pronouns, relative pronouns and definite articles, with definite articles being much more common than the other two lexical classes. This experiment and the one described in Chapter 5 were combined into a single reading task, allowing sentences from both studies to function as mutual distractors. Each carrier sentence was accompanied by an illustration designed to further clarify its meaning. The recorded target items were analyzed in terms of syllable and vowel duration. An automatic prominence tagger was used to analyze the degree of emphasis placed on the verb prefixes in relation to their immediate context. The tagger examined vowels in terms of pitch movements as well as relative vowel duration, spectral emphasis, and overall intensity. As a third measure, spectral similarity values were determined for segmentally identical prefix syllables produced by the same speaker.

Chapter 7 reports on the main study of this thesis, in which effects of syllable frequency were examined. To avoid interactions with other factors, frequent and rare target syllables were matched for syllable structure and bigram frequency. They were presented in bisyllabic nonsense words in word-initial, lexically stressed position. Target words were embedded in two sets of four different carrier sentences, where they appeared as fictive surnames. The order in which the syllables were presented to the participants was randomized using a Latin square design. Syllable and vowel boundaries were determined on the basis of a manually corrected automatic segmentation. Frequent and rare syllables were analyzed in terms of syllable and vowel duration, acoustic prominence, as well as spectral similarity and variability of duration and prominence values across contexts. In addition, the number of speech errors made for frequent and rare syllables was reported, and results based on corrected and uncorrected syllable and vowel annotations were compared. A general discussion of the results can be found in Chapter 8, while Chapter 9 presents a brief summary of the conclusions drawn from the findings. The test and distractor sentences from the three production experiments along with translations into English can be found in appendices A to D. As an example for the illustrations used in the first two experiments, Appendix E contains the pictures which accompanied the carrier sentences for one of the target verbs in Experiment 1. Appendix F contains tables with mean estimates of acoustic prominence and its two subcomponents force accent and pitch accent for the three production experiments. For frequent and rare syllables in each quadruple in the third experiment, Appendix G shows results of grouped comparisons of segmentally identical syllables produced by the same speaker in different contexts in terms of mean variability of frequent and rare syllables in different contexts. Variability values are presented for syllable and vowel duration as well as prominence, force accent, and pitch accent measures, and mean estimates of spectral similarity are given on the level of syllables, vowels, onsets, and codas.

The research for this thesis has led to several separate publications. Samlowski et al., 2011 is a summarized version of Chapters 3 and 4 and examines five of the six corpora analyzed in this thesis. Samlowski et al., 2012 reports preliminary results from the experiment in Chapter 5 based on 10 of the 30 participants. Preliminary results of the experiment in Chapter 6, based on 20 of the 30 participants, are presented in Samlowski et al., 2013, while Samlowski et al., 2014 is an adapted version of chapters 5 and 6.

2 The Role of the Syllable in Speech Production

Syllables are commonly regarded as intrinsic speech units, fulfilling an important role as the components of metrical feet and potential carriers of stress (Hayes, 1988). Speakers have an intuitive understanding of syllabic units and can use them in various types of word games, such as reversing a word's syllables (Schiller et al., 1997). In phonology, syllable structure is often used as a convenient way of explaining phonotactic rules and phonological processes such as final devoicing (Hulst/Ritter, 1999; Goldsmith, 2009), although some approaches (e.g. Brockhaus, 1999) see no need for including syllables in their models. Among theories of speech production, the exact nature and function of the syllable is still controversial.

2.1 Syllables as Possible Production Units

Initial evidence that syllables might play a role in speech production was mainly derived from analyses of speech errors in various languages (see Meyer, 1992 for an overview). Although such errors rarely pertained to syllables as a whole, they seemed to be affected by syllable structure, i.e. whether the segments in question belonged to the syllable onset, the nucleus, or the coda. Exchanges of individual segments usually concerned identical constituents (onsets exchanged with other onsets etc.), complex constituents tended to move as blocks, and changes affecting two consecutive segments rarely occurred across syllable borders. Exchanged segments often adapted to their surroundings by way of coarticulation and context-sensitive variation, which suggests that such errors occur before the final planning stages of speech encoding. Based on such error analyses, speech encoding models by Dell (1986) and Shattuck-Hufnagel (1983) assumed that during speech production, segments are retrieved individually from the mental representations of phonological words, and that these representations include information about syllable structure to which the selection process is sensitive. In later studies, however, Shattuck-Hufnagel questioned the supposed effect of syllable structure, as she found that the error patterns might be better explained by a combination of the factors word onset and lexical stress (Shattuck-Hufnagel, 1987; 1992).

Several production studies have been carried out to investigate which stages of speech production might involve syllables. Based on the assumption that priming effects would be sensitive to syllable boundaries if the mentally stored phonological representation of words included syllable structure, a number of studies have analyzed the effect of visually presented primes on the production of words with segmentally identical beginnings but different syllable structures (see Cholin et al., 2009 for an overview). Primes matching the beginning of the target words either corresponded exactly to the first syllable or were one segment longer or shorter. Although in a few cases the priming effect did prove to be strongest if the prime matched the first syllable of the target word (Ferrand et al., 1996 for French, Ferrand et al., 1997 for English, Chen et al., 2003; You et al., 2012 for Mandarin Chinese), several other studies found that longer primes caused shorter production latencies regardless of their syllable structure (e.g. Schiller, 1998 for Dutch, Schiller, 1999; 2000; Schiller/Costa, 2006 for English, Schiller et al., 2002 for Spanish, Brand et al., 2003 and Perret et al., 2006 for French).

This, however, does not necessarily mean that syllables play no part at all during speech production in the languages which did not show priming effects. It may also indicate that syllable boundaries are determined at a later production stage than the one accessible through passively viewed primes. One reason why it might be impractical to store syllabified words is that due to phonological processes such as cliticization, inflection, or fusion, in several languages the syllable boundaries in connected speech may differ from those in individually produced words (Levelt, 1992; Cholin et al., 2009). A Dutch corpus investigation showed that nearly 40% of the examined word tokens could potentially be affected by phonological rules modifying their pronunciation or syllabification (Schiller et al., 1996). In implicit priming studies for Dutch, where speakers had to associate a number of target words with certain prompts and produce these words when the prompts were shown, production latencies were shorter if the first syllable was identical for all target words than if one of them had the same initial segments, but a different syllable structure (Cholin et al., 2004; Cholin et al., 2009). A series of Dutch experiments using a similar investigation method to analyze effects of various kinds of prior knowledge concerning the target items found evidence that while words were processed sequentially segment by segment from left to right, speakers seemed to have an additional advantage if they knew the first syllable of the word they were going to produce (Meyer, 1991). Even though an implicit priming study showed no effect of abstract consonant vowel patterns on production latencies in Dutch (Roelofs/Meyer, 1998), there is some evidence that abstract syllable structure can be primed if the primes themselves as well as the target items are spoken out loud by the participants (Sevald et al., 1995 for English, Ferrand/Segui, 1998 for French, Costa/Sebastian-Gallés, 1998 for Spanish).

2.2 The Idea of a Mental Syllabary

In a discussion of various types of speech errors and processes which might explain why particular error types occur or do not occur, Crompton (1981) suggested that syllables are stored in a mental library, where they function as the main articulatory programming units of speech, providing an interface between phonemic representations and their corresponding articulatory routines. One consequence of such a mental syllabary might be that frequent syllables are accessed more quickly than rare syllables. In fact, investigations of production latencies in reading or word association tasks showed that frequent syllables were produced more quickly than rare syllables (Levelt/Wheeldon, 1994; Cholin et al., 2006; 2009 for Dutch, Carreiras/Perea, 2004 for Spanish, Brendel et al., 2008 for German, Cholin et al., 2011 for English). A further study for English found a slight tendency for frequent syllables to be produced more quickly than similar non-existent syllables, but in this case the results did not prove to be significant (Croot/Rastle, 2004). While the effect of syllable frequency on production latencies was found to be independent of word frequency (Levelt/Wheeldon, 1994) and lexical stress (Carreiras/Perea, 2004), there tended to be interactions with the position of the syllable within the word. In some cases, effects on production latencies on bisyllabic nonsense words were found only if the first syllable was manipulated, indicating that participants started to speak as soon as the first syllable had been retrieved (Carreiras/Perea, 2004; Cholin et al., 2006). Although Brendel and colleagues (2008) did not investigate effects of syllable frequency on word-final syllables, they confirmed that the frequency of initial syllables influenced how quickly bisyllabic nonsense words were produced. On the other hand, however, an investigation of real words by Levelt and Wheeldon (1994) showed a frequency effect for the second, but not the first syllable, and Cholin and colleagues (2011) discovered influences of syllable frequency in investigations of the first as well as the second syllable of bisyllabic nonsense words. This suggests that in some cases, participants may wait until the whole word is encoded before actually producing it. Possible reasons for these different findings may be a more holistic treatment of words in comparison with nonsense words, less clear syllable boundaries for English compared with other languages, and influences of different stress patterns (Cholin et al., 2011).

Another unclear point is the amount of interaction between frequency and complexity of the target syllables. Consistent with the idea that the pronunciation of syllables is not computed segment by segment, but rather retrieved as a holistic gestural score, Levelt and Wheeldon's study (1994) showed no additional effect of second syllable complexity on production latencies of words with frequent and rare second syllables. Brendel and colleagues (2008), however, found that nonsense words beginning with simple frequent syllables were produced significantly more quickly than nonsense words beginning with complex frequent syllables, with no such complexity effect occurring for rare syllables. While the authors suggest that maybe not all syllables defined as frequent were actually stored in the mental syllabary and that complexity effects for rare syllables might be masked by frequency differences, there is further evidence from German language studies that the stored motor plans themselves actually include information about their internal structure. Both syllable frequency and syllable structure have been found to influence error rates of patients suffering from apraxia of speech, a disorder affecting the ability to coordinate the various muscle movements necessary for speech production (Aichert/Ziegler, 2004; Staiger/Ziegler, 2008). It was also shown that apraxic speakers were able to improve their pronunciation of complex syllables by training on simple related syllables (Aichert/Ziegler, 2008). These findings indicate that speech production is organized in a way which makes motor programs for frequent and simple syllables less susceptible to brain damage than programs for rare and complex syllables and which allows speakers to synthesize motor plans for unfamiliar syllables by using fragments from related syllables with which they are more familiar. Ziegler (2009) uses a binary model of subsegmental gestures to explain articulatory accuracy. The probability of phoneme sequences being rendered correctly by apraxic speakers is calculated as a weighted product of the probability of correctly producing each individual gesture. While Ziegler does not include any syllable frequency effects, he proposes that irrespective of unit size, frequently co-occurring gestures may be more resistant to errors compared with less frequent gesture combinations.

2.3 Timing Issues

Several of the studies mentioned in this chapter provide evidence on the question of when during the production process syllables might come into play. To recapulate, analyses of speech errors indicate that a division of utterances into syllabic frames, or at least word onsets and rest words, takes place before phonetic encoding, i.e. before allophonic variation and coarticulation effects are specified (Meyer, 1992). While some priming experiments suggest that syllable boundaries are already included in the mentally stored phonological words, a number of other studies contradict this finding (Cholin et al., 2009). The conflicting results may in part be due to some languages having a greater tendency to resyllabify syllables than others, as it makes little sense to store syllable boundaries when the syllabification of phonological words often does not correspond to the syllabification of the same words in connected speech (Levelt/Wheeldon, 1994). Additional evidence that planned utterances are segmented into syllables comparatively early during speech planning comes from a monitoring investigation in which participants had to silently translate words from English to Dutch and decide whether the Dutch words contained certain sound sequences (Wheeldon/Levelt, 1995). Reaction times proved to be shorter if the sequence to be monitored corresponded to the first syllable of the target word than if it coincided with the word beginning, but was longer or shorter than the first syllable. Since participants were able to monitor for single segments even while counting out loud simultaneously, and since such articulatory suppression tasks have been found to inhibit phonetic but not phonological encoding, the results suggest that phonological word plans are sensitive to syllable structure.

Effects of syllable frequency, on the other hand, seem to come into play at a later point during speech production. A series of experiments for French showed that effects of syllable frequency in picture naming and nonsense word reading tasks disappeared if participants were given time to prepare their utterances, but were retained if they had to perform a subarticulation task while waiting for the signal to produce the target items (Laganaro and Alario, 2006). This indicates that syllable frequency effects occur during the phonetic encoding stage of speech production, i.e. after phonological word encoding, which is not sensitive to subarticulation, but before the actual articulation, which would not be affected by additional preparation time. Further evidence that speakers can prepare syllable articulation in advance was found in a Dutch investigation of syllable frequency effects combined with effects of implicit priming (Cholin et al., 2009). Here, the effect of priming was larger for rare syllables than for frequent syllables, with the consequence that the effect of syllable frequency disappeared if participants already knew the first syllable of the word they had to produce next. The findings lend support to the theory that syllables are mentally stored as complete motor plans, that the retrieval of these motor plans during phonetic encoding is sensitive to syllable frequency, and that the phonetic encoding can be planned in advance and kept in memory as long as there is no interference from an articulation task.

2.4 Influences of Syllable Frequency on Pronunciation

Different assembly methods for frequent and rare syllables may affect pronunciation as well as production latencies, since stored routines would be expected to have a closer coordination of articulatory movements than motor plans concatenated spontaneously from smaller units. There is, indeed, some evidence that frequent syllables show a greater amount of coarticulation than similar rare syllables (Croot/Rastle, 2004; Herrmann et al., 2008 for English, Flechsig, 2006; Benner et al., 2007 for German). Although some studies also found differences concerning duration values of frequent and rare syllables (Levelt/Wheeldon, 1994; Croot/Rastle, 2004; Flechsig, 2006; Herrmann et al., 2008), the existence, and even the direction, of such frequency effects on duration values remain unclear, as there were often confounding influences from the number of segments appearing in frequent and rare syllables.

The dual route hypothesis, a theory related to the idea of a mental syllabary, assumes that frequent words and syllables are stored as complete motor plans, but that they can also be produced via an indirect route which reconstructs the auditory representation of these items from smaller units (Whiteside/Varley, 1998a; Varley et al., 2006). While the direct assembly method from stored routines would generally be more economical, the lowered degree of coarticulation gained through the indirect route might be preferable under certain circumstances, e.g. when holding a speech or when pronouncing rare or complicated words (Whiteside/Varley, 1998a). Word repetition studies in English showed that frequent monosyllabic words tended to have longer production latencies and duration values than comparable rare words when produced by healthy speakers or speakers with dysarthria or aphasia, but that no such differences were found for speakers with apraxia (Varley/Whiteside, 1998; Whiteside/Varley, 1998b; Varley et al., 2000). The authors postulate that unlike healthy speakers or speakers with disorders which do not affect motor planning, apraxic speakers have lost their access to stored articulatory routines (although some frequently used motor plans may still be intact) and that instead they need to resort to indirect methods, which, however, have also been impaired by their disorder.

In the classic mental syllabary theory, each stored syllable type exists in the form of a single, comparatively abstract gestural score which can be adapted to different contexts in terms of acoustic prominence and precise articulatory movements (Levelt/Wheeldon, 1994; Levelt et al., 1999). Exemplar-theoretic models, however, presume that syllables, words, and even phrases are stored as clouds of highly detailed individual memories, with frequent items being represented by more exemplars than rare items (Pierrehumbert, 2001). According to this theory, for frequent syllables there would be a large number of slightly different representations available, each of which could serve as a target for production, while rare syllables would be represented by fewer exemplars and might even need to be concatenated from smaller units. As a consequence, frequent syllables would be expected to show a greater amount of variation between different contexts than rare syllables. Schweitzer and Möbius (2004) analyzed the variability of frequent and rare syllables in perceptual space in a German corpus study. They determined the perceptual space by calculating z-score duration values for syllables and individual segments based on the mean duration and standard deviation of the corresponding syllables and segment types in the corpus. While for rare syllables a strong correlation was found between the z-score value of the syllable and the mean z-score value of its segments, for frequent syllables relative segment length was less indicative of how long a syllable would be in relation to other realizations of the same syllable type. This tendency was reproduced in a model in which duration values for rare syllables were calculated from exemplars of individual segments whereas duration values for frequent syllables were determined on the basis of syllable exemplars (Walsh et al., 2007). In terms of pitch accent variability, corpus investigations on the similarity between realizations of different pitch accent types in German found no effect from the frequency of the accent type or the frequency of the syllable on which the accent appeared (Walsh et al., 2008). In an English corpus study, however, an influence was shown concerning the relative frequency of pitch accent types for a given word (Schweitzer et al., 2010). For words which were predominantly produced with a particular pitch accent type, the variability in realizing the pitch accent was lower than for words which tended to occur with other accent types.

The present thesis aims to investigate possible effects of syllable frequency on syllable and vowel duration in a German production experiment. As previous studies have shown that duration effects were easily influenced by other factors, great care was taken to ensure that there were as few interfering influences as possible. To examine the possibility that participants may have emphasized rare syllables more strongly than frequent syllables, acoustic prominence values were calculated for the target syllables in relation to their immediate context. In addition, exemplar-theoretic hypotheses were explored by analyzing spectral similarity as well as duration and prominence variability of segmentally identical syllables produced by the same speaker in different contexts.

3 Extracting Syllable Frequencies from Orthographic Texts

Languages which allow complex syllable onsets and codas, such as English and German, often have a highly uneven distribution of syllable frequencies. A small number of the possible syllable types occur with high frequency, while the rest are very rare (Möbius, 2001). This makes it difficult to obtain reliable frequency estimates for all syllables. Especially spoken language corpora are often too small to adequately represent the rarer syllable types. In addition, manual phonetic transcription and syllable tagging are highly time-consuming tasks and not feasible on a large scale. By applying technology from speech synthesis systems, it is possible to automatically extract syllable frequencies from written language corpora and thereby gain access to large amounts of data. Using this method, several German written and spoken language databases were analyzed and their syllable frequencies compared. During data processing, several potential sources of error and ambiguity were discovered. Through additional scripts, it was possible to reduce the influence of such processing errors, although many had to remain unresolved. This chapter gives a detailed overview of the analyzed corpora, the processing steps used to transform orthographic text to phonetic syllables, and the potential sources of error and ambiguity discovered for each processing step (see also Samlowski et al., 2011).

3.1 Background

Syllable frequencies play an important role in various fields of linguistic study. In psycholinguistics, for instance, both the "mental syllabary" theory and exemplar-theoretic approaches propose that high-frequency syllables are stored in the brain as whole articulatory routines whereas rare or unknown syllables need to be assembled from smaller units (Levelt/Wheeldon, 1994; Walsh et al., 2007). Syllable frequency has, indeed, been shown to have an effect on production latencies (Carreiras/Perea, 2004; Cholin et al., 2006; Brendel et al., 2008), visual word recognition (Conrad/Jacobs, 2004; Hutzler et al., 2004), pronunciation (Croot/Rastle, 2004; Schweitzer/Möbius, 2004; Flechsig, 2006; Benner et al., 2007), as well as fluency and error rates of patients with speech apraxia (Aichert/Ziegler, 2004; Staiger/Ziegler, 2008). Such findings can have implications on language acquisition and speech therapy (Alario et al., 2010; Schoor et al., 2012). Further applications exist in the field of speech technology, where there is evidence that automatic speech recognition and synthesis can be improved by including syllable-like units in addition to the conventional word and phoneme level components (Möbius, 1998; Greenberg, 1999; Raghavendra et al., 2008; Mertens/Schneider, 2009).

Many studies focusing on effects of syllable frequencies in German (e.g. Aichert/Ziegler, 2004; Conrad/Jacobs, 2004; Herrmann et al., 2008; Schoor et al., 2012) calculate syllable frequencies on the basis of CELEX, a lexical database of lemmas and inflected words containing, among other things, information on pronunciation and frequency (Baayen et al., 1993; Baayen et al., 1995; Gulikers et al., 1995). The database was created by adapting lemma transcriptions from the Duden pronunciation dictionary and expanding them to all possible inflected forms. In the 1995 version of the corpus, syllabified phonetic transcriptions were determined for over 360 000 different words containing almost 11 000 different syllables. By analyzing a 5.4 million word corpus of written texts as well as a 600 000 word database of spoken language, it was possible to obtain frequency information for more than 84 000 word types and nearly 7 000 syllable types. In this way, around 83% of the corpus tokens were assigned to a syllabified transcription. Tokens which had to remain untranscribed included numbers, proper names, spelling errors, as well as other words which were not part of the transcription dictionary.

As CELEX offers syllable frequency values only for those syllable types which appear in the analyzed corpus and can be attributed to the lemmas in the pronunciation dictionary, it fails to provide frequency information for many actually existing syllable types (Schweitzer/Möbius, 2004). For this reason, several studies (e.g. Mayer et al., 2003; Schweitzer/Möbius, 2004; Benner et al., 2007; Walsh et al., 2008) base their frequency information on probabilities gained through multi-dimensional clustering, a method which makes it possible to calculate theoretical probabilities even for very rare syllables. Müller and colleagues developed this method by adapting an EM-based clustering algorithm to build multidimensional syllable models (Müller et al., 2000; Müller, 2002). They constructed a three-dimensional model based on syllable onset, nucleus, and coda, as well as a five-dimensional model in which they added information about the position of the syllable within the word and its lexical stress status. The models were initiated with data from a newspaper corpus containing 31 million running words. By looking up the word transcriptions in the CELEX database, the authors determined frequencies for over 9 000 syllable types, which amounted to more than 16 000 types when stress status and word position were distinguished. Syllable frequencies were modeled by determining a number of different syllable classes with independent value probabilities for each of the model's dimensions and summing up the probability of a particular syllable according to each class. Twelve such classes were used by the three-dimensional model, while 50 classes were created for the five-dimensional model. The models were able to calculate probabilities for close to 42 000 different syllables (Schweitzer/Möbius, 2004) and for a total of more than 330 000 syllable types including information about stress status and word position (Flechsig, 2006).

For the present thesis, syllable frequencies were extracted from text corpora as well as from orthographic transcriptions of spoken language databases using speech synthesis technology. This means that the calculated syllable frequencies include appearances in numbers, proper names, abbreviations, and other out-of-vocabulary items. As syllable frequencies were examined in several databases, it was possible to compare individual frequencies and frequency ranks across various contexts and modalities.

3.2 Material

Syllable frequencies from six German language databases were analyzed and compared. Three of the databases tended to have a planned, formal style while the other three contained comparatively spontaneous speech. Syllable frequencies were extracted from orthographic texts. This section describes the six databases.

3.2.1 DeWaC

With 1.5 billion running words, the DeWaC corpus is the largest of the databases investigated here (Baroni/Kilgarriff, 2006). It contains texts which were automatically gathered from the World Wide Web using the Heritrix crawler (Mohr et al., 2004). Even though it is a written language corpus, DeWaC is stylistically diverse. Due to the heterogeneity of the World Wide Web, its texts range from newspaper articles to forum discussions and chat messages. The web pages used to initiate the crawling were found through Google searches for keywords that were specifically chosen to elicit informal personal texts as well as articles written for a large public (Baroni/Kilgarriff, 2006). A number of criteria were used to filter out duplicate texts, tables, and long lists, e.g. by discarding very small or large files, duplicate texts, and files containing a disproportionately high amount of HTML structuring. Additionally, the texts were required to contain a minimum ratio of function words to be included in the corpus, and a keyword list was used to remove pornographic files. These text-internal criteria only marginally restricted the topics and vocabulary of the corpus. The remaining data were stripped of their HTML markup, segmented into sentences, and automatically annotated with part-of-speech tags. A header was created for each text to indicate the web address where it was found. The final corpus is distributed in the form of nineteen subcorpora of roughly similar size.

3.2.2 Leipzig

The Leipzig corpus, provided for this study by the University of Leipzig, contains articles from online newspapers as well as from a selection of other web sources (Quasthoff et al., 2006). With about 170 million running words, it is the second largest database analyzed here. The language style of this corpus is formal and strongly planned. As newspapers are read on a daily basis by large numbers of people, frequency counts from this type of data can have a bearing on mental syllable frequencies at least in language perception. Like the DeWaC corpus, the material for this database was subjected to a filtering process before it was published (Quasthoff et al., 2006). Elements of HTML markup were removed, the texts were automatically divided into sentences, and duplicate sentences were deleted. Several criteria had to be fulfilled for a sentence to be included in the corpus. Sentences were filtered out if they did not begin with a capital letter and end with a punctuation mark, if they contained long sequences of capital or single letters, if they ended in multiple question marks or exclamation points, or if there was a disproportionately large number of commas, periods, empty spaces, or special characters. Additionally, a list of the 10 000 most frequent German words was used to identify and remove non-German sentences. Due to copyright reasons, sentence order was randomized for the final corpus. Sentences were numbered consecutively and placed in a table without any further annotation.

3.2.3 Europarl

The Europarl database includes orthographic transcripts of European Parliament proceedings in several languages (Koehn, 2005). For this study, the German language part of the corpus (release v5) was used. With 37 million running words, it is the largest collection of originally spoken language analyzed here. The speaking style is formal and planned. Given the nature of the corpus, vocabulary specific to European politics is comparatively overrepresented. Transcripts for this corpus were gathered automatically in a crawling process from web pages of the European Parliament (Koehn, 2005). The pages were stripped of HTML tags and annotated before they were included in the corpus. Annotations consist of uniformly structured tags detailing chapter and paragraph breaks as well as speaker turns and identities.

3.2.4 Subtlex

The Subtlex database is not itself a corpus, but a word frequency list condensed from a 25 million word corpus of German movie subtitles (Brysbaert et al., 2011). Although the frequency information is based on invented, pseudo-spontaneous dialogues which often deal with unusual situations and are sometimes translations from other languages, evidence indicates the Subtlex word frequencies predict results from psycholinguistic studies better than word frequencies derived from considerably larger written language corpora (Brysbaert et al., 2011). A reason for this may be that screenwriters and translators strive to imitate real-life conversations and that their introspective judgments are not influenced by linguistic considerations. Moreover, as many people watch movies on a regular basis, the speaking style found there may even influence vocabulary and structure in real-life speech. The Subtlex database is available in two versions. The first version consists of a raw word frequency table including information about whether or not words were recognized by an automatic spelling checker. In the second, processed version, there is an additional column naming the frequencies with which the word forms appear in the 6 billion word Google Books corpus. In addition, words were deleted if they did not appear at all in the Google Books corpus or if they did not begin with a letter. For the present study, the unprocessed version was used to maintain better comparability with the other corpora, none of which had been subjected to a filtering of individual word types.

3.2.5 GF

The GF database comprises around 450 000 words and was provided for this study by the Institute for the German Language IDS. As a subset of their corpus "Gespräche im Fernsehen" (television conversations), it contains real-life dialogues from television interviews and discussion shows (Schütte, 2014). Since the shows address a wide range of topics, the corpus features a comparatively broad range of vocabulary. The language style is spontaneous, but formal. However, the knowledge that the conversations were recorded and broadcast to a large audience may have affected the natural behavior of the speakers. Also, moderators and politicians are often so accustomed to speaking in public that even their impromptu utterances tend to sound like planned speech. The transcripts, which were created specifically for the corpus, follow a detailed set of guidelines (Klein/Schütte, 2004). In general, German orthography was used, although all words were written in lower-case letters and the standard spelling was sometimes modified to reflect deviations from standard pronunciation. All spoken utterances were transcribed, including disfluencies and speech overlap. The corpus was heavily annotated with information about speakers and speaker turns as well as comments on background events or the behavior of the dialogue partners. Various special characters were added to indicate prosodic phenomena such as accentuation, boundary tones, hesitations, and changes in loudness or speaking rate.

3.2.6 Verbmobil

The Verbmobil corpus was created as test and training material for the eponymous research project, the aim of which was to develop an automatic system for interpreting between different languages (Wahlster, 2000). Discussion topics were restricted to dialogues about appointment scheduling and travel arrangements. The system was designed to automatically recognize and translate utterances of the dialogue partners, and then synthesize the translation into the other language. Corpus data was collected by asking participants to act out spontaneous appointment-making dialogues in role play (Burger et al., 2000). Information about their fictive time schedules and the framework of the appointment was given to them beforehand. The transcripts of the German sub-corpora used here comprise around 300 000 running words. Because of its original purpose, the corpus is severely limited in its vocabulary. Words and syllables related to appointment making and travel planning are highly overrepresented, while terms typical for other topics occur rarely or not at all. However, as the dialogues are spontaneous and comparatively natural, the Verbmobil database can serve as an additional source of information about syllables characteristic for spontaneous conversations. The orthographic transcriptions of the corpus were annotated with a variety of additional information (Burger, 1997). Phenomena which were marked included speaker turns, pauses, disfluencies, speech overlap, various types of noise, spelled-out words, foreign terms, proper names, and numbers. While slips of the tongue and variations due to dialect or colloquial language were indicated, the standard spelling of the respective words was given as well.

3.3 Corpus Processing and Error Analysis

A number of processing steps needed to be taken in order to extract syllable frequencies from the orthographic corpus texts. First, annotations were removed and the encoding style was unified. Unless the corpora had been specifically created for linguistic purposes, they had to be submitted to a text normalization process. During this process, numbers, special characters and abbreviations were transformed into the words that would be used when reading the text aloud. Orthographic word frequency lists were then computed. As the automatic phonetic transcription and syllable segmentation did not require context information, it was possible to save processing time by performing these steps on the basis of these word type lists instead of analyzing the entire corpora. Finally, syllable frequencies were calculated from the syllabified word transcriptions and the frequencies of the corresponding orthographic words.

As a complex process involving several different automatic programs, corpus processing was prone to various types of errors. Often, errors resulted in syllables which violated phonotactic conventions, e.g. by containing long consonant clusters. When such syllables were traced back to the words and sentences from which they originated, it was possible to detect various sources of error. However, not all preprocessing errors were apparent in the form of structurally atypical syllable types. In some cases, the correct syllable was replaced by one which already existed in other contexts, leading to a slight distortion of the frequency counts for both the correct syllable and its replacement. The following sections describe the different processing phases and typical errors occurring at each stage.

3.3.1 Original Corpus Text

Incorrect syllables do not always originate from faulty processing. In some cases, they may stem from incorrect or unconventional spelling in the original corpus texts. The misspelling of the German word "Stiftung" (foundation) as "Stftung", for instance, produced the syllable [stftʊŋ], appearing nine times in the DeWaC database and twice in the Leipzig corpus. Typing errors do not necessarily lead to new syllable types. The incorrect spelling of the word "haushoch" (high as a house) as "hasuhoch", which occurred once each in the DeWaC and the Leipzig corpus, resulted in the syllable [haʊs] being replaced by the syllables [ha] and [zuː], both of which legitimately occur in other words. Common typing errors also include missing spaces between words or superfluous spaces within words. While the phonetic transcription program was often able to correctly interpret word sequences which were not separated by spaces, a blank space within a word usually resulted in an incorrect transcription or syllable segmentation. Even if the added space happened to occur at a syllable boundary, the word in question would be analyzed as two words, meaning that word boundary effects such as final devoicing could potentially prompt changes in pronunciation.

Sometimes, deviations from German orthography were intentional. The transcripts of the GF corpus, for instance, used altered spelling to reflect peculiarities of dialectal or colloquial speech without offering orthographically correct versions in their annotations. In web texts, letter repetitions were used in words or interjections as a sign of emphasis (e.g. "Aaarrghh"). Such unconventional spellings were less likely to be correctly interpreted during the automatic phonetic transcription. Foreign words or expressions not included in the pronunciation dictionary tended to result in incorrect transcriptions as well. Especially in the web-based corpora, sequences of special characters or x's were sometimes used to structure paragraphs or to block out personal data or objectionable language. Special characters tended to be replaced with the character name by the text normalization program. Sequences of x's were either spelled out during phonetic transcription or rendered as [ks] sequences if the word in question contained a vowel.

3.3.2 Removal of Annotations

The first step of the corpus preprocessing involved the removal of annotations from the corpora. As each corpus had its own annotation structure, individual removal scripts were created for this task. This was comparatively easy for the first four databases. In DeWaC, Leipzig, and Subtlex, annotations were listed in separate tabular columns, while in Europarl they were marked with XML tags. For GF and Verbmobil, annotation removal was more complex. Several layers of additional information had been integrated into the transcriptions, often in the form of special characters or punctuation marks. Here, the detailed transcription manuals served as a guideline for disentangling the actual text from the annotations (Burger, 1997; Klein/Schütte, 2004).

Most of the annotations were clearly distinguishable from the corpus texts, making it possible to remove them without introducing any bias into the frequency counts. Some cases, however, were more ambiguous. The Europarl corpus, for instance, contained several short phrases with background information including agenda points, parliament actions such as voting procedures, and reactions from parliament or audience members, e.g. applause or heckling.

These phrases did not form part of the actual speeches and discussions. As they were not distinctly marked, however, it was not possible to remove them together with the other annotations. For both GF and Verbmobil, annotations of hesitation and feedback utterances were removed. Even though their inclusion might have made frequency counts more realistic, they would have decreased the comparability with other databases where such disfluencies were only sporadically marked. Moreover, the transcription program would have spelled out purely nasal feedback utterances such as "hmm" or "mhm" letter by letter instead of rendering them correctly. Incomplete words presented a similar difficulty for the transcription program, as these were sometimes spelled out as well. Especially Verbmobil contained several instances of incomplete words which were cut off or interrupted due to technical reasons. To reduce the number of misinterpretations, word fragments consisting of only one letter were removed from the corpus texts.

3.3.3 Adjustment of Encoding Formats

The programs used in this study for text normalization and phonetic transcription required the Western European encoding format ISO-8859-1, one of the fifteen encoding formats belonging to the ISO-8859 standard. Eight binary digits are used to encode each character, leading to an inventory of 256 characters per format. Half of these characters remain the same for all ISO-8859 formats, while the other half is tailored to individual requirements of different language groups (Jendryschik, 2008). For the Europarl database, the conversion process was complicated by the fact that it contained several foreign names and terms including characters which had no counterpart in the target format. Originally, this corpus was encoded in UTF8, a format which uses one byte each for internationally known characters and two or more bytes for language-specific characters (Jendryschik, 2008). In this way, it is possible to represent almost all possible characters in a single format. To allow a smooth conversion from UTF-8 to ISO-8859-1, a script was written to replace the problematic characters with forms that existed in the target encoding. Diacritics were removed and typographical punctuation marks were changed to more basic forms. Rare instances of Greek or Cyrillic writing were transliterated individually as whole phrases. For the other corpora, the process of unifying the encoding formats was less complicated. No processing was required for DeWaC, Leipzig and Subtlex, as they were already encoded in ISO-8859-1. The GF corpus was originally encoded in UTF-16 to depict the various special characters used for annotation. After these annotations were removed, it was possible to transform the encoding without difficulty. The Verbmobil corpus only used characters which had the same binary code in all ISO-formats. German umlauts were indicated by a preceding quotation mark, with the character combinations "a, "o, "u, and "s standing for ä, ö, ü, and ß. These characters were converted to their standard form with a script. Provided that ISO-8859-1 was specified as the output encoding, no further processing was necessary.

Corpora assembled from a large number of different web sources, such as DeWaC, are prone to encoding errors, since encoding formats of web pages are often incorrectly or insufficiently documented. Especially characters which are specific to certain languages tend to be incorrectly displayed if the wrong format is used to decode such files. Because identifying and correcting encoding errors is a very laborious task, it was only possible to address a few types of errors. The character combination "©¬" was recognized as an incorrect depiction of "ß" and corrected accordingly. If HTML expressions were used to depict German umlauts, "ß", or the euro sign (e.g. "ä" for "ä"), these were replaced by the corresponding characters. Other instances of expressions beginning with ampersand and ending with a semicolon were deleted. Many further errors had to remain uncorrected. One problem which was only discovered after the frequency analysis had been completed concerns the encoding formats "ISO-8859-1" and "ISO-8859-15". While they are largely identical, only the latter format can depict the euro sign "€". In the former, the corresponding binary code is rendered as "¤". In DeWaC and Leipzig, both of which are encoded in ISO-8859-1, this symbol often appears adjacent to numbers representing prices, presumably as an incorrect depiction of the euro sign.

3.3.4 Text Normalization

Syllable frequencies extracted from written texts are supposed to be based on the way the texts would be pronounced when read aloud. However, the link between written text and the corresponding spoken words are often ambiguous (Taylor, 2009, p. 44f). Especially numbers, abbreviations, and special characters tend to be read out differently depending on context and meaning. The sequence "1998", for instance, can be pronounced either as "one thousand nine hundred ninety eight", "nineteen hundred ninety eight", or "one nine nine eight", depending on whether it refers to an amount, a year, or part of a telephone number. The process whereby such cases are disambiguated and transformed into appropriate word forms is referred to as text normalization. Of the corpora investigated in this study, only Dewac, Leipzig, and Europarl contained numbers and abbreviations which needed to be disambiguated. The Subtlex database, which was only available as a word frequency list, contained no context information on which to base a text normalization process. Since the transcriptions for GF and Verbmobil had been created for linguistic purposes, care was taken from the start not to include any such ambiguities.

As text normalization is one of the steps required for automatic speech synthesis, it was possible to resort to an existing program for the task. In this study, normalization was performed with the help of the German version of the speech synthesis system Festival (Möhler et al., 2000). The German preprocessing modules of this synthesis system divide the input text into sentences and analyze them. Complex rules deal with numbers as well as acronyms, abbreviations, Roman numerals, measuring units, web pages, and email addresses (Schunk, 2004). Numeric strings are interpreted and processed according to their context as time or date information, sequences of single digits, cardinal numbers, ordinal numbers, decimal numbers, or fractions. As the analysis of large files disproportionately reduced program runtime, corpora were not processed as a whole, but sent through the analysis in sections of 50 lines each. This made it possible to maintain a slow but constant runtime speed of 3.6 kb per second. For each block of 50 lines, modules dealing with text preprocessing were called up in a script, and the resulting sentences added to the output file. Occasionally, sentences contained phenomena which resulted in an internal program error that prevented them from being analyzed. In these cases, the original sentence was printed out instead. In preparation of the normalization phase, the Festival text processing was tested on a small database to identify and possibly counteract potential sources of error. The test database was extracted from a 100 000 word corpus of German newspaper articles from 2005 (Quasthoff et al., 2006). In a first step, sentences were automatically selected from the corpus if they contained numbers, periods in non-final sentence position, or sequences of capital letters. From this subset, 100 sentences were then manually chosen for testing with the aim of including a large variety of of different acronyms, abbreviations, and numerical expressions.

Before and after the corpus texts were processed, they were edited to avoid some of the most common errors. The most noticeable problem with the Festival normalization was that it replaced certain non-German characters, such as accented vowels, with the word "unknown". Because of this, characters which would have been rendered as unknown were replaced with recognizable forms before the processing took place. In the same step, brackets and quotation marks were removed, as they did not add substantial information and tended to cause errors. Different spellings of the word "E-Mail", which tended to cause internal program errors, were unified as "Email". As the Festival analysis did not retain the paragraph structure of the input, the beginning of each line of text was specifically marked. In this way, it was possible to reconstruct the original paragraphs and thereby facilitate the monitoring of the normalization process. As a further measure against potential errors, words which Festival had separated into single characters were reassembled. Even though in several cases the words were in fact abbreviations which should have been spelled out, on the whole the phonetic transcription program described in the next chapter was better equipped than the text normalization program to distinguish abbreviations from normal words due to its pronunciation dictionary and morphological analysis. Moreover, during the automatic transcription, the letters I, V, and X were interpreted as Roman numerals if standing by themselves. It was not possible to preclude all errors during text normalization, however. Misinterpretations are inevitable in an automatic analysis of so many different types of abbreviations, numerical expressions, and combinations of special characters. Even after editing, a few characters remained in the corpora which were interpreted as "unknown", e.g. acute accents that had been used in place of apostrophes. The labeling of sentence beginnings actually introduced a further source of error. Especially for sentence-final short words starting with a capital letter, it increased the possibility that the following period was not recognized as sentence-final and instead produced as a separate word.

3.3.5 Phonetic Transcription

The phonetic transcription of orthographic texts forms an important part of text-to-speech synthesis systems along with text normalization, making it possible to use an existing tool for the task. For this study, the phonetic transcription was performed with the program txt2pho, which was developed for the synthesis system Hadifix (Portele et al., 1995; Portele, 1999) and featured a pronunciation dictionary with about 80 000 entries as well as rules for morphological analysis and simple grapheme-to-phoneme conversion. Further conversion tables dealt with numerical expressions, special characters, and certain acronyms and abbreviations. As output, the transcription program produced a vertical list of characters in SAMPA, a computer-readable phonetic alphabet (Wells, 1997). Pauses and glottal stops were marked with an underscore and each phonetic character was followed by a series of numbers supplying information about pitch and duration for further processing with the speech synthesis system Mbrola (Dutoit et al., 1996).

As no context information was required by the programs for automatic transcription and segmentation into syllables, it was possible to save processing time by performing both steps on the basis of orthographic word frequency lists created for each corpus. Words were transcribed individually and then edited, changing the vertical list of SAMPA characters and numbers into a one-line transcription separated by free spaces. Underscores were removed or, if followed by a vowel, changed into question marks, the SAMPA symbol for glottal stops. Transcriptions were written line by line into a new file, making it possible to link them to the frequency counts of the orthographic words.

The automatic transcription process is particularly sensitive to deviations from standard German orthography. Typing errors can lead to phonotactically implausible phone sequences. Intentional letter repetitions for emphasis, onomatopoetic effect, or censoring purposes were either (partially) spelled out or resulted in long consonant sequences. The exclamation "Aaarrghh" (3 appearances in DeWaC, 1 appearance in Subtlex) was transcribed as [ʔaːʔaʁkha], the word "Schatzzz" ("preciousss" from "Lord of the Rings", 14 appearances in DeWaC, 3 appearances in Subtlex) was rendered as [ʃatststs], and the expletive Schxxxxe ("shxt", 2 appearances in DeWaC) was transcribed as [ʃkskskskskε]. Foreign words constitute another example of how unusual spellings can lead to errors. Even though common Anglicisms are included in the pronunciation lexicon of the transcription program, there still remain several instances where foreign terms do not match German conventions of letter to sound mapping. For instance, the anglicism "Highschool" (1622 appearances DeWaC, 204 appearances in Leipzig, and 360 appearances in Subtlex) was transcribed as [hɪçhaʃoːl].

A further difficulty for the transcription system arises from the uncertainty whether or not words should be interpreted as abbreviations or acronyms. Long compounds tended to be spelled out when written all in capital letters even though the same word was transcribed correctly if it was written in lower case letters or with only the initial letter capitalized. Very long words led to internal errors and were not transcribed at all, especially if the program was not able to segment them into morphological components. On the other hand, actual abbreviations were sometimes interpreted as words. The common abbreviation "usw", for instance, (15 023 appearances in DeWaC, 44 appearances in Leipzig, 6 appearances in Europarl, and 89 appearances in Subtlex) was expanded by the Festival preprocessing to "und so weiter" ("and so forth") only when it was followed by a period. Otherwise it was rendered as [ʔʊsv] during phonetic transcription. Problems can also arise if there is no space between an abbreviation and the previous or following word, e.g. "DieUSA" ("TheUSA", 34 appearances in DeWaC, 31 appearances in Leipzig), which is transcribed as [diːuːzaː]. Finally, the internal text normalization of the transcription program sometimes led to unwanted results. A number of abbreviations and acronyms were automatically expanded without regard to context. This potentially led to errors if the abbreviation was ambiguous. The abbreviation "AA", for instance, (5800 appearances in DeWaC, 1078 appearances in Leipzig, 14 appearances in Europarl, 41 appearances in Subtlex) was generally expanded to "Auswärtiges Amt" ("Foreign Office") by the program, even though in some cases it might stand for "Anonyme Alkoholiker" ("Alcoholics Anonymous") or refer to a conventional battery size.

3.3.6 Segmentation into Syllables

Syllable boundaries were determined with the help of a statistical tagger (Schmid et al. 2007). As input the tagger took a list of word transcriptions in SAMPA, with free spaces separating the individual phoneme characters. The output consisted of the same list with vertical lines added between syllables. A further list of transcribed and segmented words was required as training material. Based on this list, sequences of up to five phonemes were analyzed, estimating for each phoneme the probability of a following syllable boundary. Instead of assuming syllable boundaries between words, the tagger added a special dummy character at the end of each word, thereby capturing the comparatively high probability of consonant clusters in word-final positions. Additional rules in the model ensured that, if possible, syllables contained at least one and no more than two (consecutive) vowels.

The training corpus was created from the BOMP pronunciation dictionary, which also served as the basis for the txt2pho transcription (Portele et al., 1995; Stock et al., 2001). The version of the dictionary used here (FEST-Bomp) contained nearly 150 000 entries. As each entry consisted of a bracket structure with the word's spelling, lexical class, phonetic transcription, and stress structure, the data had to be simplified and brought into the format required by the syllable tagger. Even though the txt2pho transcription system was based on an older edition of the BOMP dictionary used for training the syllable tagger, there were a few systematic differences where the training material needed to be adapted to fit the phonetic transcriptions of the corpus data. Specifically, affricates were split into two phonemes, the sequence [əʁ,] as well as the phoneme [ʁ] following a long vowel, was changed to the vocalized form [ɐ], [j] was replaced by [iː] when not in syllable-initial position, the symbol [~] indicating nasalization was changed to [^], and the short tense vowels [e], [i], [o], [u], [ø], and [y] were changed to long vowels. After training the tagger with the adapted dictionary, it was possible to analyze the phonetic word lists from each corpus. Syllable frequencies were then calculated based on the tagger output and the frequency information from the word lists.

Evaluation experiments have shown the tagger to be largely free of errors, revealing an accuracy rate of over 98% (Schmid et al., 2007). However, accuracy rates may be somewhat lower for the syllable segmentation in the present study, as the transcription conventions employed by txt2pho differ in several respects from the conventions used for the evaluation of the tagger. In the CELEX transcriptions, which formed the basis of the evaluation experiments, no glottal stops were included, affricates were viewed as one phoneme, and no distinction was made between vocalized and non-vocalized forms of /r/. While the annotation of glottal stops may actually improve segmentation results (Schmid et al., 2007), the other differences are inclined to increase ambiguity. Vocalized /r/ functions both as a vowel and as a consonant, and affricates are not distinguishable from the same consonant combinations appearing across syllables. In addition, in its original form the tagger allowed only one vowel per syllable, whereas the version used for this study allowed sequences of two vowels, e.g. [na.tsiːoːn]. A general source of error for the segmentation program is that it cannot analyze morpheme boundaries, which means that sounds from the end of the first part of a compound word are sometimes assigned to the beginning of the second. For instance, the word [[wurm]$_{noun}$[[zer]$_{prefix}$[[fress]$_{verb\ stem}$[en]$_{ending\ (past\ participle)}$]$_{adjective}$ ("worm-eaten", 2 appearances in DeWaC), is segmented as [vʊʁ.mtsɛʁ.fʁɛ.sən] instead of [vʊʁm.tsɛʁ.fʁɛ.sən], resulting in the incorrect syllable [mtsɛʁ]. As the segmentation program works on the basis of phonological word forms, it is also impossible to distinguish between words which are pronounced the same but have a different morphological structure. One example for this phenomenon is the phonetic word [ʔeːɐ], which can be monosyllabic or bisyllabic depending on whether it refers to the German pronoun "er" ("he", 3 943 139 appearances in DeWaC, 607 579 appearances in Leipzig, 42 497 appearances in Europarl, 121 852 appearances in Subtlex, 1457 appearances in GF, 21 appearances in Verbmobil) or the less common German word "eher" ("earlier", 305 305 appearances in DeWaC, 34 125 appearances in Leipzig, 4328 appearances in Europarl, 2208 appearances in Subtlex, 121 appearances in GF, 55 appearances in Verbmobil). In the first case, the final [ɐ] is a consonant forming the syllable coda, while in the second it is a vowel forming an additional syllable. The syllable tagger uniformly segments the word into two syllables ([ʔeː.ɐ]).

3.4 Discussion

Most studies on the effects of syllable frequencies base their frequency estimates on a single database. The approach described here aims to gain information on how strongly syllable frequencies vary across different domains. Two written language corpora and orthographic transcripts of four spoken language databases were automatically analyzed with respect to syllable frequencies. A series of programs were used to remove annotations, unify character encodings, disambiguate numbers and abbreviations, create orthographic word lists, transcribe them phonetically, divide them into syllables, and finally determine syllable frequencies. With this technique it was possible to quickly syllabify large amounts of data and compare results from various domains and modalities. On the other hand, the automatic analysis to some degree disguised the variability in actual speech, and new sources of error were introduced during each processing step.

As the automatic transcription system is based on canonical pronunciation and each word type is transcribed and syllabified individually, the transcriptions largely disregard coarticulation effects, phonological processes, resyllabification across words and differences due to dialect or personal style. The same is true for the CELEX frequencies and the multidimensional clustering data, two of the most popular sources of syllable frequency information in German. Considering that word and perhaps syllable frequency may influence the degree of pronunciation variability and coarticulation (Pierrehumbert, 2001; Croot/Rastle, 2004; Schweitzer/Möbius, 2004; Pluymaekers et al., 2005; Benner et al., 2007; Walsh et al., 2007), this simplification can actually be seen as an advantage, as it means that variations of a canonical syllable form are not divided into separate syllable types, each with their own frequency count, but are rather subsumed under a single type. Transcription errors or inconsistencies also tended to occur when the corpus texts themselves employed deviations from standard orthography to reflect particular pronunciation variabilities. The syllables [ʔɪs] and [nɪç], for instance, are strongly overrepresented in the GF corpus not necessarily because they are particularly typical of television discussion shows, but because the deletion of the final "t" in the words "ist" and "nicht" was marked orthographically in the GF database, but not in the other corpora investigated. On the other hand, resyllabification has been cited as an explanation for why it seems that at least in some languages syllable boundaries are not stored together with phonological word pronunciations in the mental lexicon (Levelt, 1999; Cholin et al., 2004). In a study on lexeme and speech syllables in Dutch, a newspaper corpus was first automatically transcribed and separated into syllables based on canonical word-level pronunciation, and in a second step submitted to a number of sentence-level phonological rules in order to simulate effects of connected speech (Schiller et al., 1996). Sentence-level rules needed to be applied to over one third of all word tokens and resulted in a considerable increase in possible syllable types and overall structural complexity. As correlations between syllable frequencies in both investigations remained very high, however, the authors suggest that lexeme syllables can provide a good approximation of syllable frequencies in connected speech.

Unlike the syllable frequencies in the CELEX database, the corpus investigations described here were not restricted to words appearing in a pronunciation lexicon. A downside of such an automatic analysis is that it is highly error-prone. Although efforts were taken in the corpus analysis by Schiller and colleagues to purge the corpus text of numbers, proper names, and foreign words, the transcription resulted in a number of syllables which clearly violated phonotactic conventions (Schiller et al., 1996). Even after these were removed, the transcribed corpus contained a suspiciously large amount of different syllable types in comparison with a similar-sized sample of the Celex database. These additional types tended to be at the rare end of the frequency spectrum and together made up less than 5% of all syllable tokens. Further sources of error stem from the different requirements of speech synthesis systems and syllable frequency analysis. While the latter requires the transcriptions to reflect the way in which the text would normally be produced when read aloud, the main goal for speech synthesis is that the text be understood by its user. This is why in unclear cases the programs for text normalization and automatic transcription tended to spell out words letter by letter and to read punctuation marks or special characters by name. As a consequence, frequency values for syllables which can be found in the names of letters, punctuation marks, or special characters may be overrepresented. Sometimes an ambiguous structure in a written text can be resolved in different ways which are equally correct and can be used interchangeably. Acronyms, for example, can be spelled out or produced as a single word, leaving their meaning unexplained, or they can be expanded to their non-abbreviated form. Which of the variants is used in actual speech depends on the speaker's personal style, the degree of formality, and the audience's knowledge of the subject in question. As automatic systems have to deal with these structures in a uniform manner, a certain amount of variability is lost in terms of pronunciation and vocabulary usage.

In spite of these errors and inaccuracies, analyzing syllable frequencies from text corpora by using speech synthesis technology is a good way to counteract the difficulties presented by uneven frequency distributions. When choosing experiment stimuli on the basis of an automatic syllable frequency analysis, it can be important to ensure that rare syllable types are not merely artifacts of errors during corpus processing. Apart from examining the word types in which they appear, it can be advisable to impose additional search conditions to increase the probability of finding syllable types which actually exist, e.g. by imposing phonotactic constrictions, by demanding that the syllable types appear in more than one database, or by defining a minimum amount of different words in which the syllable has to appear. For the present study on effects of syllable frequency on pronunciation (see Chapter 7), it was mainly important that the rare target syllables were technically plausible and not inherently more difficult to produce than the frequent syllables with which they were compared. This was achieved by imposing a comparatively simple consonant-vowel structure (CVC) and by having frequent and rare syllables consist of the same bigrams.

4 Comparing Syllable Frequencies in Different Corpora

An important question to consider when investigating effects of syllable frequency is how best to classify frequent and rare syllables. In the literature dealing with effects of syllable frequency there is no consensus on where the boundaries for frequent and rare syllables should be set or whether the decision should be based on relative frequencies (e.g. Levelt/Wheeldon, 1994; Schweitzer/Möbius, 2004; Cholin et al., 2006) or on frequency ranks (e.g. Aichert/Ziegler, 2004; Benner et al., 2007; Brendel et al., 2008). Some investigations subdivide syllable frequencies of segmentally identical syllables depending on the syllable's position within the word or its lexical stress status, especially if the counts are based on multivariate clustering (e.g. Levelt/Wheeldon, 1994; Conrad/Jacobs, 2004; Benner et al., 2007). A few studies also take into account the type frequency of syllables, i.e. the number of word types in which a particular syllable appears (e.g. Croot/Rastle, 2004; Cholin et al., 2006). Although type and token frequency are generally strongly correlated, they may differ with respect to the effects they produce. A lexical recognition study for Spanish designed to disentangle the influence of these two factors showed an inhibitory effect of token frequency, but a facilitatory effect of type frequency (Conrad et al., 2008). Conrad/Jacobs (2004) point out, however, that type frequencies might be less informative for German than for other languages due to German's tendency to form compound words which are not divided by free space.

A comparison of different corpora in terms of syllable frequencies can help answer questions such as whether frequency ranks or relative frequencies are more stable across different contexts, which syllables are comparatively over- or underrepresented in a particular database, how close different corpora are to each other in terms of syllable frequency, and what information type frequencies can provide in addition to token frequencies. Comparing corpora is not a straightforward task, however. Due to their highly uneven frequency distributions, measures of relative vocabulary size or density tend to change with corpus size (Remus/Bank, 2012). Moreover, it is often unclear to what extent differences between small and large databases are due to actual differences in vocabulary and language style rather than a side effect of the fact that smaller corpora tend to have less exact frequency estimates than larger corpora. Another point to consider is that the measures used to determine corpus similarity or extract characteristic keywords often exaggerate differences in certain frequency areas while downplaying differences in others (Chujo/Utiyama, 2006). For the present thesis, various methods of word frequency were adapted to examine and compare the databases described in the previous chapter in terms of syllable frequency (see also Samlowski et al., 2011).

4.1 Background

Methods of corpus analysis and comparison have mainly been used to investigate word frequencies, although they can also be applied to other units such as parts of speech (Rayson/Garside, 2000) and bigrams (Dunning, 1993; Rose/Haddock, 1997; Liebscher, 2003). One aim of corpus comparison can be to quantify the distance between two databases by determining the similarity of individual frequencies across corpora. If subsections of a single corpus are investigated in this way, it is possible to determine how homogenous its data is. Summed Chi square statistics (Kilgarriff, 2001; Liebscher, 2003; Parsons et al., 2009) and summed log likelihood values (Rose/Haddock, 1997; Cavaglià, 2002) are two methods that have been proposed for this task. As both measures contrast the frequencies in each of the two corpora with the frequencies that would occur if both corpora were combined into a single database, the influence of frequency differences is weighted according to corpus size. This means that comparisons between small corpora or between a large and a small database tend to result in higher similarity estimates than comparisons between large databases. Although such a bias in favor of large corpora is helpful for avoiding frequency differences due to inexact estimates, the corpus pairs analyzed need to be somewhat similar in size for the measures to remain comparable. An alternative method for determining corpus similarity, which is independent of corpus size, is to calculate Spearman rank correlations between frequencies in the databases to be compared (Rose/Haddock, 1997; Kilgarriff, 2001; Huffaker et al., 2006). While Spearman rank correlations were outperformed by summed Chi square values in the evaluation of similarity measures conducted by Kilgarriff (2001), they still proved to be comparatively accurate at reconstructing corpus similarity. Kilgarriff suggests that only high-frequency words should be used for the calculation, as rare words often have large differences in frequency rank despite being very similar in actual frequency.

Similarity measures give an overview of how close different corpora are in terms of syllable frequency, but they offer no information on which syllables in particular are responsible for the similarities and differences. Several techniques have been developed to analyze this by computing lists of characteristic keywords which are overrepresented in one database in comparison with another (Dunning, 1993; Rayson/Garside, 2000; Kilgarriff, 2001; Chujo/Utiyama, 2006). As word and syllable frequencies are not randomly distributed in natural language, statistical tests such as the Mann-Whitney Rank test, log likelihood values, or Chi square statistics tend to find significant differences between any two corpora as long as they are large enough (Kilgarriff, 2005). The test measures can, however, be used to rank word or syllable types according to their degree of unexpectedness given their frequency in the two corpora to be compared. A simpler method for discovering over- and underrepresented types consists of calculating ratios between the relative frequencies of the corpora under comparison (Kilgarriff, 2009; 2012). By adding a constant value to the frequency counts beforehand, it is possible to avoid divisions by zero, and to zoom in on various frequency ranges.

Appearances of individual types are not evenly distributed throughout a corpus. Especially if words and syllables are closely associated with a specific topic, their frequency in a database depends more on whether the articles or discussions dealing with the topic are included than on the general corpus domain. This is why small corpora, which are more susceptible to the topics they feature, tend to offer less exact frequency estimates of the language type they are supposed to represent, complicating comparisons between the two. Various methods have been suggested to determine how evenly individual word types are dispersed throughout a corpus, either by comparing frequencies in subcorpora or by computing the distance between individual appearances of a word type (see Gries, 2008 for an overview). Gries developed a *dispersion measure* which is flexible with respect to the number and size of the sections into which the corpus is divided.

Several of these techniques for corpus comparison and analysis were adapted to examine syllable frequencies in the six databases described in Section 3.2. This chapter aims to investigate how steeply syllable frequencies are distributed in the individual corpora, which measures should best be used when determining groups of frequent and rare syllables, whether written language databases can be used to reliably determine syllable frequencies in spoken language, how stable individual syllable types are across databases, and to what extent their variability is related to the number of different words in which they appear on the one hand and the size of the corpora on which the frequency estimates were based on the other. Several of the analysis techniques described in this section were adapted to determine the extent to which syllable frequencies varied across different corpora.

4.2 Methods

Frequencies were visualized and compared in various ways using the statistics program R (R Core Team, 2014) and its module zipfR (Evert/Baroni, 2007). In a first step (Section 4.3.1), the total number of tokens, different types, and hapax legomena (types appearing only once throughout the corpus) were presented on word and syllable level. Type and token definitions were contingent on the preprocessing described in Section 3.3. Corpus texts were split into word tokens at space boundaries, and different word types were defined according to orthography, while syllable types and tokens were determined on the basis of the results from the automatic phonetic transcription and the syllable tagger. To compare frequency distributions as well as individual types, syllable frequency ranks and relative syllable frequencies were calculated separately for each corpus. Frequency ranks were computed by sorting the types in descending order according to their frequency and numbering them consecutively. Types with identical frequency were given successive rank numbers. Relative frequencies were determined by dividing the number of tokens for each type by the total amount of tokens in the corresponding corpus. The distribution of syllable frequencies in the six databases was examined by plotting absolute frequencies against their frequency ranks, using logarithmic scales for both axes. Syllable frequency distributions were also contrasted with similarly calculated distributions on the level of orthographic words. Syllable frequency distributions were also investigated in terms of cumulative relative frequencies by summing up relative frequencies in each corpus sequentially in order of their frequency rank. Both relative frequencies and frequency ranks were contrasted for individual syllable types, using the unweighted mean of the six values as a reference point (Section 4.3.2). In this way it was possible to investigate how robust these measures were across databases and frequency areas.

In Section 4.3.3, similarities and differences between the six corpora are analyzed more closely. To quantify the similarity between the six databases, Spearman rank correlations were computed for all syllables which appeared among the 500 most frequent types in at least one of the corpora. A qualitative investigation of the syllable types particularly characteristic of each database was performed by determining frequency ratios (Kilgarriff, 2009; 2012). Syllable frequencies were divided by the mean of the six individual frequency values, making it possible to compare all databases at the same time. Relative frequencies were not weighted according to corpus size, thereby allowing an equal influence of all six databases on the mean values. Differences among the more common syllables were emphasized by adding a constant of 1000 to the individual values per million syllable tokens as well as to their unweighted mean before dividing the two.

Section 4.3.4 examines the relation between syllable and word types. Relative syllable (token) frequencies were plotted against relative type frequencies, i.e. the number of phonological words in which a particular syllable appeared divided by the total number of syllable types in the corpus. By calculating the ratio of relative token frequencies to relative type frequencies, a measure was gained which indicates to what extent syllable frequencies were dominated by the frequency of individual words. Finally, a subcorpora-based dispersion measure developed by Gries (2008) was used to investigate how evenly syllable types were distributed across the six databases as well as across the nineteen subcorpora of DeWaC (Section 4.3.5). Frequencies in the individual subcorpora were normalized by their total in the combined databases, and distances were calculated between these values and the percentage of the subcorpora themselves in relation to the whole corpus. These differences were then summed up for each word and divided by two. This equation results in a value ranging from zero to one, with zero signifying a completely smooth dispersion and one a highly uneven distribution.

$$dp_w = \frac{1}{2} \sum_{i=1}^{K} \frac{f_w(i)}{f_w} - \frac{N(i)}{N}$$

dp_w: dispersion measure for word w
K: number of subcorpora
$f_w(i)$: frequency of w in subcorpus i
f_w: total frequency of w
$N(i)$: number of tokens in subcorpus i
N: total number of tokens

4.3 Results

4.3.1 General Statistics and Frequency Distributions

This section presents some overall information on the distribution of word and syllable frequencies in the six corpora. For each database, Table 1 gives an overview of the number of different words and syllable types (V_W, V_S), the total amount of running syllables and words (N_W, N_S), as well as the number of Hapax Legomena ($V_W(1)$, $V_S(1)$). In general, the number of different word and syllable types increases with corpus size. The Europarl corpus is an exception in that even though it is larger in size than the Subtlex database, it contains fewer word and syllable types. Possible reasons for this may be that Europarl is comparatively restricted in its domain, or that its transcripts are more closely monitored for errors than Subtlex.

	Words			Syllables		
	N_w Tokens	V_w Types	$V_w(1)$ Hapax Legomena	N_s Tokens	V_s Types	$V_s(1)$ Hapax Legomena
DeWaC	1 529 416 366	8 956 162	4 494 024	3 029 111 173	151 003	55 873
Leipzig	172 668 580	1 648 328	824 771	337 754 854	42 549	12 034
Europarl	37 370 082	307 333	145 201	76 310 287	12 439	2 528
Subtlex	25 397 500	368 394	164 120	40 211 581	21 666	5 058
GF	457 803	30 432	16 702	740 151	6 315	1 772
Verbmobil	270 530	6 588	2 786	421 569	2 698	653

Table 1: Number of tokens, types, and hapax legomena for the six analyzed corpora on the level of orthographic words and phonetic syllables

Figures 1 and 2 depict rank frequency distributions for the six investigated corpora on the level of syllables and orthographic words on a double logarithmic scale. According to George Zipf's eponymous law, which states that for natural language the product of word frequency and a power of its frequency rank is approximately constant, data points in such diagrams would be expected to follow a straight line (Zipf, 1968). At the edges of the distribution, however, word frequencies often deviate from this law (Baroni, 2009). Actual frequency curves tend to be slightly convex, with a slower decline among the high-frequency types, and, especially for large databases, a steeper curve among the very low frequencies. Here, the downward curvation for frequent types proved to be considerably more pronounced for syllable frequencies than for word frequencies. This means that comparatively more high frequency and fewer low frequency syllables were found than words. Verbmobil showed comparatively similar curves for word and syllable frequencies, which is probably an effect of its restricted vocabulary and of its comparatively high number of monosyllabic words (there are, on average, 1.56 syllable tokens per word token in Verbmobil). Even though Subtlex has fewer word and syllable tokens than Europarl, the frequency distribution curves for the two databases overlap, which means that there are fewer rare types and more tokens for the high-frequency types in Europarl than in Subtlex.

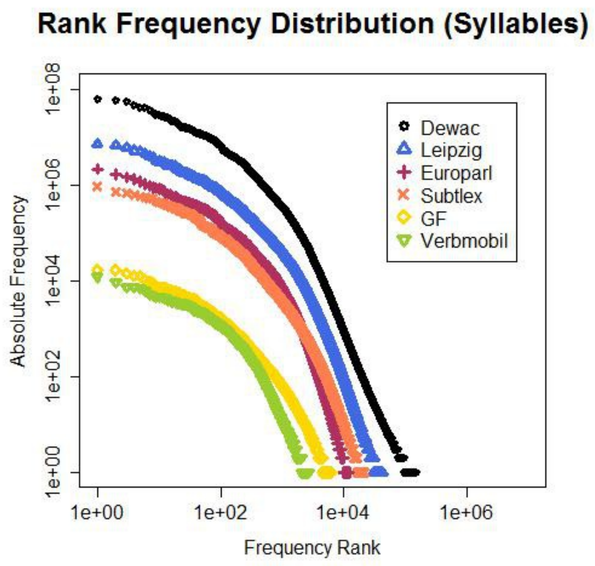


Figure 1: Rank frequency distribution for phonetic syllables in the six analyzed corpora

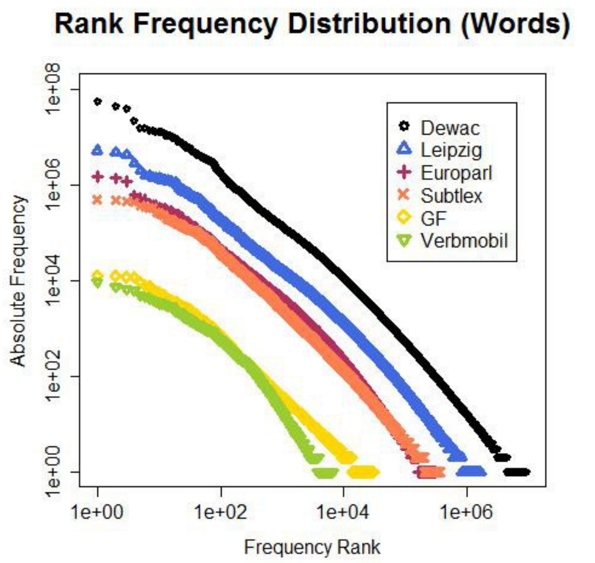


Figure 2: Rank frequency distribution for orthographic words in the six analyzed corpora

The distribution of syllable frequencies in the investigated databases is highly uneven, with the 500 most frequent syllable types of each corpus making up at least 80% of its syllable tokens (see Figure 3). Mere corpus size, however, does not completely explain the differences between the distribution curves. Even though DeWaC is much larger than the Leipzig database, both corpora have very similar frequency distributions. The distribution curves for

Europarl and GF are almost overlapping as well, even though the former is more than eighty times the size of the latter. This is another indication that Europarl has a comparatively small syllable vocabulary. The steepest increase in cumulative syllable frequency by far was found for Verbmobil, which is strongly restricted in its domain as well as the smallest of the databases analyzed.

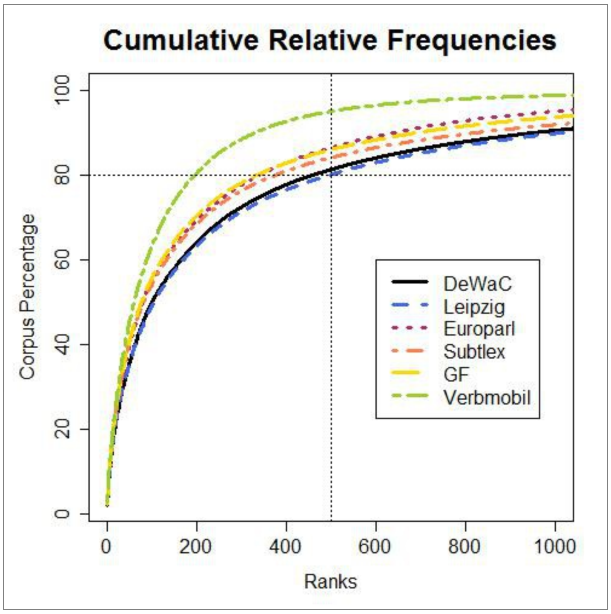


Figure 3: Cumulative relative syllable frequency for the six analyzed corpora

4.3.2 Relative Frequencies versus Frequency Ranks

Although frequency distribution plots are useful for visualizing the number of frequent and rare types in a database, and can, under certain circumstances, give an indication of vocabulary size and diversity, they present no information on the extent to which frequencies of individual types vary across corpora. Relative frequencies and frequency ranks are two measures with which it is possible to directly compare syllable types in differently sized corpora. Figure 4 indicates the relative frequency values for each syllable in the six corpora on the vertical axis, while the horizontal axis represents the unweighted mean of these values. Large differences in relative frequency became apparent especially among the frequent syllables, while the low-frequency syllables were largely clustered in the lower left corner of the graph. Less variation was found for the two largest corpora, DeWaC and Leipzig, than for the other databases, while Verbmobil in particular contained several syllable types which deviated strongly from the mean values.

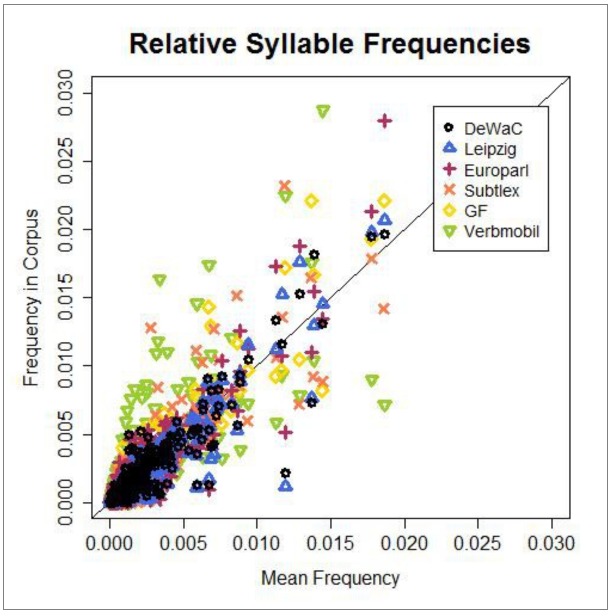


Figure 4: Relative frequencies in individual corpora plotted against their unweighted mean

Figure 5 depicts frequency ranks in the individual corpora on the vertical axis and unweighted mean rank values on the horizontal axis. As the larger corpora contained a considerably greater amount of syllable types and, consequently, frequency ranks than the smaller corpora, a logarithmic scale was used for both axes. Especially considering that in this representation, differences among high-frequency syllable types were emphasized, the diagram indicates that frequency ranks become more varied across databases as syllable frequencies decrease. Again, Verbmobil showed the largest amount of variation from the mean, while the syllable ranks in DeWaC stayed comparatively close to the mean values. One difficulty in comparing frequency ranks is the question of how to deal with syllable types which do not appear at all in a database. While these can justifiably be assigned a frequency value of zero, frequency ranks cannot be defined in a straightforward way. Even if the missing types were regarded as part of the frequency lists and given consecutive ranks, the resulting values would depend on the size of the corpus under investigation and the number of additional types documented in the databases used for comparison rather than on any positive evidence in the corpus itself. In Figure 5, missing values were not depicted and rank means were calculated only on the basis of the corpora which contained the syllable types in question. For this reason, the diagram seems to show that the very rare syllables in DeWaC tend to deviate less from the mean ranks as rank values increase. These syllables often had no counterparts in the smaller databases, which meant that there was a comparatively larger influence of the corpora containing them.

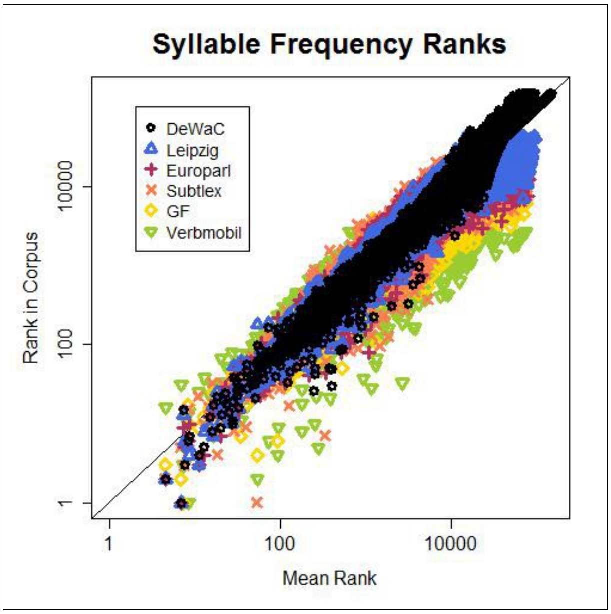


Figure 5: Frequency ranks in the individual corpora plotted against their unweighted mean on a double logarithmic scale

Frequent syllable types tended to be comparatively robust across different corpora in terms of frequency ranks. A total of 186 syllable types appeared among the 100 most frequent syllables in at least one of the databases. Of these, forty-eight types were among the 100 most frequent syllables in all corpora. For the 186 syllable types, Figure 6 presents differences in syllable ranks in the individual corpora subtracted by rank means. Syllables were sorted according to their mean frequency rank. As in the preceding figure, a tendency was visible for rank differences to increase as syllables become rarer. Of the 186 syllable types, 141 had rank differences of less than 250, 93 types deviated from the mean by less than 100, and 57 types had differences of less than 50 ranks. Only the 7 types [tsɪçs], [miːn], [haːp], [hast], [ʔʊnts], [ɪ] and [pɛː] had rank differences of over 1000 in some of the databases and are therefore not completely represented in the diagram. The syllables [ɪ] and [pɛː], which form part of the word "europäischen" (European) and as such were among the 100 most frequent in Europarl, did not occur at all in Verbmobil.

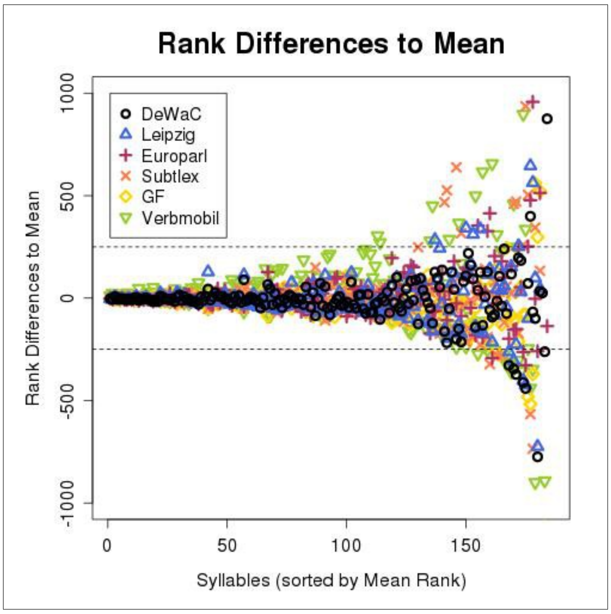


Figure 6: Differences between individual frequency ranks and their unweighted mean for all syllables appearing among the 100 most frequent types in at least one of the six corpora

4.3.3 Overall Corpus Similarity and Qualitative Analysis

In this section, the corpora were further investigated to gain an impression of their overall degree of similarity and to discover which syllable types are particularly over- or underrepresented in one database as compared with the others. Figure 7 depicts Spearman rank correlations for the six databases, taking only those syllables into account which appear among the 500 most frequent types in at least one of the corpora. According to this measure, two groups of databases show correlations greater than 0.8: on the one hand the three corpora of planned language DeWaC, Leipzig, and Europarl, and on the other, the two comparatively spontaneous corpora Subtlex and GF. Verbmobil showed the least similarity to the other databases, possibly because it was the smallest of the databases and restricted to a specific domain. However, it proved to be a bit closer to GF and Subtlex than to DeWaC, Leipzig, and Europarl. These three databases appeared to have a greater resemblance to GF than to the other spontaneous speech corpora Subtlex and Verbmobil. This might be explained by the fact that the television talk shows found in GF often cover political topics which are also dealt with in newspaper articles and internet articles or parliamentary debates.

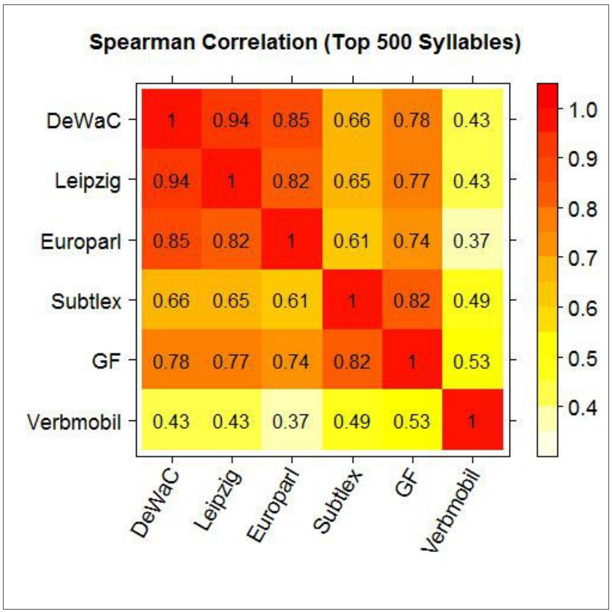


Figure 7: Spearman Correlation based on relative syllable frequencies in the six analyzed corpora, analyzing only syllables which appear among the 500 most frequent types in at least one of the databases

Key differences between the corpora in terms of syllable frequency were analyzed with the help of frequency ratios. Figure 8 depicts how strongly individual syllable types were overrepresented in one corpus as compared with the others. While all databases contained syllables particularly characteristic to them, only a few obvious cases appear where there are systematic differences between corpora of written and spoken language. As already shown in the correlation analysis, Europarl tended to be more similar to the written databases DeWaC and Leipzig than to Subtlex, GF, and Verbmobil, which contained comparatively spontaneous speech. A few examples of syllables overrepresented in spoken dialogues corresponded to the German singular and plural first person pronouns [ʔɪç] and [viːɐ], the German affirmative [jaː], and [das], which can appear in German as a definite article, a demonstrative pronoun, a relative pronoun, or a conjunction.

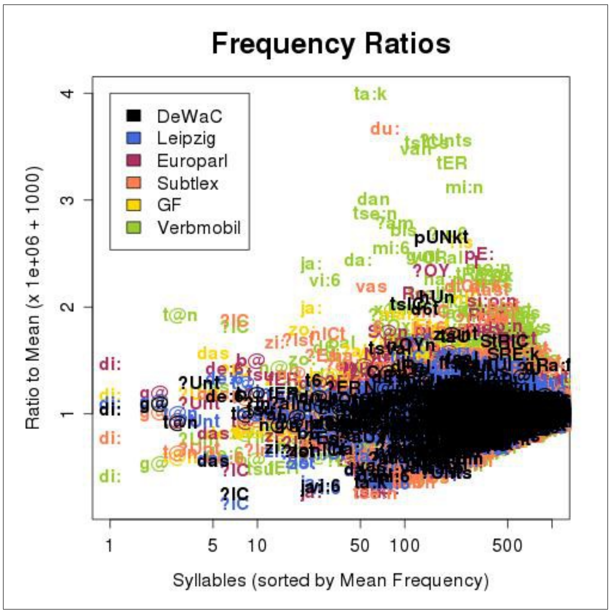


Figure 8: Ratios between relative syllable frequency per million tokens in the six analyzed corpora and their unweighted mean, with a constant of 1000 added to divisor and dividend

A closer examination of words featuring the most overrepresented syllables in each corpus (see Table 2) demonstrates that these can often be attributed to characteristics of the individual corpus domain rather than to general differences between planned and spontaneous speech. This was particularly apparent for the Europarl corpus, where five of the ten most typical syllables can be traced back to the German word for "European". As Subtlex is the only database analyzed which contains a high amount of informal dialogues, its most salient syllables were dominated by words referring to the second person singular, including the informal "you" in various grammatical cases as well as second person singular forms for the verbs "to have" and "to be". Verbmobil contained several key syllables that can be traced back to terms used for making appointments, such as numbers which may refer to dates, the words for "day" and "date" as well as prepositions commonly used with temporal information. Some of the most characteristic syllables in the GF database resulted from the transcription convention of using spelling to reflect colloquial pronunciation, such as final [t] deletion in the words "is(t)" (German for "is") and "nich(t)" (German for "not") or the shortened indefinite article "(ei)ne". Finally, DeWaC and Leipzig both contained a comparatively high number of tokens for syllables associated with the numbers "hundred", "thousand" and "twenty", perhaps due to a higher amount of dates in written texts. DeWaC also contained a few overrepresented syllables referring to punctuation marks and special characters which occur in web addresses, while a few of the syllables characteristic for the Leipzig database appeared in words and expressions common in news reports, such as Germany's capital city Berlin and its government as well as the conjunctive for "to be" which is used in reported speech.

	Characteristic Syllables	Associated Words	Word Translations
DeWaC	pʊŋkt, hʊn, tsɪç, dɐt, taʊ, zənt, ʃtʁɪç, nɔʏn, tsvaɪ, ʃʁɛːk	[pʊŋkt], ['hʊn.dɐt], ['tsvan.tsɪç], ['taʊ.zənt], ['ʃʁɛːk.ʃtʁɪç], ['nɔʏn.tseːn], [tsvaɪ]	dot/period, hundred, twenty, thousand, slash, nineteen, two
Leipzig	zaɪ, taʊ, dəs, tsɪç, zənt, bʊn, tsɛnt, bɛʁ, hʊn, lɐ	[zaɪ], ['taʊ.zənt], ['mɪn.dəs.təns], ['tsvan.tsɪç], ['bʊn.dəs.ʁeː.giː.ʁʊŋ], [pʁoː.'tsɛnt], [bɛʁ.'liːn], ['hʊn.dɐt], [ʔa.lɐ.'dɪŋs]	was (conjunctive), thousand, at least, twenty, federal government, percent, Berlin, hundred, however
Europarl	pɛː, ɪ, ʔɔʏ, ʁoː, siːoːn, ʁɪçt, oːn, paʁ, prɛ, ʃən	[ʔɔʏ.ʁoː.'pɛː.ɪ.ʃən], [kɔ.mɪ.'siːoːn], [bə.ʁɪçt], [ʔuː.niː.'oːn], [paʁ.la.'mɛnt], [prɛ.ziː.'dɛnt]	European, committee, report, union, parliament, president
Subtlex	duː, dɪç, diːɐ, vas, hast, bɪst, mɪç, daɪ, ʔiːn, ʔiːɐ	[duː], [dɪç], [diːɐ], [vas], [hast], [bɪst], [mɪç], ['daɪ.nə], [ʔiːn], [ʔiːɐ]	you (nominative), you (accusative), you (dative), what, (you) have, (you) are, me, your, him, your/her
GF	ʔɪs, nɪç, jaː, man, jɛtst, zoː, fʁaʊ, nɛ, ʔal, zaːkt	[ʔɪs], [nɪç], [jaː], [man], [jɛtst], [zoː], [fʁaʊ], [nɛ], [ʔal.zoː], [gə.zaːkt]	is, not, yes, you/one, now, as/so/like this, Mrs/woman, a(n), so/then, (has) said
Verbmobil	taːk, ʔʊnts, tsɪçs, van, tɛʁ, miːn, dan, tseːn, ʔam, bɪs	[taːk], [tsvaɪ.ʔʊnts.'van.tsɪçs.tən], ['tsvan.tsɪçs.tən], [van], [tɛʁ.'miːn], [dan], [fiʁ.tseːn], [ʔam], [bɪs]	day, twenty-second, twentieth, when, date, then, fourteen, on the, until

Table 2: Top ten syllables most characteristic for each of the six corpora according to frequency ratios, the most frequent words including them, and English translations of these words

4.3.4 Syllable Type Frequencies

In the preceding sections, syllable frequency was analyzed in terms of token frequency, i.e. the sum of all word tokens in which a particular syllable occurs. Syllable type frequencies are determined by counting how many different words contain the syllable in question. As this measure shows the number of words on which the syllable token frequencies depend, it can give an indication of how susceptible token frequencies are to the influence of individual word forms. Syllable type frequencies were calculated on the basis of the phonetic word forms gained during the automatic transcription. This meant that homophonous words were counted as a single type. As larger corpora have a greater number of different words in which the syllables might appear, syllable type frequencies were normalized with respect to corpus size by dividing them by the total number of word types in each of the corpora. For each syllable, Figure 9 shows relative type frequencies on the horizontal axis and relative token frequencies on the vertical axis. Type and token frequencies are necessarily correlated, since token frequencies become higher the more different word types a syllable occurs in. Especially among the more common syllables, however, data points seem to be clustered along two lines (the line at the bottom of the diagram is due to a few very long syllable types stemming from transcription errors). Frequent syllables with low type frequency levels tended to have highly specialized roles as high-frequency monosyllabic words such as the definite articles [deːɐ], [diː], and [das], the first person pronouns [ʔɪç] (singular) and [viːɐ] (plural) or the conjunction [ʔʊnt] ("and"), while syllables with high token as well as type frequencies usually formed productive prefixes such as [gə], [bə], and [fɛʁ], or inflectional endings such as [tə] or [tən].

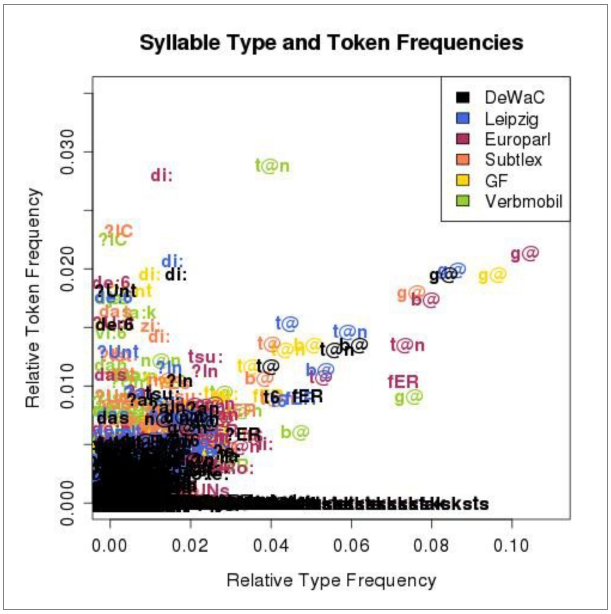


Figure 9: Relative syllable token frequencies in the six databases plotted against their relative type frequencies

One way of determining how versatile individual syllable types are is by dividing relative token frequencies by relative type frequencies, i.e. by computing the number of syllable tokens per word type in relation to the average number of syllable tokens per word type in the corpus. Especially among the higher frequencies, a low token type ratio can indicate that syllables do not stem from artifacts of the corpus preprocessing and are not overly influenced by one particular word. Some of the highest token type ratios were found for syllables stemming from foreign terms ([slɛʃ] for "/" or "Slash"), proper names ([dʒʔɔdʒ] for "George"), unexpanded abbreviations ([bəkl] for "Bekl(agte)", the German word for "defendant"), or incorrect syllable segmentation ([myːʊ] in [bə.'my(.)ʊ.ŋən], the German word for "efforts"). An analysis of token type ratios for the overrepresented syllable types identified in Section 4.3.3 shows that their high frequency was not necessarily dominated by a few isolated words (see Figure 10). While [dan] and [ʔam] in the Verbmobil corpus were highly dependent on the words "dann" (then) and "am" (on), the token type ratios for other syllables typical for one of the databases were spread across a wide range of values.

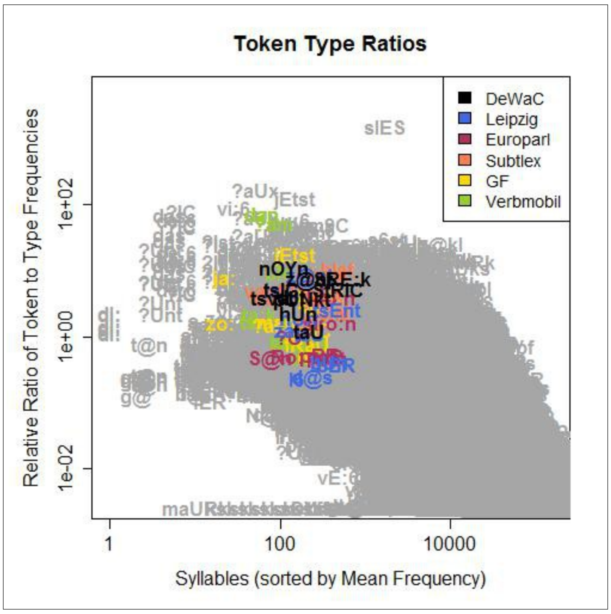


Figure 10: Relative token frequencies for syllables in the six databases divided by their relative type frequencies, with the ten most characteristic syllables of each database according to frequency ratios marked in color

4.3.5 Dispersion Within and Across Databases

To a greater or lesser extent, words can be specific to certain subject matters. This means that a corpus which by chance contains texts dealing with these topics will overestimate certain word and syllable frequencies in comparison to corpora which do not contain such texts. As an indication of how smoothly frequencies are distributed across a corpus, Gries (2008) proposed a subcorpus-based dispersion measure which is flexible with respect to the number and size of the sections into which the corpus is divided. If the six corpora are treated as parts of the same database, this measure can be used to calculate the variability of syllable frequencies while at the same time accounting for differences in corpus size. As rare types have a higher probability of being unevenly distributed than frequent types, the dispersion measure is dependent on syllable frequency. Values need to be compared with those for syllables with similar frequencies to allow conclusions with respect to their dispersion level. Figure 11 depicts dispersion values for the six corpora. Syllables were sorted in descending order according to their frequency in the combined database. The syllables named in Table 2 (see Section 4.3.3) were color coded according to the corpus for which they are most characteristic. Here, it can be seen that syllables particularly characteristic of one of the corpora tended to have higher dispersion values than their immediate surroundings. Generally, the syllables found to be particularly characteristic of one of the databases in an analysis of frequency ratios also showed comparatively high dispersion levels. For several of the syllable types which were typical of the smaller databases, however, differences between corpora were not confirmed in an analysis which takes corpus size into consideration.

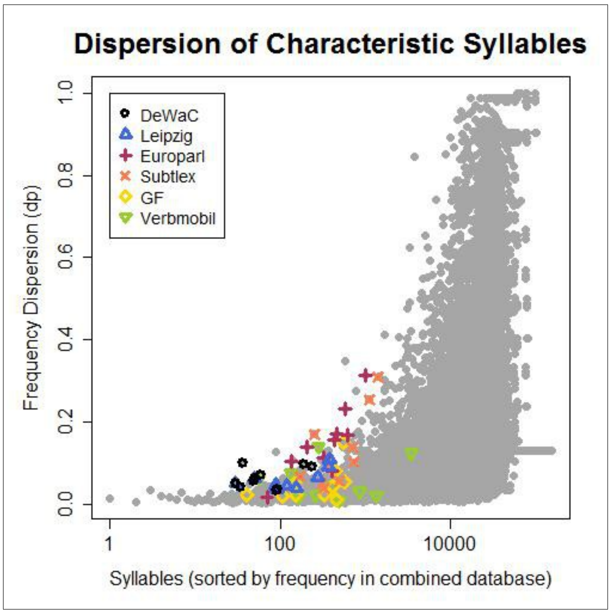


Figure 11: Dispersion level of syllables in the six analyzed corpora with the ten most characteristic syllables of each database according to frequency ratios marked in color

As the DeWaC database is divided into nineteen subcorpora, it can be used to investigate how syllable frequencies vary within one particular domain. Of these subcorpora, eighteen contain around 84 million running words, while the last one has about thirteen million. A dispersion analysis on the basis of syllable frequencies shows that in spite of the broad genre of the corpus (web articles), frequencies in the subcorpora are highly similar (see Figure 12). For the 2500 most frequent syllables in the Dewac corpus, dispersion values stay below 0.1. Spearman correlations between subcorpora remained over 0.95 even when taking into consideration all syllables which appeared among the 10 000 most frequent types in one of the sections. Only 5116 of the total of 151 003 syllable types in the Dewac corpus do not fulfill this criterion. Their relative frequency in the nineteen subcorpora varies between 0 and 6e-07. When these syllables were added to the calculation, correlation values dropped to values of 0.64 to 0.68. The 5116 types made 0.04% of the total number of tokens and varied in relative frequency between 0 and 6e-07.

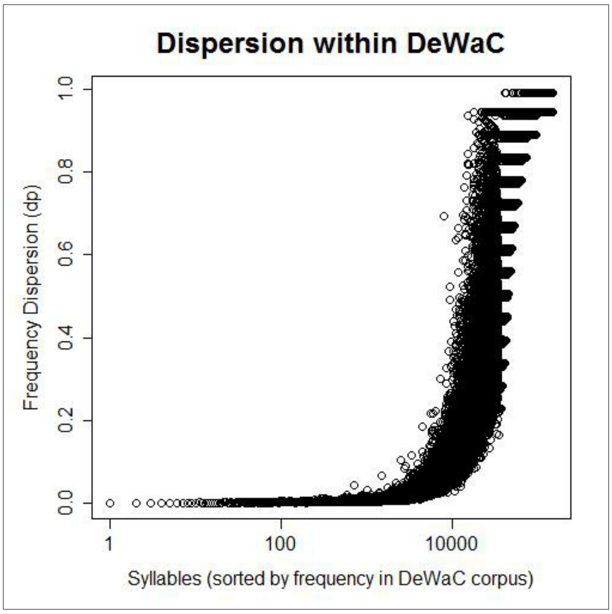


Figure 12: Dispersion level of syllables in the nineteen subcorpora of DeWaC

4.4 Discussion

Before possible influences of syllable frequency can be analyzed, it is necessary to define groups of frequent and rare syllables. However, while spoken language corpora are often too small to adequately represent all existing syllable types, databases of written language may differ in style and vocabulary from spontaneous everyday speech, the usually intended target of linguistic investigations. In general, relative frequencies derived from a single database can be regarded as mean values which are in need of some sort of dispersion measures to qualify them (Gries, 2010). In order to determine how stable frequency values remained across different contexts and to discover to what extent written language corpora can be used to estimate frequencies in spontaneous speech, the six databases from Chapter 3 were compared and contrasted in terms of syllable frequency. Due to their uneven frequency distribution, it is difficult to compare language corpora in a straightforward way, especially if they differ in size. For this reason, a number of different methods were employed to gain a rounded perspective on similarities and differences. In a first step, some general corpus statistics were calculated, rank frequency plots were investigated on word and syllable level, and cumulative syllable frequencies were examined relative to the total number of syllable tokens in each corpus. Both syllables and words had highly uneven frequency distributions, with syllables showing more high frequency and fewer low frequency types than words. Of the over 150 000 different syllables found, the 500 most frequent types in each corpus covered at least 80% of all its tokens. In Section 3.3 it was shown that especially among the rare syllables, many types stemmed from errors present in the original texts or introduced during corpus processing. As opportunities for such errors increase with corpus size, this makes it difficult to estimate the total number of actual syllable types in German. Even though the statistics and distribution plots were to a large extent dependent on corpus size, it was possible to derive some conclusions by comparing the results for different corpora. Europarl in particular appeared to have a comparatively restricted word and syllable vocabulary, especially when contrasted with the smaller Subtlex database.

As long as frequency counts are based on a single database, it makes no difference whether the classification into frequent and rare syllables is based on relative frequencies (e.g. Levelt/Wheeldon, 1994; Schweitzer/Möbius, 2004; Cholin et al., 2006) or on frequency ranks (e.g. Aichert/Ziegler, 2004; Benner et al., 2007; Brendel et al., 2008), since there is an unambiguous assignment between the two. In a comparison of several databases, however, the robustness of both measures across databases proved to be highly dependent on the frequency level of the syllable type under investigation. While the largest differences between corpora in terms of relative syllable frequencies appeared for the more frequent types, frequency rank differences grew as syllables became rarer. As many of the rarer syllable types were assigned different frequency ranks even though they appeared the same number of times, frequency ranks also became indicative of frequency changes as syllable types became rarer. A good procedure when classifying syllables into frequency groups on the basis of different databases may be to use frequency ranks for determining high frequency types and relative frequencies for defining rare syllables, thereby maximizing the robustness of values across databases.

Both in terms of relative frequencies and frequency ranks, Verbmobil tended to show most deviations from mean values. This was confirmed in an analysis of Spearman rank correlations using syllables which appeared among the 500 most frequent in at least one of the databases. The two written language corpora appeared to share more similarities than the spoken language databases. Possibly due to its formal, planned speech style, Europarl was closer to the written than to the spoken language corpora. As the investigated databases differed in size as well as content, syllable frequency differences may stem from inexact estimates on the part of the smaller corpora as well as actual differences between the databases in terms of style and topics. With the help of a qualitative analysis using ratios between individual frequencies and mean frequency values, syllable types were identified which were particularly overrepresented in one database as compared with the others. A constant value added to both measures before computing the ratios was chosen so as to focus the investigation on syllables with a comparatively high frequency. The ten most characteristic syllable types in each database tended to appear in words which were linked to the subject matter and language style of the corresponding corpus. Although there were a few consistent differences in syllable frequency between formal written texts on the one hand and spontaneous dialogues on the other, these tended to be overshadowed by differences due to individual corpus characteristics. The results show that while syllable frequencies in written language can mostly be taken as a rough estimate of frequencies in spoken language, it is still advisable to rely on more than just one database when selecting a set of frequent and rare syllable types for research purposes. By taking into account information from several different corpora, it can be ensured that the frequency of the chosen syllables remains stable over various contexts and modalities.

The ten most characteristic syllables of each database were further investigated in terms of how strongly they were influenced by the frequency of individual words and how smoothly their appearances were distributed across the combined database of the six corpora. An analysis of ratios between relative token and type frequencies showed that although the overrepresented syllable types could be connected with specific word forms, their frequency was not necessarily dominated by those words. However, when syllable frequencies in the individual corpora were compared with the frequencies that would be expected if the corpora belonged to a single database, several of the syllables shown to be characteristic of a particular database were, indeed, found to be unevenly distributed in the combined database. Nonetheless, especially the smaller corpora also contained some types which appeared to be strongly overrepresented in an analysis of frequency ratios but were not confirmed to be unequally distributed across databases when corpus size was taken into consideration. The results indicate that both type frequency and dispersion measures can be used to qualify the findings from the frequency ratio analysis, showing how strongly the syllable frequencies are connected to certain words and to what extent the apparent overrepresentedness of certain syllables may simply be due to the small size of the corpora containing them. In an investigation of the nineteen subcorpora of DeWaC, most of which contained around 84 million word tokens, syllable frequencies proved to be highly similar to each other, revealing that corpora of this size are more than sufficient to adequately estimate frequencies of most existing syllable types even for a very broad genre such as web texts.

Frequent and rare syllables in the experiment on syllable frequency effect (see Chapter 7) were defined on the basis of the corpus study described in this chapter. Frequent syllables were determined according to frequency ranks, while rare syllables were ascertained on the basis of relative frequencies. To ensure that the target syllables were consistently frequent or rare across different contexts, it was postulated that the conditions (being among the top 750 syllable types for frequent syllables and having a maximum relative frequency of 5e-6 for rare syllables) had to be fulfilled in all six databases. To prevent the target syllables from being overly dominated by certain word forms, it was further stipulated that the ratio of relative token frequencies to relative type frequencies had to remain above three in all six databases.

5 Experiment 1: Effects of Word Stress, Sentence Stress, and Syntactic Boundaries

Spoken syllables can vary strongly in the way they are pronounced, even when in canonical pronunciation they are segmentally identical. One source of variation is prominence, i.e. the comparative degree of emphasis which is placed on syllables and with which they are perceived. Strong emphasis may be realized, among other things, through pitch accents, higher duration and intensity values, and overall larger articulatory effort (Wagner, 2002). Differences in syllable prominence can serve various functions. In languages which employ lexical stress, stressed syllables are differentiated from their unstressed neighbors (Dogil/Williams, 1999). On the level of the sentence, prominence differences are used to distinguish the new and important parts of an utterance from information which is less important or which has already been established (Wolters/Wagner, 1998). Duration in particular is also used as a cue for prosodic boundaries, for instance in the form of final lengthening (Turk/Shattuck-Hufnagel, 2007). The experiment described in this chapter uses a reading task to examine effects of word and sentence stress as well as different types of syntactic boundaries on the pronunciation of German verb prefixes (see Samlowski et al., 2012; Samlowski et al., 2014).

5.1 Background

A number of languages have prosodic minimal pairs, i.e. words which are segmentally identical but differ in meaning depending on which of their syllables carries the lexical stress. In perception studies where acoustic recordings of such minimal pairs were manipulated in terms of various possible indicators of stress, duration was shown to have a strong effect on the meaning attributed to the target words (e.g. Fry, 1955; 1958 for English, Ortega-Llebaria et al., 2007 for Spanish, and Ortega-Llebaria et al., 2010 for Catalan, Heuven/Jonge, 2011 for Dutch, and for German). The studies found smaller influences for vowel quality (Ortega-Llebaria et al., 2010; Heuven/Jonge, 2011; Kohler, 2012), intensity (Fry, 1958; Ortega-Llebaria et al., 2007; Ortega-Llebaria et al., 2010), and spectral tilt, i.e. the intensity in higher frequency bands compared with the intensity in lower frequency bands (Ortega-Llebaria et al., 2010). It was also shown that fundamental frequency levels and pitch movements in particular may affect the perception of word stress and, under certain circumstances, even override contradictory duration information (Fry, 1958 for English, Kohler, 1987 for German).

Production experiments have established duration, vowel quality, and fundamental frequency as correlates of word stress in German, while only small differences were shown for intensity levels (Dogil/Williams, 1999; Kleber/Klipphahn, 2006). Various voice quality measures also proved to be affected by changes in word stress (Marasek, 1996; Claßen et al., 1998; Schneider/Möbius, 2007; Lintfert, 2010). Duration differences between stressed and unstressed syllables were found to be considerably stronger for tense than for lax vowels (Kleber/Klipphahn, 2006; Miglio/Chun, 2008). Corpus studies of child-directed speech confirmed duration and voice quality parameters as robust signals of word stress, with duration being used more consistently than voice quality (Schneider/Möbius, 2006; 2007; Lintfert, 2010). Stress-related differences were also found for formant values and fundamental frequency, although the influence of the latter decreased with the age of the child in question (Schneider/Möbius, 2006; Lintfert, 2010). In a corpus analysis of read speech, duration values were correlated with word stress more strongly than measures of intensity and fundamental frequency (Mengel, 1997). Here, the integral of intensity over the rhyme duration proved to be a stronger predictor of lexical stress for bisyllabic words than the duration values themselves.

In most of these production studies, effects of word stress were confounded with influences of sentence stress. Corpus studies did not differentiate between accented and unaccented words and the target items of the experiments were either placed in the focus of the carrier sentence or presented as lists of individual words. Dogil and Williams (1999) performed an experiment which specifically investigated the effects of word stress in relation to sentence stress by having participants read the word "Liliputaner" [li.li.pu.'ta:.nɐ] as well as a reiterant [da.da.da.'da.da] in accented as well as unaccented sentence positions. While the accented condition showed higher fundamental frequency levels and stronger frequency changes for lexically stressed syllables than the unaccented condition, the differences between the two remained insignificant. A syntagmatic comparison of the five syllables showed significant differences between stressed and unstressed syllables in terms of duration but no effects of fundamental frequency or intensity. For the reiterant word, there was also a difference in vowel quality between syllables with and without lexical stress. The experiment material was also used in an automatic classification task in which the classifier was given various combinations of available cues (Rapp, 1995). Again, duration proved to be the most important indicator of lexical stress. In a separate investigation of words in focused and unfocused conditions using a number of different fundamental frequency parameters, no significant improvement of recognition results was found for focused words. For word reiterations, the model actually performed better when the reiterant [da.da.da.'da.da] was not in the focus of the sentence.

Evidence from other languages points towards interactions between word stress and sentence stress (Campbell/Beckman, 1997; Okobi, 2006; Cho/Keating, 2009; Plag et al., 2011 for English, Sluijter/Heuven, 1996 for Dutch, Cambier-Langeveld/Turk, 1999 for Dutch and English, Ortega-Llebaria/Prieto, 2011 for Spanish and Catalan, Sadeghi, 2011 for Persian). Differences in intensity and fundamental frequency tended to be strongly reduced for words which were not in the focus of the sentence (Campbell/Beckman, 1997; Ortega-Llebaria/Prieto, 2011; Plag et al., 2011), and the influence of sentence stress was not necessarily limited to lexically stressed syllables (Sluijter/Heuven, 1996; Cambier-Langeveld/Turk, 1999; Cho/Keating, 2009; Sadeghi, 2011). While several studies established duration as a strong signal of word stress operating independently of sentence accent (Sluijter/Heuven, 1996; Okobi, 2006; Cho/Keating, 2009; Ortega-Llebaria/Prieto, 2011; Sadeghi, 2011), Plag (2011) found no effect at all of word stress on duration and Campbell and Beckmann (1997) discovered stress-related duration differences only in one of the two unaccented contexts examined. For English and Dutch, the parameter spectral tilt appeared to be another robust correlate of word stress in accented as well as unaccented contexts (Sluijter/Heuven, 1996; Okobi, 2006; Plag et al., 2011).

Pronunciation and syllable prominence is affected by prosodic boundaries as well as word and sentence stress. Phrase-final lengthening has been demonstrated for German in controlled production experiments (Féry/Kentner, 2010; Kentner/Féry, 2013) as well as studies based on corpora of spontaneous speech (Streck, 2004; Peters et al., 2005). Prosodic boundaries have been statistically modeled in recognition tasks using information about fundamental frequency, duration, and intensity (Batliner et al., 1995; Strom/Widera, 1996). While Kohler (1983) suggests that utterance boundaries affect the whole word preceding them, with a particular lengthening of the last syllable as well as the one carrying the main stress, Streck (2004) did not find any evidence of lengthening effects extending beyond the final syllable. His study showed, however, that accented final syllables tended to be longer than unaccented final syllables. For English and Dutch, studies do not provide a uniform picture concerning the scope of final lengthening and possible interactions with sentence accent. In English, phrase-final lengthening appeared to affect the final syllable as well as the one carrying lexical stress (Shattuck-Hufnagel/Turk, 1998), while for Dutch, lengthening effects seemed to extend to the penultimate syllable only when the word in question ended with a schwa vowel (Cambier-Langeveld et al., 1997). Moreover, Cambier-Langeveld (1999) found that differences in sentence accent were neutralized for utterance-final words in Dutch, but not in English. Finally, there is also evidence of word boundary effects on pronunciation, with consonants being strengthened and coarticulation reduced at word boundaries, particularly word-initially (Fougeron/Keating, 1997; Cho, 2004), and syllables being shortened as word length increases (Turk/Shattuck-Hufnagel, 2000; White/Turk, 2010).

The present experiment employed German verbs which formed prosodic minimal pairs to investigate effects of implied word stress and sentence focus on the verb prefixes. In addition, effects of word and sentence boundaries were examined by comparing the word prefixes with segmentally identical function words and by placing the verbs in inflected forms where the prefix is separated and placed at the end of the sentence. The target items were analyzed in terms of syllable and vowel duration. An automatic prominence tagger based on various acoustic measures was used to determine the level of emphasis placed on the target items in comparison to their immediate surroundings. Finally, segmentally identical syllables produced by the same speaker were investigated through a measure of spectral similarity to discover how strongly they resembled each other in pronunciation.

5.2 Methods

Sentences from this experiment and the one described in Chapter 6 were combined into one reading task, allowing sentences from both studies to function as mutual distractors.

5.2.1 Participants

Thirty participants took part in the experiment (15 men, 15 women, ages ranging from 19 and 47). All were native speakers of German. Participants were compensated for their time.

5.2.2 Material

Certain German verbs can differ in meaning depending on whether their lexical stress falls on the prefix or the verb stem. For example, the word [ʔʊn.tɐ.ʃtɛ.lən] ([unter]$_{\text{prefix}}$-[[stell]$_{\text{stem}}$-[en]$_{\text{ending}}$]$_{\text{verb}}$), literally "to underput") means "to store / take shelter" when stressed on the prefix, but "to insinuate" when stressed on the stem. This ambiguity is not visible in all inflectional forms, however. In most finite forms, lexically stressed prefixes are separated from the verb and placed at the end of the clause. As the verb prefixes in this experiment are segmentally identical to prepositions or conjunctions, it was possible to use them to investigate effects of word and sentence boundaries as well as influences of word and sentence stress. Target items consisted of the four German verb prefixes "um" ([ʔʊm] – "around"), "unter" (['ʔʊn.tɐ] – "under"), "über" (['ʔyː.bɐ] – "over"), and "durch" ([dʊʁç] – "through") combined with two different verb stems each. The phonetic transcriptions given here are canonical. In actual speech the glottal stop preceding onset vowels may be omitted or realized through vowel glottalization, and the [ʁ] in "durch" is commonly rendered as [ɐ]. Each of the eight resulting verbs was placed in seven different carrier sentences. The resulting 56 sentences are listed in Appendix A.

Category	**Canonical Word Stress**	**Additional Semantic Contrast**	**Right-Hand Boundary**
w+s+	**yes**	**no**	(morpheme)
w+s-	**yes**	**yes**	(morpheme)
w-s+	**no**	**no**	(morpheme)
w-s-	**no**	**yes**	(morpheme)
sb	(undefined)	(no)	**sentence**
mb	(no)	(no)	**morpheme**
wb	(undefined)	(no)	**word**

Table 3: Word and sentence stress status and succeeding syntactic boundary for the target items in each of the seven sentences

In each group of seven sentences, word and sentence stress were varied in sentences 1 to 4, while sentences 5 to 7 compared different types of syntactic boundaries (see Table 3). As a way of further clarifying the intended word meaning and fixing the participants' attention on the content of what they had to read, illustrations were created for each sentence with the help of the text-to-scene conversion program WordsEye (Coyne/Sproat, 2001). Some example illustrations together with the sentences they accompanied are presented in Appendix E. In addition to the test sentences, 48 similarly illustrated sentences were used as distractors (see Chapter 6 for details).

Word and sentence stress differences were not elicited in a uniform manner. As participants needed to be able to infer the verb meaning and intended stress pattern from the sentence context, a different set of carrier sentences was created for each verb. While sentences belonging to the categories "w+s+" and "w-s+" were formulated so as to imply a broad focus, sentences in categories "w+s-" and "w-s-" contained elements designed to attract a contrasting focus and thereby move the sentence stress away from the main verb. Among the strategies used for this were the inclusion of two contrasting objects, topic fronting, and the addition of an emphasized modifier. A deliberate decision was made against using underline, font style, or a question-answer structure to indicate lexical stress and sentence focus, as this might attract the participants' attention to the intended reading and thereby evoke exaggerated responses. As a result, the context in which the target items appeared was not controlled across verbs and only up to a limited degree within each set of sentences. For the first four sentences within one set, the half syllable preceding and following the prefix were kept constant. Sentence 5 ("sb") contained a separated prefix in sentence-final position which was preceded by the same half-syllable as the prefix in the first four sentences. While the prefix in sentence 6 ("mb") fulfilled the same conditions as in sentence 3 ("w-s+"), its target sentence was formulated so that the preceding and following half-syllables matched those of the identical prepositions or conjunctions in sentence 7 ("wb"). For the sake of brevity the first four sentence categories will be referred to in terms of stressed and unstressed prefixes ("w+" vs. "w-") in accented and unaccented conditions ("s+" vs. "s-").

It is important to note, however, that the categories do not reflect the actual stress patterns used by the participants. Instead they describe potential differences in word and sentence stress due to different word meanings and the presence or absence of an additional motivation for deaccentuating the verb. This experiment aims to discover the extent to which these conceptual differences were realized in the acoustic production of the target syllables.

5.2.3 Procedure

The experiment took place in a sound-treated chamber at Bielefeld University. Acoustic recordings were performed using a Sennheiser TLM 103 microphone. Although electroglottographic recordings were made as well, they were not analyzed for this thesis. Test sentences for this experiment and the experiment described in Chapter 6 were put into quasi-random order, ensuring that there were no direct sequences of sentences using the same verb. Three additional illustrated sentences were placed at the beginning of the experiment to allow participants to get used to the setting (see Appendix D). The order of sentences was not varied between participants, who proceeded through the experiment in a self-paced manner. One sentence at a time was presented on a computer screen along with its illustration. Participants were asked to look at each sentence and the accompanying picture and then read the sentence out loud, using their regular speaking/reading style. Before the experiment, they were instructed to repeat sentences if they noticed any mistakes or slips of the tongue. Sentences in which participants corrected themselves or hesitated noticeably were omitted from the analysis, as were sentences where the target items or their immediate context was impaired through speech errors, noise, or strongly reduced speech. The remaining stimuli were analyzed acoustically in terms of syllable and vowel duration, acoustic prominence, and spectral similarity.

For the duration analysis, syllable and vowel boundaries of the target items were manually annotated with the program Praat (Boersma, 2001), using auditory cues as well as the visual representation of the spectrogram and the sound waveform to determine boundary positions. Based on the annotations, syllable and vowel duration values were calculated in seconds, rounding up to the third decimal place. Acoustic prominence was investigated by means of an automatic prominence tagger which analyzed annotated syllable nuclei in terms of pitch movement, duration, intensity, and spectral emphasis, a measure describing the energy increase in higher areas of the frequency spectrum (Tamburini/Wagner, 2007). Values for the last three parameters were normalized via z-scores across all investigated syllables in the utterance. Weights for force accents (duration, spectral emphasis) and pitch accents (pitch movement, intensity) as well as alignment parameters for pitch movements were calculated so as to model perceptual ratings of German prominence. For the present study, only the syllable immediately preceding and following the target items were taken as a context for the tagger. To determine syllable nuclei, vowel boundaries for the target items and their immediate context were annotated in Praat. If the vowel of a context syllable tended to be elided, the vowel of the preceding/following syllable was used as context instead. Output values for force and pitch accent were collected separately and rounded up to the third decimal place. Overall prominence ratings were calculated by combining these two values using the weights determined for German by Tamburini and Wagner (2007). Sometimes, the tagger did not produce an output due to internal segmentation errors. In these cases force and pitch accent values were extracted from the log output of the program. Spectral similarity between pairs of segmentally identical syllables produced by the same speaker was investigated with a method developed by Wade and Möbius (2007) and Lewandowski (2011). Using a sampling rate of 500 Hertz, amplitude envelopes for four contiguous frequency bands were computed for the items to be compared. Frequency bands were equally spaced on a logarithmic scale ranging from 80 to 7800 Hertz. Spectral similarity was calculated by cross-correlating envelope pairs for each frequency band separately, adding the resulting vectors together, and determining the maximum value. Values were rounded up to the fifth decimal place.

5.3 Results

Of the 1680 sentences recorded (4 prefixes × 2 verbs × 7 sentences × 30 participants), 113 had to be discarded, either because they contained hesitations or self-corrections, or because the target items themselves or their immediate context contained speech errors, noise or unexpected reduction phenomena. As two of the four prefixes used were bisyllabic, investigations were based on a total of 2278 syllables. The results were analyzed in terms of sentence category ("w+s+","w+s-","w-s+","w-s-","sb","mb","wb") and syllable identity ([ʔʊm], [ʔʊn], [tɐ], [ʔyː], [bɐ], [dʊʁç]). As ANOVA residuals were not normally distributed, Wilcoxon rank sum tests were used to compare individual sentence categories as well as combinations of sentence category and syllable identity. To confirm the duration and prominence findings, pairwise comparisons using Wilcoxon signed rank tests were performed for items produced by the same speaker in the same word context. Significance values were adjusted for multiple comparisons via the Bonferroni correction. Data visualization and statistical analysis were performed with the statistics program R (R Core Team, 2014).

5.3.1 Duration

Figure 13 gives an overview of vowel duration results for the seven sentence categories examined. When placed in a sentence-final position ("sb") prefixes and their vowels were considerably longer than in the other conditions. A small influence of word stress was also visible, with syllables and vowels being slightly longer in lexically stressed than in lexically unstressed prefixes. Sentence stress had little influence on word or syllable duration, even though lexically stressed prefixes tended to have minimally lower duration values when the verbs were not in the focus of the sentence ("w+s+" vs. "w+s-"). There was also a slight tendency for syllables in bound prefixes to be shorter than syllables in segmentally identical prepositions or conjunctions ("mb" vs. "wb").

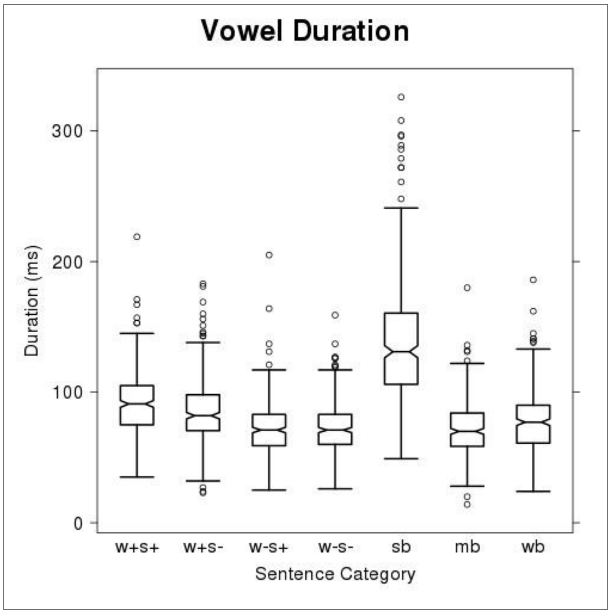


Figure 13: Vowel duration values across syllables in the seven sentence categories

A Wilcoxon rank sum analysis of sentence categories (corrected for 21 comparisons) showed significant effects of word stress ("w+s+" vs. "w-s+", "w+s-" vs. "w-s-") as well as of sentence boundary ("sb" vs. others), both in terms of syllable and vowel duration (W>62 000, p<.0001). While an effect of sentence stress for lexically stressed prefixes was found only in a comparison of vowel duration values ("w+s+" vs. "w+s-", W=59 074, p<.05), syllable and vowel duration differences were significant in comparisons between unstressed prefixes and free function words in similar contexts ("wb" vs. "mb", W>59 000, p<.05). In direct comparisons paired for speaker and word identity, Wilcoxon signed rank tests (corrected for 21 comparisons) confirmed influences of word stress ("w+s+" vs. "w-s+","w+s-" vs. "w-s-", V>33 000, p<.0001), sentence stress for lexically stressed prefixes ("w+s+" vs. "w+s ", V>26 000, p<.0001), sentence boundary ("sb" vs. others, V>41 000, p<.0001), and word boundary ("wb" vs. "mb", V>25 000, p<.001) on syllable and vowel duration. Syllable and vowel duration results were also analyzed with Wilcoxon tests for combinations of syllable identity and sentence category (corrected for 861 comparisons, see Table 4 for mean values).

Category	[ʔʊm]	[ʔʊn]	[tɐ]	[ʔy]	[bɐ]	[dʊʁç]	All
"w+s+"	147.6 72.2	142.4 88.6	132.1 88.0	90.0 90.0	138.0 87.5	199.1 114.4	140.0 91.7
"w+s-"	151.4 77.1	130.7 79.4	125.6 79.2	84.8 84.8	136.6 85.5	194.7 110.2	137.0 85.6
"w-s+"	143.4 71.6	112.6 66.1	115.5 74.7	59.9 59.9	126.3 75.7	149.0 81.2	117.0 71.3
"w-s-"	140.5 66.5	112.5 65.5	114.5 73.0	66.9 66.9	127.0 77.2	161.9 88.7	120.7 72.9
"sb"	261.0 128.6	185.2 106.3	205.5 137.8	132.3 132.3	207.1 147.1	414.2 188.0	228.7 140.1
"mb"	127.0 61.7	116.3 71.1	119.5 76.4	59.6 59.6	121.9 74.1	146.9 83.5	114.5 71.2
"wb"	140.2 71.9	127.8 82.5	124.5 76.5	71.4 71.4	127.8 76.9	152.8 84.3	124.4 77.3

Table 4: Mean duration values in milliseconds for syllables (above) and vowels (below) in the seven sentence categories.

Results for an analysis of word-stress difference ("w+s+" vs. "w-s+", "w+s-" vs. "w-s-") are given in Table 5. For [ʔʊn], [ʔyː], and [dʊʁç] the differences in both vowel and syllable duration were highly significant. In accented words, even the lexically unstressed syllables [bɐ] and [tɐ] showed slight differences depending on whether the prefix they appeared in had lexical stress or not, although the effect was less stable. For [bɐ] there was a significant effect of vowel duration, but not of syllable duration. The syllable [tɐ], on the other hand, was only marginally significant in terms of vowel duration (p=.06), but did show an effect of syllable duration. In unaccented conditions ("w+s-" vs. "w-s-"), only [ʔʊn] and [ʔyː] proved to have significant differences in terms of syllable and vowel duration. The syllable [dʊʁç] showed a significant effect of vowel duration, but differences were only marginally significant in terms of syllable duration (p=.054). Despite this reduction of significance, no effect of sentence stress itself ("w+s+" vs. "w+s-", "w-s+" vs. "w-s-") was visible for any of the syllables.

Category	[ʔʊm]	[ʔʊn]	[tɐ]	[ʔy]	[bɐ]	[dʊʁç]
"w+s+" vs. "w-s+"	1868.0 (n.s.) 1756.5 (n.s)	2485.0**** 2423.5****	2214.5* 2177.5 (n.s.)	2687**** 2687****	2264.5 (n.s.) 2365.5**	2063.5**** 2199****
"w+s-" vs. "w-s-"	1743.0 (n.s.) 1870.5 (n.s.)	2353.5* 2309.5*	2150.0 (n.s.) 2116.5 (n.s.)	1921* 1921*	1670.5 (n.s.) 1674 (n.s.)	1925.5 (n.s.) 1945.4*

*Table 5: W values with significance levels concerning effects of word stress on syllable duration (above) and vowel duration (below), n.s.: p>=.05, *: p<.05, **: p<.01, ****: p<.0001*

Considering word and sentence boundary effects, all investigated syllables were significantly longer when appearing in separated utterance-final prefixes than when appearing in other contexts ("sb" vs. others, W>2400, p<.0001). Differences in vowel duration were significant for all syllables except [ʔʊn], where comparisons with lexically stressed prefixes in accented positions ("sb" vs. "w+s+") failed to reach significance, and comparisons with free prepositions or conjunctions ("sb" vs. "wb") were significant on a lower level (W=2386, p<.01) than the other comparisons (W>2200, p<.0001). None of the syllables showed significant differences for bound, unstressed prefixes as compared with free prepositions or conjunctions ("mb" vs. "wb").

5.3.2 Acoustic Prominence

Mean values for prominence estimates as well as for the two subcomponents force accent and pitch accent are given in Appendix F. An investigation of acoustic prominence ratings across syllables showed a tendency for syllables in lexically stressed prefixes to receive higher values than syllables in unstressed prefixes ("w+s+" vs. "w-s+", "w+s-" vs. "w-s-"). Lexically stressed prefixes in unaccented conditions were slightly less prominent than stressed prefixes in accented conditions ("w+s+" vs. "w+s-"). Separated, sentence-final prefixes were particularly high in prominence, while no differences appeared between bound, unstressed prefixes and free function words in similar contexts ("mb" vs. "wb"). Wilcoxon rank sum tests for different sentence categories across syllables (corrected for 21 comparisons) showed significant effects of word stress in accented as well as unaccented conditions ("w+s+" vs. "w-s+", W=72 651, p<.0001; "w+s-" vs. "w-s-, W=59 677, p<.01). Sentence stress effects were significant for lexically stressed prefixes ("w+s+" vs. "w+s-", W=63 503, p<.0001). Syllables in separated, sentence-final prefixes differed significantly from all other conditions ("sb" vs. others, W>76 000, p<.0001). There was no visible effect of word boundary ("mb" vs. "wb"). Wilcoxon signed rank tests paired for speaker and word identity (corrected for 21 comparisons) confirmed significant effects of sentence boundary ("sb" vs. others, V>36 000, p<.0001), word stress ("w+s+" vs. "w-s+", "w+s-" vs. "w-s-", V>26 000, p<.0001), as well as sentence stress for lexically stressed prefixes ("w+s+" vs. "w+s-", V=29 709, p<.0001).

In an examination of prominence levels for individual syllables (see Figure 14), [ʔʊm], [tɐ], and [ʔyː] showed a falling tendency from from stressed and accented ("w+s+") over stressed, but unaccented ("w+s-") to unstressed conditions ("w-s+" and "w-s-"). In the case of [bɐ] and [dʊʁç], however, prominence values in stressed but unaccented prefixes ("w+s-") tended to be lower than in one or both of the lexically unstressed conditions ("w-s+" and "w-s-"). The syllable [ʔʊn] received higher prominence values in lexically stressed conditions as compared with unstressed conditions ("w+s+" vs. "w-s+", "w+s-" vs. "w-s-"), but no clear difference due to sentence stress ("w+s+" vs. "w+s-").

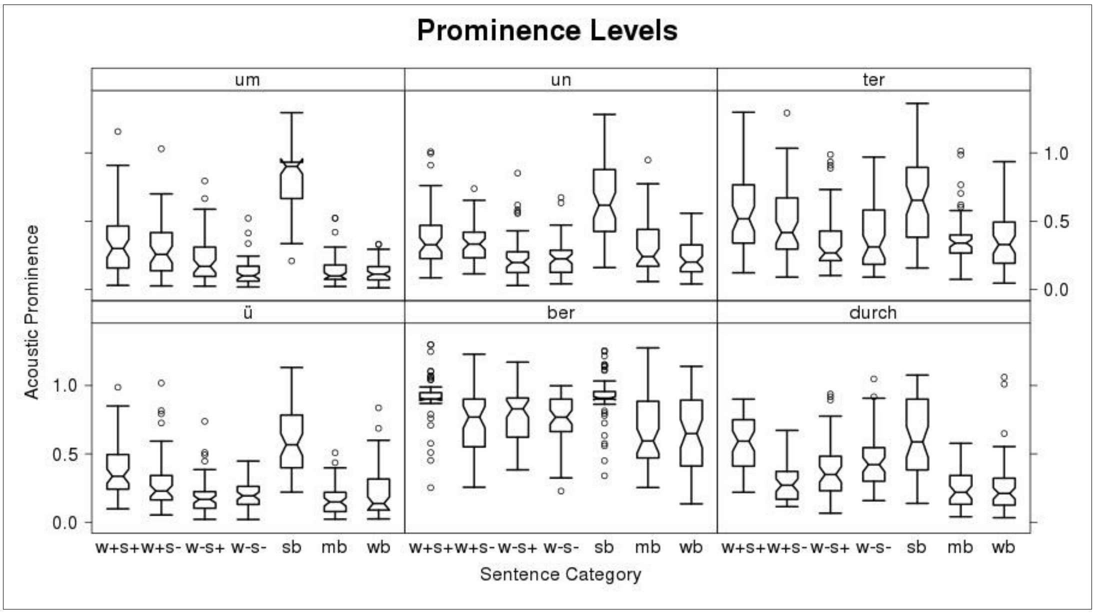


Figure 14: Acoustic prominence values for individual syllables in the seven sentence categories

In Wilcoxon rank sum tests for combinations of sentence categories and syllables (corrected for 821 comparisons) differences related to word and sentence stress mostly failed to reach significance (see Table 7). Word stress effects were found for [ʔyː] and [dʊʁç] in accented conditions ("w+s+" v. "w-s+") as well as for [ʔʊm] and [ʔʊn] in unaccented conditions ("w+s-" vs. "w-s-"). The syllables [ʔʊn] and [tɐ] showed marginally significant influences of word stress in accented conditions (p<.071, "w+s+" vs. "w-s+"). Only lexically stressed [dʊʁç] appeared to be significantly affected by changes in sentence stress ("w+s+" vs. "w+s-", W=2189, p<.0001). In separated, sentence-final prefixes, syllables often received significantly higher prominence values than in the other categories ("sb" vs. others, W>1900, p<.05). Exceptions for this last tendency were found for [tɐ] ("sb" vs. "w+s+", "sb" vs. "w+s-"), [bɐ] ("sb" vs. "w+s+", "sb" vs. "w-s+"), and [dʊʁç] ("sb" vs. "w+s+", "sb" vs. "w-s-"). No significant differences appeared between bound prefixes and corresponding prepositions or conjunctions ("mb" vs. "wb") or between unstressed prefixes in accented and unaccented conditions ("w-s+" vs. "w-s-").

Categories	[ʔʊm]	[ʔʊn]	[tɐ]	[ʔyː]	[bɐ]	[dʊʁç]
"w+s+" vs. "w-s+"	1922.5 (n.s.) 1954.5 (n.s.)	2171.5 (n.s.) 1763.5 (n.s.)	2174 (n.s.) 2212*	2669.5**** 2687.5****	2085 (n.s.) 2290*	1977** 2057***
"w+s-" vs. "w-s-"	2357.5**** 2377****	2307.5* 2209.5 (n.s.)	1985 (n.s.) 2037.5 (n.s.)	1664.5 (n.s.) 1653.5 (n.s.)	1240.5 (n.s.) 1283.5 (n.s.)	741 (n.s.) 700.5*
"w+s+" vs. "w+s-"	1514.5 (n.s.) 1489.5 (n.s.)	1570 (n.s.) 1128.5 (n.s.)	1764.5 (n.s.) 1765 (n.s.)	1925.5 (n.s.) 1854 (n.s.)	1981.5 (n.s.) 2123.5***	2189**** 2244****

*Table 6: W values with significance levels concerning effects on acoustic prominence (above) and force accent (below) for word stress in accented and unaccented conditions as well as sentence stress for lexically stressed prefixes, n.s.: p>=.05, *: p<.05, **: p<.01 ***: p<.001, ****: p<.0001*

When the force accent part of the prominence calculation (computed from vowel duration and spectral emphasis) was regarded separately, tendencies were very similar to the general prominence results. Across syllables, Wilcoxon rank sum tests (corrected for 21 comparisons) showed effects of word stress ("w+s+" vs. "w-s+", W=71 730.5 , p<.0001, "w-s+" vs. "w-s-", W=59 343.5, p<.05), sentence stress for lexically stressed prefixes ("w+s+" vs. "w+s ", W=61 651, p<.001), and sentence boundary ("sb" vs. others, W>76000, p<.0001). In tests for combinations of sentence category and syllable identity (corrected for 861 comparisons, see Table 7), word stress proved to have a significant effect for all syllables except [ʔʊm] and [ʔʊn] in accented conditions ("w+s+" vs. "w-s+"). In unaccented conditions ("w+s-" vs. "w-s-"), prominence values were significantly higher for lexically stressed than for unstressed prefixes only in the case of [ʔʊm], while [dʊʁç] actually showed a significant tendency in the opposite direction. Effects of sentence stress for lexically stressed prefixes were found for [bɐ] and [dʊʁç] ("w+s+" vs. "w+s-"). For syllables in sentence-final separated prefixes, the results were similar to those for overall prominence ratings ("sb" vs. others, W>2000, p<.01), but with an additional significant difference between conditions "sb" and "w-s+" for [bɐ]. Again, no significant differences appeared between lexically unstressed prefixes in accented and unaccented contexts ("w-s+" vs. "w-s-") or between unstressed prefixes and free function words in similar contexts ("mb" vs. "wb").

The pitch accent part of the prominence calculation (computed from overall intensity and changes in fundamental frequency) proved to have a highly uneven distribution, with most values being close to zero. Across syllables, distributions were less steep for the sentence categories "w+s+" and "sb" than for the other contexts. Unstressed prefixes in accented conditions ("w-s+") were slightly more variable in their pitch accent values than stressed or unstressed prefixes in unaccented conditions ("w+s-" and "w-s-"). Wilcoxon rank sum tests (corrected for 21 comparisons) showed no significant results apart from a few comparisons involving sentence-final prefixes ("sb" vs. "w-s-", "sb" vs. "mb", "sb" vs. "wb" , W>64 000, p<.0001). Comparisons between sentence-final prefixes and prefixes in lexically stressed but unaccented conditions were marginally significant ("sb" vs. "w+s-", W=59 140.5, p=.056). In Wilcoxon signed rank tests paired for speaker and word identity (corrected for 21 comparisons), differences between sentence final prefixes and some of the other conditions were confirmed ("sb" vs. "w-s-", "sb" vs. "mb", "sb" vs. "wb", V>20 000, p<.001). A significant difference was also found between stressed prefixes in accented conditions and unstressed prefixes in unaccented conditions ("w+s+" vs. "w-s-", V=16 825, p<.001). Differences between lexically stressed prefixes in accented and unaccented contexts were marginally significant ("w+s+" vs. "w+s-", V=14 067, p=.063).

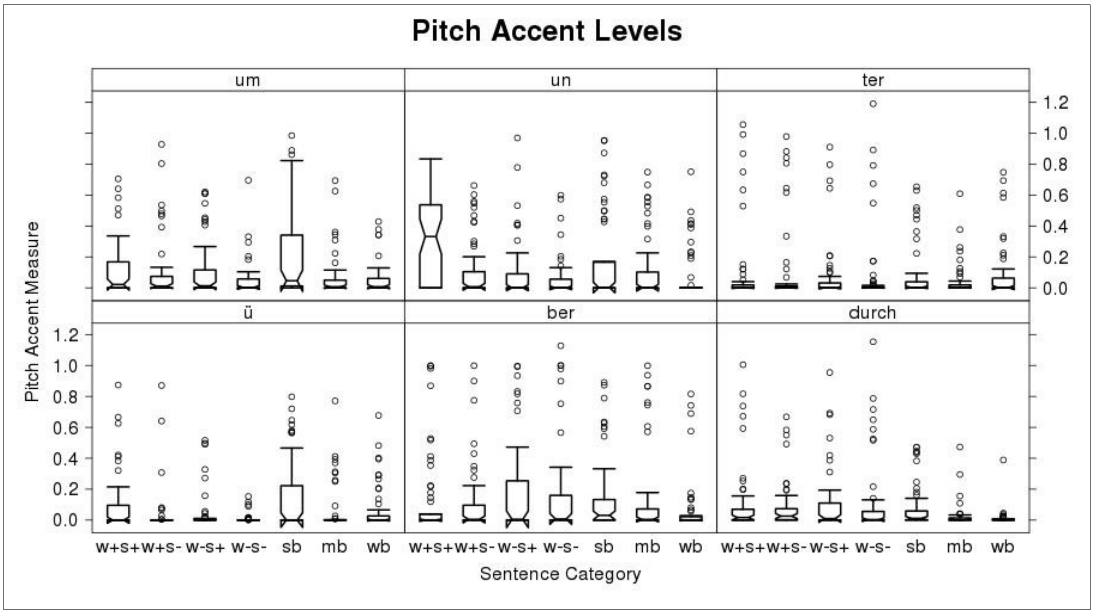


Figure 15: Pitch accent estimates for individual syllables in the seven sentence categories

While pitch accent values inclined towards zero for all syllables (see Figure 15), there were no consistent trends with respect to sentence condition. Accented and lexically stressed [ʔʊn] tended to have higher pitch-accent values than other combinations of syllables and sentence conditions, while for [tɐ] and [dʊʁç] values remained comparatively close to zero throughout conditions. Both [ʔʊm] and [?yː] showed the highest variation of pitch accent values for sentence-final prefixes ("sb"), followed by lexically stressed prefixes in accented condition ("w+s+"). The syllable [bɐ], on the other hand, actually tended to have higher pitch accent values in lexically unstressed than in stressed prefixes. Especially the syllables [ʔʊm], [bɐ], and [dʊʁç] had a higher variation of pitch accent values for lexically unstressed syllables in accented positions ("w-s+") than for stressed or unstressed syllables in unaccented positions ("w+s-", "w-s-"). Wilcoxon rank sum tests (corrected for 861 comparisons) mostly failed to show effects of word stress, sentence stress, or sentence boundary (with the exceptions of "w+s+" vs. "w-s-" for [ʔʊn] and "sb" vs. "wb" for [dʊʁç], W>2200, p<.01).

5.3.3 Spectral Similarity

For target syllables which were produced by the same speaker in the same sentence category, the level of similarity between their realization in both verb contexts was computed. In a comparison of these similarity estimates across syllables, the strongest similarity was found between syllables in lexically stressed and accented prefixes ("w+s+") as well as between syllables in separated, sentence-final prefixes ("sb"). Similarity values were lowest for syllables in accented verbs with lexically unstressed prefixes ("w-s+", "mb"). In Wilcoxon rank sum tests (corrected for 21 comparisons), differences between syllable similarity in stressed and unstressed prefixes were found for verbs in accented conditions ("w+s+" vs. "w-s+", "W= 16 183.5, p<.01, mean values: 0.889 vs. 0.848), while sentence stress differences in stressed prefixes only had a marginally significant effect ("w+s+" vs. "w+s-", W=14 189.5, p=.052, mean values: 0.889 vs. 0.864). Syllables in separated, sentence-final prefixes ("sb", mean value: 0.897) received significantly higher similarity values (p<.001) than syllables in all examined contexts except for stressed and accented prefixes ("w+s+"). Effects were most pronounced for the syllables [ʔyː], [dʊʁç], and to a lesser extent, [ʔʊn], although results failed to reach significance in Wilcoxon rank sum tests for combinations of syllable identity and sentence category (corrected for 861 comparisons).

In an analysis of spectral similarities between sentence categories for syllables produced by the same speaker in the same verb context, comparisons involving syllables in separated, sentence-final prefixes ("sb") tended to result in lower similarity values than comparisons between the other sentence categories. This effect was found to be significant (W>57 000, p<.0001) in Wilcoxon rank sum tests for combinations of sentence categories (corrected for 210 comparisons). Lexically stressed prefixes proved to be significantly closer to sentence-final prefixes than prefixes without lexical stress ("w+s+" and "sb" vs. "w-s+" and "sb, "w+s-" and "sb" vs. "w-s-" and "sb", W>64 000, p<.001, mean values: 0.794 vs. 0.748 and 0.799 vs. 0.764). As was the case for similarities within sentence categories, effects were most clearly visible for [ʔyː] and [dʊʁç]. Wilcoxon rank sum analyses for similarity between the sentence-final prefixes and the other categories ("sb" vs. others) combined with syllable identity (corrected for 630 comparisons) showed significant differences due to lexical stress for [ʔyː] in accented conditions ("w+s+" and "sb" vs. "w-s+" and "sb", $W = 2552$, $p<.0001$, mean values: 0.816 vs. 0.709).

5.4 Discussion

Apart from Dogil and Williams (1999), few studies have specifically investigated the interaction of word and sentence stress in German. The experiment described here examines the extent to which canonical word stress differences and additional semantic contrasts triggered differences in the word and sentence stress patterns which in turn were visible in the acoustic realization of the target syllables. Based on German language corpus studies as well as evidence from other Germanic languages, lexically stressed syllables were expected to be longer than unstressed syllables in accented as well as unaccented conditions. It was also assumed that acoustic prominence levels relative to the immediate surroundings would be influenced by the changes in the word and sentence stress implied in the carrier sentences. Although spectral parameters have been shown to be affected by stress, there were no clear expectations as to how word and sentence stress might influence similarity across and within sentence categories. While a clear effect of final lengthening on separated, sentence-final prefixes was expected, predictions concerning duration differences between unstressed prefixes and similarly placed free function words were less straightforward. On the one hand, a longer duration of free words might be expected due to effects of word-final lengthening (e.g. Beckman/Edwards, 1990) or polysyllabic shortening (e.g. Turk/Shattuck-Hufnagel, 2000; White, 2002), since bound prefixes were not followed by a word boundary, and therefore appeared in words that were longer than the corresponding prepositions or conjunctions. Also, bisyllabic items had lexical stress on the first syllable as free words, but not as bound prefixes. On the other hand, there might have been counteracting influences of word frequency and accentual lengthening, as the verbs used were generally less frequent than the matching function words and tended to attract sentence focus.

The present study did, indeed, show a significant influence of lexical stress on syllable and vowel duration. While there was a tendency for lexically stressed prefixes to be higher in prominence than unstressed syllables, a separate analysis of the force and pitch parts of the calculation revealed that this effect was due mainly to differences in relative vowel duration and spectral emphasis rather than to changes in intensity or the amount of pitch movement. In terms of spectral similarity, stressed syllables were more similar to syllables in sentence-final prefixes than unstressed syllables. When no deaccentuation cues were given, lexically stressed syllables were also more similar across verb contexts than unstressed syllables. In a syllable-by-syllable analysis, duration effects on word stress remained significant for all prefixes except [ʔʊm]. Effects on prominence estimates were less stable and tended to occur either for accented or for unaccented conditions, while differences in spectral similarity were mostly not significant in an investigation of individual syllables. One reason for the small size of the word stress effects might have been that all investigated syllables except [ʔyː] contained lax vowels, which have been found to show a considerably reduced effect of lexical stress on duration in comparison to tense vowels (Mooshammer et al., 1999; Kleber/Klipphahn, 2006).

Across syllables, a direct influence of sentence stress was visible for lexically stressed prefixes in terms of prominence, force accent, and vowel duration. In a comparison paired for speaker and word identity, syllable duration differences proved to be significant as well. While pitch accent values of lexically stressed syllables were less biased towards zero in accented conditions than in unaccented conditions, the effect was marginally significant only in a paired comparison. Pitch accent values for lexically unstressed syllables in accented positions appeared to be slightly higher than values for stressed or unstressed syllables in unaccented positions. Although this might be an indication of overall larger pitch movements for verbs in accented contexts, the tendency did not prove to be significant. Sentence stress may have had an indirect influence on word stress effects, as tests for stressed versus unstressed prefixes in accented conditions often resulted in lower p-values than similar tests for prefixes in unaccented conditions. When sentences were given a broad focus, even the lexically unstressed second syllables of the prefixes [ʔʊn.tɐ] and [ʔyː.bɐ] were affected by the stress level of the prefix in terms of duration and force accent levels. This result may be explained by accentual lengthening of the word carrying sentence stress, as there is evidence that in English and Dutch this effect is stronger to the right of the lexically stressed syllable than to the left (Cambier-Langeveld/Turk, 1999). Although the data was not analyzed perceptually, auditory impressions suggest that participants often placed a secondary accent on the target verb in unaccented conditions – perhaps because they wanted to better clarify the intended word meaning or because the given cues were not strong enough. The unusually strong effect of sentence stress on prominence levels for [dʊʁç] may have been due to the fact that ['dʊʁç.ʃaʊ.ən] was one of the few verbs where the potentially contrasting sentence stress in the unaccented condition would actually fall on the syllable used as preceding context by the prominence tagger.

As was to be expected, a large effect of sentence boundary on syllable and vowel duration was observed. All examined syllables, including the first syllables of the prefixes ['ʔʊn.tɐ], and ['ʔyː.bɐ], were considerably lengthened when appearing in sentence-final, separated prefixes. The results confirm findings by Kohler (1983) and Silverman (1990), who concluded that sentence-final lengthening extends beyond the final syllable. Effects of sentence boundary were also found for prominence, force accent, pitch accent, and spectral similarity, although not all syllables were affected equally. Even though the target words in the first four sentence categories were mostly in clause-final positions and therefore subject to final lengthening, effects of word and sentence stress were not neutralized. Particularly in the case of ['dʊʁç.ʃaʊ.ən] ("to look through") and ['ʊm.faː.ʁən] ("to run over"), however, differences in sentence position might have canceled out possible sentence stress effects, as these verbs were sentence-final in the unaccented, but not in the accented conditions. Syllables in bound prefixes tended to be slightly shorter than when they occurred in segmentally identical prepositions or conjunctions, with the first syllable of the bisyllabic ['ʔʊn.tɐ] and ['ʔyː.bɐ] being affected more strongly than the second syllable. This suggests that possible effects of polysyllabic shortening and word-final lengthening were stronger than opposing effects of word frequency and accentual lengthening. No influence was found for prominence and similarity values, and the word boundary effect was not significant in a separate investigation of the individual target syllables.

In view of the study on syllable frequency effects (see Chapter 7), this experiment showed that it was necessary to control for effects of word stress and information structure of the carrier sentence as well as the position of the target syllable within the word and sentences. Moreover, as acoustic prominence measures seemed to be easily affected by the context syllables comparison, it was also found to be important to keep the surroundings of the target items syllables constant in order to gain conclusive measurements.

6 Experiment 2: Effects of Lexical Class and Lemma Frequency

Speakers modify their pronunciation depending on the ease with which the word in question can be deduced from its context. Frequent or predictable words are enunciated less clearly than unusual words, and function words tend to have lower prominence values than content words (Bell et al., 2009). The experiment described here analyzes the pronunciation of German words which can be used as segmentally identical demonstrative pronouns, relative pronouns, and definite articles. As the three lexical classes also differ in their frequency of occurrence, effects of lemma frequency as well as grammatical function may play a role.

6.1 Background

When speaking, people constantly have to navigate between their own wish to minimize muscular effort and their need to remain intelligible to their communication partner (Lindblom, 1990). Content words, which are vital for the understanding of the sentence, tend to be emphasized more strongly than function words. The distinction is not a binary one, however, as prominence differences can be found between various classes of content or function words (Sereno/Jongman, 1995 for English, Chamonikolasová, 2000 for English and Czech). In German speech synthesis, lexical class has been used as an important indicator for predicting prominence levels (Widera et al., 1997; Windmann et al., 2011). Another strategy for economizing speech effort is to enunciate statistically improbable words and word combinations more carefully than common collocations. English language studies of read speech as well as spontaneous dialogues have shown that, independent of segmental length, frequent words tend to have smaller duration values than rare words (Aylett/Turk, 2004; Baker/Bradlow, 2009; Bell et al., 2009). However, a large part of the effect of word frequency can also be explained by other factors such as accentuation and prosodic boundaries (Aylett/Turk, 2004) or conditional probability given the immediate context (Bell et al., 2009). Content words which were already mentioned in the conversation also tend to be reduced in comparison with new occurrences (Baker/Bradlow, 2009; Bell et al., 2009). The exemplar theoretic approach integrates various influences on pronunciation into one model according to which numerous instances of words and word sequences are mentally stored in the form of concrete and highly detailed representations, the so-called exemplars (Pierrehumbert, 2001).

Frequency information is encoded by having more such exemplars stored for frequent than for rare words and word combinations. Recent exemplars are assumed to be more strongly activated than older exemplars. It is also possible that accented versus unaccented versions of a word and even different types of pitch accent are distinguished in this way. In German, words occurring predominantly with one particular type of pitch accent have been shown to be less varied in their pitch trajectories than words where the accent type is less predictable (Schweitzer et al., 2010).

As function words are generally more frequent than content words, it is not easy to disentangle effects of word frequency on the one hand and grammatical class on the other. However, studies indicate that both factors seem to play a role. In an investigation by Bell and colleagues (2009), English function and content words were found to differ in the extent to which they were affected by effects of frequency and predictability. Pluymakers and colleagues (2005) discovered word frequency effects irrespective of lexical class by examining Dutch affixes attached to different words belonging to the same grammatical category. Effects of lexical class and lemma frequency in English were investigated by Jurafsky and colleagues (2002), who compared words which were segmentally identical but served different grammatical functions. This approach allowed them to explore whether frequency distinctions are made between different uses of the same word, i.e. whether there is not only an effect of word frequency but also of lemma frequency. Three of the four investigated word forms varied significantly in their degree of reduction depending on the grammatical role they played. Pure lemma frequency, however, did not prove to be a good predictor of these differences. A large part of the variance could equally well be explained by differences in predictability based on the surrounding word forms. Also, differences in lemma frequency did not necessarily correlate with how strongly words were reduced.

To examine whether lexical class and lemma frequency can influence the pronunciation and the degree of emphasis placed on German function words, the present study used a setup similar to the one used by Jurafsky and colleagues (2002). In a controlled production experiment, segmentally identical demonstrative pronouns, relative pronouns, and definite articles were analyzed in terms of syllable and vowel duration, acoustic prominence as determined by an automatic tagger, and spectral similarity between segmentally identical syllables produced by the same speaker.

6.2 Methods

Sentences from this experiment and the one described in Chapter 5 were combined into a single reading task, allowing sentences from both studies to function as mutual distractors.

6.2.1 Participants

Participants were the same as in the Experiment 1 (see Section 5.2.1).

6.2.2 Material

The German function words "der", "die", "das", "dem", and "den" most commonly appear as definite articles, but can also function as relative or demonstrative pronouns. Which of the five words is used depends on its grammatical case as well as the gender and number of the item to which it refers. While "der" is either the nominative case for masculine singular or the dative case for feminine singular (as a definite article, it is also the genitive plural as well as the feminine genitive singular form), "die" appears as the nominative and accusative case for plural as well as feminine singular nouns. The word "das" is the nominative and accusative case for neuter singular nouns, "dem" is the dative case for masculine or neuter singulars, and "den" is the accusative case for masculine singular nouns. There are considerable frequency differences between definite articles on the one hand and demonstrative and relative pronouns on the other. According to the DeWaC corpus (Baroni/Kilgarriff, 2006), a 1.5 billion word database of German internet articles which was automatically tagged for lexical classes, the words "der" ([deːɐ]), "die" ([diː]), "das" ([das]), "dem" ([deːm]), and "den" ([deːn]) were used as definite articles 89.8% of the time, while 7.2% of their appearances were classified as relative pronouns, and only 3% were demonstrative pronouns.

To examine whether these differences in frequency of occurrence have an influence on pronunciation, the realization of these function words in different grammatical roles was analyzed and compared (see Table 7). Sentences containing relative and demonstrative pronouns were formulated so as to match definite articles already appearing in one of the other carrier sentences from the present experiment or the one described in Chapter 5. As each of the lexical classes required different types of surrounding grammatical structure, only the half-syllables preceding and following the target word were held constant across each group of three sentences. Within each sentence group, items were matched in terms of number and, in the case of singular forms, gender. Three sentence groups each were assembled for masculine and feminine "der", singular and plural "die", "das", masculine and neuter "dem", and "den". A total of forty-eight new sentences containing relative and demonstrative pronouns were constructed for this experiment. The 24 definite articles matching these pronouns were taken from 20 of the sentences used in the present experiment or the one described in the preceding chapter. The seventy target items along with their carrier sentences are presented in Appendix B.

<table>
<tr><th>Orthographic Form</th><th>Phonetic Transcription</th><th>Gender / Number / Case</th><th>No. of Stimuli</th></tr>
<tr><td rowspan="2">der</td><td rowspan="2">[deːɐ]</td><td rowspan="2">masculine singular nominative or
feminine singular dative</td><td>3x3</td></tr>
<tr><td>3x3</td></tr>
<tr><td rowspan="2">die</td><td rowspan="2">[diː]</td><td rowspan="2">feminine singular nominative/accusative or
masculine/feminine/neuter plural nominative/accusative</td><td>3x3</td></tr>
<tr><td>3x3</td></tr>
<tr><td>das</td><td>[das]</td><td>neuter singular nominative/accusative</td><td>3x3</td></tr>
<tr><td rowspan="2">dem</td><td rowspan="2">[deːm]</td><td rowspan="2">masculine singular dative or
neuter singular dative</td><td>3x3</td></tr>
<tr><td>3x3</td></tr>
<tr><td>den</td><td>[deːn]</td><td>masculine singular accusative</td><td>3x3</td></tr>
</table>

Table 7: Orthographic form, phonetic transcription, grammatical description, and number of stimuli used for each target item

6.2.3 Procedure

The procedure was the same as in Experiment 1 (see Section 5.2.3).

6.3 Results

Of the 2160 items recorded (8 words × 3 lexical classes × 3 contexts × 30 participants), 310 had to be omitted from the analysis due to disfluencies in the target sentences or impairment of the target words themselves or their immediate context. Results are based on the remaining 1850 items, which were investigated with regard to the factors lexical class ("dp", "rp", "da") and word identity ("der masc.", "der fem.", "die sg.", "die pl.", "das", "dem masc.", "dem neut.", "den"). Prominence estimates and spectral similarity values did not lend themselves to ANOVA investigations, as residuals did not follow a normal distribution. They were analyzed with Wilcoxon rank sum tests instead. While word and syllable duration values did show a normal distribution of ANOVA residuals, variances were not equal across groups. For this reason, Wilcoxon Rank sum tests were used as a follow-up of ANOVA investigations in order to validate the findings. In addition, duration and prominence values were investigated for pairwise comparisons of words produced by the same speaker in the same context using Wilcoxon signed rank tests. Significance values of the Wilcoxon rank sum and signed rank analyses were Bonferroni-corrected for multiple comparisons. The statistics program R was used to analyze and visualize the data (R Core Team, 2014).

6.3.1 Duration

In terms of syllable and vowel duration, demonstrative pronouns tended to be slightly longer than segmentally identical definite articles, with relative pronouns usually falling somewhere in between. Significant word and vowel duration differences between all three word classes ("dp" vs. "rp", "dp" vs. "da", "rp" vs. "da", corrected for 3 comparisons) were shown in Wilcoxon rank sum tests ($W>214\,000$, $p<.0001$) and confirmed with Wilcoxon signed rank tests paired for speaker and context ("dp" vs. "rp", "dp" vs. "da", "rp" vs. "da", $V>83\,000$, $p<.0001$). With regard to individual words, the overall trend was especially noticeable for feminine "der" as well as masculine and neuter "dem". Differences for "den", masculine "der", and singular "die" were less pronounced, while hardly any changes were observed for "das" and plural "die" (see Table 8 for mean duration values).

Lexical Class	der (masc.)	der (fem.)	die (sg.)	die (pl.)	das	dem (masc.)	dem (neut.)	den	All
"dp"	150.0 96.9	193.4 139.4	124.7 79.4	126.7 71.6	186.7 72.7	223.5 84.2	185.5 85.6	200.8 82.7	175.0 89.8
"rp"	124.6 78.3	144.5 101.8	126.8 68.0	117.1 64.4	186.8 60.5	212.9 76.7	172.2 76.5	189.1 73.9	158.0 74.8
"da"	118.2 74.7	104.3 81.9	100.0 65.5	112.3 66.7	182.0 71.6	161.9 45.8	141.9 61.6	150.7 58.6	133.6 65.8

Table 8: Mean duration values of demonstrative pronouns (dp), relative pronouns (rp) and definite articles (da) in milliseconds for syllables (above) and vowels (below).

Two-way ANOVAs for log-transformed syllable and vowel duration both showed highly significant effects ($p<.0001$) for the two factors word identity (syllable duration: $F_{7,1824}=147.8$, vowel duration: $F_{7,1824}=61.1$) and lexical class (syllable duration: $F_{2,1824}=123.2$, vowel duration: $F_{2,1824}=109.3$) as well as for their interaction (syllable duration: $F_{14,1824}=8.6$, vowel duration: $F_{14,1824}=11.0$). Tukey's HSD tests were used to further investigate the data (see Table 9). In terms of both syllable and vowel duration, significant differences were found between masculine and neuter "dem" as well as between masculine and feminine "der", but not between singular and plural "die". Concerning the interaction between lexical class and word identity, all words except "das" and plural "die" had significantly longer syllable duration values when occurring as demonstrative pronouns than when appearing as definite articles. For vowels, differences between these two lexical classes reached significance for all words except "das" and singular and plural "die", with singular "die" being marginally significant (p=.061). A significant difference between relative and demonstrative pronouns in terms of syllable duration was only found for masculine and feminine "der". In terms of vowel duration only "das" and feminine "der" showed a significant effect, while for masculine "der" the effect was marginally significant (p=.075). Finally, relative pronouns and definite articles differed significantly in syllable duration for all items except "das", masculine "der", and plural "die". As far as vowel duration was concerned, differences were only significant for "den" and masculine and neuter "dem". In the case of "das" and feminine "der", vowel duration differences were marginally significant (p<.076).

Lexical Classes	der (masc.)	der (fem.)	die (sg.)	die (pl.)	das	dem (masc.)	dem (neut.)	den
"dp" vs. "rp"	p<.05 n.s.	p<.0001 p<.0001	n.s. n.s.	n.s. n.s.	n.s. p<.05	n.s. n.s.	n.s. n.s.	n.s. n.s.
"dp" vs. "da"	p<.01 p<.01	p<.0001 p<.0001	p<.001 n.s.	n.s. n.s.	n.s. n.s.	p<.0001 p<.0001	p<.0001 p<.0001	p<.0001 p<.0001
"rp" vs. "da"	n.s. n.s.	p<.0001 n.s.	p<.0001 n.s.	n.s. n.s.	n.s. n.s.	p<.0001 p<.0001	p<.01 p<.01	p<.01 p<.05

Table 9: Adjusted p-values (Tukey HSD) for comparisons of syllable duration (above) and vowel duration (below) between demonstrative pronouns (dp), relative pronouns (rp) and definite articles (da).

The Tukey HSD results were mostly confirmed by Wilcoxon rank sum tests for combinations of word identity and lexical class (corrected for 271 comparisons, see Table 10). Word duration differences lost their significance for comparisons between demonstrative pronouns and relative pronouns in the case of masculine "der" as well as between demonstrative/relative pronouns and definite articles in the case of singular "die". Vowel duration differences between "den" as a relative pronoun and as a definite article also failed to reach significance in Wilcoxon rank sum analyses. On the other hand, the tests showed an additional effect for feminine "der", with definite articles having shorter vowels than relative pronouns, as well as an effect in the opposite direction in the case of "das", with longer vowels for definite articles than for relative pronouns.

Lexical Classes	der (masc.)	der (fem.)	die (sg.)	die (pl.)	das	dem (masc.)	dem (neut.)	den
"dp" vs. "rp"	3547.5 (n.s.) 3651.0 (n.s.)	4800.0 **** 4877.0 ****	2602.0 (n.s) 3198.0 (n.s.)	3289.0 (n.s.) 3200.5 (n.s.)	3472.0 (n.s.) 4712.5 ***	2971.0 (n.s.) 2921.5 (n.s.)	3598.0 (n.s.) 3694.5 (n.s.)	2778.0 (n.s.) 2948.5 (n.s.)
"dp" vs. "da"	3740.5 * 3872.5 **	5982.5 **** 5761.0 ****	3835.5 (n.s.) 3655.0 (n.s.)	3605.5 (n.s.) 3013.5 (n.s.)	3235.5 (n.s.) 3233.5 (n.s.)	4749.0 **** 4967.0 ****	4592.5 **** 4434.5 ****	4399.0 **** 4252.5 ****2
"rp" vs. "da"	3475.0 (n.s.) 3675.5 (n.s.)	4477.0 **** 4013.5 **	4126.0 (n.s.) 3340.5 (n.s.)	3790.0 (n.s.) 3223.0 (n.s.)	3322.5 (n.s.) 1984.5 **	4192.5 **** 4602.0 ****	4770.5 **** 4506.5 **	3785.5 ** 3460.5 (n.s.)

*Table 10: W values with significance levels concerning effects of lexical class on syllable/word duration (above) and vowel duration (below) for individual function words, n.s.: p>=.05, *: p<.05, **: p<.01, ****: p<.0001*

6.3.2 Acoustic Prominence

Mean values for prominence estimates as well as for the two subcomponents force accent and pitch accent are given in Appendix F. In an overall comparison, prominence values were higher for demonstrative pronouns than for relative pronouns and definite articles. Relative pronouns were minimally less prominent than definite articles. While Wilcoxon rank sum tests (corrected for 3 comparisons) confirmed differences between demonstrative pronouns and the other two lexical classes ("dp" vs. "rp", "dp" vs. "da", W>254 000, p<.0001), only a marginally significant effect was found for differences between definite articles and relative pronouns ("da" vs. "rp", W=211 209, p=.098). In Wilcoxon signed rank tests paired for speaker, word, and context, all three comparisons showed significant results ("dp" vs. "rp", "dp" vs. "da", W>102 000, p<.0001, "da" vs. "rp", W=86 710.5, p<.01). Two different trends became visible in a word-by-word comparison (see Figure 16). For the words "das", masculine "der", and singular and plural "die", relative pronouns were lowest in prominence, with demonstrative pronouns having values similar to, or slightly higher than, those for definite articles. In the case of feminine "der", masculine "dem" and, to a lesser degree, "den", however, values for relative pronouns tended to be lower than those for demonstrative pronouns, but higher than those for definite articles. For neuter "dem" there were no clear tendencies either way. Wilcoxon rank sum tests for combinations of lexical class and word identity (corrected for 276 comparisons, see Table 11) showed significant differences between demonstrative and relative pronouns for all items except "den" and masculine and neuter "dem". The difference between demonstrative pronouns and definite articles was significant for masculine "dem" as well as masculine and feminine "der". In a comparison between relative pronouns and definite articles, all items except "den", neuter "dem", and feminine "der" showed significant differences.

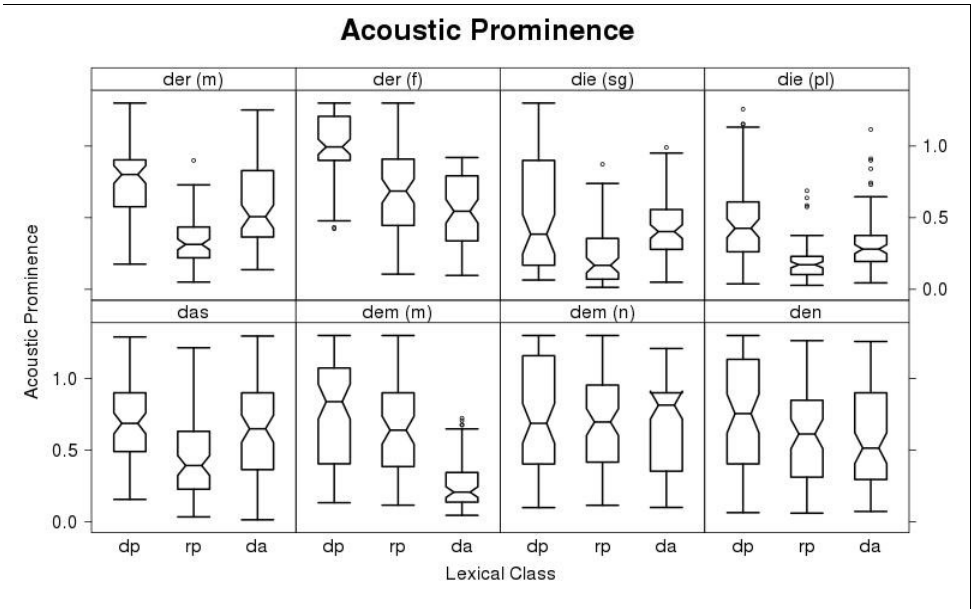


Figure 16: Acoustic prominence values for individual function words used as demonstrative pronouns (dp), relative pronouns (rp), and definite articles (da)

In a separate examination of the two factors of the prominence calculation, the force accent parameter (computed from vowel length and spectral emphasis) generally followed the trends discovered for overall prominence. Across words, differences between all three lexical classes (corrected for 3 comparisons) were significant in Wilcoxon rank sum tests ("dp" vs. "rp", "dp" vs. "da", $W>244\,000$, $p<.0001$, "da" vs. "rp", $W=215\,419$, $p<.05$) as well as in Wilcoxon signed rank tests paired for speaker, word, and context ("dp" vs "rp", "dp" vs. "da", "rp" vs. "da, $V>81\,000$, $p<.0001$). Investigations concerning individual function words in different lexical classes (corrected for 276 comparisons, see Table 11) also led to results that were very similar to those for the overall prominence estimates, but with an additional significant difference between plural "die" as a demonstrative pronoun and a definite article.

Lexical Classes	der (masc)	der (fem)	die (sing)	die (plur)	das	dem (masc)	dem (neut)	den
"dp" vs. "rp"	4922.5 **** 4968.5 **** 2984.0 (n.s.)	5067.0 **** 4745.5 **** 4490.0 **	3935.5 *** 3891.0 ** 3731.5 **	4934.5 **** 4940.0 **** 3379.5 (n.s.)	4840.5 **** 4972.5 **** 1921.5 ***	3089.0 (n.s.) 2918.0 (n.s.) 3121.0 (n.s.)	3286.5 (n.s.) 2948.5 (n.s.) 3315.0 (n.s.)	3061.5 (n.s.) 3055.5 (n.s.) 2911.0 (n.s.)
"dp" vs. "da"	3851.0 ** 3809.0 * 3219.0 (n.s.)	5689.5 **** 5354.0 **** 5187.5 ****	3099.5 (n.s.) 2950.0 (n.s.) 4135.5 ***	4022.5 (n.s.) 4126.5 * 3189.0 (n.s.)	3353.5 (n.s.) 3487.5 (n.s.) 2328.5 (n.s.)	5054.5 **** 5009.0 **** 4463.5 ****	3388.0 (n.s.) 2810.0 (n.s.) 4100.5 ***	3687.0 (n.s.) 3319.5 (n.s.) 4192.5 ****
"da" vs. "rp"	4937.5 **** 4960.0 **** 2952.5 (n.s.)	2135.5 (n.s.) 2274.5 (n.s.) 1638 ***	4739.0 **** 4728.0 **** 2836.5 (n.s.)	5144.4 **** 5044.5 **** 3749.0 (n.s.)	4566.0 ** 4504.5 * 2830.5 (n.s.)	793.0 **** 769.5 **** 1547.5 ***	2887.5 (n.s.) 3170.0 (n.s.) 1728.0 ****	2567.5 (n.s.) 2815.5 (n.s.) 1762.0 *

Table 11: W values with significance levels concerning effects of lexical class on estimates of acoustic prominence (above), stress accent (middle), and pitch accent (below) for individual function words, n.s.: p>=.05, *: p<.05, **: p<.01, ***: p<.001, ****: p<.0001

The pitch accent parameter (calculated from intensity and changes in fundamental frequency), on the other hand, proved to have a highly uneven distribution, with most values clustered around zero (see Figure 17). While results for "das", masculine "der", and plural "die" did not vary across grammatical categories, the function words "den", feminine "der", and masculine and neuter "dem" revealed a large variety of values for demonstrative pronouns and a slightly less skewed distribution for relative pronouns than for definite articles. Singular "die" showed results around zero for relative pronouns and definite articles, while values tended to be slightly higher for demonstrative pronouns. Wilcoxon rank sum tests for pitch accent estimates across words (corrected for 3 comparisons) showed higher values for demonstrative pronouns as compared with relative pronouns or definite articles, while relative pronouns had significantly higher values than definite articles ("dp" vs. "rp", "dp" vs. "da", "rp" vs. "da", W>200 000, p<.0001). These tendencies were confirmed with Wilcoxon signed rank tests paired for speaker and context ("dp" vs. "rp", "dp" vs. "da", "rp" vs. "da", V>55 000, p<.0001, corrected for 3 comparisons). In Wilcoxon rank sum tests for individual words (corrected for 276 comparisons, see Table 11), demonstrative pronouns proved to have higher pitch accent values than relative pronouns in the case of feminine "der" and singular "die", while "das" showed a significant tendency in the opposite direction. For "den", feminine "der", and masculine and neuter "dem", definite articles had significantly lower values than demonstrative or relative pronouns. The difference between definite articles and demonstrative pronouns was also significant for singular "die".

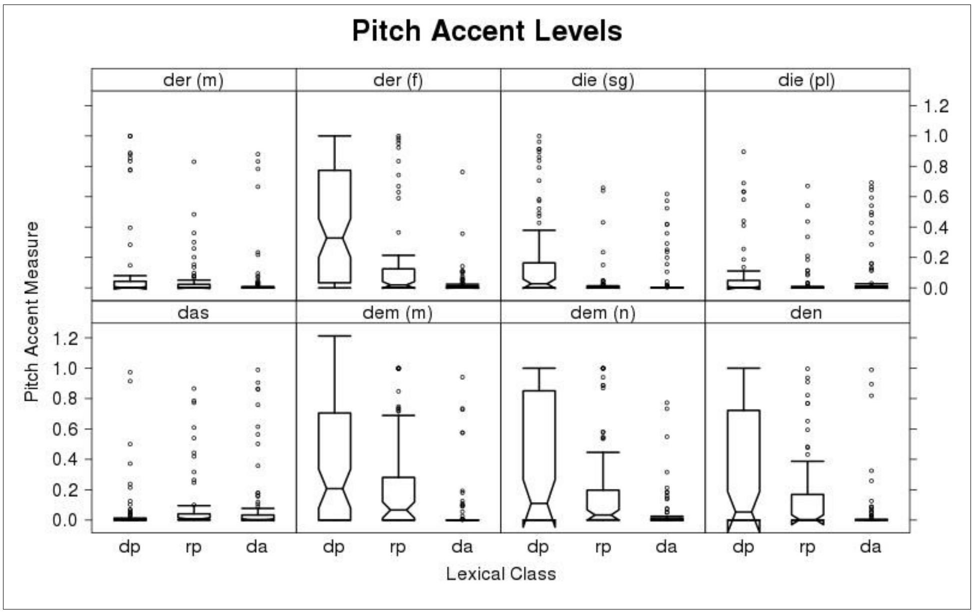


Figure 17: Pitch accent estimates for individual function words used as demonstrative pronouns (dp), relative pronouns (rp), and definite articles (da)

6.3.3 Spectral Similarity

Similarity levels were computed for segmentally identical items belonging to the same lexical class and produced by the same speaker in different contexts. Across words, definite articles (mean value: 0.814) appeared to be minimally less consistent in their pronunciation than demonstrative or relative pronouns (mean values: 0.823, 0.823). The difference, however, was only significant in Wilcoxon rank sum tests ($W>739\,000$, $p<.05$, corrected for 3 comparisons) when similarities were calculated without regard to gender or class. No effects were found when word identity as well as segmental identity was controlled (corrected for 3 comparisons), or when lexical classes were compared separately for individual word identities (corrected for 276 comparisons). Another investigation concerned similarity levels between words belonging to different lexical classes (paired for speaker, word identity, and context). Here, significant differences were found between similarity measures of relative and demonstrative pronouns on the one hand and relative pronouns and definite articles on the other (mean values: 0.861 vs. 0.850, $W=155\,459.5$, $p<0.05$, corrected for 3 comparisons). In separate comparisons for individual word identities (corrected for 276 comparisons), this tendency was only confirmed for masculine "dem" (mean values: 0.893 vs. 0.831, W=2651, p<.001).

6.4 Discussion

Relative and demonstrative pronouns were expected to have larger duration values than segmentally identical definite articles due to effects of frequency and predictability. Not only are they much less common then definite articles, their carrier sentences were specifically constructed to mirror the phonetic context of definite articles occurring elsewhere in the experiment, which probably increased their artificiality and reduced the predictability of the pronouns in these contexts. According to exemplar-theoretic approaches, definite articles might also be more strongly adapted to their surroundings, which would lead to lowered spectral similarity values across contexts. However, differences in pronunciation cannot always be explained by lemma frequency, and lexical classes may vary in the degree to which they can be emphasized. Jurafsky and colleagues (2002) found that although the English word "that" was most commonly used as a demonstrative pronoun, it tended to be longer in this function than when it was produced as a segmentally identical relative pronoun, complement, or determiner. They hypothesized that this might be explained by a higher tendency for demonstrative pronouns to receive pitch accents. In order to monitor for differences in emphasis, the target words' level of acoustic prominence in this investigation was measured in relation to their immediate context.

Significant differences in terms of syllable and vowel duration were discovered between all three lexical classes. Although these differences were not contradictory to lemma frequency effects, they did not mirror the fact that in German, frequency differences between the two types of pronouns are minimal compared with their difference from definite articles. The particularly high duration of demonstrative pronouns was probably due to their semantic role, as it is their function to point out and emphasize the entity to which they refer. Results for acoustic prominence confirmed that participants tended to emphasize demonstrative pronouns more strongly than relative pronouns or definite articles. Concerning differences between relative pronouns and definite articles, prominence results were inconclusive when calculated across words. No significant effect was found for overall prominence, and the two parameters of the prominence calculation showed tendencies in opposite directions, with relative pronouns receiving lower force accent values, but higher pitch accent values than definite articles. Although spectral similarity was slightly lower for definite articles than for the other lexical classes across items, and comparisons between lexical classes showed demonstrative and relative pronouns to be closer in pronunciation than relative pronouns and definite articles, effects were minimal and not consistent across individual words.

A closer examination of the data revealed that the individual target words varied in the ways and extent to which they were affected by changes in lexical class. For all examined words except "das", demonstrative pronouns were more strongly emphasized than definite articles with respect to one or more of the examined variables. While for this comparison, differences in overall prominence were generally accompanied by differences in word and vowel duration, comparisons including relative pronouns sometimes showed prominence differences with no corresponding differences in duration. Comparisons between demonstrative and relative pronouns in terms of prominence proved to be significant for all items except "dem" and masculine and neuter "den". However, only feminine "der" proved to be longer as a demonstrative pronoun than as a relative pronoun in terms of both syllable and vowel duration. And although relative pronouns tended to be longer than definite articles, prominence and force accent showed tendencies in the opposite direction for "das", masculine "der", and singular and plural "die". Duration results for this comparison were supported by pitch accent measures, as both variables proved to have higher values for relative pronouns than for definite articles in the case of "den", masculine and neuter "dem", and feminine "der". Only masculine "dem" showed significantly higher prominence estimates as well as duration values for relative pronouns than for definite articles. These conflicting findings may have resulted from the difficulty in controlling the context of the target items. As relative pronouns are generally used to introduce relative clauses, the syllables preceding them tended to be clause-final and therefore subject to final lengthening. It is very likely that relative pronouns received particularly low prominence ratings by the tagger due to their relatively prominent preceding context, especially as relative pronouns were actually more strongly emphasized than definite articles in terms of duration and the pitch accent part of the prominence calculation. In the case of feminine "der", masculine and neuter "dem", and one sentence used for "den", possible context lengthening was avoided by placing the relative pronouns in prepositional phrases. For these words, there was, indeed, no tendency for relative pronouns to be less prominent than definite articles, and prominence differences were supported by differences in syllable and vowel duration.

As the experiment demonstrated that both lexical class and lemma frequency may influence syllable duration, it was important to consider both factors when designing the study on syllable frequency effects (see Chapter 7). For this reason, the target syllables in this study were embedded in nonsense words used as fictive proper names. In this way, the lemma frequency of the carrier words was consistently set to zero and there were no possible confounding influences due to differences in lexical class. The experiment also reinforced the need to control for the surrounding context and the structure of the carrier sentences. To make certain that there were no interfering influences of these factors, the same four target sentences were used for both frequent and rare syllables.

7 Experiment 3: Effects of Syllable Frequency

According to several related theories of speech production (mental syllabary theory, dual route hypothesis, exemplar theory), frequent syllables can be retrieved as whole articulatory routines from the mental lexicon, while rare or unknown syllables have to be assembled from smaller units. The experiment described in this chapter analyzes possible effects of syllable frequency on error rates, duration, prominence levels, spectral similarity, and the performance of automatic alignment systems in a highly controlled production study using fictive surnames.

7.1 Background

The concept of mentally stored syllabic units has been corroborated through analyses of production latencies (Levelt/Wheeldon, 1994; Cholin et al., 2006; Brendel et al., 2008; Cholin et al., 2009) as well as through studies on the behavior of patients suffering from apraxia of speech (Aichert/Ziegler, 2004; Staiger/Ziegler, 2008). An extensive overview of various theories and literature on the role of syllables in speech production is presented in Chapter 2 of this thesis. The present section focuses on the possible implications that the mental syllabary theory and related approaches have for the pronunciation of frequent and rare syllables.

According to the classic mental syllabary theory, frequent syllables stored as whole articulatory routines would be expected to show a smoother interaction of individual articulatory gestures and consequently a higher degree of coarticulation than rare syllables. The double route hypothesis qualifies this expectation by stating that frequent syllables may also be reassembled from subsyllabic units if, for example, particularly clear speech is required or if the retrieval of larger units is in some way impaired (Whiteside/Varley, 1998b; Varley et al., 2000; Varley et al., 2006). Exemplar-theoretic approaches propose that units of various sizes (syllables, words, morphemes, phrases) are mentally stored in the form of highly detailed representations (Bybee, 2001; Pierrehumbert, 2001). The more frequent a unit is, the more such representations or exemplars would be available according to this theory, which means that frequent syllables would be expected to be more closely adapted to their surroundings than rare syllables, and therefore to show a greater variability. However, Pierrehumbert (2001) suggests that there may be an opposite tendency towards entrainment, which would cause articulatory movements of frequent items to become more similar through increased practice, and would thereby prevent them from becoming so variable in their pronunciation that they are no longer recognizable. As, by definition, frequent syllables appear more often than rare syllables, they would also be more susceptible to an accumulative general bias towards reduction.

There is, indeed, some evidence that syllable frequency has an effect on the acoustic realization of speech in terms of articulatory effort and degree of coarticulation. A reading experiment for English showed that while there were no significant effects on second formant measures, differences in the third formant values and the intensity of fricative noise indicated a higher amount of coarticulation for frequent syllables than for phonotactically legal but nonexistent syllables which were matched for vowel and either onset or coda identity (Croot/Rastle, 2004). In a repetition task using English monosyllabic words matched for onset identity, segmental complexity, and, as far as possible, for the voicing status of the coda, differences were discovered in terms of locus equations (Herrmann et al., 2008). When second formant measures at the beginning and midpoint of the vowel were plotted against each other, frequent syllables showed a steeper slope and a better fit to the regression line than rare syllables. No influences were found for absolute formant differences between vowel onsets and midpoints. For German, significant differences in normalized formant measures designed to locate vowels in perceptual space were revealed in a reading experiment in which bisyllabic nonsense words were matched for segmental complexity and embedded in carrier sentences (Benner et al., 2007). In corpus studies which compared frequent and rare syllables differing in lexical stress status, but otherwise segmentally identical and produced by the same speaker in the same word position, it was found that frequent syllables tended to have a more centralized pronunciation, quicker formant transitions, and smoother curves for voicing probability than rare syllables (Flechsig, 2006; Benner et al., 2007).

Results for duration are less conclusive, especially as it is difficult to control for interactions with segmental influences. In a Dutch word association task, Levelt and Wheeldon (1994) found frequent syllables to have shorter duration values as well as shorter production latencies than rare syllables. Segmental syllable length was not completely controlled for in this study, however, as rare syllables tended to have a slightly higher mean number of segments than frequent syllables. Hermann et al. (2008) confirmed a tendency for rare syllables to be longer than frequent syllables, but did not find any effect on the formant transition duration, calculated from vowel onsets to vowel midpoints. The study by Croot and Rastle (2004) showed a duration effect in the opposite direction, with segmentally identical onsets or rhymes being longer for frequent than for rare syllables. However, this effect may have been due to syllable pairs which were not matched with respect to their total number of segments, as the tendency disappeared when these pairs were removed. The corpus study by Flechsig (2006) also showed an overall tendency for frequent syllables to have longer duration values than rare syllables. When lexically stressed and unstressed syllables were examined separately, however, differences in the opposite direction became visible. The influence of syllable frequency on duration only proved to be marginally significant. A further German corpus study examining z-score duration values for syllables and individual segments demonstrated that z-scores of rare syllables were more dependent on the z-scores of their segments than were z-scores of frequent syllables (Schweitzer/Möbius, 2004). This result was explained by an exemplar-theoretic model, according to which frequent syllables are retrieved as whole entities, whereas rare syllables need to be assembled from smaller segments (Walsh et al., 2007). In this model, differences from mean segment values tended to cancel each other out for rare syllables, leading to more stable total syllable duration values, while frequent syllables had a stronger tendency to act as independent units.

The main goal of the present study was to examine effects of syllable frequency on duration while controlling for other influences such as the segments contained in the analyzed syllables, their lexical stress status, and the word class and lemma frequency of the words in which they appeared. To determine whether duration differences may have been caused by differences in the level of accentuation, acoustic prominence values were computed for the target syllables and their immediate context. Exemplar-theoretic predictions were investigated by comparing duration and prominence differences across sentence contexts for frequent and rare syllables, and by calculating the degree of spectral similarity between realizations of the same syllable. Finally, frequent and rare syllables were analyzed in terms of the number of speech errors participants made while producing them and the extent to which the automatic segmentation led to different results than the manually corrected annotations.

7.2 Methods

7.2.1 Participants

Thirty-two participants took part in the experiment (7 men, 25 women). All were native speakers of German, although three participants stated that they were raised bilingually. In ages they ranged from 20 to 49 (mean age: 25.4). Participants were compensated for their time.

7.2.2 Material

For the selection of syllable types, a paradigm was used which was originally developed by Cholin and colleagues (2004). Quadruples consisting of frequent and two rare CVC syllables were assembled where each of the frequent syllables formed a minimal pair with both of the rare syllables, differing once in the onset and once in the coda (see Table 12). Within each quadruple, the same opposition in onset and coda was accompanied by frequency differences in opposite directions. In this way it was possible to counteract effects of segmental duration as well as bigram frequency.

Syllable No.	Frequency	Structure	Onset difference	Coda difference	Example
1	frequent	$C_1V_1C_3$	to syllable 2	to syllable 3	[mʊs]
2	rare	$C_2V_1C_3$	to syllable 1	to syllable 4	[tʊs]
3	rare	$C_1V_1C_4$	to syllable 4	to syllable 1	[mʊŋ]
4	frequent	$C_2V_1C_4$	to syllable 3	to syllable 2	[tʊŋ]

Table 12: Structuring of syllable quadruples

Eight groups of four syllables were assembled accordingly, with no repetitions of syllables across quadruples (see Table 13). Frequent syllables appeared among the top 750 syllable types in each of the six investigated corpora, while rare syllables appeared less than five times per million syllable tokens in all databases. A further condition concerned the type frequency of the syllables, i.e. the relative amount of word types in which a syllable had to appear in each corpus. To avoid syllables which were predominantly used in one particular word, it was stipulated that for each of the syllables chosen for the experiment, the token frequency (relative to the total number of syllable tokens) divided by the type frequency (relative to the total number of word types) had to remain below three in all corpora. A second, lexically unstressed, syllable was added to each of the target syllables to create a bisyllabic nonsense word. As care had to be taken that none of the words existed in the German language, the second syllable was kept constant within quadruples, but not across them. The 32 words were embedded in two different groups of four carrier sentences, where they appeared as fictive surnames (see Appendix C). Half of the participants were given the first set of carrier sentences, while the other half received the second set of sentences.

Quadruple No.	1	2	3	4	5	6	7	8
1st syllable	[mʊs] [tʊs] [mʊŋ] [tʊŋ]	[tɪç] [hɪç] [tɪn] [hɪn]	[zɔl] [kɔl] [zɔm] [kɔm]	[naːx] [maːx] [naːl] [maːl]	[taɪl] [kaɪl] [taɪt] [kaɪt]	[fɪn] [lɪn] [fɪç] [lɪç]	[kan] [fan] [kal] [fal]	[zaːm] [taːm] [zaːt] [taːt]
2nd syllable	[lɐ]	[təl]	[bɐ]	[tɐs]	[nɐ]	[təl]	[dɐs]	[gɐ]
Carrier word	Mussler Tussler Mungler Tungler	Tichtel Hichtel Tinntel Hinntel	Sollber Kollber Sommber Kommber	Naachters Maachters Naalters Maalters	Teilner Keilner Teitner Keitner	Finntel Linntel Fichtel Lichtel	Kannders Fannders Kallders Fallders	Saamger Taamger Saatger Taatger

Table 13: Quadruple syllables and carrier words

The order in which the stimuli were presented to each participant was determined by using an 8×8 Latin square design. Frequent and rare syllables were chosen alternatively for each quadruple according to the numbers indicated by the Latin square. In this way, each row of the square served to create four individual orders (see Table 14 for details). This manner of randomization ensured that the 32 target words appeared once in each position throughout the experiment and that syllables belonging to the same quadruple were maximally far apart from each other.

Order	**1-8**	**9-16**	**17-24**	**25-32**	**33-64**	**65-96**	**97-128**
Quadruple	12345678 | 24136857 | 31427586 | 43218765 | 56781234 | 68572413 | 75863142 | 87654321						
Syllable	1313.. | 4242.. | 3131.. | 2424..	4242.. | 1313.. | 2424.. | 3131..	3131.. | 2424.. | 1313.. |4242..	2424.. | 3131.. | 4242.. | 1313..	see 1-32		
Context	12341234				2341 2341	3412 3412	4123 4123

Table 14: Latin square design employed for sentence randomization

For each group of four syllables belonging to the same quadruple and produced by the same participant, the four context sentences appeared in the experiment in a consistent order. Test sentences were interspaced with 128 distractor sentences. Three additional distractor sentences were presented at the beginning of the experiment, one of which contained a fictive surname which was structurally similar to the target words. All 131 distractor sentences are listed in Appendix D. Their order was not varied across participants.

7.2.3 Procedure

The experiment took place in a sound-treated chamber at Bielefeld University. Test and distractor sentences were presented on a computer screen to the participants, who proceeded through the experiment in a self-paced manner. The participants were instructed to look at each sentence and then read it out loud, using their normal reading style. They were also told to correct themselves if they noticed any slips of the tongue or other reading errors. Their speech was recorded with a Sennheiser TLM 103 microphone. Although electroglottographic recordings were made as well, they were not evaluated for this thesis.

Syllable and nucleus boundaries were annotated for the target syllables as well as the syllables immediately preceding them. Post-vocalic /r/ ([ɐ]) and [l] preceded by [ə] were both counted as part of the syllable nucleus. To enhance the level of consistency, the annotations were based on an analysis with the automatic segmentation and labelling program WebMAUS (Kisler et al., 2012). Audio files for each sentence and identically named text files containing orthographic transcriptions were uploaded to the program website as input for processing. From this information, the program generated a canonical phonetic transcription and automatically aligned it with the audio file. In a second step, the annotations were manually inspected with the program Praat (Boersma, 2001) to correct obvious errors in the segmentation of the target items or their immediately preceding and following syllables. The acoustic analysis was mainly based on the corrected annotations, although the original versions were investigated as well and used for comparison.

Analysis methods were similar to those described in Experiment 1 (see Section 5.2.3). Sentences were omitted from the acoustic analysis if they contained self-corrections by the participants or if the target syllables or their immediate context was impaired through errors, strong reduction, or noise. However, pronunciation errors in the target and context syllables were counted in order to examine whether there was an effect of syllable frequency on error rate. Based on the corrected WebMAUS annotations, duration values in seconds were computed for the target syllables as well as the immediately preceding and following syllables. For the target syllables, duration measures were subdivided into onset, vowel, and coda duration. The results were rounded up to the third decimal place. Acoustic prominence values were calculated up to the sixth decimal place with the help of an automatic tagger developed by Tamburini and Wagner (2007). Target syllables as well as the syllables immediately preceding and following them were analyzed in terms of overall prominence as well as the two subcomponents force accent (computed from z-score values of duration and spectral emphasis) and pitch accent (computed from z-score intensity values and pitch movement measures). Parameters for weighting these components and determining the pitch accent positions were determined on the basis of a corpus analysis of German perceptual prominence ratings (Tamburini/Wagner, 2007). In addition to the investigations of duration and acoustic prominence, pairwise comparisons of spectral similarity were performed by computing amplitude envelopes of four contiguous frequency bands which were equally spaced on a logarithmic scale ranging from 80 to 7800 Hertz, using a sampling rate of 500 Hertz (Wade/Möbius, 2007; Lewandowski, 2011). Envelope pairs for each frequency band of the two audio files to be compared were cross-correlated separately, the resulting vectors were added together, and the maximum value was determined. In this way, similarity measures were calculated for segmentally identical syllables produced by the same speaker in different contexts. Syllable similarity was calculated as well as onset, vowel, and coda similarity. Finally the results based on the manually corrected segmentations were compared with the results based on the original, uncorrected MAUS segmentations.

7.3 Results

In total, 4096 sentences were recorded (4 syllables × 8 quadruples × 4 sentences × 32 participants). Of these, 225 were discarded, leaving 1935 sentences with frequent target syllables and 1936 sentences with rare target syllables. The prominence tagger was not able to compute values for one further sentence containing a rare syllable. As ANOVA residuals were not normally distributed, statistical analyses were performed with Wilcoxon rank sum tests. Wilcoxon signed rank tests were used to investigate direct comparisons between frequent and rare syllables paired for speaker, quadruple, and context. Bonferroni corrections were applied when investigating multiple comparisons such as combinations of frequency level and quadruple identity. The statistics program R was used for data analysis and visualization (R Core Team, 2014).

7.3.1 Speech Errors

There were 99 cases where the target syllables themselves were impaired through errors. Slightly more low-frequency than high-frequency syllables were affected (56 vs. 43). An additional 25 errors concerned the syllables immediately preceding or following the target syllables. Of these, 13 appeared in the context of frequent and 12 in the context of rare syllables. Variations in vowel duration and tenseness were not counted as errors, as it was often difficult to decide whether the speaker intended to produce a long or a short vowel. Especially quadruples 4, 7, and 8 were ambiguous in this respect, as there are hardly any differences in vowel quality between short and long /a/ in German. A certain amount of variability was also allowed in the vowel of the second syllable of the target words. Some participants tended to pronounce the second syllable vowel [ɐ] as [eːɐ] due to the local regiolect, and the [ə] in the syllable [təl] was often merged with the following [l] or elided completely.

7.3.2 Duration

Mean values for syllable and vowel duration of frequent and rare syllables in each quadruple are given in Table 15. For all quadruples except 5 and 6 there was a slight tendency for frequent syllables to be shorter than rare syllables. A Wilcoxon rank sum test showed a significant effect of syllable frequency across quadruples ($W=1\,760\,773$, $p<.01$). In an investigation of combinations of quadruples and frequency level, results only proved to be significant for quadruple 2 ($W=22\,399.5$, $p<.05$, corrected for 120 comparisons). Vowel duration differences between frequent and rare syllables were less consistent. Hardly any differences in mean values appeared for quadruples 4 and 5, and for quadruple 3 there was actually a tendency in the opposite direction, with rare syllables tending to have shorter vowels than frequent syllables. Even so, the effect remained significant across quadruples ($W=1\,799\,040$, $p<.05$). No duration effects were visible for syllable onsets and codas or for either of the two context syllables immediately preceding and following the target items. Investigations of combinations of quadruple identity and frequency level showed a significant frequency effect on vowel duration of quadruple 6 ($W=18\,601$, $p<.0001$, corrected for 120 comparisons).

Quadruple	1	2	3	4	5	6	7	8	All
Frequent syllables	245.8 66.0	215.7 61.2	254.5 76.0	270.8 123.3	294.0 138.8	233.5 60.3	264.3 72.0	299.3 138.8	259.5 91.8
Rare syllables	249.4 69.5	227.0 64.6	260.1 71.5	275.9 124.6	294.4 138.5	232.5 68.8	271.8 75.7	306.2 142.3	265.0 94.8

Table 15: Mean duration in milliseconds for syllables (above) and vowels (below) by quadruple and frequency category

Although differences between frequent and rare syllables were very small, they remained visible even in a direct comparison of items. For each error-free group of four syllables produced by the same speaker in the same context, duration values for the two frequent syllables were subtracted from duration values for the two rare syllables.

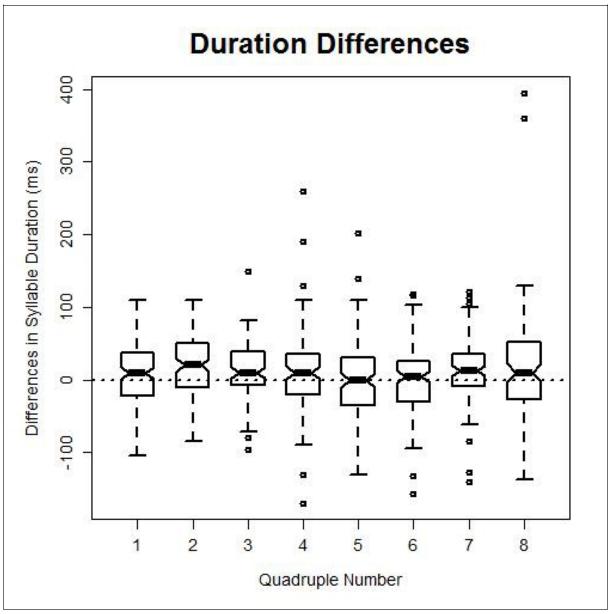


Figure 18: Duration difference between frequent and rare syllables in milliseconds, grouped by participant, quadruple, and context

In Figure 18, results from a comparison of syllable duration values are presented as a boxplot diagram. Wilcoxon signed rank tests showed significant deviations from zero on the level of the syllable (V=210 810, p<.0001) as well as for vowels (V=205 467.5, p<.0001) and codas (V=170 626, p<.05), but not for onsets. In such direct comparisons, the lengthening effect of rare syllables seemed to extend to the duration of the following context syllable (V=195 667.5, p<.001), while the preceding context remained unaffected.

Both frequent and rare syllables tended to be longer in the first appearance of the target words than in subsequent renditions. An interaction plot between frequency level and order of occurrence (Figure 19) shows that although duration differences decreased as participants became more familiar with the target items, mean values for rare syllables remained above those for frequent syllables. In Wilcoxon rank sum tests for combinations of frequency level and order of appearance, frequency effects for the same order position remained below significance level, as did order effects for the same frequency level (corrected for 28 comparisons). Only the duration difference between the first and last appearance of rare syllables proved to be marginally significant (W=125 886.5, p=.067, mean values: 273.1 ms vs. 260.6 ms).

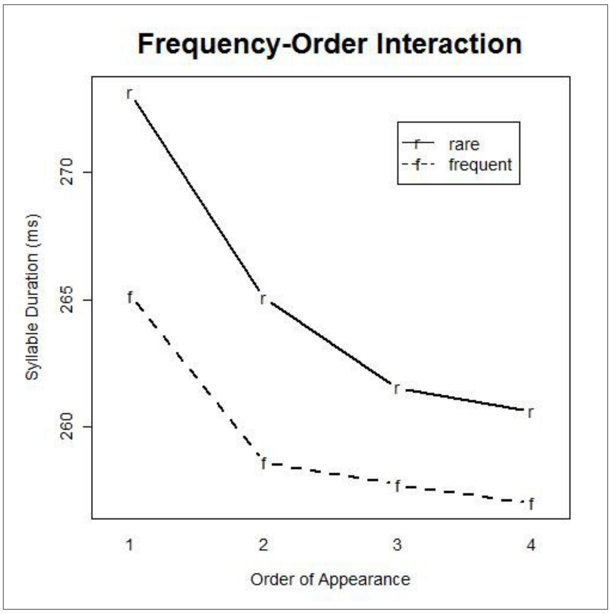


Figure 19: Mean syllable duration by order of appearance for frequent and rare syllables in milliseconds

To investigate duration variability, absolute duration differences were calculated between segmentally identical syllables spoken by the same participant in different contexts. Mean values for syllable and vowel variability are given in Appendix G. In this examination, vowel duration appeared to be more variable across contexts for rare than for frequent syllables. Wilcoxon rank sum tests confirmed this effect (W=3 576 768, p<.0001). While no significant tendencies were visible for syllable or onset duration, codas of rare syllables were marginally more variable than codas of frequent syllables (W=3 716 340, p<.097, mean values: 186.8 ms vs. 172.1 ms). An analysis of the context syllables themselves revealed that they tended to show higher absolute duration differences from one context to the next when surrounding a rare syllable than when appearing together with a frequent syllable (W>3 950 000, p<.05, mean values preceding syllable: 53.5 ms vs. 50.9 ms, mean values following syllable: 48.3 ms vs. 44.7 ms). To confirm the findings, distance values for frequent syllables were subtracted from those for rare syllables for each context comparison. Figure 20 presents differences in vowel variability for individual quadruples. Analyses of differences in variability using Wilcoxon signed rank tests were significant in terms of vowel duration (V=329 791.5, p<.0001), coda duration (V=305 409.5, p<.01), and the duration of the following context syllable (V=325 542, p<.0001). In this analysis, the effect on the preceding context syllable was marginally significant (V=305 453.5, p=.06). When examining vowel variability for combinations of quadruple identity and frequency levels (corrected for 120 comparisons), marginally significant frequency effects appeared for quadruples 4 (W=51 503.5, p=0.079) and 6 (W=49 293, p=0.068). Combinations of quadruple identity and frequency level showed no significant differences in terms of onset, coda, or syllable variability, or the variability of context syllables (corrected for 120 comparisons).

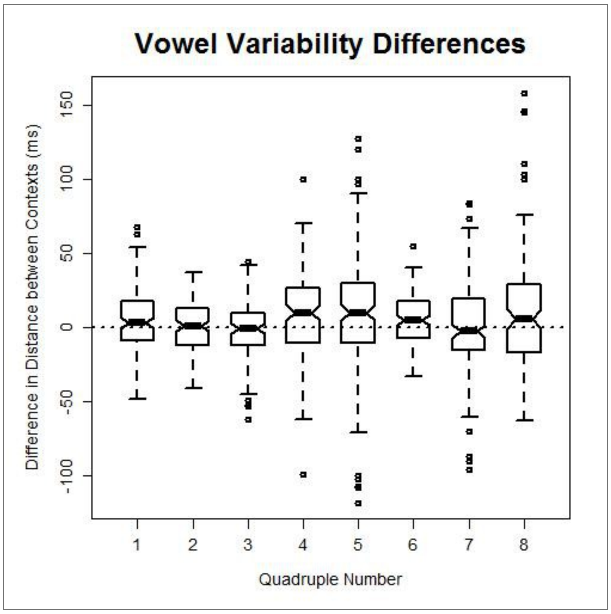


Figure 20: Vowel variability difference between frequent and rare syllables in milliseconds, grouped by participant, quadruple and context comparison

7.3.3 Prominence

Hardly any effects of syllable frequency were found for acoustic prominence or the two subcomponents force accent (calculated from duration and spectral emphasis) and pitch accent (calculated from intensity and pitch changes). Mean values for these measures are given in Appendix F. Wilcoxon rank sum tests showed no significant differences between frequent and rare syllables or their immediate context in terms of prominence, force accent, or pitch accent values. When results were grouped according to speaker, quadruple, and carrier sentence, subtracting values for frequent syllables from those for rare syllables, the preceding context syllables tended to receive a higher force accent value when followed by a frequent syllable than when followed by a rare syllable. This grouped comparison proved to be marginally significant in a Wilcoxon signed rank test (V=156 678, p=.068). Similar tests for prominence, stress accent, and pitch accent of the target syllables and the following context syllables revealed no significant effects.

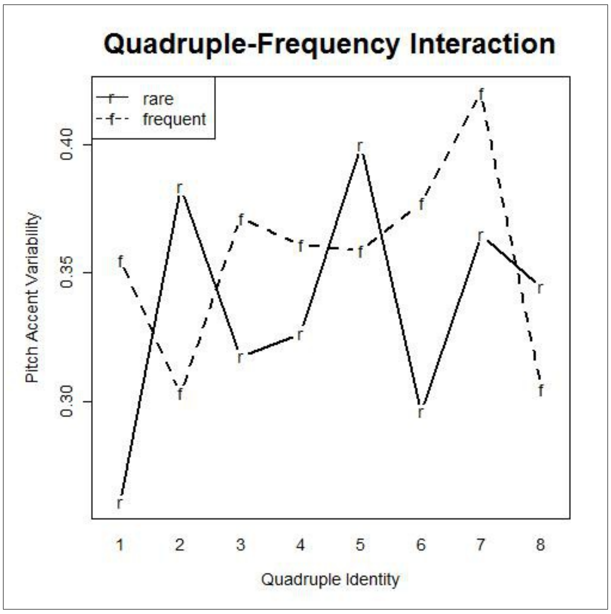


Figure 21: Mean absolute pitch accent distance between segmentally identical syllables produced by the same speaker for frequent and rare syllables in different quadruples

To examine the variability of prominence, force accent, and pitch accent values across contexts, absolute differences were calculated between segmentally identical syllables produced by the same speaker. Appendix G lists mean values for the variability of prominence, force accent, and pitch accent values across contexts grouped by quadruple and frequency level. In this investigation, pitch accent variability tended to be higher for frequent than for rare syllables. Although a difference in the opposite direction appeared for quadruples 2, 5, and 8 (see Figure 21), the overall effect was proved to be significant in a Wilcoxon rank sum test (W=3 997 747, p<.01). The effect was confirmed in a Wilcoxon signed rank test where distance values for sentences with frequent syllables were subtracted from those with rare syllables within each quadruple and context comparison (V=266 291, p<.05). Wilcoxon rank sum tests for combinations of pitch accent variability and quadruple identity (corrected for 120 comparisons) showed significant differences between frequent and rare syllables for quadruple 1 (W=83 690, p<.0001) and quadruple 6 (W=67 759, p<.05), while a marginally significant effect in the opposite direction was found for quadruple 2 (W=46 465, p=.08). Wilcoxon rank sum and signed rank tests showed no overall tendencies for frequency effects concerning prominence or force accent variability of the target syllables. Only in the case of quadruple 4 did frequent syllables tend to have smaller force accent differences across contexts than rare syllables (W=55 724, p<.05, corrected for 120 comparisons). An analysis of the syllables immediately surrounding the target syllable showed larger absolute prominence differences for syllables following frequent syllables than for syllables following rare syllables (W=3 933 934, p<.05, mean values: 0.235 vs. 0.225). This effect on the prominence variability of the following context syllable proved to be marginally significant in a Wilcoxon signed rank test grouped for quadruple, speaker, and context comparison (V=270 263, p=.088). No other syllable frequency effects were found on the variability of context syllables in terms of prominence, pitch accent, or force accent values.

7.3.4 Spectral Similarity

Spectral similarity values were calculated for segmentally identical syllables produced by the same speaker in different contexts. For frequent and rare syllables in each quadruple, Appendix G shows mean similarity values of syllables, onsets, vowels, and codas. A Wilcoxon rank sum test comparing syllable similarity suggests that frequent syllables were less consistent in their pronunciation than rare syllables (W=3 642 957, p<.01). However, a similar analysis for syllable onsets produced in the same syllable by the same speaker in different carrier sentences revealed an effect in the opposite direction, with frequent syllables showing less variation than rare syllables (W=3 942 244, p<.05). Wilcoxon signed rank tests examining comparisons paired for quadruple identity, speaker and context comparison, in which values for frequent syllables were subtracted from those for the corresponding rare syllables, confirmed the influence of syllable similarity (V=330 032, p<.0001), but found no significant effects concerning onset, vowel, and coda similarity.

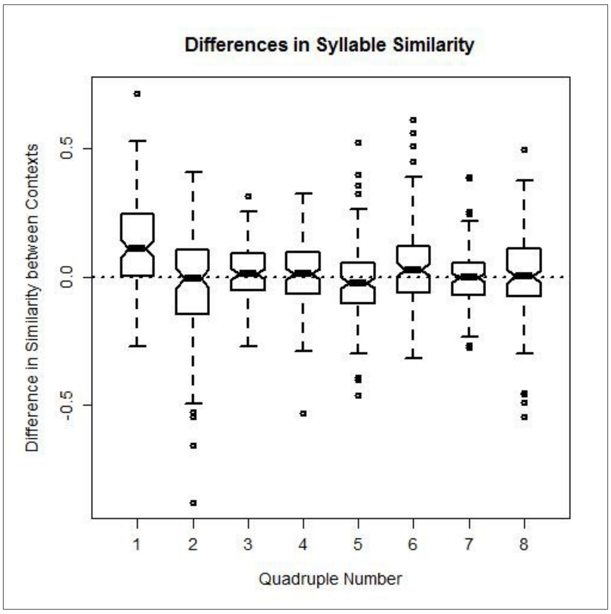


Figure 22: Differences in syllable similarity between frequent and rare syllables, grouped by participant, quadruple, and context comparison

A boxplot diagram of syllable similarity differences for individual quadruples (see Figure 22) shows that while quadruple 6 also exhibits a slight tendency for frequent syllables to be less consistant in their pronunciation than rare syllables, the effect is mainly due to quadruple 1. In investigations of combinations of quadruple identity and frequency level (corrected for 120 comparisons), a frequency effect on syllable similarity was discovered for quadruple 1 (W=49 315.5, p<.0001), while analyses of onset, nucleus, and coda similarity for individual quadruples revealed no significant effects.

7.3.5 Comparison with Original Annotations

In 1882 of the 3871 sentences analyzed, the original segmentation of the target syllables performed by the MAUS system was manually corrected, with 981 of the corrected syllables being frequent and 901 rare. For 946 additional sentences, changes were made concerning the duration of the two context syllables (452 sentences with frequent syllables, 494 with rare syllables). As there were also several changes made concerning the context vowels used for prominence calculation (e.g. the /l/ in the syllable [təl] was counted as part of the vowel in the corrected annotations), only 267 sentences with frequent syllables and 315 sentences with rare syllables remained completely unrevised. In spite of the large number of corrections made, most effects found in the present study were also visible when analyses were based on the original, uncorrected MAUS segmentations (see Table 16). There were a few differences, however. Effects of syllable frequency on vowel duration, syllable duration of quadruple 2, prominence of the following syllable, and onset similarity across contexts were no longer significant when results were based on automatic segmentations, while additional effects appeared with regard to the force accent levels of the preceding context syllable, the pitch accent variability of the following context syllable, and the force accent variability of the target syllables of quadruple 1. There were also some cases where differences were insignificant in one analysis and marginally significant in the other.

Investigations of absolute differences between original and corrected results with Wilcoxon rank sum tests showed a few significant effects of syllable frequency. In all cases, frequent syllables tended to have larger absolute differences between results from corrected versus uncorrected segmentations than rare syllables. In terms of duration, only a marginal effect for coda duration distances was visible (W=1 916 504, p=.074, mean values: 5.4 ms vs. 4.9 ms). There were, however, significant tendencies with regard to pitch accent values of the target syllable (W=1 945 699, $p<.01$, mean values: 0.036 vs. 0.027) and the preceding context syllable (W=1 940 859, $p<.05$, mean values: 0.027 vs. 0.024), prominence levels of the following syllable (W=1 954 656, $p<.05$, mean values: 0.094 vs. 0.08), as well as for spectral similarity results concerning the target syllables' vowels (W=3 925 562, $p<.05$, mean values: 0.0308 vs. 0.0307) and codas (W=3 954 416, $p<.01$, mean values 0.043 vs. 0.034). Marginally significant differences appeared for distances in prominence levels of the preceding context syllable (W=1 927 876, p=.094, mean values: 0.054 vs. 0.053), force and pitch accent levels of the following context syllable ($W>1\,926\,000$, $p<.098$, mean values force accent: 0.101 vs. 0.089, mean values pitch accent: 0.042 vs. 0.038), and spectral similarity results for the target syllables (W=3 916 931, p=.061, mean values: 0.028 vs. 0.026).

Parameter	Statistical Test	Corrected Annotations	Automatic Annotations
duration values (lower values for frequent syllables)	Wilcoxon rank sum tests	syllable: W=1760773** **vowel: W=1799040***	syllable: W=1751866***
	Wilcoxon rank sum tests (individual quadruples)	**syllable (q2): W=22399.5*** vowel (q6): W=18601****	vowel (q6): W=22137***
	Wilcoxon signed rank tests (grouped by speaker, quadruple, and context)	syllable.: V=210810**** vowel: V=205467.5**** coda: V=170626* follow. syl.: V=195667.5***	syllable: V=202310**** vowel: V=184541**** coda: V=170685.5*** follow. syl.: V=188522.5**
duration variability (lower values for frequent syllables)	Wilcoxon rank sum tests	vowel: W=3576768**** *coda: W=3716340 .* preced. syl.: W = 3669458* follow. syl.: W = 3663160*	vowel: W=3612274*** *coda: W=3712059 .* preced. syl.: W = 3658347** follow. syl.: W= 3692333*
	Wilcoxon rank sum tests (individual quadruples)	***vowel (q4): W=51503.5 .*** vowel (q6): W=49293 .	***syllable (q6): W=67031.5 .*** vowel (q6): W=49561 .
	Wilcoxon signed rank tests (grouped by speaker, quadruple, and context comparison)	vowel: V=329791.5**** coda: V=305409.5** *preced. syl: V=305453.5 .* follow. syl: V=325542****	vowel: V=299869*** coda: V=289419.5** *preced. syl.: V=290031 .* follow. syl.: V=316141****
prominence values (higher values for frequent syllables)	Wilcoxon signed rank tests (grouped by speaker, quadruple, and context)	*force accent preced. syl.: V=156678 .*	**force accent preced. syl.: V=153952***
prominence variability (higher values for frequent syllables in terms of pitch accent, but lower values in terms of force accent and prominence)	Wilcoxon rank sum tests	pitch accent: W=3997747** **prominence follow. syl: W=3933934***	pitch accent: W=4041135**** **pitch accent follow. syl.: W = 3999149****
	Wilcoxon rank sum tests (individual quadruples)	force accent (q4): W=55724* pitch accent (q1): W=83690**** ***pitch accent (q2): W=46465 . (oppos. direct.)*** pitch accent (q6): W=67759*	**force accent (q1): W=55313*** stress accent (q4): W=55204** pitch accent (q1): W=82238**** pitch accent (q6): W=72239****
	Wilcoxon signed rank tests (grouped by speaker, quadruple, and context comparison)	pitch accent: V=266291* ***prominence follow. syl: V= 270263 .***	***stress accent: V=187381 .*** pitch accent: V=263725* ***pitch accent preced. syl.: V=271907 .*** **pitch accent follow. syl. V=264307***
spectral similarity (lower values for frequent syllables, opposite tendency for onset similarity)	Wilcoxon rank sum tests	syllable: W=3642957** **onset: W=3942244***	syllable: W=3673212* onset: W=3919599 .
	Wilcoxon rank sum tests (individual quadruples)	syllable (q1): W=49315.5****	syllable (q1): W=48732****
	Wilcoxon signed rank tests (grouped by speaker, quadruple, and context comparison)	syllable: V=330032****	syllable: V=321500**

*Table 16: Syllable frequency effects based on original MAUS segmentations and manually corrected data, differences in bold print, marginally significant effects (p<.1) in cursive, .: p<.1, *: p<.05, **: p<.01, ***: p<.001, ****: p<.0001*

7.4 Discussion

The mental syllabary theory proposes that frequent syllables are stored in the brain as whole articulatory routines, whereas rare syllables need to be assembled from smaller units. According to this theory, frequent syllables should be more resistant to pronunciation errors and show a stronger coordination of articulatory movements than rare syllables. Due to their higher degree of coarticulation, frequent syllables would be expected to have shortened duration values in comparison with similar rare syllables. A stronger degree of coarticulation for frequent than for rare syllables might also affect an automatic segmentation of audio files, with frequent syllables requiring more and larger manual corrections. Previous studies suggest that potential shortening effects of syllable frequency can easily be overridden or reversed by differences with respect to the lexical stress status of the syllables or the segments they contain (Croot/Rastle, 2004; Flechsig, 2006; Herrmann et al., 2008). Even though in the present experiment frequent and rare syllables were closely matched for segmental content, stress status, and carrier sentences, there may still have been confounding influences of sentence stress realization and the amount of emphasis placed on the carrier words. Participants may have accentuated names containing unfamiliar rare syllables more strongly than names with frequent syllables. To control for this, the target syllables were analyzed in terms of acoustic prominence relative to their immediate surroundings. An exemplar-theoretic approach to speech production would predict an influence of syllable frequency on duration and pronunciation variability, as it assumes that frequent syllables have more possible exemplars available to chose from than rare syllables and would therefore be more closely adapted to their surroundings than rare syllables (Schweitzer/Möbius, 2004; Walsh et al., 2007). Thus, frequent syllables would be expected to show larger absolute differences in duration and prominence across contexts and receive lower values in investigations of spectral similarity between individual realizations than rare syllables.

There were a few more speech errors directly affecting the rare target syllables in comparison with those which were made while pronouncing the frequent target syllables. Although the difference between the two counts was in the expected direction, it was too small to lend itself to any conclusive interpretation. Across quadruples and participants, there was a small but consistent influence of syllable frequency on syllable and vowel duration, with frequent syllables being shorter than rare syllables. The effect remained significant even in a direct comparison of syllables belonging to the same quadruple and produced by the same speaker in the same carrier sentence. In such direct comparisons, the shortening effect of syllable frequency was also found for syllable codas. Analyses of syllables duration values in different order positions seemed to show a shortening effect due to syllable repetitions throughout the experiment, with a stronger influence of the order of appearance for rare than for frequent syllables. Although this tendency would be consistent with the idea of mentally stored syllabic routines, with frequent syllables already being firmly anchored in the brain and therefore benefiting less from a practice effect than rare syllables, duration differences between frequent and rare syllables in different order positions remained below significance level.

Frequent syllables did not vary more strongly across contexts in their duration than rare syllables. On the contrary, vowel variability was found to be smaller for frequent than for rare syllables, particularly for syllables with long vowels or diphthongs. Effects on absolute duration distances of vowels and syllable codas were confirmed in analyses grouped for speaker, quadruple, and context comparison. Although unexpected, these findings are not necessarily incompatible with an exemplar-theoretic point of view. Even if there are more syllable exemplars available for the frequent target syllables than for the rare syllables, the highly controlled contexts in which they appear provides little motivation for duration differences. The target items always appeared as the first, lexically stressed, syllable of bisyllabic surnames and were embedded in four different carrier sentences which were all formulated so as to imply a broad focus. And although the exemplar-theoretic model developed by Walsh and colleagues (2007) predicts a lower duration variability for rare syllables, it does not state that this is due to a low variability on a segmental level. Rather, it is proposed that duration variations among individual segments cancel each other out for rare syllables, while frequent syllables tend to behave as a single unit. It seems plausible that a large number of exemplars are stored for the subunits from which rare syllables are concatenated, which might explain the higher degree of variability of rare syllable nuclei and codas found in the present study.

There were no significant differences between frequent and rare target syllables in terms of acoustic prominence or its two subcomponents force accent (calculated from relative duration and spectral emphasis) and pitch accent (calculated from pitch differences and relative intensity). There was, however, an effect on the variability of pitch accent values across contexts, with frequent syllables showing greater distances between estimates for segmentally identical syllables produced by the same speaker in different target sentences. Although three of the examined quadruples showed a tendency in the opposite direction, differences between frequent and rare syllables remained significant in an analysis grouped for speaker, quadruple, and context comparison. An exemplar-theoretic explanation for this effect might be that the exemplars stored for frequent syllables offer a broader range of possible pitch contours, whereas rare syllables either have fewer exemplars available or need to be concatenated from smaller units, in which case an appropriate pitch contour may have to be calculated separately. While a German corpus study on the variability of particular pitch accent types for frequent and rare syllables revealed no clear effects of syllable frequency (Walsh et al., 2008), a corpus study for English was able to demonstrate an influence of how often words were associated with a particular pitch accent (Schweitzer et al., 2010). Here, pitch contours of a certain accent type were more variable for words which tended to appear with a range of different pitch accent types rather than for words which were mainly accompanied by the accent type under investigation. However, these findings are not directly comparable to the results of the present study, which did not distinguish different pitch accent types and investigated a single parameter calculated from pitch movements and relative overall intensity rather than describing pitch contours in terms of four-dimensional vectors.

Spectral similarity between frequent syllables in different contexts was lower than between rare syllables, which again fits in with exemplar-theoretic approaches and seems to run contrary to the findings for vowel duration, which showed a greater variation among rare syllables than among frequent syllables. For similarity on the level of the syllable onset, there was a tendency in the opposite direction, with lower similarity values for frequent than for rare syllables. However, this tendency did not prove to be significant in investigations paired for speaker, quadruple, and context comparison. Even though the effect of syllable frequency on similarity was significant across stimuli, and its direction was consistent with an exemplar-theoretic model, the findings proved to be inconclusive in a more detailed analysis. A closer examination of the data revealed that the effect was almost completely due to the behavior of the first quadruple, with quadruple 6 showing a minimal tendency in the same direction.

The frequency level of the target syllables also seemed to have had an effect on the pronunciation and variability of the syllables immediately surrounding them. The following context syllable tended to be less variable in terms of acoustic prominence and duration if preceded by a frequent syllable, with the influence on duration variability being confirmed in an analysis grouped for speaker, quadruple, and context comparison. Grouped comparisons also revealed an effect on the duration values themselves, with syllables following frequent items being shorter than syllables following rare items. The preceding context syllables tended to be less variable in duration if they were followed by a frequent syllable, with a marginally significant increase in force accent values. Neither the effect on duration variability or influence on force accent values of the preceding context syllable was confirmed in a grouped comparison, however. It is not obvious why syllable frequency should affect the surrounding syllables in this way. One explanation may be that names containing rare syllables were more likely to be surrounded by slight hesitations or have phrase boundaries inserted after them than were names containing frequent syllables. As long as this tendency was not particularly strong and did not affect all items to the same extent, it might have led to an increased variability in duration and prominence in the context of rare syllables.

While a comparison between results based on the original MAUS segmentations and the manually corrected files showed that many of the effects were visible in both of the two versions, there were also a number of differences between the two. As a result of the corrections, effects of syllable frequency on vowel duration and onset similarity appeared, while several effects concerning prominence measures and their variability across contexts were lost. In some cases, new effects were only marginally significant or moved from marginally significant to a significance level of $p<.05$. Distances between results from original and corrected segmentations tended to be larger for frequent syllables than for rare syllables, indicating that, possibly because of higher levels of coarticulation, these were more susceptible to segmentation errors. Most of these effects occurred for investigations of acoustic prominence and spectral similarity rather than for duration results. Although the automatic annotations captured duration and duration variability accurately enough to reveal even subtle effects of syllable frequency, they became less reliable if the segments were further analyzed in terms of fundamental frequency or spectral characteristics. It seems that while duration effects were comparatively robust with respect to boundary placement as long as boundary positions were assigned in a consistent manner, acoustic investigations were more sensitive as to where segmentation boundaries are positioned, since this determined which parts of the audio files were actually analyzed. In the present study, frequent syllables tended to be more strongly affected by the manual corrections than rare syllables, especially in terms of the prominence and spectral similarity data. It may be that frequent syllables and their immediate context were produced with a greater amount of coarticulation and enunciated less carefully than comparable rare syllables, leading to an increase in segmentation errors and thereby indirectly causing some of the differences in frequency effects for corrected and uncorrected data.

8 Summary and General Discussion

Languages such as German and English tend to have a highly uneven distribution of syllable frequencies, with comparatively few highly frequent syllables and a large number of rare types (Möbius, 2001). It seems plausible that at least the frequent syllables are not assembled from smaller segments each time they need to be produced, but that their pronunciation is stored in a "mental syllabary" in the form of complete articulatory programs (Levelt/Wheeldon, 1994; Levelt et al., 1999). A review of the literature dealing with the general role of syllables in speech production, and the possibility of stored syllabic units in particular, showed that evidence for the mental syllabary theory mostly stems from investigations of production latencies and the behavior of patients suffering from apraxia of speech (see Chapter 2). There is, however, also some evidence that frequent syllables tend to show a higher degree of interaction between articulators, and consequently a greater amount of coarticulation than rare syllables (Croot/Rastle, 2004; Flechsig, 2006; Benner et al., 2007; Herrmann et al., 2008). Although frequent syllables would also be expected to be shorter than rare syllables as a consequence of their storage as whole units and higher degree of coarticulation, studies show conflicting results with regard to possible syllable frequency effects on duration values (Levelt/Wheeldon, 1994; Croot/Rastle, 2004; Flechsig, 2006; Herrmann et al., 2008).

Many of the effects supporting a mental syllabary may also be explained by an exemplar-theoretic approach to speech production, which proposes that frequent syllables, words, and even phrases are mentally stored in the form of a large number of highly-detailed exemplars (Walsh et al., 2007). In contrast to the classic mental syllabary theory, which predicts a comparatively lowered degree of coarticulation between individual syllabic routines, exemplar-theoretic approaches expect frequent syllables to be more closely adapted to their surroundings than rare syllables and to show a greater amount of variability across different contexts. However, a different possibility might be that articulatory movements become entrenched through frequent practice, resulting in a greater stability and similarity of individual realizations (Pierrehumbert, 2001). A German corpus study supports the hypothesis that frequent syllables are more variable than rare syllables by showing that their average duration tends to be less dependent on the average duration of their individual segments than is the case for rare syllables (Schweitzer/Möbius, 2004). The aim of the present thesis was to investigate whether effects of syllable frequency can be found in terms of duration, acoustic prominence, and pronunciation variability across contexts.

8.1 Corpus Analysis

In order to determine groups of consistently frequent and rare syllables for the main experiment of the thesis, syllable frequencies were computed for two corpora of written language (Baroni/Kilgarriff, 2006; Quasthoff et al., 2006) as well as for orthographic transcriptions of four spoken language corpora (Wahlster, 2000; Koehn, 2005; Brysbaert et al., 2011; Schütte, 2014). The corpora, the process of extracting syllable frequencies, and the potential errors connected with each processing step are described in Chapter 3 (see also Samlowski et al., 2011). Gaining frequencies for phonetic syllables from orthographic texts was a complex process involving several steps. Automatic programs were used to remove annotations, unify text encoding formats, disambiguate numbers and abbreviations (Möhler et al., 2000), count orthographic words, transcribe them phonetically (Portele, 1999), segment the transcriptions into syllables, and finally calculate syllable frequencies (Schmid et al., 2007). An advantage of this procedure was that it was possible to quickly analyze large amounts of data. This was especially important as, due to the highly uneven distribution of syllable frequencies, large corpora were needed to determine adequate estimates for rare syllable types. By investigating frequencies in different corpora the advantages in the large size of written language corpora were combined with the insight which smaller spoken-language databases can offer into the vocabulary used in spontaneous, everyday speech. The resulting frequency lists also provided the opportunity to examine the extent to which individual syllable types varied in their usage across contexts and topics. At the same time, however, automatically calculating syllable frequencies necessarily led to a large number of errors and inaccuracies as well as to a general loss of variability. The fact that words were consistently transcribed according to their standard pronunciation is actually an advantage, since frequent syllables are prevented from being split up into various different possible pronunciations. Potential variability was also lost in the ways in which numbers and abbreviations were rendered. Whenever there were several equally valid ways to produce these items, syllable counts were biased towards the possibility chosen by the normalization program. A further difficulty was that errors frequently led to new syllables which inflated the total number of types in the corpus. These syllables were often in conflict with phonotactic rules, and tended to accumulate among the less frequent types, making it difficult to identify bona fide rare syllables. In the syllable frequency study described in Chapter 7, these difficulties were addressed by stipulating that all syllables had to follow the same consonant-vowel structure, and by postulating that frequent syllables had to have a minimum frequency and rare syllables a maximum frequency in all analyzed corpora. In this way it was ensured that frequent syllables were consistently frequent and that even if rare syllables were partly or mostly due to errors, they remained easily pronounceable.

Syllable frequencies in the six corpora were extensively analyzed and contrasted, using a number of methods suitable for comparing highly uneven frequency distributions in differently sized databases (Kilgarriff, 2001; Baroni, 2009). The analysis results are presented in Chapter 4 of this thesis (see also Samlowski et al., 2011). Within each of the corpora, syllable frequencies were found to be highly unevenly distributed, albeit to a lesser extent than word frequencies. The 500 most frequent syllable types of each database comprised at least 80% of its tokens. In a comparison of frequency ranks and relative frequencies for individual types, both measures proved to be strongly dependent on the frequency level of the syllable under investigation. Frequency ranks were more robust than relative frequencies for high-frequency syllables, whereas rare syllables tended to be less varied in terms of relative frequency than in terms of frequency ranks. Spearman correlations for frequencies of the more common syllable types revealed a comparatively high level of similarity between corpora of planned speech on the one hand and databases containing more spontaneous speech on the other. While an analysis of overrepresented key syllables using frequency ratios (Kilgarriff, 2009) confirmed a few systematic differences between these two groups, many variations in syllable frequency were found to be due to individual characteristics of the different corpora and the topics they tended to cover. The analysis suggests that while written language corpora can provide a rough estimate of frequencies in spoken language, it is still important to check how stable syllable frequencies are across different language styles and topics. One potentially important factor when investigating syllable frequency effects is their type frequency, i.e. the number of different words in which a particular syllable appears (Conrad et al., 2008). Syllables can be frequent either because they appear in a large number of word types or because they form part of a few high-frequency words. Although syllables dominated by few word types might be more susceptible to vocabulary differences arising from differences in corpus topics, an examination of the databases showed that syllable types which were particularly overrepresented in one corpus in comparison with the others did not necessarily have high token to type frequency ratios. Most of the key syllables identified for each corpus were confirmed to be comparatively overrepresented in an investigation of dispersion levels which took into account the differing corpus sizes (Gries, 2008). For a few syllable types, however, frequency differences between corpora were not strong enough to rule out the possibility that they may have simply been caused by inexact estimates due to the small size of the database in which they were overrepresented. When DeWaC (Baroni/Kilgarriff, 2006), the largest of the investigated databases, was examind in terms of the syllable frequency dispersion across its nineteen subcorpora, results indicated that even for a broad genre such as web articles it is possible to obtain very precise estimates of syllable frequencies on the basis of a 184 million word corpus. Consistent with the implications of the corpus analysis summarized here, the syllable frequency study presented in Chapter 7 defined frequent syllables on the basis of frequency ranks and rare syllables on the basis of relative frequencies. To avoid syllables which were overly influenced by the frequency of certain word forms, a further criterion limited the maximum ratio between type and token frequencies. Target syllables were only accepted if they fulfilled the conditions in all six databases.

8.2 Effects of Prosody and Grammar

Syllable pronunciation is influenced by a variety of interacting factors apart from syllable frequency, such as word and sentence stress (e.g. Sluijter/Heuven, 1996; Dogil/Williams, 1999), word boundaries (e.g. Cho, 2004; White/Turk, 2010), phrase boundaries (e.g. Cambier-Langeveld/Turk, 1999; Kentner/Féry, 2013), lexical class (Chamonikolasová, 2000; Bell et al., 2009), and word or lemma frequency (Jurafsky et al., 2002; Pluymaekers et al., 2005; Baker/Bradlow, 2009). As it is important to determine which influences need to be taken into consideration when conducting an experiment on syllable frequency effects, a reading study was carried out to examine various possible influences on syllable and vowel duration, acoustic prominence, and spectral similarity in German. In this way, it was also possible to compare the effect sizes of the influencing factors and to better assess the impact of a possible syllable frequency effect. The study consists of two experiments which were combined into one reading task. This meant that sentences for both experiments were produced by the same participants, thereby functioning as mutual distractors. In the first of the two experiments, effects of implied word and sentence stress as well as word and sentence boundaries were examined by investigating German verb prefixes and segmentally identical function words (see Chapter 5, see also Samlowski et al., 2012; Samlowski et al., 2014). The second experiment concerned effects of lexical class and lemma frequency on the pronunciation of segmentally identical German demonstrative pronouns, relative pronouns, and definite articles, with definite articles occurring much more frequently than the other two lexical classes (see Chapter 6, see also Samlowski et al., 2013; Samlowski et al., 2014). Only the immediate context of the target syllables was controlled, as carrier sentences for each item needed to be composed in a way that would allow the participants to identify the intended word meaning. In addition, the information structure of the sentences was manipulated to encourage certain sentence stress patterns. In the first experiment, sentences designed for verb prefixes in accented positions were formulated so as to imply a broad focus, while prefixes in potentially unaccented positions contained an additional contrast which was supposed to shift the focus away from the main verb. In the second experiment, sentences were generally formulated so as not to encourage a particular emphasis of the target pronouns and articles. Each sentence was presented along with an illustration created with the help of a text-to-scene program (Coyne/Sproat, 2001). These illustrations were intended to further clarify the sentence meaning and focus the participants' attention on what they were reading. Syllable and vowel duration of the target items was calculated on the basis of manual annotations. The strongest influence found for these variables in the two experiments was that of sentence boundary. Separated, sentence-final verb prefixes proved to be considerably longer than their bound counterparts or segmentally identical free function words. In the case of bisyllabic prefixes both syllables were affected by final lengthening. There was a small, but consistent influence of word stress, with syllables and vowels being longer for stressed than for unstressed prefixes. Effects of lexical class on syllable and vowel duration were similar in size to influences of lexical stress, although considering that different stimuli were used, the results were not directly comparable. Demonstrative pronouns tended to be longer than definite articles, with duration values of relative pronouns lying between the two. In terms of vowel duration, relative pronouns were closer to definite articles than to demonstrative pronouns, while in terms of syllable duration they were more similar to demonstrative pronouns than to definite articles. As relative and demonstrative pronouns do not differ strongly in frequency, differences between these two lexical classes are probably a result of their semantic functions, with demonstrative pronouns tending towards a higher degree of emphasis. Effects of lemma frequency may, however, help to explain why definite articles tended to be shorter than the other lexical classes. Differences between relative pronouns and definite articles in terms of syllable duration may also have been due to the fact that relative pronouns were often preceded by a syntactic clause boundary, which may have led to an increased closure duration of the initial plosives. Sentence stress only proved to have a minimal effect on vowel duration of lexically stressed verb prefixes, but may have had an indirect influence on word stress differences, which were more pronounced for words in accented positions. There was also a minimal influence of word boundary, with function words being slightly longer than segmentally identical bound prefixes. Effects on syllable and vowel duration due to sentence boundary, word stress, lexical class, sentence stress for lexically stressed prefixes, and word boundary were confirmed in comparisons paired for speaker and word context.

Since sentence stress was only manipulated indirectly through the information structure of the carrier sentences, it was important to examine the prosodic structure actually realized by the participants. An automatic prominence tagger (Tamburini/Wagner, 2007) was used to determine the extent to which absolute duration differences of target items across sentences coincided with differences in emphasis in relation to the immediate surroundings within each sentence. Based on manual vowel annotations, the tagger analyzed syllable nuclei in the target items as well as in the preceding and following context syllables in terms of force accent (calculated from relative duration and spectral emphasis) and pitch accent (calculated from pitch movements and relative intensity). Overall prominence was determined as a weighted product of these two measures. While the investigation showed clear differences between lexical classes in the second experiment, relative pronouns were found to be lowest in prominence. This result, which was contrary to the tendencies for absolute duration, may be interpreted as an artifact of sentence structure. As relative pronouns were often preceded by clause-final syllables, lengthening effects for these syllables probably led to comparatively lower prominence levels for the pronouns themselves. Relative pronouns which were embedded in prepositional phrases, and therefore did not immediately follow clause boundaries, tended to have higher prominence levels than definite articles. Demonstrative pronouns were generally slightly higher in prominence than definite articles, which is consistent with the theory that their high duration values may have partly been caused by higher levels of emphasis. Probably as a consequence of final lengthening, there was a clear influence of sentence boundary, with sentence-final prefixes receiving higher values than bound prefixes or segmentally identical free words. Investigations across syllables also showed significant effects for word stress in accented and unaccented positions as well as an effect of sentence stress on lexically stressed prefixes, although these differences mostly failed to reach significance in separate investigations of individual syllables. In comparisons paired for speaker, context, and syllable identity, significant differences in terms of acoustic prominence were confirmed for effects of lexical class, sentence boundary, word stress, and sentence stress of stressed prefixes. No prominence differences were found between unstressed bound prefixes and free function words in similar contexts.

In a separate investigation of the two components of the prominence measure, force accent values tended to follow patterns similar to those of the general prominence levels. In comparisons paired for speaker, context, and syllable identity, both measures revealed effects for the same factors. Pitch accent measures, on the other hand, were mostly clustered around zero. However, boxplot diagrams revealed a less steep distribution with a larger variety of possible values for demonstrative pronouns and, to a lesser extent, for relative pronouns, sentence-final prefixes and lexically stressed syllables in accented positions. Paired comparisons of pitch accent values showed significant differences between lexical classes, between sentence-final prefixes and a few of the other prefix conditions, and between accented stressed and unaccented unstressed prefixes. The results indicate that effects on absolute duration were often accompanied by relative differences in vowel duration and spectral emphasis in comparison with immediate context syllables. Although pitch movement and relative intensity also played a role in signaling demonstrative and relative pronouns, as well as in indicating sentence boundary and sentence stress, this was more in the form of an optional possibility that was only sporadically utilized, than as a reliable tendency towards pitch accents.

According to exemplar theory, syllable frequency effects may have an influence on the variability of their pronunciation across different contexts (Walsh et al., 2007). As grammatical and prosodic structure might influence the degree of pronunciation variability as well, the acoustic analysis included a measure of spectral similarity within and across categories (Lewandowski, 2011). Pairs of segmentally identical syllables produced by the same speaker were analyzed by cross-correlating amplitude envelopes for four contiguous frequency bands equally spaced on a logarithmic scale from 80 to 7800 Hertz. In the first experiment, lexically stressed prefixes were found to be more similar across verb contexts than unstressed prefixes, as were separated, sentence-final prefixes. Comparisons between sentence categories showed that sentence-final prefixes were more similar to lexically stressed than to unstressed prefixes belonging to the same verb. No consistent tendencies appeared for the second experiment, although in an analysis across target syllables definite articles were found to be minimally less similar across contexts than demonstrative and relative pronouns. In an overall comparison of different lexical classes, definite articles appeared to be slightly less similar to relative pronouns than were demonstrative pronouns. A comparison of results suggests that, on the whole, syllables tended to be closer to each other in pronunciation if they were longer and acoustically more prominent. A reason for this may be that when syllables are spoken more slowly, they tend to be enunciated more carefully and the articulators have a better chance of reaching their target positions, while quickly produced syllables tend to be reduced in pronunciation and more strongly coarticulated with their surroundings.

The two production experiments demonstrated that sentence boundaries, lexical class, word stress status, word boundaries, and implied sentence stress may all play a part in how segmentally identical syllables are pronounced. Duration in particular was found to be affected by nearly all of the investigated factors. Prominence results proved to be valuable indicators of the emphasis placed on the target items relative to their surroundings, but at the same time were highly sensitive to differences in the structure of the target sentences. Spectral similarity values seemed to be comparatively less easy to influence and only reacted to some of the effects analyzed. These findings clearly establish the importance of taking into account the factors analyzed in the two experiments when investigating syllable frequency effects.

8.3 Effects of Syllable Frequency

The experiment on syllable frequency was carefully designed in order to control for potentially confounding influences (see Chapter 7). Eight quadruples consisting of two frequent and rare syllables each were constructed in a way to balance effects of bigram identity and frequency of bigrams and individual segments. They were embedded as the first, lexically stressed syllable of bisyllabic nonsense words, thereby eliminating interactions with word stress effects and influences of lemma frequency. All words were placed in the same eight different carrier sentences, where they were used as fictive surnames. This way there were no confounding influences of sentence structure or lexical class. By dividing the eight carrier sentences into two groups and presenting each group to half of the speakers, it was possible to investigate the pronunciation of the syllables in a variety of contexts without overly straining the participants. Finally, the test sentences were interspaced with filler sentences and presented to each participant in a unique order to control for effects of learning or fatigue.

Sentences were omitted from the analysis if they contained self-corrections by the participants or if the target syllables or their immediate context was impaired in some way. Nonetheless, error rates for frequent and rare syllables were compared to ascertain whether there was any effect of syllable frequency. Syllable and vowel boundaries for the target syllables as well as the preceding and following context syllable were determined on the basis of an automatic phonetic annotation process (Kisler et al., 2012). The annotation files were manually edited to correct obvious segmentation errors. As was the case for the preceding experiments, target syllables were analyzed with respect to syllable and vowel duration, acoustic prominence (Tamburini/Wagner, 2007), and spectral similarity (Lewandowski, 2011). To investigate whether syllable frequency affected the pronunciation of the immediate surroundings, duration and prominence measures were analyzed for the preceding and following context syllable as well. In addition to the spectral similarity measures, variability of duration and prominence across contexts was investigated by calculating absolute differences between values for segmentally identical syllables produced by the same speaker in different contexts. Finally, the results from the manually corrected annotation files were compared with results based on a fully automatic analysis. On the one hand, this comparison allowed an evaluation of the alignment system used for the automatic annotation. On the other, absolute differences between results based on automatic and corrected annotations were compared for frequent and rare syllables in order to determine whether a potentially higher amount of coarticulation among the frequent syllables resulted in more extensive manual corrections.

Frequent syllables were found to be shorter than rare syllables in syllable and vowel duration. The effect was very subtle and similar in size to the effect of sentence stress differences in the first experiment. The duration differences were confirmed in a direct comparison of syllables belonging to the same quadruple and produced by the same speaker in the same context. There were no accompanying differences in syllable prominence, which indicates that the duration differences were not a result of systematic changes in the prosodic structure of the carrier sentences. Frequent syllables were also slightly less prone to production errors. A comparison between results based on the original automatic alignment and results based on the manually corrected annotations showed a few effects of syllable frequency on prominence and spectral similarity estimates, with larger differences appearing for frequent syllables as compared with rare syllables. This may be an indication that the automatic segmentation of the frequent syllables was hampered by a higher level of coarticulation. No such frequency effect was found in terms of the size of duration differences between the two annotation versions. In comparisons of the syllable frequency effects visible for automatic versus corrected annotations, divergences between findings also mostly affected prominence and spectral similarity estimates rather than duration results. A possible explanation might be that while the automatic annotations were, for the most part, sufficiently consistent to reveal duration differences between frequent and rare syllables, there were still errors in the automatic alignment which potentially impaired a further acoustic analysis of the segments. On the whole, the results support a mental syllabary model according to which frequent syllables are expected to show a close coordination between articulatory movements along with a decreased overall duration and a greater resistance to speech errors (Levelt/Wheeldon, 1994).

The experiment provided no definite evidence for or against an exemplar-theoretic model of speech production according to which frequent syllables should show a stronger variation across different realizations (Pierrehumbert, 2001; Walsh et al., 2007). Comparisons between segmentally identical syllables produced by the same speaker in different contexts revealed that vowel duration values were actually less variable for frequent than for rare syllables, especially for syllables containing long vowels. However, exemplar-theoretic approaches did not make any predictions concerning the variability of individual syllable segments and no effect in either direction was found for the duration variability of whole syllables. Investigations of spectral similarity were inconclusive as well. Even though the data seemed to show a tendency for frequent syllables to be less similar across contexts than rare syllables, the effect was mainly due to the behavior of one particular quadruple, and an examination of onset similarities revealed a slight trend in the opposite direction. For five of the eight quadruples, frequent syllables did show larger absolute differences between contexts than rare syllables where the pitch accent part of the prominence calculation was concerned. This might indicate that pitch accents were more varied when placed on frequent syllables than on rare syllables. Syllable frequency effects on the duration and variability of immediate surroundings of the target syllables suggest that rare syllables may, at least occasionally, have provoked a lengthening or prosodic strengthening of the following context syllable.

Duration differences between frequent and rare syllables were very small and tended to be visible only when the two frequent syllables in each quadruple were compared with the two rare syllables, thereby canceling out any segmental effects. Because of the small effect size, it seems probable that the influences found are not produced or perceived as deliberate cues to the listener to enable better understanding of rare syllables, but that they are rather a consequence of motor plans for frequent syllables being slightly better coordinated due to a higher amount of training. Although rare syllable types may not be represented as whole routines at all, effects of syllable frequency do not necessarily reflect a binary distinction between syllables which are mentally stored and those which are not. Articulatory coordination of syllable routines would be expected to become stronger through frequent repetition, and a larger number of stored exemplars would result in a greater choice of possible pronunciation variations. On the other hand, according to the dual route hypothesis (Whiteside/Varley, 1998a), frequent syllables may not always be produced via stored motor plans. Sometimes, a subsyllabic assembly route may prove to be more appropriate if the context or the situation requires a clearly enunciated production.

Considering that syllable frequency effects can be gradual, and that the method of assembling the articulatory programs may depend on the conditions under which speech is produced, the experiment setting was not entirely favorable to strong effects of syllable frequency. Due to the highly restricted number of contexts in which each target syllable was placed, there was little reason for a high variability across sentences in terms of duration, prominence, and spectral characteristics. In addition, participants did not take part in an actual communication situation where they needed to make themselves understood to a listener. Rather, they had to read aloud random, unconnected sentences with a highly repetitive structure. This may have further discouraged variation in the manner of production. As the target names were unfamiliar and, if mentioned in a real-life situation, would provide important and unpredictable information, participants may have been prompted to reduce the degree of coarticulation, e.g. by assembling the syllable pronunciation from smaller units. For some participants there may even have been interferences from syllable frequencies in other languages. Three of the people who took part in the study had a bilingual background and several others stated that they had had close contact to foreign languages. Finally, the experiment did not control for the conditional frequency of the target syllables in lexically stressed, word-initial positions. A further division of the frequency counts according to word stress or position within the word might have provided a stronger cue for syllable frequency effects. It seems noteworthy that significant syllable frequency effects prevailed over these potentially detrimental influences. By choosing target syllables which proved to be consistently frequent or rare across various databases, carefully matching them in terms of segmental content, and presenting them in a highly controlled environment, it was possible to neutralize many different influences on syllable duration, prominence, and similarity across contexts which might otherwise have overshadowed the comparatively minimal influence of syllable frequency.

9 Conclusions and Future Work

This thesis presents evidence in support of the mental syllabary theory, according to which frequent syllables are stored in the brain in the form of complete articulatory routines (Levelt/Wheeldon, 1994). In a highly controlled German-language reading study investigating frequent and rare syllables in terms of duration, acoustic prominence (Tamburini/Wagner, 2007), and spectral similarity of across contexts (Lewandowski, 2011), frequent syllables proved to be slightly shorter than comparable rare syllables. Only weak support was found for the related exemplar-theoretic approach, which proposes that frequent syllables are stored in the form of many detailed exemplars and would consequently show a greater variation across different contexts than rare syllables (Walsh et al., 2007). While frequent syllables did tend to be to be more variable than rare syllables in terms of pitch accent values (a prominence measure combining pitch movements and overall intensity in relation to the immediate context), vowel duration was actually less variable for frequent syllables than for rare syllables. A slight trend towards lower similarity for frequent syllables across contexts in terms of their spectral characteristics was discovered to be mainly due to differences found for one of the eight syllable groups analyzed. To gain annotations of syllable and vowel boundaries for the target syllables and their immediately preceding and following context syllables, the audio files were automatically segmented with a forced alignment system (Kisler et al., 2012) and then manually corrected. Comparisons with results based on the original, uncorrected annotations revealed that, apart from a few exceptions, the same duration effects appeared in both versions. The automatic annotation was less reliable, however, when the segmented syllables and vowels were analyzed in terms of prominence or spectral similarity. This suggests that while the automatic alignment system can offer reliable information to discover duration effects, it may be problematic to use the automatic segmentation to extract parts of the audio files for further analysis without first checking for errors. In the present investigation, frequent syllables tended to show larger differences than rare syllables in a comparison of results based on the automatic annotations with results based on manually corrected annotations. A reason for this might be that segments in frequent syllables were more strongly coarticulated and therefore less easy to separate automatically.

The syllable frequency information on which the experiment was based was collected through an extensive analysis of two written language corpora and orthographic transcriptions of four spoken language databases. The texts were automatically processed in order to gain syllabified phonetic transcriptions from which to compute syllable frequencies. A comparison of frequencies in the different corpora suggested that despite a few systematic differences between corpora of planned language and spoken language corpora, written language databases can generally serve as a rough estimate of syllable frequencies in spontaneous speech. Nonetheless, it is advisable to analyze more than one database in order to determine how variable individual syllable frequencies are across different contexts and modalities. For the production experiment on syllable frequency in this thesis, all six databases were investigated in order to ensure that the stimuli used were consistently frequent or rare across different contexts and modalities. Rare syllables were defined on the basis of relative frequencies, whereas for frequent syllables, rank position were used as they proved to be more stable from one database to another.

Two further production experiments were carried out in order to analyze the influence of various other factors on syllable pronunciation in German, determine the size of their effect, and discover to what extent they need to be taken into consideration when examining influences of syllable frequency. Sentences from both experiments were mixed together and presented in a single reading task. The first experiment used verb prefixes to examine interactions between implied word and sentence stress as well as influences of word and sentence boundaries, while the second experiment analyzed influences of lexical class and lemma frequency on segmentally identical words functioning as demonstrative pronouns, relative pronouns, and definite articles. Nearly all examined parameters proved to have a significant influence on syllable or vowel duration. These duration differences were often, but not always, accompanied by similar effects on acoustic prominence values. The factors word stress and sentence boundary in particular also influenced measures of spectral similarity across different contexts. The results indicated that experiments investigating syllable frequency effects on duration have to take into account confounding influences of word stress, information structure and implied sentence stress, word and sentence boundaries, lexical class, and lemma frequency. The fact that acoustic prominence measures were found to be highly sensitive to differences regarding the context syllables used as a reference for comparing the degree of relative emphasis reinforced the need to ensure that frequent and rare stimuli are closely matched in terms of their surrounding context.

Despite the highly controlled setting of the main study in this thesis, effects of syllable frequency on duration did not prove to be highly significant. Further experiments are needed to confirm whether or not the subtle duration differences between frequent and rare syllables are merely due to chance. It may also be worthwhile to place the target items in a larger number of different contexts, varying factors such as the syllables' stress status, the lexical class of the carrier word, or the syntactic boundary following the target syllable, while at the same time still closely matching frequent and rare syllables. In this way, it may be possible to increase the overall duration variability of the target items and perhaps elicit syllable frequency effects which may or may not support an exemplar-theoretic model of speech production. Another point where further research would be advisable is the definition of frequent and rare syllables. For the present thesis syllable counts were based on canonical pronunciations of individual words. Syllable frequencies in connected speech may differ from these counts due to phonological processes such as assimilation, lenition, or elision. A corpus investigation comparing speech and lexeme syllable frequencies in German in a similar manner as the Dutch study by Schiller and colleagues (1996) could help to discover key differences between actually produced syllables and the more theoretical syllable types normally used in experiments on syllable frequency effects. By conducting syllable frequency studies specifically for syllables which are frequent in connected speech but less frequent in canonical production of individual words and vice versa, it may be possible to explore how abstract the mentally stored syllable types are, and whether or not pronunciation variations of the same canonical syllable type are stored and counted separately. Finally, further research is needed on the various effects of prosody and grammar on syllable production in German, especially with respect to the interaction of the different influencing factors. Although the production experiments in this thesis showed that word stress, sentence stress, word and sentence boundaries, lexical class, and lemma frequency all may play a role in determining syllable duration, much of the variability remains unexplained. Here, a larger study with closely matched stimuli in a greater number of contexts and a separate quasi-random order of sentences for each speaker might lead to more robust findings. Especially for investigations of word and sentence stress, it can be useful to confirm the actual realization of the test sentences with a perception study.

10 References

Aichert, I., and Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language 88:1*, 148-159.

Aichert, I., and Ziegler, W. (2008). Learning a syllable from its parts: Cross-syllabic generalisation effects in patients with apraxia of speech. *Aphasiology 22:11*, 1216-1229.

Alario, F.-X., Goslin, J., Michel, V., and Laganaro, M. (2010). The functional origin of foreign accent: Evidence from the syllable-frequency effect in bilingual speakers. *Psychological Science 21:1*, 15-20.

Aylett, M.P., and Turk, A.E. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech 47:1*, 31-56.

Baayen, H.R., Piepenbrock, R., and Gulikers, L. (1995). "The CELEX Lexical Database (CD ROM)". University of Pennsylvania: Linguistic Data Consortium. URL: https://catalog.ldc.upenn.edu/LDC96L14

Baayen, H.R., Piepenbrock, R., and van Rijn, H. (1993). "The CELEX Lexical Database (CD ROM)". University of Pennsylvania: Linguistic Data Consortium.

Baker, R.E., and Bradlow, A.R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech 52:4*, 391-413.

Baroni, M. (2009). "Distributions in text," in A. Lüdeling & M. Kytö, *Corpus Linguistics: An International Handbook*, Berlin: Mouton de Gruyter, 803-821.

Baroni, M., and Kilgarriff, A. (2006). "Large linguistically-processed web corpora for multiple languages," in: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, Trento, 87-90.

Batliner, A., Kompe, R., Kießling, A., Nöth, E., Niemann, H., and Kilian, U. (1995). "The prosodic marking of phrase boundaries: Expectations and results," in A.J. Rubio Ayuso & J.M. Lopez Soler, *Speech Recognition and Coding: New Advances and Trends*, Berlin; Heidelberg; New York: Springer, 89-92.

Beckman, M.E., and Edwards, J. (1990). "Lengthenings and shortenings and the nature of prosodic constituency," in J. Kingston & M.E. Beckman, *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*, Cambridge: Cambridge University Press, 152-178.

Bell, A., Brenier, J.M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language 60:1*, 92-111.

Benner, U., Flechsig, I., Dogil, G., and Möbius, B. (2007). "Coarticulatory resistance in a mental syllabary," in: *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, Saarbrücken, 485-588.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10*, 341-345.

Brand, M., Rey, A., and Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory and Language 48*, 435-443.

Brendel, B., Ziegler, W., Erb, M., Riecker, A., and Ackermann, H. (2008). "Does our brain house a "mental syllabary"? An fMRI study," in: *Proceedings of the 8th International Seminar on Speech Production (ISSP 2008)*, Strasbourg, 73-76.

Brockhaus, W. (1999). "The syllable in German: Exploring an alternative," in H.v.d. Hulst & N.A. Ritter, *The Syllable: Views and Facts*, Berlin: Mouton de Gruyter, 169-218.

Brysbaert, M., Buchmeier, M., Conrad, M., and Jacobs, A.M. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology 58:5*, 412-424.

Burger, S. (1997). "Transliteration spontansprachlicher Daten - Lexikon der Transliterations-konventionen". Munich University: Institute of Phonetics and Speech Processing. URL: http://www.bas.uni-muenchen.de/forschung/Verbmobil/ VMtrlex2d.html

Burger, S., Weilhammer, K., Schiel, F., and Tillmann, H. (2000). "Verbmobil data collection and annotation," in W. Wahlster, *Verbmobil: Foundations of speech-to-speech translation*, Berlin; New York: Springer, 539-549.

Bybee, J.L. (2001). *Phonology and Language Use.* Cambridge; New York: Cambridge University Press.

Cambier-Langeveld, T., Nespor, M., and Heuven, V.J.v. (1997). "The domain of final lengthening in production and perception in Dutch," in: *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, Rhodes, 931-934.

Cambier-Langeveld, T., and Turk, A.E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics 27*, 255-280.

Campbell, N., and Beckman, M. (1997). "Stress, prominence, and spectral tilt," in: *Proceedings of the ESCA Workshop "Intonation: Theory, Models and Applications"*, Athens, 67-70.

Carreiras, M., and Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language 90:1-3*, 393-400.

Cavaglià, G. (2002). "Measuring corpus homogeneity using a range of measures for inter-document distance," in: *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas.

Chamonikolasová, J. (2000). On the capacity of different word classes to signal prosodic prominence: A comparative study of Czech and English. *Brno Studies in English 26*, 5-12.

Chen, J.-Y., Lin, W.-C., and Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology 45:1*, 107-120.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics 32*, 141-176.

Cho, T., and Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics 37:4*, 466-485.

Cholin, J., Dell, G.S., and Levelt, W.J.M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition 37:1*, 109-122.

Cholin, J., Levelt, W.J.M., and Schiller, N.O. (2006). Effects of syllable frequency in speech production. *Cognition 99:2*, 205-235.

Cholin, J., Levelt, W.J.M., and Schiller, N.O. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. *Language and Cognitive Processes 24:5*, 662-684.

Cholin, J., Schiller, N.O., and Levelt, W.J.M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language 50:1*, 47-61.

Chujo, K., and Utiyama, M. (2006). Selecting level-specific specialized vocabulary using statistical measures. *System 34:2*, 255-269.

Claßen, K., Dogil, G., Jessen, M., Marasek, K., and Wokurek, W. (1998). Stimmqualität und Wortbetonung im Deutschen. *Linguistische Berichte 174*, 202-245.

Conrad, M., Carreiras, M., and Jacobs, A.M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes 23:2***,** 296-326.

Conrad, M., and Jacobs, A.M. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes 19:3***,** 369-390.

Costa, A., and Sebastian-Gallés, N. (1998). Abstract phonological structure in language production: Evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition 24:4***,** 886-903.

Coyne, B., and Sproat, R. (2001). "WordsEye: An Automatic Text-to-Scene Conversion System," in: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, Los Angeles, 487-496.

Crompton, A. (1981). Syllables and segments in speech production. *Linguistics 19:7-8***,** 663-716.

Croot, K., and Rastle, K. (2004). "Is there a syllabary containing stored articulatory plans for speech production in English?," in: *Proceedings of the 10th Australian International Conference on Speech Science and Technology (SST 2004)*, Sydney, 376-381.

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review 93:3***,** 283-321.

Dogil, G., and Williams, B. (1999). "The phonetic manifestation of word stress," in H. van der Hulst, *Word Prosodic Systems in the Languages of Europe*, Berlin; New York: Mouton de Gruyter.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics 19:1***,** 61-74.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and van der Vrecken, O. (1996). "The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes," in: *Proceedings of the 4th International Conference on Spoken Language (ICSLP 1996)*, vol. 3, 1393-1396.

Evert, S., and Baroni, M. (2007). "zipfR: Word frequency distributions in R," in: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, 29-32.

Féry, C., and Kentner, G. (2010). "The prosody of embedded coordinations in German and Hindi," in: *Proceedings of the 5th International Conference on Speech Prosody (SP 2010)*, Chicago, vol. 100237, 1-4.

Ferrand, L., and Segui, J. (1998). The syllable's role in speech production: Are syllables chunks, schemas, or both? *Psychonomic Bulletin & Review 5:2*, 253-258.

Ferrand, L., Segui, J., and Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language 35*, 708-723.

Ferrand, L., Segui, J., and Humphreys, G.W. (1997). The syllable's role in word naming. *Memory and Cognition 25:4*, 458-470.

Flechsig, I. (2006). *An investigation of coarticulation in rare and frequent syllables with regard to (the existence of) a mental syllabary.* Diploma Thesis, University of Stuttgart.

Fougeron, C., and Keating, P.A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America 101:6*, 3728-3740.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America 27*, 765-768.

Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech 1*, 126-152.

Goldsmith, J. (2009). "The syllable," in J.A. Goldsmith, J. Riggle & A.C.L. Yu, *The Handbook of Phonological Theory*, Oxford: Wiley-Blackwell.

Greenberg, S. (1999). Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation. *Speech Communication 29*, 159-176.

Gries, S.T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics 13:4*, 403-437.

Gries, S.T. (2010). "Useful statistics for corpus linguistics," in A. Sánchez & M. Almela, *A Mosaic of Corpus Linguistics - Selected Approaches*, Frankfurt am Main: Peter Lang, 269-291.

Gulikers, L., Rattink, G., and Piepenbrock, R. (1995). "German linguistic guide". University of Pennsylvania: Linguistic Data Consortium. URL: https://catalog.ldc.upenn.edu/docs/LDC96L14/gug_a4.pdf

Hayes, B. (1988). "Metrics and phonological theory," in F. Newmeyer, *Linguistics: The Cambridge Survey*, Cambridge: Cambridge University Press, 220-249.

Herrmann, F., Whiteside, S.P., and Cunningham, S. (2008). "An acoustic investigation into coarticulation and speech motor control: high vs. low frequency syllables," in: *Proceedings of the Meetings on Acoustics - 155th Meeting of the Acoustical Society of America (ASA)*, Paris, vol. 4, 060007.

Heuven, V.J.v., and Jonge, M.d. (2011). Spectral and temporal reduction as stress cues in Dutch. *Phonetica 68*, 120-132.

Huffaker, D., Jorgensen, J., Iacobelli, F., Tepper, P., and Cassell, J. (2006). "Computational measures for language similarity across time in online communities," in: *Proceedings of the ACL Workshop on Analyzing Conversations in Text and Speech*, New York City, 15-22.

Hulst, H.v.d., and Ritter, N.A. (1999). "Theories of the syllable," in H.v.d. Hulst & N.A. Ritter, *The Syllable: Views and Facts*, Berlin: Mouton de Gruyter.

Hutzler, F., Bergmann, J., Conrad, M., Kronbichler, M., Stenneken, P., and Jacobs, A.M. (2004). Inhibitory effects of first syllable-frequency in lexical decision: an event-related potential study. *Neuroscience Letters 372:179-184*.

Jendryschik, M. (2008). "Zeichenkodierung," in: *Einführung in XHTML, CSS und Webdesign.*).

Jurafsky, D., Bell, A., and Girand, C. (2002). "The role of the lemma in form variation," in C. Gussenhoven & N. Warner, *Laboratory Phonology 7*, Berlin; New York: Mouton de Gruyter, 3-34.

Kentner, G., and Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review 30:2***,** 277-311.

Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics 6:1***,** 97.

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory 1:2***,** 263-275.

Kilgarriff, A. (2009). "Simple maths for keywords," in: *Proceedings of the Corpus Linguistics Conference (CL 2009)*, Liverpool.

Kilgarriff, A. (2012). "Getting to know your corpus," in: *Proceedings of the 15th International Conference on Text, Speech and Dialogue (TSD 2012)*, Brno, 3-15.

Kisler, T., Schiel, F., and Sloetjes, H. (2012). "Signal processing via web services: The use case WebMaus," in: *Proceedings of the Digital Humanities 2012 (DH 2012)*, Hamburg, 30-34.

Kleber, F., and Klipphahn, N. (2006). An acoustic investigation of secondary stress in German. *Arbeitsberichte Institut für Phonetik Kiel 37***,** 1-18.

Klein, W., and Schütte, W. (2004). "Transkriptionsrichtlinien für die Eingabe in DIDA". Mannheim: Institut für deutsche Sprache (IDS). URL: http://agd.ids-mannheim.de/download/dida-trl.pdf

Koehn, P. (2005). "Europarl: A parallel corpus for statistical machine translation," in: *Proceedings of the 10th Machine Translation Summit (MT Summit X)*, Phuket, 79-86.

Kohler, K.J. (1983). Prosodic boundary signals in German. *Phonetica 40:2*, 89-134.

Kohler, K.J. (1987). "The linguistic function of F0 peaks," in: *Proceedings of the 11th International Congress on Phonetic Sciences (ICPhS 1987)*, Talinn, vol. 3.

Kohler, K.J. (2012). "Segment duration and vowel quality in German lexical stress perception," in: *Proceedings of the 6th International Conference on Speech Prosody (SP 2012)*, Shanghai, vol. 2, 633-636.

Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition 42:1-3*, 1-22.

Levelt, W.J.M. (1999). "Producing spoken language: a blueprint of the speaker," in C.M. Brown & P. Hagoort, *The Neurocognition of Language*, New York: Oxford University Press.

Levelt, W.J.M., Roelofs, A., and Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences 12*, 1-75.

Levelt, W.J.M., and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition 50:1-3*, 239-269.

Lewandowski, N. (2011). *Talent in nonnative phonetic convergence.* PhD dissertation, University of Stuttgart.

Liebscher, R. (2003). New corpora, new tests, and new data for frequency-based corpus comparisons. *CRL Newsletter 15:2*, 3-14.

Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H&H theory," in W.J. Hardcastle & A. Marchal, *Speech Production and Speech Modelling*, Dordrecht; Boston: Kluwer Academic Publishers, 403-439.

Lintfert, B. (2010). *Phonetic and phonological development of stress in German.* PhD Dissertation, University of Stuttgart.

Marasek, K. (1996). "Glottal corellates of the word stress and the tense/lax opposition in German," in: *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 1996)*, Philadelphia, vol. 3, 1573-1576.

Mayer, J., Ackermann, H., Dogil, G., Erb, M., and Grodd, W. (2003). "Syllable retrieval vs. online assembly: fMRI examination of the syllabary," in: *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, 2541-2544.

Mengel, A. (1997). "Das akustische Korrelat des deutschen Wortakzents," in: *Proceedings of the 8. Konferenz Elektronische Sprachsignalverarbeitung (ESSV 1997)*, Cottbus, 271-277.

Mertens, T., and Schneider, D. (2009). "Efficient subword lattice retrievala for German spoken term detection," in: *Proceedings of the 34th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, Taipei.

Meyer, A.S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language 30*, 69-89.

Meyer, A.S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition 42*, 181-211.

Miglio, V., and Chun, D. (2008). An acoustic study of stress in L2 production of German and Spanish. *Journal of the Canadian Acoustical Association 36:4*, 44-45.

Möbius, B. (1998). "Word and syllable models for German text-to-speech synthesis," in: *Proceedings of the 3rd ESCA/COCOSDA Workshop on Speech Synthesis (SSW3-1998)*, Jenolan Caves.

Möbius, B. (2001). Rare events and closed domains: Two delicate concepts in speech synthesis. *International Journal of Speech Technology 6:1*, 57-71.

Möhler, G., Schweitzer, A., Breitenbücher, M., and Barbisch, M. (2000). "IMS German Festival" (Version: 1.2-os). University of Stuttgart: Institut für maschinelle Sprachverarbeitung (IMS). URL: http://www.ims.uni-stuttgart.de/phonetik/synthesis/festival_opensource.html

Mohr, G., Kimpton, M., Stack, M., and Ranitovic, I. (2004). "Introduction to Heritrix, an archival quality web crawler," in: *Proceedings of the 4th International Web Archiving Workshop (IWAW 2004)*, Bath.

Mooshammer, C., Fuchs, S., and Fischer, D. (1999). "Effects of stress and tenseness on the production of CVC syllables in German," in: *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 1999)*, San Francisco, 409-412.

Müller, K. (2002). *Probabilistic syllable modeling using unsupervised and supervised learning methods.* PhD Dissertation, University of Stuttgart.

Müller, K., Möbius, B., and Prescher, D. (2000). "Inducing probabilistic syllable classes using multivariate clustering," in: *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*, Hong Kong, 225-232.

Okobi, A.O. (2006). *Acoustic correlates of word stress in American English.* PhD Dissertation, Massachusetts Institute of Technology.

Ortega-Llebaria, M., and Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech 54:1*, 73-97.

Ortega-Llebaria, M., Prieto, P., and Vanrell, M.d.M. (2007). "Perceptual evidence for direct acoustic correlates of stress in Spanish," in: *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, Saarbrücken, 1121-1124.

Ortega-Llebaria, M., Vanrell, M.d.M., and Prieto, P. (2010). Catalan speakers' perception of word stress in unaccented contexts. *Journal of the Acoustical Society of America 127:1*, 462-471.

Parsons, K., McCormac, A., and Butavicius, M. (2009). "Human dimensions of corpora comparison: An analysis of Kilgarriff's (2001) approach". Edinburgh South Australia: Defense Science and Technology Organization (DSTO). URL: www.dtic.mil/dtic/tr/fulltext/u2/a506585.pdf

Perret, C., Bonin, P., and Méot, A. (2006). Syllabic priming effects in picture nameing in French - Lost in the sea! *Experimental Psychology 53:2*, 95-104.

Peters, B., Kohler, K.J., and Wesener, T. (2005). "Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache". University of Kiel: Institite of Phonetics and Digital Speech Processing.

Pierrehumbert, J.B. (2001). "Exemplar dynamics: Word frequency, lenition and contrast," in J.L. Bybee & P.J. Hopper, *Frequency Effects and the Emergence of Linguistic Structure*, Amsterdam: John Benjamins Publishing Company, 137-157.

Plag, I., Kunter, G., and Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics 39:3*, 362-374.

Pluymaekers, M., Ernestus, M., and Baayen, H.R. (2005). Lexical frequency and acoustic recuction in spoken Dutch. *Journal of the Acoustical Society of America 118:4*, 2561-2569.

Portele, T. (1999). "TXT2PHO - a TTS front end for the German inventories of the MBROLA project" (Version: 0.9.4). Bonn University. URL: http://www.sk.uni-bonn.de/forschung/phonetik/sprachsynthese/txt2pho

Portele, T., Krämer, J., and Stock, D. (1995). "Symbolverarbeitung im Sprachsynthesesystem Hadifix," in: *Proceedings of the 6. Konferenz Elektronische Sprachsignalverarbeitung (ESSV 1995)*, Wolfenbüttel, 97-104.

Quasthoff, U., Richter, M., and Biemann, C. (2006). "Corpus portal for search in monolingual corpora," in: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genua, 1799-1802.

R Core Team (2014). "R: A language and environment for statistical computing" (Version: 3.1.1). Vienna, Austria: R Foundation for Statistical Computing. URL: http://www.r-project.org/

Raghavendra, E.V., Yegnanarayana, B., and Prahallad, K. (2008). "Speech synthesis using approximate matching of syllables," in: *Proceedings of the 2008 IEEE Spoken Language Technologies Workshop (SLT 2008)*, Goa.

Rapp, S. (1995). Maschinelles Lernen von Aspekten des deutschen Wortakzents. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS) 2:2*, 151-240.

Rayson, P., and Garside, R. (2000). "Comparing corpora using frequency profiling," in: *Proceedings of the ACL Workshop on Comparing Corpora*, Hong Kong, vol. 9, 1-6.

Remus, R., and Bank, M. (2012). "Textual Characteristics of different-sized corpora," in: *Proceedings of the 5th Workshop on Building and Using Comparable Corpora (BUCC 2012)*, Istanbul, 148-152.

Roelofs, A., and Meyer, A.S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition 24:4*.

Rose, T., and Haddock, N. (1997). "The effects of corpus size and homogeneity on language model quality," in: *Proceedings of the 5th ACL Workshop on Very Large Corpora*, Hong Kong, 178-191.

Sadeghi, V. (2011). "Acoustic correlates of lexical stress in Persian," in: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)*, Hong Kong, 1738-1741.

Samlowski, B., Möbius, B., and Wagner, P. (2011). "Comparing syllable frequencies in corpora of written and spoken language," in: *Proceedings of the Interspeech 2011*, Florence, 637-640.

Samlowski, B., Möbius, B., and Wagner, P. (2014). Phonetic detail in German syllable pronunciation: Influences of prosody and grammar. *Frontiers in Psychology 5:500*.

Samlowski, B., Wagner, P., and Möbius, B. (2012). "Disentangling lexical, morphological, syntactic and semantic influences on German prosody - Evidence from a production study," in: *Proceedings of the Interspeech 2012*, Portland, 2406-2409.

Samlowski, B., Wagner, P., and Möbius, B. (2013). "Effects of lexical class and lemma frequency on German homographs," in: *Proceedings of the Interspeech 2013*, Lyon, 597-601.

Schiller, N.O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language 39*, 484-507.

Schiller, N.O. (1999). Masked syllable priming of English nouns. *Brain and Language 68*, 300-305.

Schiller, N.O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition 26:2*, 512-528.

Schiller, N.O., and Costa, A. (2006). Activation of segments, not syllables, during phonological encoding in speech production. *The Mental Lexicon 1:2*, 231-250.

Schiller, N.O., Costa, A., and Colomé, A. (2002). "Phonological encoding of single words: In search of the lost syllable," in C. Gussenhoven & N. Warner, *Laboratory Phonology 7*, Berlin: Mouton de Gruyter, 35-59.

Schiller, N.O., Meyer, A.S., Baayen, H.R., and Levelt, W.J.M. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantative Linguistics 3:1*, 8-28.

Schiller, N.O., Meyer, A.S., and Levelt, W.J.M. (1997). The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech 40:2*, 103-140.

Schmid, H., Möbius, B., and Weidenkaff, J. (2007). "Tagging syllable boundaries with joint n-gram models," in: *Proceedings of the Interspeech 2007*, Antwerp, 2857-2860.

Schneider, K., and Möbius, B. (2006). "Production of word stress in German: Children and adults," in: *Proceedings of the 3rd International Conference on Speech Prosody (SP 2006)*, Dresden.

Schneider, K., and Möbius, B. (2007). "Word stress correlates in spontaneous child-directed speech in German," in: *Proceedings of the Interspeech 2007*, Antwerp, 1394-1397.

Schoor, A., Aichert, I., and Ziegler, W. (2012). A motor learning perspective on phonetic syllable kinships: How training effects transfer from learned to new sllables in severe apraxia of speech. *Aphasiology 26:7*, 880-894.

Schunk, C. (2004). "Token-Wort-Konvertierung in der Text-To-Speech-Synthese". University of Stuttgart: Institut für Maschinelle Sprachverarbeitung. URL: http://www.ims.uni-stuttgart.de/lehre/studentenarbeiten/fertig/Studienarbeit_Schunk.pdf

Schütte, W. (2014). "Korpus Gespräche im Fernsehen (GF)". Mannheim: Institut für Deutsche Sprache (IDS), Abteilung Pragmatik: Archiv für Gesprochenes Deutsch. URL: http://agd.ids-mannheim.de/download/Korpus_GF_extern.pdf

Schweitzer, A., and Möbius, B. (2004). "Exemplar-based production of prosody: Evidence from segment and syllable durations," in: *Proceedings of the Speech Prosody 2004 (SP 2004)*, Nara, 459-462.

Schweitzer, K., Calhoun, S., Schütze, H., Schweitzer, A., and Walsh, M. (2010). "Relative frequency affects pitch accent realisation: Evidence for exemplar storage of prosody," in: *Proceedings of the 13th Australasian International Conference on Speech Science and Technology (SST 2010)*, Melbourne, 62-65.

Sereno, J.A., and Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech 38***,** 57-76.

Sevald, C.A., Dell, G.S., and Cole, J.S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language 34***,** 807-820.

Shattuck-Hufnagel, S. (1983). "Speech errors as evidence for a serial-order mechanism in sentence production," in P.F. MacNeilage, *The Production of Speech*, New York: Springer.

Shattuck-Hufnagel, S. (1987). "The role of word-onset consonants in speech production planning: New evidence from speech error patterns," in E. Keller, *Motor and Sensory Processes of Language*, Hillsdale, N.J.: Erlbaum, 17-51.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition 42***,** 213-259.

Shattuck-Hufnagel, S., and Turk, A.E. (1998). The domain of phrase-final lengthening in English. *Journal of the Acoustical Society of America 103:5***,** 1235-1236.

Silverman, K.E.A. (1990). "The separation of prosodies: comments on Kohler's paper," in J. Kingston & M.E. Beckman, *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*, Cambridge: Cambridge University Press, 139-151.

Sluijter, A.M.C., and Heuven, V.J.v. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America 100:4***,** 2471-2485.

Staiger, A., and Ziegler, W. (2008). Syllable frequency and syllable structure in the spontaneous speech production of patients with apraxia of speech. *Aphasiology 22:11***,** 1201-1215.

Stock, D., Braun, M., Bernhard, H., Portele, T., Krämer, J., Rauth, M., Breuer, S., and Bröggelwirth, J. (2001). "Bonn Machine-Readable Pronunciation Dictionary (BOMP, HADI-BOMP, BOSS-BOMP, FEST-BOMP)". University of Bonn: Institut für Kommunikationsforschung und Phonetik (IKP). URL: http://www.ikp.uni-bonn.de/forschung/phonetik/sprachsynthese/bomp/bomp/

Streck, T. (2004). *Finale Dehnung im Deutschen: Eine kontrastive Untersuchung zu den städtischen Varietäten von Mannheim und Hamburg.* Master Thesis, Freiburg University.

Strom, V., and Widera, C. (1996). "What's in the "pure" prosody?," in: *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 1996)*, Philadelphia, vol. 3.

Tamburini, F., and Wagner, P. (2007). "On automatic prominence detection for German," in: *Proceedings of the Interspeech 2007*, Antwerp, 1809-1812.

Taylor, P.A. (2009). *Text-to-Speech Synthesis.* Cambridge; New York: Cambridge University Press.

Turk, A.E., and Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics 28*, 397-440.

Turk, A.E., and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics 35*, 445-472.

Varley, R.A., and Whiteside, S.P. (1998). "Evidence of dual-route phonetic encoding from apraxia of speech: Implications for phonetic encoding models," in: *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP 1998)*, Sydney.

Varley, R.A., Whiteside, S.P., and Donovan, M. (2000). "A dual-route model of speech control: The need for convergent evidence from motor speech disorders," in: *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST-Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling (SPS5)*, Kloster Seeon.

Varley, R.A., Whiteside, S.P., Windsor, F., and Fisher, H. (2006). Moving up from the segment: A comment on Aichert and Ziegler's Syllable frequency and syllable structure in apraxia of speech, Brain and Language, 88, 148–159, 2004. *Brain and Language 96:2*, 235-239.

Wade, T., and Möbius, B. (2007). "Speaking rate effects in a landmark-based phonetic exemplar model," in: *Proceedings of the Interspeech 2007*, Antwerp, 402-405.

Wagner, P. (2002). *Vorhersage und Wahrnehmung deutscher Betonungsmuster.* PhD Dissertation, Bonn University.

Wahlster, W. (2000). "Mobile speech-to-speech translation of spontaneous dialogs: An overview of the final Verbmobil system," in W. Wahlster, *Verbmobil: Foundations of Speech-to-Speech Translation*, Berlin: Springer, 3-21.

Walsh, M., Schütze, H., Möbius, B., and Schweitzer, A. (2007). "An exemplar-theoretic account of syllable frequency effects," in: *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, Saarbrücken, 1173-1176.

Walsh, M., Schweitzer, K., Möbius, B., and Schütze, H. (2008). "Examining pitch-accent variability from an exemplar-theoretic perspective," in: *Proceedings of the Interspeech 2008*, Brisbane, 877-880.

Wells, J.C. (1997). "SAMPA - computer readable phonetic alphabet," *Handbook of Standards and Resources for Spoken Language Systems*, Berlin; New York: Mouton de Gruiter, 684-732.

Wheeldon, L.R., and Levelt, W.J.M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language 34*, 311-334.

White, L. (2002). *English speech timing: A domain and locus approach.* PhD Dissertation, University of Edinburgh.

White, L., and Turk, A.E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics 38:3*, 459-471.

Whiteside, S.P., and Varley, R.A. (1998a). "Dual-route phonetic encoding: Some acoustic evidence," in: *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP 1998)*, Sydney.

Whiteside, S.P., and Varley, R.A. (1998b). A reconceptualisation of apraxia of speech: A synthesis of evidence. *Cortex 34:2*, 221-231.

Widera, C., Portele, T., and Wolters, M. (1997). "Prediction of word prominence," in: *Proceedings of the 5th European Conference on Speech Communication and Technologies (Eurospeach 1997)*, Rhodes, 999-1002.

Windmann, A., Jauk, I., Tamburini, F., and Wagner, P. (2011). "Prominence-based prosody prediction for unit selection speech synthesis," in: *Proceedings of the Interspeech 2011*, Florence, 325-328.

Wolters, M., and Wagner, P. (1998). "Focus perception and prominence," in: *Proceedings of the 4. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 1998)*, Bonn, 227–236.

You, W., Zhang, Q., and Verdonschot, R.G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PlOS ONE 7:10*.

Ziegler, W. (2009). Modelling the architecture of phonetic plans: Evidence from apraxia of speech. *Language and Cognitive Processes 24:5*, 631-661.

Zipf, G.K. (1968). *The Pysycho-Biology of Language: An Introduction to Dynamic Philology.* Cambridge: The MIT Press.

11 Appendices

11.1 Appendix A: Stimuli for Experiment 1

This appendix contains the test sentences for the production experiment on effects of word stress, sentence stress, and syntactic boundaries (see Chapter 5). Sentences are sorted according to the target verbs and the categories represented by the carrier sentences. The English translations that are given for each sentence are meant to help understand the design of the target items and the carrier sentences and therefore tend to be literal rather than idiomatic. The tables also include information on the order in which the sentences were presented, naming for each sentence the position in which it appeared within the experiment.

Table A1: Sentences for "umfahren" –['ʔʊm.faː.ʁən] ("to drive over") vs. [ʔʊm.'faː.ʁən] ("to drive around")

Category	Sentence + Translation	Position
w+s+	Die Kinder waren sehr traurig, als der Mann nicht bremsen konnte und ihre Katze **umfahren** musste. *The children were very sad when the man couldn't brake and had to **run over** their cat.*	073
w+s-	Geländewagen können kleinere Sträucher überrollen, aber keine Bäume **umfahren**. *SUVs can navigate over small bushes but cannot **drive over** trees.*	057
w-s+	Der Mann war in den Gegenverkehr geraten, als er die Katze **umfahren** wollte. *The man swerved into the oncoming traffic as he tried to **drive around** the cat.*	047
w-s-	Das Boot muss sowohl die rote als auch die grüne Boje **umfahren**. *The boat has to **drive around** the red as well as the green buoy.*	008
sb	Auf dem Weg nach Hause **fuhr** der Mann die Katze **um**. *On his way home the man **ran over** the cat.*	078
mb	Die roten Bojen sollte man **umfahren**, wenn man sich nicht in Wasserpflanzen verheddern möchte. *You should **drive around** the red buoys if you do not want to get tangled in seaweed.*	019
wb	Vor allem sollte man, **um Fahren** zu lernen, viel Geduld und Aufmerksamkeit mit sich bringen. *What you need above all, **in order to** learn to **drive**, is a lot of patience and concentration.*	005

Table A2: Sentences for "umstellen" – ['ʔʊm.ʃtɛ.lən] ("to move around") vs. [ʔʊm.'ʃtɛ.lən] ("to surround")

Category	Sentence + Translation	Position
w+s+	Nach dem Umzug werden wir die Möbel vollkommen **umstellen**. *After the move we are going to completely **re-arrange** the furniture.*	095

w+s-	Die Computer springen von selbst auf Sommerzeit um – ich muss nur noch die Uhren **umstellen**. *The computers switch to daylight-saving time automatically – I only have to* ***re-set*** *the clocks.*	091
w-s+	Die Polizei konnte den Entführer vollkommen **umstellen**. *The police managed to completely* ***surround*** *the kidnapper.*	059
w-s-	Die Polizei konnte das Haus, aber nicht den Garten **umstellen**. *The police managed to* ***surround*** *the house but not the garden.*	086
sb	Nach dem Umzug **stellte** er seine Möbel vollkommen **um**. *After the move, he completely* ***re-arranged*** *the furniture.*	061
mb	Die Polizei positionierte sich so, dass der Einbrecher sich **umstellt** fühlen musste. *The police positioned themselves in a way which let the burglar feel himself* ***surrounded****.*	029
wb	Es handelt sich **um stellenweise** undichtes Material. *It's* ***about partially*** *permeable material.*	022

Table A3: Sentences for "unterschlagen" – ['ʔʊn.tɐ.ʃla:.gən] ("to fold in") vs. [ʔʊn.tɐ.'ʃla:.gən] ("to embezzle")

Category	**Sentence + Translation**	**Position**
w+s+	Man muss die Butter **unterschlagen** und etwa drei Minuten verrühren, bis der Teig cremig wird. *You have to* ***fold in*** *the butter and mix for three minutes until the dough is creamy.*	039
w+s-	Zunächst muss man Eier und Milch verrühren und anschließend den Zucker **unterschlagen**. *First you have to mix the eggs with the milk and then* ***fold in*** *the sugar.*	026
w-s+	Sie hatte in dem Rezept aus Versehen Backpulver und Zucker **unterschlagen**. *In the recipe, she accidentally* ***left out*** *the baking powder and the sugar.*	093
w-s-	Die Täterin hatte nicht nur zahlreiche Banken ausgeraubt, sondern auch Steuergelder **unterschlagen**. *The culprit had not only robbed several banks but had also* ***embezzled*** *tax money.*	030
sb	Anschließend schlägt man Eier und Zucker **unter**. *Then you* ***fold in*** *the eggs and the sugar.*	065
mb	Er kann sich die teuren Autos leisten, weil er regelmäßig Gelder **unterschlägt**. *He can afford the expensive cars because he regularly* ***embezzles*** *money.*	100
wb	Schließlich gab er die Gelder **unter Schlägen** dem Räuber. ***Under blows****, he finally gave the money to the robber.*	096

Table A4: Sentences for "unterstellen" – ['ʔʊn.tɐ.ʃtɛ.lən] ("to store/ take shelter") vs. [ʔʊn.tɐ.'ʃtɛ.lən] ("to insinuate")

Category	**Sentence + Translation**	**Position**
w+s+	Wir wollten uns **unterstellen**, weil es so stark regnet. *We wanted to* ***take shelter*** *because it is raining so hard.*	033
w+s-	Sie können nicht nur Fahrräder, sondern auch Autos bei uns **unterstellen**. *You can* ***store*** *not only bicycles but also cars with us.*	043

w-s+	Der Kellner wollte uns **unterstellen**, dass wir nicht bezahlt hätten. *The waiter wanted to* ***insinuate*** *that we hadn't paid.*	098
w-s-	Tierquälerei haben sie uns **unterstellt** – nur, weil unser Hund ihre Katze auf den Baum gejagt hat. *They* ***accused*** *us of animal cruelty – just because our dog chased their cat up the tree.*	070
sb	Da es gerade stark regnete, **stellten** wir uns **unter**. *Because it was raining hard we* ***took shelter****.*	103
mb	Man kann **unterstellen**, dass Ratten von vielen als Ungeziefer gesehen werden. *One can* ***assume*** *that rats are commonly seen as vermin.*	007
wb	Man kann **unter Ställen** oft Mäuse und Ratten finden. *One can often find mice and rats* ***under sheds****.*	037

Table A5: Sentences for "überlaufen" – ['ʔyː.bɐ.laʊ.fən] ("to spill over") vs. [ʔyː.bɐ.'laʊ.fən] ("crowded")

Category	Sentence + Translation	Position
w+s+	So sehr ich auch aufpasse – immer wird das Nudelwasser ein wenig **überlaufen**. *As careful as I am – the pasta water will always* ***spill over*** *a bit.*	016
w+s-	Wenn du beim Eingießen acht gibst, wird nur wenig **überlaufen**. *If you pour carefully, only a little will* ***spill over****.*	051
w-s+	Am Samstag Nachmittag ist die Stadt immer ein wenig **überlaufen**. *On Saturday afternoon the town is always a little* ***crowded****.*	035
w-s-	Der Strand ist überall sehr voll, aber in diesem Abschnitt verhältnismäßig wenig **überlaufen**. *The beech is very full everywhere, but this part is relatively less* ***crowded****.*	069
sb	Wenn es stark gießt, **läuft** die Regentonne immer ein wenig **über**. *When it rains hard, the rain barrel always* ***spills over*** *a little.*	023
mb	Joggen gehen sollte man dort eher abends, da der Park vorher ein wenig **überlaufen** ist. *You should preferably go there in the evening to jog, since before then the park is a little* ***crowded****.*	006
wb	Wenn man joggen möchte, sollte man sich vorher ein wenig **über Laufen** als Sportart informieren. *If you want to go jogging, you should learn a little* ***about running*** *as a sport beforehand.*	087

Table A6: Sentences for "überziehen" – ['ʔyː.bɐ.tsiː.ən] ("to put on") vs. [ʔyː.bɐ.'tsiː.ən] ("to overdraw")

Category	Sentence + Translation	Position
w+s+	Es ist kalt draußen – du solltest dir etwas **überziehen**. *It is cold outside – you should* ***put*** *something* ***on****.*	021
w+s-	Nimm dein neues Jackett – bei so einem Fest sollte man sich nicht einfach irgendwas **überziehen**. *Wear your new suit jacket – at a party like this you shouldn't* ***put on*** *just anything.*	010
w-s+	Es gibt drei Tage Kulanzzeit – Sie können also etwas **überziehen**. *There is a grace period of three days – you can* ***overstep*** *(the deadline) a little.*	082

w-s-	Er würde sich nicht stark verschulden, sondern sein Konto nur etwas **überziehen**. *He wouldn't go way into debt but just **overdraw** his account slightly.*	032
sb	Bevor er in die Kälte hinausging, **zog** er sich noch etwas **über**. *Before he went out into the cold, he **put** something **on**.*	090
mb	Muss man sein Konto häufig **überziehen**, sollte man auf günstige Zinsen achten. *If you need to frequently **overdraw** your account, you should look for favorable interest rates.*	041
wb	Er klagte häufig **über ziehende** Schmerzen in den Unterschenkeln. *He frequently complained **about twinging** pains in his lower legs.*	050

Table A7: Sentences for "durchschauen" – ['dʊʁç.ʃaʊ.ən] ("to examine") vs. ['dʊʁç.'ʃaʊ.ən] ("to see through")

Category	**Sentence + Translation**	**Position**
w+s+	Der Lehrer hofft, dass er sich in den Ferien alle Arbeiten **durchschauen** kann. *The teacher hopes that he can **look through** all the exams over the holidays.*	079
w+s-	Ich habe den Text fast ganz durchgesehen, ich muss mir nur noch die letzten Seiten **durchschauen**. *I read almost the whole text; I only have to **look through** the last few pages.*	076
w-s+	Der Lehrer hofft, dass er die Plagiate in den Arbeiten **durchschauen** kann. *The teacher hopes that he can **see through** plagiarism in the exams.*	013
w-s-	Der Junge ahnte, dass sein Vater ihn getäuscht hatte, konnte aber nicht den Grund für seine Lügen **durchschauen**. *The boy suspected that his father had tricked him but couldn't **figure out** the reason for his lies.*	083
sb	Der Lehrer **schaute** sich in den Ferien die Arbeiten **durch**. *During the holidays, the teacher **looked through** the exams.*	049
mb	Der Wissenschaftler wollte den Kandidaten nur **durchschauen** und so seine Methoden herausfinden. *The scientist only wanted to **figure out** the candidate and discover his methods.*	066
wb	Der Kandidat hat gewettet, dass er über 50 Teesorten nur **durch Schauen** unterscheiden kann. *The candidate made a bet that he could distinguish 50 brands of tea merely **by looking** at them.*	054

Table A8: Sentences for "durchstreichen" – ['dʊʁç.ʃtʁaɪ.çən] ("to cross out") vs. [dʊʁç.'ʃtʁaɪ.çən] ("to wander through")

Category	**Sentence + Translation**	**Position**
w+s+	In dem Formular muss man alle leer gebliebenen Felder **durchstreichen**. *On the form you should **cross out** any fields you left blank.*	052
w+s-	Bei der Korrektur sollte man nicht ganze Sätze, sondern nur die jeweiligen Fehler **durchstreichen**. *During revision you should not **cross out** entire sentences but just the mistakes.*	068
w-s+	Auf ihrer Reise wollten sie Flüsse durchwaten und Felder **durchstreichen**. *On their journey they wanted to wade through rivers and **wander through** fields.*	044

w-s-	Sie wollten keine Bergwanderung machen, sondern nur ein wenig die Felder **durchstreichen**. *They didn't want to go hiking in the mountains, but just wanted to **wander through** the fields a little.*	072
sb	**Streichen** Sie in dem Antrag bitte alle nicht ausgefüllten Felder **durch**. *Please **cross out** any fields you left blank in the application.*	074
mb	**Durchstreichen** Sie die zahlreichen Wälder oder genießen Sie die Aussicht vom Burgturm aus. ***Wander through** the many forests or enjoy the view from the castle tower.*	060
wb	**Durch Streichen** wird der Zaun vor Wind und Wetter geschützt. ***By painting,** you protect the fence from wind and weather.*	056

11.2 Appendix B: Stimuli for Experiment 2

This appendix contains the test sentences for the production experiment on effects of word class and lemma frequency (see Chapter 6). Sentences are sorted according to the target words, the phonetic context in which the target words were placed, and the grammatical role they fulfilled within the sentence. The English translations that are given for each sentence are meant to help understand the design of the target items and the carrier sentences and therefore tend to be literal rather than idiomatic. The tables also include information on the order in which the sentences were presented, naming for each sentence the position in which it appeared within the experiment.

Table B1: Sentences for "der (m)" – [deːɐ] (masculine singular nominative)

Context	Lexical Class	Sentence+Translatiion	Position
1	Demonstr. Pron.	So, wie dieser Typ angezogen ist, hatte ich mir schon gedacht, dass **der** ein Feuerwehrmann ist. *The way that guy was dressed, I already guessed that **that** was a fireman.*	075
	Rel. Pron.	Seitdem ihre Katze überfahren wurde, verzehrte sie ein Hass, **der** eines Tages ausbrechen würde. *Since her cat was run over, she was consumed by a hatred **that** would some day explode.*	089
	Def. Art.	Die Polizei positionierte sich so, dass **der** Einbrecher sich umstellt fühlen musste. *The police positioned themselves in a way which had to make **the** burglar feel himself surrounded.*	029
2	Demonstr. Pron.	Kann ich den 845er Bus nehmen? Früher fuhr **der** manchmal über Lüftelberg. *Can I take bus number 845? **That** (bus) sometimes used to go via Lüftelberg.*	003

	Rel. Pron.	Der Fahrer des Wagens hinter der Müllabfuhr, **der** Maler von Beruf war, fand einfach keine Möglichkeit zum Überholen. *The driver of the car behind the garbage truck, **who** was a painter, simply could not find a chance to pass it.*	046
	Def. Art.	Auf dem Weg nach Hause fuhr **der** Mann die Katze um. *On the way home **the** man ran over the cat.*	078
3	Demonstr. Pron.	Der Flug hat Verspätung – es dauert noch eine ganze Weile, bis **der** Thailand erreicht. *The flight is delayed – it will take some time before **that** (flight) reaches Thailand.*	058
	Rel. Pron.	Die Schlange hat einen Biss, **der** teilweise lähmend wirkt. *The snake has a bite **which** is partially paralyzing.*	081
	Def. Art.	Man muss die Butter unterschlagen und etwa drei Minuten verrühren, bis **der** Teig cremig wird. *You have to fold in the butter and mix for three minutes until **the** batter is creamy.*	039

Table B2: Sentences for "der (f)" – [deːɐ] (feminine singular dative)

Context	**Lexical Class**	**Sentence+Translation**	**Position**
1	Demonstr. Pron.	Die Eisdiele ist sehr teuer - bei **der** kostet eine Kugel 1,20 Euro. *The ice cream parlor is very expensive – at **that** (place) one scoop costs 1.20 euros.*	088
	Rel. Pron.	Vor der Fahrt gibt es eine Routineuntersuchung, bei **der** kontrolliert wird, ob alle Gurte fest sitzen *Before the ride there is a routine inspection, during **which** they check to see whether all the belts are securely fastened.*	099
	Def. Art.	Bei **der** Korrektur sollte man nicht ganze Sätze, sondern nur die jeweiligen Fehler durchstreichen. *During (**the**) revision you should not cross out entire sentences, but just the mistakes.*	068
2	Demonstr. Pron.	Für ein kräftiges Rosa muss man eine Brühe aus roter Beete kochen – in **der** färbt man dann die Eier. *For bright pink you need to make a dye from of red beets – in **that** (dye) you then color the eggs.*	045
	Rel. Pron.	Sie mochte diese Geschichte, in **der** ferne Länder und fremdartige Kulturen beschrieben wurden. *She liked this story, in **which** far-away countries and strange cultures were described.*	015
	Def. Art.	In **der** Ferne ist ein Kahn zu sehen – auf dem weht eine amerikanische Fahne. *In **the** distance you can see a boat – on that (boat) an American flag is waving.*	002
3	Demonstr. Pron.	Die Kutsche darf die Kreuzung nicht als erste überqueren – vor **der** fahren zunächst der Käfer und dann der LKW. *The horse-drawn carriage may not cross the intersection first – before **that** (vehicle), the VW and then the truck have right of way.*	034

	Rel. Pron.	Es gibt eine Hütte neben dem Haus, vor **der** Fahrräder abgestellt werden können. *There is a shed alongside the house, in front of* ***which*** *bicycles can be stored.*	080
	Def. Art.	Vor **der** Fahrt gibt es eine Routineuntersuchung, bei der kontrolliert wird, ob alle Gurte fest sitzen. *Before* ***the*** *ride there is a routine inspection, during which they check to see whether all the belts are securely fastened.*	099

Table B3: Sentences for "die (sg)" – [di:] (feminine singular nominative/accusative)

Context	**Lexical Class**	**Sentence+Translation**	**Position**
1	Demonstr. Pron.	Die Frau kam ihm vertraut vor, aber er hatte keine Ahnung, woher er **die** kannte. *The woman seemed familiar to him, but he had no idea from where he knew* ***that*** *(woman).*	017
	Rel. Pron.	Das Mädchen mochte die Geschichte sehr, **die** Kaninchen, Ponys und Drachen enthielt. *The girl very much liked the story,* ***which*** *contained rabbits, ponies, and dragons.*	001
	Def. Art.	Der Mann war in den Gegenverkehr geraten, als er **die** Katze umfahren wollte. *The man swerved into the oncoming traffic as he tried to drive around* ***the*** *cat.*	047
2	Demonstr. Pron.	Das Auto hat die Ampel überfahren, obwohl **die** rot war. *The car went through the traffic light even though* ***that*** *(light) was red.*	063
	Rel. Pron.	Da war eine Schnecke im Kohl, **die** rosa und schleimig zwischen den Blättern hervorlugte. *There was a snail in the cabbage,* ***which*** *peeped out pink and slimy from between the leaves.*	036
	Def. Art.	Das Boot muss sowohl **die** rote als auch die grüne Boje umfahren. *The boat has to drive around* ***the*** *red as well as the green buoy.*	008
3	Demonstr. Pron.	Diese Suppe ist sehr vielseitig – man kann **die** kalt oder warm genießen. *The soup is very versatile – you can enjoy* ***that*** *(soup) cold or warm.*	048
	Rel. Pron.	Eine neue Ära begann, **die** Kathedralenbau überall in Europa revolutionieren sollte. *A new era began* ***which*** *would revolutionize the building of cathedrals everywhere in Europe.*	011
	Def. Art.	Auf dem Weg nach Hause fuhr der Mann **die** Katze um. *On the way home the man ran over* ***the*** *cat.*	078

Table B4: Sentences for "die (pl)" – [di:] (masculine/feminine/neuter plural nominative/accusative)

Context	Lexical Class	Sentence+Translation	Position
1	Demonstr. Pron.	Ich finde die Brötchen nicht – haben wir **die** möglicherweise vergessen? *I cannot find the rolls – might we have forgotten **those**?*	040
	Rel. Pron.	Da sind ein paar Noten auf dem Klavier, **die** möglicherweise dir gehören. *There are some notes on the piano **which** might belong to you.*	101
	Def. Art.	Nach dem Umzug werden wir **die** Möbel vollkommen umstellen. *After the move we are going to completely re-arrange **the** furniture.*	095
2	Demonstr. Pron.	Aus den Äpfeln würde er Kompott machen – zwar wollte der Koch **die** ursprünglich für einen Kuchen verwenden, doch das Mehl war ausgegangen. *He would make stew out of the apples – the cook originally wanted to use **those** for a cake, but there was no flour left.*	067
	Rel. Pron.	Er besaß die Fische immer noch, **die** ursprünglich seiner Schwester gehört hatten. *He still owned the fishes **which** had originally belonged to his sister.*	085
	Def. Art.	Die Computer springen von selbst auf Sommerzeit um – ich muss nur noch **die** Uhren umstellen. *The computers switch to daylight-saving time automatically – I only have to re-set **the** clocks.*	091
3	Demonstr. Pron.	Eine exakte Aussprache ist wichtig für Sänger von Opernarien – **die** artikulieren jede Silbe überdeutlich. *An exact pronunciation is important for singers of operatic arias – **those** (singers) exaggerate their articulation of every syllable.*	062
	Rel. Pron.	Panik wurde ausgelöst durch Bären, **die** arglos in die Dörfer eindrangen, um Mülltonnen zu durchstöbern. *Panic was caused by bears **which** naively entered the villages to search through garbage cans.*	027
	Def. Art.	Der Lehrer schaute sich in den Ferien **die** Arbeiten durch. *During the holidays, the teacher looked through **the** exams.*	049

Table B5: Sentences for "das" – [das] (neuter singular nominative/accusative)

Context	Lexical Class	Sentence+Translation	Position
1	Demonstr. Pron.	Es handelte sich um ein Piratenschiff – er erkannte **das** hauptsächlich an der schwarzen Flagge mit dem Totenkopf. *It was a pirate ship – he recognized **that** mainly from the black flag with the skull and crossbones.*	042
	Rel. Pron.	Sie wusste, dass sie dem Tier nicht zu nahe kommen sollte, **das** hauptsächlich wegen seines Stachels gefürchtet war. *She knew that she should not come too close to the animal **which** was mainly feared because of its sting.*	014
	Def. Art.	Die Polizei konnte **das** Haus, aber nicht den Garten umstellen. *The police managed to surround **the** house but not the garden.*	086

2	Demonstr. Pron.	Egal, wie viele Luftballons der Junge hat, kann er damit nicht fliegen – er wird **das** nur aus Erfahrung lernen können. *Regardless of how many balloons the boy has, he won't be able to fly with them – he will learn **that** only through experience.*	012
	Rel. Pron.	Das ist die Fliege, die um das Flusspferd schwirrt, **das** nur seine Ruhe haben möchte. *That is the fly which is buzzing around the hippopotamus **that** only wants to be left in peace.*	053
	Def. Art.	So sehr ich auch aufpasse – immer wird **das** Nudelwasser ein wenig überlaufen. *As careful as I am – **the** pasta water will always spill over a bit.*	016
3	Demonstr. Pron.	Der Junge lässt ein großes Blatt von der Brücke fallen, um **das** flussabwärts treiben zu sehen. *The boy drops a large leaf from the bridge in order to see **that** (leaf) float downstream.*	018
	Rel. Pron.	In dem Haus geht ein Gespenst um, **das** fluchtartig davonhuscht, sobald sich ein Mensch in der Nähe blicken lässt. *The house is haunted by a ghost **that** hastily disappears as soon as a human being approaches.*	024
	Def. Art.	Das ist die Fliege, die um **das** Flusspferd schwirrt, das nur seine Ruhe haben möchte. *That is the fly which is buzzing around **the** hippopotamus that only wants to be left in peace.*	053

Table B6: Sentences for "dem (m)" – [deːm] (masculine singular dative)

Context	Lexical Class	Sentence+Translation	Position
1	Demonstr. Pron.	Sie hatte ihren Fahrradschlüssel verlegt und wollte sich hier nach **dem** umsehen. *She lost her bicycle key and wanted to look for **that** (key) here.*	104
	Rel. Pron.	Der Junge beschrieb der Polizei seinen Entführer, nach **dem** umgehend eine Großfahndung eingeleitet wurde. *The boy gave the police a description of the kidnapper, after **whom** a man hunt was immediately organized.*	055
	Def. Art.	Nach **dem** Umzug werden wir die Möbel vollkommen umstellen. *After **the** move we are going to completely re-arrange the furniture.*	095
2	Demonstr. Pron.	In der Ferne ist ein Kahn zu sehen – auf **dem** weht eine amerikanische Fahne. *In the distance you can see a boat – on **that** (boat) an American flag is waving.*	002
	Rel. Pron.	Er lief den Strand entlang, auf **dem** wegen des anhaltenden Regens kaum jemand zu sehen war. *He ran along the beach, on **which**, due to the continuing rain, hardly anybody could be seen.*	102
	Def. Art.	Auf **dem** Weg nach Hause fuhr der Mann die Katze um. *On **the** way home the man ran over the cat.*	078
3	Demonstr. Pron.	Er hatte sich für seine Forschungsergebnisse einen Safe gekauft – niemand ahnte, dass er in **dem** andere Dinge aufbewahrte als Bargeld. *He had bought a safe for his research results – no one suspected that he kept other things in **that** (safe) than cash.*	071

	Rel. Pron.	Es gab einen gesonderten Karton für Pralinen und einen kleinen Korb, in **dem** andere Süßigkeiten aufbewahrt wurden. *There was a separate box for chocolates and a small basket in **which** other sweets were kept.*	031
	Def. Art.	Streichen Sie in **dem** Antrag bitte alle nicht ausgefüllten Felder durch. *Please cross out any fields you left blank in **the** application.*	074

Table B7: Sentences for "dem (n)" – [deːm] (neuter singular dative)

Context	**Lexical Class**	**Sentence+Translation**	**Position**
1	Demonstr. Pron.	Sie hatte sich ein gelbes Notizbuch gekauft und wollte in **dem** Rezepte aufschreiben. *She had bought a yellow notebook and wanted to write down recipes in **that** (notebook).*	038
	Rel. Pron.	Die Apotheke hat ein abgetrenntes Zimmer, in **dem** rezeptpflichtige Arzneimittel aufbewahrt werden. *The pharmacy has a separate room, in **which** prescription drugs are stored.*	004
	Def. Art.	Sie hatte in **dem** Rezept aus Versehen Backpulver und Zucker unterschlagen. *In **the** recipe, she had inadvertently left out the baking powder and the sugar.*	093
2	Demonstr. Pron.	Der Mann lief aus der Bank zu seinem Auto und fuhr in **dem** fort, ohne den zweiten Wagen zu bemerken, der ihm unauffällig folgte. *The man ran from the bank to his car and drove away in **that** (car) without noticing the second car which followed him inconspicuously.*	084
	Rel. Pron.	Hier ist das Delphinbecken, in **dem** Forschungen zum Verhalten von Meeressäugern durchgeführt werden. *Here is the dolphin tank, in **which** research is conducted on the behavior of marine mammals.*	094
	Def. Art.	In **dem** Formular muss man alle leer gebliebenen Felder durchstreichen. *On **the** form you should cross out any fields you left blank.*	052
3	Demonstr. Pron.	Von der Verpflegung her ist unser Kaninchen sehr unkompliziert – wir geben **dem** hauptsächlich Gras und Heu. *Our rabbit is very uncomplicated as far as its diet is concerned – we mainly give **that** (rabbit) grass and hay.*	097
	Rel. Pron.	Auf dem Grundstück, neben **dem** haushoch ein Baum steht, ist ein Spielplatz. *On the property, alongside **which** there is a tree as high as a house, there is a playground.*	092
	Def. Art.	Es gibt eine Hütte neben **dem** Haus, vor der Fahrräder abgestellt werden können. *There is a shed alongside **the** house, in front of which bicycles can be stored.*	080

Table B8: Sentences for "den" – [de:n] (masculine singular accusative)

Context	**Lexical Class**	**Sentence+Translation**	**Position**
1	Demonstr. Pron.	Der Hund hatte den Hasen schon so oft weglaufen sehen - er wollte **den** endlich einmal fangen. *The dog had seen the hare run away so often – he finally wanted to catch **that** (hare).*	077
	Rel. Pron.	Es war deutlich, dass der Fuchs den See beobachtete, **den** Enten als ihre Heimat gewählt hatten. *It was clear that the fox was watching the lake **which** ducks had chosen as their home.*	028
	Def. Art.	Die Polizei konnte **den** Entführer vollkommen umstellen. *The police managed to completely surround **the** kidnapper.*	059
2	Demonstr. Pron.	Das Auto hat einen großen Kofferraum - in **den** geht eine ganze Menge hinein. *The car has a large luggage compartment – in **that** (compartment) you can fit a great deal.*	025
	Rel. Pron.	Zum Burgturm fährt in regelmäßigen Abständen ein kleiner Bus, in **den** Gehfaule steigen können. *At regular intervals, a small bus drives to the castle tower, into **which** those who are tired of walking can climb.*	064
	Def. Art.	Der Mann war in **den** Gegenverkehr geraten, als er die Katze umfahren wollte. *The man swerved into **the** oncoming traffic as he tried to drive around the cat.*	047
3	Demonstr. Pron.	Der neue Gasherd war ihr immer noch suspekt und sie versäumte nicht, **den** grundsätzlich auszumachen, sobald sie die Küche verließ. *She was still wary of the new gas oven and did not fail to turn **that** (oven) off as a matter of principle whenever as she left the kitchen.*	009
	Rel. Pron.	Er mochte den Lärm nicht, **den** grunzende Ferkel verursachen. *He did not like the noise **which** grunting piglets made.*	020
	Def. Art.	Der Junge ahnte, dass sein Vater ihn getäuscht hatte, konnte aber nicht **den** Grund für seine Lügen durchschauen. *The boy suspected that his father had tricked him but couldn't figure out **the** reason for his lies.*	083

11.3 Appendix C: Stimuli for Experiment 3

This appendix contains the carrier sentences for the production experiment on effects of syllable frequency (see Chapter 7). Sentences are sorted according the context in which the target syllable appeared. The English translations that are given for each sentence are meant to help understand the design of the target items and the carrier sentences and therefore tend to be literal rather than idiomatic. Each sentence includes a capital X as a placeholder for the fictive surname containing the target syllable.

Table C1: Sentences used for participant group 1 (participants 1-4, 9-12, 17-20, and 25-28)

Context	Sentence+Translation
1	Sie können Frau X am kommenden Montag sprechen. *You can talk to Ms X this coming Monday.*
2	Morgen Nachmittag ist Herr X bis 17:30 Uhr im Haus. *Tomorrow afternoon Mr. X will be in the building until 5:30 p.m.*
3	Ich werde Herrn X erst nächsten Freitag sehen. *I won't be seing Mr. X until next Friday.*
4	Heute Vormittag ist Frau X ab 8:30 Uhr im Haus. *This morning Ms X will be in the building from 8:30 a.m. onwards.*

Table C2: Sentences used for participant group 2 (participants 5-8, 13-16, 21-24, and 29-32)

Context	Sentence+Translation
1	Ich leite Sie an Frau X von der Abteilung 5C weiter. *I'll put you through to Ms X from Department 5C.*
2	Ich bedaure, Herr X ist leider nur bis 16:00 Uhr zu erreichen. *I'm sorry, Mr. X can only be reached until 4 p.m.*
3	Wenden Sie sich bitte an Herrn X von der Abteilung D8. *Please contact Mr. X from Department D8.*
4	Ich fürchte, Frau X ist heute erst ab 10:30 im Haus. *I'm afraid Ms X won't be in the building today until 10:30 a.m.*

11.4 Appendix D: Distractor Sentences

This appendix contains the three initial sentences used for Experiments 1 and 2 (see Chapter 5) as well as the distractor sentences used for Experiment 3 (see Chapter 7). Sentences are sorted according to their order of appearance in the experiments. The English translations that are given for each sentence are meant to help understand the design of the target items and the carrier sentences and therefore tend to be literal rather than idiomatic. The table for Experiment 3 includes information about the position of each of the distractor sentences within the experiment.

Table D1: Opening sentences for Experiments 1 and 2

Sentence + Translation
Dies ist der verrückte Hund, der von Katzen den Baum hochgejagt wird. *This is the crazy dog which is being chased up a tree by cats.*
Diese Pilze sind hochgiftig – die kommen in die Suppe, wenn du weiter so nervst. *These mushrooms are highly poisonous – they will land in the soup if you keep annoying me.*
Der Roboter ist eine große Hilfe im Haushalt und kann sogar bügeln. *The robot is a great help with housework and is can even do the ironing.*

Table D2: Distractor sentences for Experiment 3

Sentence + Translation	**Position**
Guten Tag, hier Firma Meisel & Co., Rezeption, was kann ich für Sie tun? *Good morning, this is the receptionist at Meisel & Co., how can I help you?*	001
Ich verbinde Sie mit Frau Hullfer von der Abteilung 7A. *I'll put you through to Ms Hullfer from Department 7A.*	002
Ich fürchte, die Abteilung 22C ist bis Ende nächster Woche geschlossen. *I'm afraid Department 22C is closed until the end of next week.*	003
Am kommenden Mittwoch können Sie um 9:30 Uhr vorbeikommen. *This coming Wednesday you can come in at 9:30 a.m.*	005
Bis zum 08.06. ist das Gebäude wegen Pfingstferien geschlossen. *Until June 8th the building is closed for the Pentecost holidays.*	007
Der Raum A42 befindet sich im 4. Stock. *Room A42 is on the 5th floor.*	009
Einen Moment bitte – ich verbinde Sie mit dem Management. *One moment please – I'll transfer you to the management department.*	011
Anfang April werden die Bauarbeiten voraussichtlich beendet sein. *By the beginning of April the construction work is expected to be finished.*	013
Sie erhalten den Passierschein A38 in der Geschäftsstelle. *You can obtain Permit A38 at the business office.*	015
Über die Feiertage ist das Sekretariat geschlossen. *Over the holidays the secretariat is closed.*	017
Wir bedanken uns für Ihr Verständnis. *We thank you for your understanding.*	019
Am Dienstagmorgen ist die Bibliothek von 8:30-10:30 Uhr geöffnet. *On Tuesday mornings the library is open from 8:30-10:30 a.m.*	021
Bis zum 30.01. sind wir leider vollkommen ausgebucht. *Unfortunately, we're completely booked out until January 30th.*	023
Der Raum H23 ist im 3. Stock des Altbaus. *Room H23 is on the 4th floor of the old building.*	025
Sie werden in wenigen Minuten mit der Buchhaltung verbunden. *You will be transferred to the accounts department in a few minutes.*	027
Im März sind Osterferien. *The Easter holidays are in March.*	029
Den Antrag G25 bekommen Sie im Sekretariat. *You can obtain the application G25 in the secretariat.*	031
Unsere Abteilung ist über Ostern geschlossen. *Our department is closed over Easter.*	033
Wir sind momentan leider nur eingeschränkt zu erreichen. *Unfortunately, we are only reachable at a limited basis at the moment.*	035
Am nächsten Samstag werden in der Abteilung B57 die Heizkörper gewartet. *Next Saturday, the radiators in Department B57 are being serviced.*	037
Bis zum 15.10. finden keine Sprechstunden statt. *There won't be any consultation hours until October 15th.*	039
Den Raum P17 finden Sie im 2. Stock des Neubaus. *You'll find Room P17 on the 3rd floor of the new building.*	041

Warten Sie einen Augenblick – ich verbinde Sie mit der Zentrale. *Just a second – I'll tranfer you to the main office.*	043
Ende September zieht der Gebäudeteil K4 um. *At the end of September section K4 of the building will be moving.*	045
Sie benötigen noch das Formular Z72 aus der Verwaltung. *You still need Form Z72 from the administration office.*	047
Das Gebäude 33D ist den nächsten Monat über geschlossen. *Building 33 D is closed until the end of next month.*	049
Wir können Ihre Anfrage zur Zeit nicht bearbeiten. *At the moment we cannot process your inquiry.*	051
Am Donnerstag ist die Bibliothek nachmittags geschlossen. *On Thursdays the library is closed in the afternoon.*	053
Bis zum 23.08. werden die Reparaturen voraussichtlich beendet sein. *Repairs are expected to be finished by August 23rd.*	055
Den Raum 21A können Sie im 1. Stock finden. *You'll find Room 21A on the 2nd floor.*	057
In wenigen Minuten werden Sie mit der Finanzabteilung verbunden. *In a few minutes you will be transferred to the finance department.*	059
Mitte Januar beginnt der Umzug der Abteilung B3. *The move of department B3 will start in mid January.*	061
Von der Buchhaltung lassen Sie sich bitte die Bescheinigung M12 ausstellen. *Please have the accounting department issue you Certificate M12.*	063
Wir haben über die Ostertage geschlossen. *We are closed over the Easter holidays.*	065
Leider sind wir zur Zeit nicht zu erreichen – bitte hinterlassen Sie eine Nachricht. *Unfortunately, we cannot be reached at the moment – please leave a message.*	067
Am kommenden Mittwoch können Sie um 13:40 Uhr vorbeikommen. *This coming Wednesday you can come in at 1:40 p.m.*	069
Bis zum 03.09. ist das Gebäude wegen Sommerferien geschlossen. *Until September 3rd the building is closed for the summer holidays.*	071
Der Raum B11 befindet sich im 1. Stock. *Room B11 is on the 2nd floor.*	073
Einen Moment bitte – ich verbinde Sie mit der Buchhaltung. *One moment please – I'll tranfer you to the accounting department.*	075
Anfang März wird der Umzug voraussichtlich beendet sein. *The move is expected to be finished by the beginning of March.*	077
Sie erhalten den Passierschein B35 im Sekretariat. *You can obtain Permit B35 from the secreatariat.*	079
Über die Feiertage ist die Bibliothek geschlossen. *The library is closed over the holidays.*	081
Wir bitten Sie um Verständnis für die Verzögerung. *We ask for your understanding for the delay.*	083
Am Dienstagmorgen ist das Sekretariat von 10:30-12:30 Uhr geöffnet. *On Tuesday mornings the secretariat is open from 10:30-12:30.*	085
Bis zum 31.05. sind wir leider vollkommen ausgebucht. *Unfortunately, we are completely booked out until May 31st.*	087
Der Raum F14 ist im 4. Stock des Neubaus. *Room F14 is on the 5th floor of the new building.*	089

Sie werden in wenigen Minuten mit der Zentrale verbunden. *You will be tranferred to the main office in a few minutes.*	091
Im Dezember sind Weihnachtsferien. *The Christmas holidays are in December.*	093
Den Antrag H44 bekommen Sie bei der Verwaltung. *You can obtain the application H44 from the administrative department.*	095
Unsere Abteilung ist über Pfingsten geschlossen. *Our department is closed over Pentecost.*	097
Wir sind montags bis freitags telefonisch von 9-17 Uhr zu erreichen. *Monday through Friday we can be reached by telephone from 9 a.m. to 5 p.m.*	099
Am nächsten Samstag werden in der Abteilung C42 die Fenster geputzt. *Next Saturday the windows are being cleaned in Department C42.*	101
Bis zum 12.11. finden keine Sprechstunden statt. *There won't be any consultation hours until November 12th.*	103
Den Raum R44 finden Sie im 3. Stock des Neubaus. *You'll Room R44 on the 4th floor of the new building.*	105
Warten Sie einen Augenblick – ich verbinde Sie mit der Finanzabteilung. *Just a second – I'll transfer you to the finance department.*	107
Ende August zieht der Gebäudeteil M3 um *At the end of August section M3 of the building will be moving.*	109
Sie benötigen noch das Formular W32 von der Buchhaltung. *You still need Form W32 from the accounting department.*	111
Das Gebäude 22C ist die nächste Woche über geschlossen. *The building 22C is closed the next week.*	113
Wir können Ihnen zur Zeit keine Termine anbieten. *At the moment we cannot offer you any appointments.*	115
Am Donnerstag ist das Sekretariat nachmittags geschlossen. *On Thursdays the secretariat is closed in the afternoon.*	117
Bis zum 25.02. werden die Reparaturen voraussichtlich beendet sein. *The repairs are expected to be finished by February 25th.*	119
Den Raum 32B können Sie im 2. Stock finden. *You'll find Room 32B on the 3rd floor.*	121
In wenigen Minuten werden Sie mit dem Management verbunden. *In a few minutes you will be transferred to the management department.*	123
Mitte Februar beginnt die Renovierung der Abteilung A7. *The renovation of Department A7 will start in mid February.*	125
Von der Geschäftsstelle lassen Sie sich bitte die Bescheinigung L13 ausstellen. *Please have the business office issue you Certificate L13.*	127
Wir haben über die Weihnachtstage geschlossen. *We are closed over the Christmas holidays.*	129
Leider sind im Moment alle Telefone besetzt – bitte haben Sie etwas Geduld. *Unfortunately, all telephones are busy at the moment – please be patient.*	131
Am kommenden Mittwoch können Sie um 10:20 Uhr vorbeikommen. *This coming Wednesday you can come in at 10:20 a.m.*	133
Bis zum 07.01. ist das Gebäude wegen Weihnachtsferien geschlossen. *The building is closed for the Christmas holidays until January 7th.*	135
Der Raum C20 befindet sich im 2. Stock. *Room C20 is on the 3rd floor.*	137

Einen Moment bitte – ich verbinde Sie mit der Zentrale. *One moment please – I'll transfer you to the central office.*	139
Anfang Juni wird die Renovierung voraussichtlich beendet sein. *The renovation is expected to be finished by the beginning of June.*	141
Sie erhalten den Passierschein C37 bei der Verwaltung. *You can obtain Permit C37 from the administration department.*	143
Über die Feiertage ist die Personalabteilung geschlossen. *Over the holidays the personnel department is closed.*	145
Wir danken Ihnen für Ihr Verständnis. *We thank you for your understanding.*	147
Am Dienstagmorgen ist die Verwaltung von 9:30-11:30 Uhr geöffnet. *On Tuesday mornings the administration department is open from 9:30-11:30.*	149
Bis zum 31.03. sind wir leider vollkommen ausgebucht. *Unfortunately, we're completely booked out until March 31st.*	151
Der Raum G62 ist im 1. Stock des Altbaus. *Room G62 is on the 2nd floor of the old building.*	153
Sie werden in wenigen Minuten mit der Finanzabteilung verbunden. *You will be tranferred to the finance department in a few minutes.*	155
Im Juli sind Sommerferien. *Summer holidays are in July.*	157
Den Antrag K12 bekommen Sie bei der Buchhaltung. *You can obtain the application K12 from the accounting department.*	159
Unsere Abteilung ist über Weihnachten geschlossen. *Our department is closed over Christmas.*	161
Wir sind über Telefon, Email, oder auch per Post für Sie zu erreichen. *You can reach us via telephone, email, or mail.*	163
Am nächsten Samstag werden in der Abteilung A23 die Wasserleitungen überprüft. *Next Sturday the water pipes in Department A23 are being serviced.*	165
Bis zum 13.12. finden keine Sprechstunden statt. *There won't be any consoltation hours until December 13th.*	167
Den Raum S32 finden Sie im 4. Stock des Altbaus. *You'll find Room S32 on the 5th floor of the old building.*	169
Warten Sie einen Augenblick – ich verbinde Sie mit dem Management. *Just a second – I'll tranfer you to the management department.*	171
Ende November zieht der Gebäudeteil Q1 um. *At the end of November Section Q1 of building will be moving.*	173
Sie benötigen noch das Formular V42 von der Geschäftsstelle. *You still need Form V42 from the business office.*	175
Das Gebäude 77B ist die nächsten Tage über geschlossen. *Building 77B will remain closed during the next few days.*	177
Wir können Ihren Anruf zur Zeit nicht entgegennehmen. *We cannot answer your call at the moment.*	179
Am Donnerstag ist die Geschäftsstelle nachmittags geschlossen. *On Thursdays the business office is closed in the afternoon.*	181
Bis zum 26.07. werden die Reparaturen voraussichtlich beendet sein. *The repairs are expected to be finished by July 26th.*	183
Den Raum 43C können Sie im 3. Stock finden. *You'll find Room 43C on the 4th floor.*	185

In wenigen Minuten werden Sie mit der Buchhaltung verbunden. *In a few minutes you will be transferred to the accounting department.*	187
Mitte Oktober beginnt der Umbau der Abteilung C4. *The reconstruction of Department C4 will start in mid October.*	189
Vom Sekretariat lassen Sie sich bitte die Bescheinigung N18 ausstellen. *Please have the secretariat issue you Certificate N18.*	191
Wir haben über die Karnevalstage geschlossen. *We are closed over the carnival days.*	193
Leider rufen Sie außerhalb unserer Geschäftszeiten an – bitte versuchen Sie es später wieder. *Unfortunately, you are calling outside of our office hours – please try again later.*	195
Am kommenden Mittwoch können Sie um 15:10 Uhr vorbeikommen. *This coming Wednesday you can come in at 15:10 p.m.*	197
Bis zum 05.04. ist das Gebäude wegen Osterferien geschlossen. *The building is closed for the Easter holidays until April 5th.*	199
Der Raum D31 befindet sich im 3. Stock. *Room D31 is on the 4th floor.*	201
Einen Moment bitte – ich verbinde Sie mit der Finanzabteilung. *One moment please – I'll transfer you to the finance department.*	203
Anfang Juli wird der Umbau voraussichtlich beendet sein. *The reconstruction is expected to be finished by the beginning of July.*	205
Sie erhalten den Passierschein D36 in der Buchhaltung. *You can obtain Permit D36 from the accounting department.*	207
Über die Feiertage ist die Finanzabteilung geschlossen. *The finance department is closed over the holidays.*	209
Wir entschuldigen uns für die Verzögerung. *We're sorry for the delay.*	211
Am Dienstagmorgen ist die Personalabteilung von 7:30-10:30 Uhr geöffnet. *Tuesday mornings the personnel department is open from 1:30-10:30.*	213
Bis zum 30.04. sind wir leider vollkommen ausgebucht. *Unfortunately, we are completely booked out until April 30th.*	215
Der Raum K77 ist im 2. Stock des Altbaus. *Room K77 is on the 3rd floor of the old building.*	217
Sie werden in wenigen Minuten mit dem Management verbunden. *You will be transferred to the management department in a few minutes.*	219
Im Oktober sind Herbstferien. *Fall holidays are in October.*	221
Den Antrag L81 bekommen Sie in der Geschäftsstelle. *You can obtain the application L81 from the business office.*	223
Unsere Abteilung ist über Karneval geschlossen. *Our department is closed over carnival.*	225
Wir sind auch über unsere Web-Adresse zu erreichen. *We can also be reached via our website.*	227
Am nächsten Samstag werden in der Abteilung D31 die Parkettböden gereinigt. *Next Saturday the parquet floors of Department D31 are being cleaned.*	229
Bis zum 11.10. finden keine Sprechstunden statt. *There won't be any consultation hours until October 11th.*	231
Den Raum U21 finden Sie im 1. Stock des Altbaus. *You'll find Room U21 on the 2nd floor of the old building.*	233

Warten Sie einen Augenblick – ich verbinde Sie mit der Buchhaltung. *Just a second – I'll transfer you to the accounting department.*	235
Ende Dezember zieht der Gebäudeteil L2 um. *At the end of December Section L2 of the building will be moving.*	237
Sie benötigen noch das Formular X22 vom Sekretariat. *You still need Form X22 from the secretariat.*	239
Das Gebäude 55A ist die nächste Zeit über geschlossen. *Building 55A is temporarily closed.*	241
Wir können Ihnen da leider nicht weiterhelfen. *I'm afraid we cannot help you in that matter.*	243
Am Donnerstag ist die Personalabteilung nachmittags geschlossen. *On Thursdays the personnel department is closed in the afternooon.*	245
Bis zum 28.11. werden die Reparaturen voraussichtlich beendet sein. *The repairs are expected to be finished by November 28th.*	247
Den Raum 54D können Sie im 4. Stock finden. *You'll find Room 54D on the 5th floor.*	249
In wenigen Minuten werden Sie mit der Zentrale verbunden. *In a few minutes you will be transferred to the central office.*	251
Mitte Juli beginnen die Bauarbeiten an der Abteilung D5. *Construction works on Department D5 will start mid July.*	253
Von der Verwaltung lassen Sie sich bitte die Bescheinigung P19 ausstellen. *Please have the administration department issue you Certificate P19.*	255
Wir haben über die Pfingsttage geschlossen. *We are closed over the Pentecost holidays.*	257
Leider haben wir jetzt Dienstschluss – bitte rufen Sie doch morgen wieder an. *I'm afraid our office hours are over for today – please call back tomorrow.*	259

11.5 Appendix E: Example Illustrations (Experiment 1 – "unterstellen")

This appendix contains the illustrations used for one of the target verbs in the production experiment on effects of word stress, sentence stress, and syntactic boundaries (see Chapter 5). For each of the illustrations, the accompanying sentence is given as well as (literal rather than idiomatic) translation into English.

Figure E1: Illustration for Sentence 1 (w+s+) – "Wir wollten uns unterstellen, weil es so stark regnet." (We wanted to take shelter because it is raining so hard.)

Figure E2: Illustration for Sentence 2 (w+s-) – "Sie können nicht nur Fahrräder, sondern auch Autos bei uns unterstellen." (You can store not only bicycles but also cars with us.)

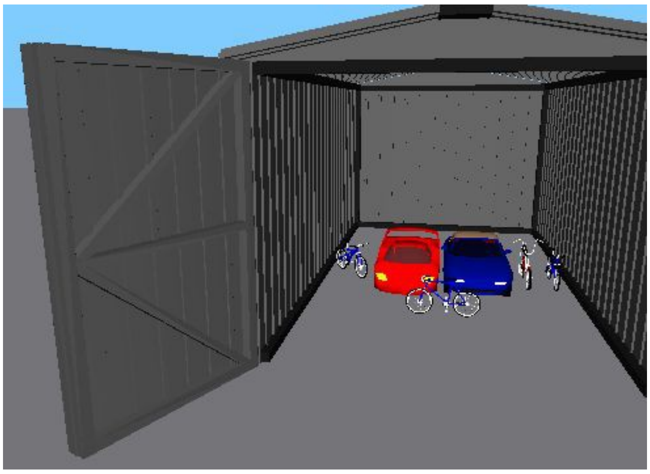


Figure E3: Illustration for Sentence 3 (w-s+) – "Der Kellner wollte uns unterstellen, dass wir nicht bezahlt hätten." (The waiter wanted to insinuate that we hadn't paid.)

Figure E4: Illustration for Sentence 4 (w-s-) – "Tierquälerei haben sie uns unterstellt – nur weil unser Hund ihre Katze auf den Baum gejagt hat." (They accused us of animal cruelty – only because our dog had chased their cat up the tree.)

Figure E5: Illustration for Sentence 5 (sb) – "Da es gerade stark regnete, stellten wir uns unter." (Because it was raining hard we took shelter.)

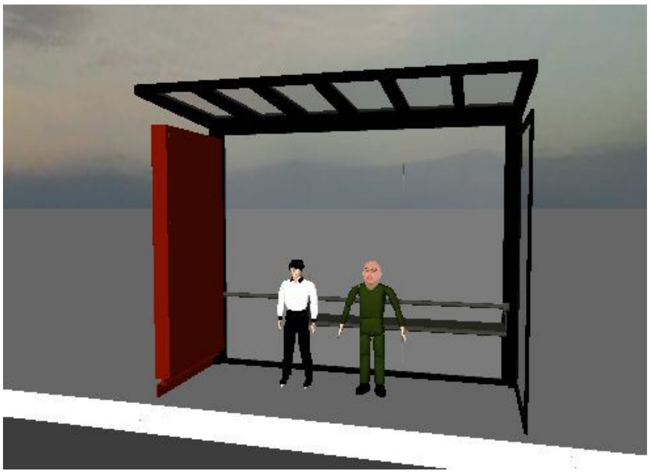
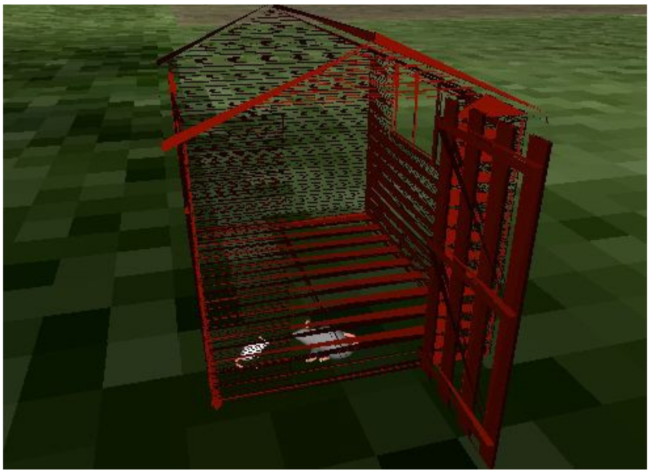


Figure E6: Illustration for Sentence 6 (mb) – "Man kann unterstellen, dass Ratten von vielen als Ungeziefer gesehen werden." (One can assume that rats are commonly seen as vermin.)

Figure E7: Illustration for Sentence 7 (wb) – "Man kann unter Ställen oft Mäuse und Ratten finden." (One can often find mice and rats under sheds.)



11.6 Appendix F: Prominence Tables

This appendix contains tables with mean values of prominence, force accent, and pitch accent estimates for the three production experiments (see Chapters 5 to 7).

Table F1: Mean estimates of prominence (above), force accent (middle), and pitch accent (below) for Experiment 1 in each of the sentence categories

Category	[ʔʊm]	[ʔʊn]	[tɐ]	[?y]	[bɐ]	[dʊʁç]	All
"w+s+"	0.342 0.330 0.114	0.378 0.277 0.322	0.558 0.575 0.100	0.401 0.407 0.088	0.907 0.955 0.120	0.570 0.584 0.109	0.529 0.524 0.142
"w+s-"	0.295 0.283 0.103	0.348 0.338 0.109	0.484 0.494 0.100	0.288 0.301 0.043	0.739 0.769 0.116	0.296 0.288 0.091	0.407 0.411 0.094
"w-s+"	0.217 0.193 0.109	0.236 0.218 0.100	0.353 0.357 0.077	0.188 0.189 0.045	0.782 0.788 0.180	0.374 0.366 0.113	0.359 0.352 0.104
"w-s-"	0.127 0.119 0.050	0.236 0.236 0.059	0.400 0.408 0.083	0.195 0.212 0.010	0.756 0.769 0.159	0.452 0.453 0.110	0.359 0.364 0.078
"sb"	0.844 0.850 0.198	0.632 0.631 0.160	0.629 0.663 0.080	0.588 0.591 0.139	0.915 0.953 0.144	0.606 0.637 0.081	0.703 0.721 0.134
"mb"	0.143 0.127 0.072	0.302 0.286 0.112	0.368 0.391 0.041	0.164 0.155 0.060	0.659 0.671 0.137	0.247 0.262 0.027	0.319 0.321 0.075
"wb"	0.135 0.126 0.053	0.228 0.220 0.077	0.348 0.349 0.085	0.213 0.206 0.068	0.636 0.675 0.070	0.260 0.284 0.012	0.303 0.310 0.060

Table F2: Mean estimates of prominence (above), force accent (middle), and pitch accent (below) for Experiment 2 in each of the lexical classes

Lexical Class	der (masc.)	der (fem.)	die (sg.)	die (pl.)	das	dem (masc.)	dem (neut.)	den	All
"dp"	0.765 0.789 0.137	1.010 0.938 0.416	0.529 0.505 0.186	0.471 0.487 0.081	0.681 0.734 0.050	0.757 0.680 0.362	0.762 0.687 0.358	0.761 0.714 0.294	0.722 0.696 0.239
"rp"	0.351 0.371 0.042	0.692 0.704 0.145	0.240 0.253 0.032	0.190 0.195 0.036	0.445 0.460 0.077	0.655 0.628 0.224	0.706 0.699 0.191	0.593 0.590 0.154	0.479 0.483 0.110
"da"	0.567 0.607 0.051	0.562 0.611 0.029	0.442 0.468 0.052	0.327 0.331 0.072	0.635 0.661 0.100	0.266 0.269 0.059	0.654 0.705 0.049	0.564 0.604 0.050	0.501 0.530 0.058

Table F3: Mean estimates of prominence (above), force accent (middle), and pitch accent (below) in Experiment 3 for frequent and rare syllables in each quadruple

Quadruple	1	2	3	4	5	6	7	8	All
Frequent syllables	0.552 0.373 0.542	0.749 0.638 0.436	0.803 0.649 0.548	1.122 0.994 0.567	1.110 0.975 0.581	0.818 0.722 0.421	0.950 0.858 0.445	1.039 0.932 0.500	0.891 0.766 0.505
Rare syllables	0.550 0.405 0.465	0.724 0.611 0.434	0.752 0.598 0.534	1.105 0.985 0.547	1.095 0.964 0.568	0.873 0.776 0.436	0.970 0.884 0.437	1.061 0.928 0.565	0.892 0.770 0.499

11.7 Appendix G: Variability across Contexts (Experiment 3)

For the frequent and rare syllables in Experiment 3 (see Chapter 7), this appendix contains mean values for absolute duration and prominence differences across contexts as well as mean values for the cross-context spectral similarity.

Table G1: Mean variability of duration values in milliseconds for syllables (above) and vowels (below) by quadruple identity and frequency level

Quadruple	1	2	3	4	5	6	7	8	All
Frequent syllables	29.5 12.3	25.4 14.0	24.5 14.5	32.7 17.4	32.9 22.5	30.0 10.1	27.3 17.7	38.0 20.0	30.0 16.0
Rare syllables	30.3 15.0	29.1 14.5	29.4 13.3	38.0 22.5	32.5 26.2	24.4 13.1	27.9 17.1	37.3 24.2	31.2 18.3

Table G2: Mean variability of prominence values (above), force accent values (middle) and pitch accent values (below) by quadruple and frequency level

Quadruple	1	2	3	4	5	6	7	8	All
Frequent syllables	0.244 0.174 0.355	0.288 0.248 0.303	0.263 0.207 0.371	0.150 0.010 0.361	0.170 0.047 0.359	0.270 0.219 0.377	0.245 0.168 0.420	0.177 0.086 0.305	0.226 0.146 0.357
Rare syllables	0.260 0.216 0.260	0.279 0.233 0.383	0.266 0.231 0.317	0.146 0.030 0.326	0.186 0.055 0.400	0.261 0.219 0.296	0.224 0.140 0.365	0.283 0.092 0.344	0.224 0.150 0.336

Table G3: Mean estimates of spectral similarity for syllables (above), onsets (second from above), vowels (second from below), and codas (below) by quadruple and frequency level

Quadruple	1	2	3	4	5	6	7	8	All
Frequent syllables	0.812 0.786 0.838 0.867	0.821 0.829 0.847 0.869	0.877 0.773 0.837 0.851	0.873 0.829 0.888 0.829	0.892 0.811 0.886 0.793	0.831 0.826 0.861 0.854	0.895 0.813 0.857 0.849	0.856 0.802 0.888 0.761	0.857 0.808 0.862 0.835
Rare syllables	0.867 0.812 0.863 0.878	0.796 0.790 0.833 0.860	0.878 0.755 0.835 0.843	0.869 0.826 0.872 0.828	0.885 0.794 0.880 0.787	0.844 0.814 0.854 0.840	0.896 0.808 0.854 0.843	0.860 0.790 0.886 0.775	0.863 0.799 0.860 0.832