Reduction in Natural Speech

Inauguraldissertation

zur Erlangung des Grades eines Doktors der Philosophie
im Fachbereich Sprach- und Kulturwissenschaften
der Johann Wolfgang Goethe-Universität
zu Frankfurt am Main

Vorgelegt von Frank Zimmerer aus Konstanz

2008
(Einreichungsjahr)
2009
(Erscheinungsjahr)

Tag der Disputation: 03. August, 2009
Gutachter:
Prof. Dr. Reetz, Prof. Dr. Gippert, Prof. Dr. Dr. Lahiri

Reduction in Natural Speech

Inauguraldissertation von
Frank Zimmerer

Abstract

Natural (conversational) speech, compared to cannonical speech, is earmarked by the tremendous amount of variation that often leads to a massive change in pronunciation. Despite many attempts to explain and theorize the variability in conversational speech, its unique characteristics have not played a significant role in linguistic modeling. One of the reasons for variation in natural speech lies in a tendency of speakers to reduce speech, which may drastically alter the phonetic shape of words. Despite the massive loss of information due to reduction, listeners are often able to understand conversational speech even in the presence of background noise.

This dissertation investigates two reduction processes, namely regressive place assimilation across word boundaries, and massive reduction and provides novel data from the analyses of speech corpora combined with experimental results from perception studies to reach a better understanding of how humans handle natural speech. The successes and failures of two models dealing with data from natural speech are presented: The FUL-model (Featurally Underspecified Lexicon, Lahiri & Reetz, 2002), and X-MOD (an episodic model, Johnson, 1997). Based on different assumptions, both models make different predictions for the two types of reduction processes under investigation. This dissertation explores the nature and dynamics of these processes in speech production and discusses its consequences for speech perception. More specifically, data from analyses of running speech are presented investigating the amount of reduction that occurs in naturally spoken German.

Concerning production, the corpus analysis of regressive place assimilation reveals that it is not an obligatory process. At the same time, there emerges a clear asymmetry: With only very few exceptions, only [coronal] segments undergo assimilation, [labial] and [dorsal] segments usually do not. Furthermore, there seem to be cases of complete neutralization where the underlying Place of Articulation feature has undergone complete assimilation to the Place of Articulation feature of the upcoming segment. Phonetic analyses further underpin these findings. Concerning deletions and massive reductions, the results clearly indicate that phonological rules in the classical generative tradition are not able to explain the reduction patterns attested in conversational speech. Overall, the analyses of deletion and massive reduction in natural speech did not exhibit clear-cut patterns. For a more in-depth examination of reduction factors, the case of final /t/ deletion is examined by means of a new corpus constructed for this purpose. The analysis of this corpus indicates that although phonological context plays an important role on the deletion of segments (i.e. /t/), this arises in the form of tendencies, not absolute conditions. This is true for other deletion processes, too.

Concerning speech perception, a crucial part for both models under investigation (X-MOD and FUL) is how listeners handle reduced speech. Five experiments investigate the way reduced speech is perceived by human listeners. Results from two experiments show that regressive place assimilations can be treated as instances of complete neutralizations by German listeners. Concerning massively reduced words, the outcome of transcription and priming experiments suggest that such words are not acceptable candidates of the intended lexical items for listeners in the absence of their proper phrasal context.

Overall, the abstractionist FUL-model is found to be superior in explaining the data. While at first sight, X-MOD deals with the production data more readily, FUL provides a better fit for the perception results. Another important finding concerns the role of phonology and phonetics in general. The results presented in this dissertation make a strong case for models, such as FUL, where phonology and phonetics operate at different levels of the mental lexicon, rather than being integrated into one. The findings suggest that phonetic variation is not part of the representation in the mental lexicon.

Acknowledgements

This dissertation has been a constant company for a noticeable part of my life. Now that this project is completed, I feel like an important period has been mastered. Such landmarks in life are always a very good occasion for looking back, evaluate the past and, most importantly, thank the people that helped you reach this point. I would like to thank

Henning Reetz, my supervisor, who is an excellent teacher with an immense and profound knowledge not only in phonetics. He never forced me to learn from him but he made me interested to ask questions. I learned and profited a lot. Henning always gave me freedom and trust to write the dissertation my way.

Aditi Lahiri, my co-supervisor. She was the person that showed me how fascinating and important linguistics in general and phonology in particular are. Her incredible competence, her wisdom, her patience, her endurance, her ingenious ideas and also her generosity have been the founding basis for my interest in linguistics. Without her, this dissertation would not have been written.

When I started to study phonology, **Mirco Ghini** was one of the reasons to continue doing so. I have never met someone who was so enthusiastic, so full of joy for phonology than he was. Discussing linguistics with him was always fun, and extremely instructive. He will not be forgotten.

Mathias Scharinger who took the same “Intro to Phonology” as I did. On the way through the time and space of linguistics, we became colleagues and close friends. He always had time discussing linguistic and extra-linguistic topics, and commented on prior versions of this dissertation. Working with him was always fun and effective.

Verena Felder, who shared office with me and Mathias in Konstanz. The atmosphere of our office was always motivating and never boring.

Willi Nagl, who helped me with pressing questions concerning statistics, even at very short notice.

The **SFB 471** and the department of linguistics at the University of Konstanz. I am very grateful having had the opportunity to learn and work in such a stimulating linguistic biotope. These institutions were a perfect academic home for me for quite some time. I am agreeing with Frans Plank, the current speaker of the SFB, who emphasized that working in the SFB is very inspiring. Colleagues never seize to ask questions, helping to think and rethink your work, improving it every time. It really is a perfect place to do linguistics, not only in Germany.

The **Johann Wolfgang Goethe-University of Frankfurt**, the **Fachbereich 9**, and the Institut für Phonetik that became my new academic home.

The **DFG**; without this institution, neither the SFB 471, nor the current SPP 1234 in which our project participates, would exist.

Peter Cartsunis who did a great job in improving my English writing.

Claas & Heike, who made me feel welcome from the first time we talked. I always felt like at home.

Thanks to my **friends** who know me and still are my friends.

Jana, Anita & Klaus, who enlarged our family.

My parents and my sister. From the beginning of my life, they were there for me. And they still are. With trust and patience my parents let me make my decisions letting me know that they support me with all their love and strength.

Manuela Lopez who makes my life better. Every day. Thank you for her creativity, her patience, her smiles and her love. I cannot express my gratitude enough for having met her and her endurance of sharing me with linguistics and Frankfurt.

Thank you so much. I know that without you all, my life and this dissertation would not be the same.

Table of contents

«…what is out of the common is usually a guide rather than a hindrance.»
Sherlock Holmes *(Arthur Conan Doyle, A Study in Scarlet)*

Chapter 1 – **Introduction**

When humans wish to convey meaning to other humans they can speak with each other. If they speak the same language, speaking and understanding usually works quite well, and seemingly also effortless. Jokes can be made, compliments can be passed, people can flirt with each other, they can tell what happened to them during the day, or students can gossip about their professors; to mention just a few examples. For the most part, successful speaking and understanding even operate subconsciously. As well as there are different topics people talk about there are also different contexts and settings in which language is spoken. Depending on these contexts, language use differs considerably (e.g. Dressler, 1972, Zwicky, 1972). The language used by a professor in a formal lecture to honor a well-known scientist is not the same as the one she uses when chatting with two friends in a bar. Language use is also influenced by sociological factors. The geographical place and social strata where one is brought up has enormous impact on how one speaks (e.g. Labov, 1966, 2001, 2006; Trudgill, 1974, Clopper & Pierrehumbert, 2008).

Although computers (i.e. e-mailing or chatting) or (cell) phones become increasingly popular for communication between humans, it is face-to-face conversation that still is the most common, and, if one likes, the most "natural" kind of language use. Therefore, conversational speech is the speech register that speakers most often produce and consequently listeners have to deal with most of the time. Both speakers and listeners do so in a very effective way. A widely accepted assumption about conversational speech is that it is characterized by a strong tendency on the speakers' side to produce speech with as little effort as possible (e.g. Lindblom, 1990, but see Kingston, 2006). However, the most extreme possibility - giving no effort at all to the production of speech – would be fatal for speech perception. One important point is that as objective of our talking we expect listeners to react, thus during a conversation we usually "[...] speak to be heard in order to be understood" (Greenberg & Fosler-Lussier, 2000, citing Jakobson, Fant, & Halle, 1963).

Lindblom's H&H Theory (Lindblom, 1990) includes exactly this factor, i.e. speakers talk to be understood, as a counter weight to the tendency of speakers to reduce effort in speech production as much as possible. Speakers – this is another basic assumption – reduce their effort by reducing speech gestures (see also Flege, 1988, or Kohler, 1990). This leads to so-called undershoot in speech productions. For example, speakers' minimization leads to an overlay of speech gestures, some target positions of articulators are not reached, segments are reduced or deleted, and vowels become more centralized. Effort minimization as single reason for reduction is not undisputed, though. It has been called in question whether effort minimization really accounts for the observed reduction patterns. Additionally, it is not clear whether there is really much effort that is minimized in reduction compared to a more careful pronunciation (Kingston, 2006). Whether or not effort-minimization is the actual driving force for reduction in natural speech, there is ample evidence for it to occur (e.g. Johnson, 2004a). Despite these reductions, listeners usually understand what has been said, even in very noisy conditions.

Yet, from a linguistic point of view, the processes underlying successful speech perception are not completely understood. Nor are issues concerning linguistic structures in the brain allowing for speech perception. It is also not understood completely to what degree speakers include expectancy about speech perception into their productions. What is clear is that an enormous amount of variation and reduction is characteristic for natural speech. Different speaking styles lead to different kinds of variation and the more natural and casual speech is, the more likely it is to be reduced. It is therefore extremely important to understand the processes and regularities that are characteristic for "natural" or conversational speech and the differences from "perfect", laboratory or canonical speech in order to understand speech production and perception.[1] Insights for linguistic theory and models of speech perception and production are not only required to describe the processes that occur in conversational speech, but they also have to predict what can be expected to occur in natural speech and what cannot.

The examination of natural speech is not unknown to linguistics, most notably in phonetic sciences.[2] Henry Sweet's treatment of tone groups, for example, illustrates intonational groups in natural language. Daniel Jones' explains why some transcription mistakes of his students possibly are not real mistakes but differences in natural pronunciation (Jones, 1967, Chapter 12). Such studies, however, had to rely mostly on impressionist data: phoneticians were listening to what speakers do while they do it (for a similar argumentation, see Byrd, 1994), and technical aspects did not allow for the use of extensive speech corpora.[3] In the late 1950's, and early 1960's linguists were also interested in processes connected with natural or rapid (fast) speech (for example see Pollack & Picket, 1963: 165 and references therein; Harris, 1969). In the following years, natural speech was

[1] The terms "natural", "conversational" speech on the one hand or "perfect", "canonical", or "laboratory" speech on the other, are often used in the literature interchangeably. This is also reflected in the use of the terms throughout this dissertation.
[2] In some sense, every speech uttered by a human is natural, by definition. However, in the sense used here, natural is defined by speech used in natural settings as opposed to speech produced in laboratories or read speech.
[3] This is not to say that those works are to be considered less valid. But as shown below, there might be some misperceptions that are only heard correctly if one is able to rehearse small bits and pieces of utterances.

studied from time to time (e.g. Dalby, 1984; Shockey, 2003; or for German Dressler et al., 1972; Kohler, 1990), however a systematic study of conversational speech was not of central interest (cf. Cutler, 1998; Johnson, 2004a). The amount of natural speech that has been investigated in phonological studies is even smaller. Phonology has had the study of *perfect* speech as a center of research focus (e.g. Cutler, 1998; Johnson, 2004a; Tucker, 2007). And even for acoustic-phonetic studies, the items that have been investigated were most often produced in laboratory speech with word lists that were read by a small number of speakers usually from a small, well-controlled sociolinguistic background (cf. Byrd, 1994). She claims that the "... limitation to carefully controlled test items may focus the speaker's attention on contrasts, thereby exaggerating them" and points also to the importance of natural speech corpora for linguistic studies (Byrd, 1994: 40). This view is also underpinned by results provided by Kessinger and Blumstein (1998). They showed that natural speech has different characteristics from the stimuli that have been used mostly in VOT studies, another argument for using natural rather than laboratory speech in linguistic experiments.

In recent years, researchers in several fields of linguistics "rediscovered" the importance of natural speech and the use and analysis of naturally spoken corpora was increased (e.g. Ernestus, 2000; Wester et al., 2001; Pitt & Johnson, 2003; Connine 2004; Johnson, 2004b; Sumner and Samuel, 2005; Snoeren et al., 2006, 2008; Raymond et al., 2006; Dilley & Pitt, 2007). Despite the growing interest linguists have for natural speech, there is still a dearth of works on conversational speech. A lack of availability of speech corpora in different languages is one reason for this. Corpora are very difficult to construct (e.g. Lamel et al., 1986; Zue et al; 1990; Umeda, 1991; Kohler et al., 1995; Ernestus, 2000; Wester et al., 2001). Making people talk in a natural way is not an easy task at all - still it is not possible to control for many factors of language use. Subsequently, it is even more laborious to transcribe what subjects said, and to do so correctly (cf. Wester et al., 2001). For many languages, this effort has not been taken yet.

Another reason for the relatively small number of corpora of spontaneous speech is connected to technical issues. Only recently, computer's power and memory size have increased immensely. Speech files necessitate huge amounts of storage on computers. Thus, only recently, it became possible to store huge amounts of speech corpora inexpensively on computers and make them accessible to other researchers at relatively low costs. It is interesting to note that one of the driving factors for large natural language corpora was and still is the interest in evaluating automatic speech recognition systems and their performance on "real" data (Byrd, 1994; Lahiri & Reetz, 2002; Van Bael et al., 2007).

Yet another important reason for the neglected role of natural speech may be found also in the linguistic theories of the (post-)SPE generative framework themselves

(Chomsky & Halle, 1968, see also Johnson, 2004a, for a similar line of argumentation). In this framework, variation in naturally spoken language has not been seen as crucial for linguistic research. It has been regarded as an issue performance that did not warrant closer examination and as not having repercussion for competence (cf. Johnson, 2004a). The processes occurring in natural speech have been labeled "fast speech phenomena", being unimportant for linguistic theory. Consequently, perfect speech has been the central focus of research. Newer works, however, have shown that this estimation is too short sighted, and that although a very valuable starting point for linguistic research, laboratory speech is not sufficient for a complete understanding of how language "works". If it were not for data from real speech, the importance of frequency of use would not have been appreciated correctly (e.g. Bybee, 2007). Thus, for a better understanding of language production and perception, phonetic and phonological aspects of conversational speech have to be studied more thoroughly. The amount of variation sets apart natural from perfectly produced speech, although even laboratory speech is already considerably variable.

This thesis will examine two kinds of reductions and study their occurrence in conversational German: Regressive place assimilations and deletions possibly create a huge amount of variation if they are produced by speakers, and subsequently it will be analyzed how (German) listeners deal with this kind of variation when they encounter it. Crucially, corpus data from speech production will be evaluated on the one hand, but also the repercussion of natural speech for speech perception will be tested, because these variations have important implications for the assumptions of linguistic theories and they have to be examined more closely. The combination of these two views has been even more neglected than the study of either corpus data or perception. In this sense, variation in this dissertation is regarded as informative for linguistic theory. Thus, for linguistic research – as for detective's work – what is out of the common (i.e. variation in natural speech) should be regarded as guide to new insights on how language works, rather than a problem that has to be kept aside or to be controlled for in every possible way.

1.1 Variation

Variation is a very broad term and it encompasses several, quite distinct processes and factors. It is not possible to examine all of them within one dissertation. Variation is not only an attribute of different registers. Many studies have shown that there is a huge amount of variation occurring within a single speech register, such as in conversational speech (e.g. Flege, 1988; Kohler, 1990; Lindblom, 1990; Byrd, 1994; Byrd & Tan, 1996;

Johnson, 1997, 2004a; Kirchner, 1998, 2001, 2004; Greenberg, 1999; Greenberg & Fosler-Lussier, 2000; Lahiri & Reetz, 2002). Due to these numerous factors, two words uttered at two different points in time, physically will not be the same. For instance, the German word *irgendwie* (‘somehow’) is uttered 8 times in the Kiel Corpus of Spontaneous Speech (henceforth: Kiel Corpus, IPDS, 1994) by 7 different speakers (3 female, 4 male) and is reported to have seven different variants. All the variants occurring in the corpus are listed in Table 1.[4]

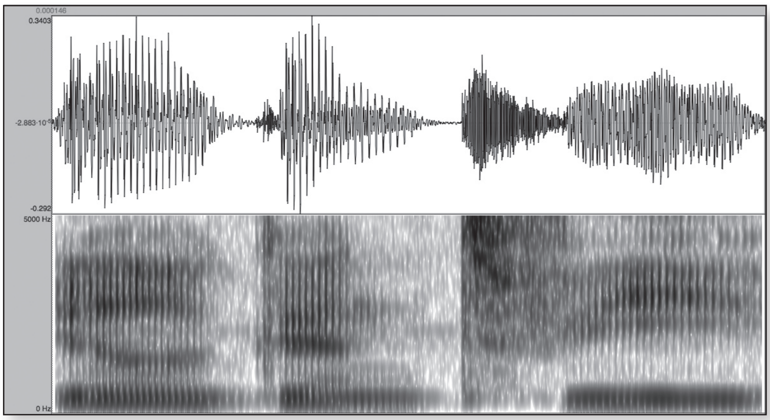
Table 1:
Variants of irgend-wie in the Kiel corpus: phonetic transcriptions from the corpus.

	[ˈʔɪɐgəntˌviː]	canonical transcription
i	[ˈɪ̰ɐmˌviː]	4 deletions, 1 assimilation, 1 glottalization
ii	[ˈʔɪɐŋtˌviː]	2 deletions, 1 assimilation
iii	[ˈɪ̰ɐgŋtˌviː]	2 deletions, 1 assimilation, 1 glottalization
iv	[ˈʔɪ̰ɐgŋtˌviː]	1 deletion, 1 assimilation, 1 glottalization
v	[ˈɪ̰ɐgŋˌviː]	3 deletions, 1 assimilation, 1glottalization
vi	[ˈɪ̰ɐŋˌviː]	4 deletions, 1 assimilation, 1 glottalization
vii	[ˈɪ̰ɐŋv]	5 deletions, 1 assimilation, 1 glottalization

There is no single word in Table 1 which exactly matches the canonical pronunciation; all the cases show three or more deviations. The Schwa which should be present, according to the “ideal” pronunciation, is *never* realized.[5] Regularly, glottalization is observable. In every example, at least one segment got deleted compared to the perfect speech. Another process that is occurring regularly in the examples is assimilation of the place of articulation (cf. Chapter 3). The CORONAL /n/ assimilates to the DORSAL place of articulation of [g] and becomes [ŋ]. Even the speaker that produced *somehow* twice did not pronounce the instances identically (v. and vi.). Processes that change features of segments, segments, syllables or even words in utterances are common in natural speech (e.g. Kohler, 1990; Jun, 1995; 1996; 2004; Cutler, 1998; Greenberg, 1999; Greenberg & Fosler-Lussier, 2000; Ernestus, 2000; Lahiri & Reetz, 2002; Gow, 2003; Johnson, 2004a; Kirchner, 2001; 2004). The ones leading to the reduction of words such as deletions and assimilation will be the main topic of this dissertation.

[5] Throughout this dissertation, German examples will be given in italics, the English translation in parenthesis and single quotes. For this dissertation, the following conventions were used for the description of letters and sounds: The sign < > is used to refer to orthography, [ ] indicates phonetic transcription, / / is used for underlying segments and { } encloses morphemes.

Three different pronunciations of *irgendwie* depicted in Figure 1 vividly illustrate the difference between perfect laboratory speech and natural conversational speech. The first instance (Figure 1a) is carefully produced laboratory speech; the second and third examples are taken from the Kiel Corpus. The second example (Figure 1b) is the one transcribed in (iv) of Table 1, whereas Figure 1(c) is (vii) from that table and very reduced. To make comparison easier, silence has been added in (ii) and (iii) making them of equal length as (i).

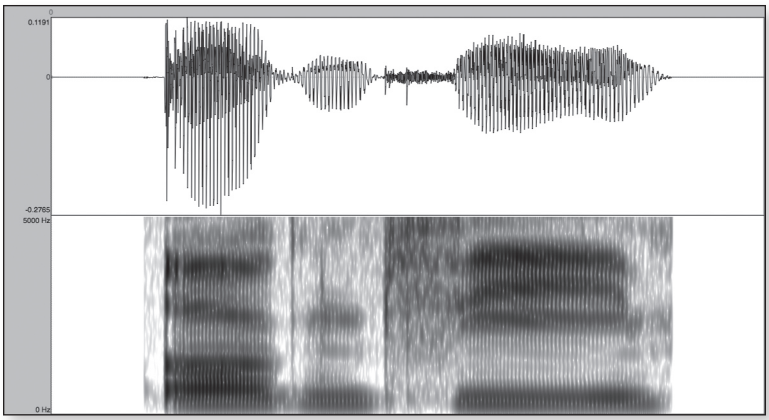


[ʔɪɐ g ə n t^{h} v iː]

Figure 1:

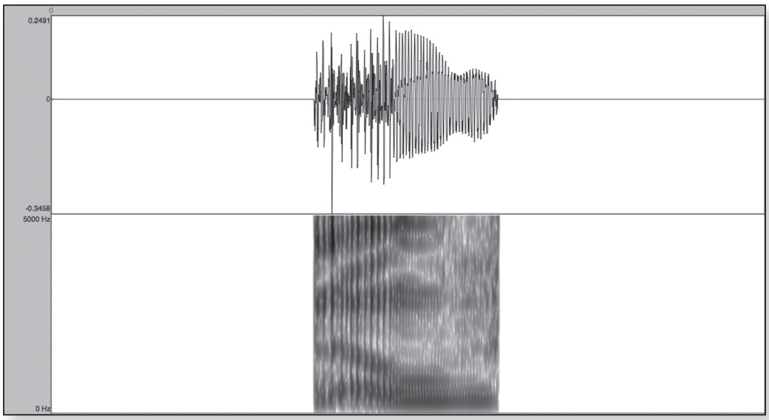
3 variants of irgendwie 'somehow' (waveform, spectrogram, and transcription)

(a) Carefully pronounced instance of irgendwie



[ʔɪ̯ɐ g ŋ tv iː]

(b) Naturally produced but well articulated instance of irgendwie



[ɪ̯ɐ ŋ v]

(c) Naturally produced but very reduced instance of irgendwie

Such an enormous range of different pronunciation has to be accounted for in linguistic theories. At the same time, the question arises how listeners are able to correctly recognize such varying exemplars of the same word and whether, for example, a token like (vii) where 5 segments are deleted – where even the syllabic structure is severely changed – is able to correctly activate *irgendwie* compared to a less reduced instance of *irgendwie* (when heard in isolation).

Depending on the point of view of different linguistic theories, such variation can be seen as the foremost hindrance to an "easy" process of language recognition. Hocket (1955) illustrated the possible problem of natural variation very vividly:

> "Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter. At a subsequent point, we have an inspector whose task it is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread-out albumen, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer."
>
> (Hocket (1955: 210), taken from Lively et al., 1994; 268)

Speech perception (i.e. the role of the inspector) seems to be a job that is not very desirable: according to Hocket's scenario, it should be very challenging and time consuming. Yet, in natural speech, listeners perform it very effectively and without the need for time. Therefore, theoretic models have to explain how the enormous variation is dealt with effectively. This can be done in several ways. Basically there are two extreme options to deal with variation. One group of models treats it as something that is not part of the mental representation itself and some form of abstraction or normalization has to be made during speech perception. The second group of models internalizes variation and makes it part of the lexical representation as such. These models are increasing the amount of stored information but at the same time, they are getting rid of the need for treating variation as problem for speech perception. A question that is linked to both points of view is whether different kinds of variation that can be identified in natural speech have also different effects on perception. Connected to these issues is the question how models explain variation in speech production. Are there variations that are more likely to occur than others? Are there patterns of variations that are more frequent than others? Depending on the architecture of the models (i.e. abstraction vs. listing of variation), these questions are answered differently and different predictions are made that will be tested in this dissertation. One important objection can also be made at this point, which will be also a guideline for the remainder of the dissertation. Although the amount of variation seems to be enormous, there is also some hope, since it is not (completely) chaotic or random (e.g. Foulkes & Docherty, 2006). One of the goals of this dissertation is also to examine in how far rules can be found that account for the variation such as reductions or assimilations in natural speech.

1.1.1 Sources of Variation

The sources of variation are almost as numerous as the possible kinds of variation in natural speech. This section of the dissertation tries to categorize the sources of variation and to elaborate more on the kinds of variation that will be examined in this dissertation. Two important factors of variation are speaker-oriented and can be broadly categorized as inter-speaker differences on the one hand, and as intra-speaker differences on the other. The third factor, where the main focus of the dissertation lies, is somewhat abstracted from speaker(s) and is concerned with phonological units, most notably with segments, syllables or features. Segments in natural speech show a range of variation independent of speakers and languages, such as contextual variation or complete changes (or their features) due to phonetic and/or phonological processes.

1.1.1.1 Inter-Speaker Variation

Every individual has certain characteristics in his voice setting him apart from other speakers. Those differences make it possible for listeners to identify different speakers very efficiently. Each speaker's voice can also be regarded as his acoustic fingerprint, which actually is used for security identification mechanisms. At the same time, this implies that these differences themselves have an impact on the amount of variation in spoken language.

Such differences between speakers are due partly to anatomical differences between speakers. The size of the speaker's oral tract, for instance, differs between men and women (and children) and has consequences for the speech signal that is produced (e.g. Stevens, 1998; Reetz, 1999b, Jongman & Reetz, 2009). Another possible source between any given two speakers due to anatomical distinctions, are for example, the size and mass of their vocal folds which influence the speech signal that they produce (cf. Stevens, 1998; Reetz & Jongman, 2009).

Besides such physiological distinctions between speakers, there are also sociological factors that influence speech production of different speakers. Speakers of different age groups, to mention one of those factors, produce speech very differently. Other factors that differ among speakers and that have an influence on speech production are age, dialectal background. All of these parameters add variation to the articulation of speech. Of course, interactions between physiological and sociological sources of variation exist and influence variation in speech (e.g. Byrd, 1994), as Byrd cites work by Labov: "For example, Labov states that 'sexual differentiation of speech often plays a major role in the mechanism of linguistic evolution' (1972, p. 303)." She also cites work by Herold (1990; in Labov, 1991), who examined the merger of *don* and *dawn* in Philadelphia. His results suggested that it was girls who substantially promoted this merger. These results indicate how the interaction of gender physiological and social factors leave their traces in natural speech and that an understanding of the linguistic principles that lead to variation has to be further promoted. As interesting as inter-speaker differences are for linguistic research, this dissertation will try to focus on variation that can be observed regardless of those factors. For a complete understanding of language production and perception, however, these factors cannot be left out completely, or have to be controlled for otherwise. In Chapter 4, for instance, results show that men delete final /t/ more often than women.

1.1.1.2 Intra-Speaker Variation

As seen already in two examples of Table 1, even when one single speaker produces the same word twice, it will physically not be the same, their pronunciations will vary considerably. Several factors can be identified that influence the amount of variation even for one and the same speaker. Some of those factors are dependent on the situation in which the speaker is talking and for what purpose the speech is produced (cf. Bell, 1984; Lahiri & Reetz, 2002; Wassink et al., 2007). In a loud environment, for example, when speakers have to make more effort to be understood and therefore try to produce clear, loud and even over-articulated speech, the pronunciation is very different compared to a silent environment. In conversations with friends, talkers usually care less about correct pronunciation, which is what they tend to do in official settings, such as giving a speech, or presentation in front of an (unknown) audience.

Age has already been identified as a source of variation between different speakers in the previous paragraph. But even for the same speaker, age is an important factor creating variation. Talkers produce language differently in different points of lifetime. Evidence for such a change in pronunciation has been presented, for example, by a longitudinal study of the Queen of England's production of [i] in her annual Christmas broadcasts over a period of 50 years (Harrington, 2006). Harrington provided evidence that the Queen's production of [i] has changed significantly, albeit less than the general trend, over that period of time (Harrington, 2006).

There are many more intra-speaker factors that add to the enormous variation that can be observed in natural speech. For example, emotional mood, or the use of drugs have demonstrable effects on speech production as well. As for inter-speaker differences, the differences for the same speaker are not the main focus of this dissertation, but they should be always kept in mind for additional explanatory power.

1.1.1.3 Segmental Variation

Shifting attention away from sources of variation attributable to speakers, either to different ones or the same ones, variation is also found due to phonetic and phonological processes that occur in natural speech. The contexts in which segments are produced have an effect on their pronunciation (e.g. Ernestus, 2000; Lahiri & Reetz, 2002; Gow, 2003; Mitterer & Ernestus, 2006). Especially for phonetic variation these effects are stronger and more frequent in casual speech than in clear speech. Some other processes even occur exclusively in natural speech and are absent when talkers produce language very clearly. These processes alter the *gestalt* of words in two different ways. Speakers can add something (i.e. feature, segment) to a canonical pronunciation of a word, or reduce it (i.e. features, segments). Despite the fact that in general, speakers tend to reduce words in conversational speech, there is also a possibility that they insert segments (or features). In natural speech, for example, talkers regularly insert stop consonants into homorganic nasal fricative clusters (e.g. Warner & Weber, 2001; Warner, 2002). Also, for instance, German speakers may produce, or "insert" word final /r/ which becomes usually /ɐ/ when preceding a vowel initial word in natural speech (cf. Kohler, 1995a).[6] More often, admittedly, speakers reduce words, segments, or features in natural speech. Reduction is often defined differently (cf. Byrd, 1994; Crosswhite, 2004). For instance, Crosswhite (2004) considers phonological (featural) neutralization as a case of reduction, whereas others define reduction more literally as a process that reduces gestures, or produces undershoot, or lenite segments (cf. Kirchner, 1998; Kingston, 2006). This dissertation examines two different reduction processes, both of which occur in natural speech. In section 1.2 of this dissertation, these processes are presented combined with a formulation of the research questions that led to the investigation of these processes.

1.2 Research Questions

Variation leading to reduction is the main topic of this dissertation. Two linguistic models will be examined. An important objective is the evaluation of their predictions and explanations for both production and perception. Two processes will be examined in more detail to gain further insights for linguistic modeling. Only when we know more about what processes occur in natural speech and how listeners deal with them, a more realistic modeling of speech perception is possible. A central quest of this dissertation is linking corpus data directly to perception experiments.

Firstly, a case of assimilation is analyzed in Chapter 3. Many studies have been conducted that have focused on assimilation, most notably on the assimilation of place of articulation (PoA) (e.g. Nolan, 1992; Jun, 1995; Gow, 2001; 2002; Coenen et al., 2001; Reetz & Lahiri, 2002; Wheeldon & Waksler, 2004; Dilley & Pitt, 2007). This dissertation does not aim at just adding yet another study to this list. Regressive place assimilation across in German is a process where linguists still disagree of its very existence (cf. Kohler, 1995a; Wiese, 1996). Thus, one of the objectives is to examine whether this

[6] This process is indicative of the fact, that it is not always the case that natural speech is more reduced compared to laboratory speech, but that there occur processes that are absent in less natural settings.

process actually occurs in natural German or not. A combination of corpus analysis, behavioral experiments and acoustic measures sheds light on this debate, showing that regressive assimilation of PoA in German is really occurring. This finding has important repercussions for phonological modeling.

The second case of variation that is examined in this dissertation concerns "massive reductions". In the literature, the term has been coined to capture several kinds of reductions that have a stronger impact on the gestalt of words (Johnson, 2004a). Massively reduced words are characterized by undergoing a combination of lenition and deletion processes (Kohler, 1990; 1995b; Johnson, 2004a). One important issue is whether massive reductions occur at random or whether there are rules that predict what is deleted. Therefore, corpus studies have to reveal what kinds of deletions are observable in conversational speech. Then it is possible to examine the effect of massive reductions for perception. Especially the latter question has important consequences for speech perception. How do listeners deal with them? Are they still able to understand what has been said without any additional "cost"? These questions will be elaborated further in Chapter 4.

1.3 Architecture of this Dissertation

After the introduction into the topic of the dissertation and the presentation of the research questions that will be examined, the remainder of this chapter will present the corpus that was used as data basis and explain the criteria for its selection. In Chapter 2, the central theoretical approaches are presented. Two rather different models are presented: X-MOD, as proposed by Johnson (1997) and the Featurally Underspecified Lexicon (FUL) model (Lahiri & Reetz, 2002, accepted). Their basic assumptions will be presented, differences and similarities will be discussed and their predictions will be elaborated. Chapter 3 will focus on regressive place assimilation across words, a process that will shed light on several important issues both for production as well as for perception of spontaneous speech. Another process that does not only change the featural instantiation of words in conversational speech, but also possibly affects segmental and/or syllabic structure of words will be the topic of Chapter 4: Reductions and deletions occurring in conversational speech. As in Chapter 3, both production and perception will be examined. Chapter 5 summarizes the findings of the preceding chapters and evaluates the findings in the light of the two different theoretical approaches that are examined in this dissertation. Future research questions that follow from the results of this dissertation are discussed and possible ways for further investigation of the processes occurring in conversational speech are outlined.

1.4 Corpora

Testing the predictions of two different phonological models of speech perception and production is only possible with adequate speech data. Speech corpora with spontaneous speech are the best source of information for this enterprise. They are useful for determining what speakers actually do, when speaking to others. But they are not only a useful source for production studies, since they contain natural (conversational) speech, words and phrases from the corpora consequently can be used also for examining perception of natural speech.

Constructing speech corpora is by no means an easy task. In order to ensure for “naturalness”, several strategies can be followed to make speakers produce speech as they do in natural circumstances (see, e.g. Kohler et al., 1995; Ernestus, 2000; Pitt et al., 2003, 2005; 2007). Making speakers speak freely is only one important goal that has to be reached with a corpus of spontaneous speech. Of similar importance is to control for what they talk about and what words are used. For example, if a word is uttered only once in a corpus, it is very hard to conclude on any regularities from this utterance. It might be exhibiting characteristics that are regularly encountered in natural speech; but it could as well be a “mispronounced” item. When using a corpus that has been created by other researchers, there is no control over what has been said and how the transcriptions were made. Especially the fact that phonetic transcriptions of corpora often have been made without prior knowledge for what purpose the transcription was to be used has repercussions for the quality of those transcriptions (see, e.g. Van Bael et al. 2007).

Keeping this in mind, there are basically two possible options how to proceed. One possibility is to create a new corpus of conversational speech. This approach has been taken by Ernestus (2000) for example, for her study of reduction in conversational Dutch. The second possible way is to use an already existing corpus. In this dissertation, a combination of both possibilities is pursued. An already existing corpus of German will be used and analyzed for a majority of the data. This corpus of choice for this dissertation is the Kiel corpus (IPDS, 1994). However, there is also a small corpus that has been created exclusively to examine a question that arose from the findings of the Kiel corpus analysis.

The Kiel corpus consists of dialogues from 42 (northern) German speakers (18 female, 24 male). Overall, the length of the corpus is about 4 hours of speech, containing almost 2000 turns of dialogues. The speakers were engaged in an appointment making task. At the time of the recording, they were naive about the goal of the corpus. They were each given different schedules and lists of appointments that had to be arranged. Their instructions were to find possible dates for the appointments. The speakers were ignorant of the schedule of their partner, and the schedules were manipulated with conflicting agendas. This was done to force the two talkers to negotiate on their future meetings, and this procedure also ensured a high degree of natural speech. Another important feature of the Kiel Corpus is the high quality of the recordings. The dialogues were recorded with the speakers placed in different sound-treated rooms wearing headsets. The talkers did not see each other. When they wanted to communicate they had to press a button, otherwise their partner would not hear them. What makes the corpus even more valuable than the sound quality is that all dialogues were transcribed and labeled by trained phoneticians. The fact that the transcriptions were made only by very trained phoneticians ensures to a certain degree the correctness of transcription. The phoneticians used visual scaleable spectrograms and oscillogram displays as well as auditory information (Kohler et al., 1995: 33) for the transcriptions. There are three different transcriptions in the corpus. Firstly, there is an orthographic transcription of the dialogues. Then, there is a canonical phonetic transcription. And finally, and most importantly, the corpus has also a phonetic transcription of what was actually pronounced. This allows for a comparison of an idealized pronunciation (i.e. canonical) transcription with the actual (i.e. phonetically transcribed) pronunciation. The idealized canonical transcription denotes how speakers should utter the words if they were talking in accordance with a careful dictionary-like pronunciation.

The nature of the task and the fact that all the pairs of talkers had to make the same appointments restricted the vocabulary on the one hand and let to a large number of utterances for other words, on the other hand. For instance, days of the week as well as dates and times occur very often. Nevertheless, since the speakers were unaware of the purpose of the recordings, the conversations were very natural and the corpus meets all the requirements that were asked for as a basis for an analysis of the processes that occur in natural speech.

There exist also other corpora with natural speech in German such as the a corpus that was created by recording the conversations of the participants from the first season of the German TV-Show “Big Brother”, or the so-called Lindenstrassencorpus (IPDS, 2007), where pairs of subjects talked about two different versions of the same episode of the German TV series “Lindenstrasse”. However, the Kiel Corpus was chosen as the basis for this dissertation, since it is the corpus of German conversational speech that best fit the expectations for the purpose of this dissertation. Note that the results reported here rely on a great deal on what has been said by the speakers of the corpus and on the correctness of the transcription provided with the corpus. Personal reports and knowledge of the procedures for transcription make it plausible, that the corpus is indeed reliable.

In Chapter 4, where final /t/ deletion is examined, additional data became necessary. Neither the Kiel corpus nor the Lindenstrassen corpus allowed for extracting enough verbs in the second person singular. Therefore, a production task was created that allowed for rather natural speech, and had a strict control over what subjects produced. Thus, the combination of an existing corpus with an excellent reputation and where needed the creation of a smaller corpus that served exactly the purpose of a special question are the basis for an investigation of reduction processes occurring in German and how listeners perceive them. But first, the theoretical frameworks that will be evaluated in this dissertation are presented in the next chapter.

«…when you have eliminated the impossible, whatever remains, however improbable, must be the truth»
Sherlock Holmes *(Arthur Conan Doyle, The Sign of the Four)*

Chapter 2 – **Theoretical Assumptions**

2.1 Introduction

Human auditory speech recognition is extremely complex and yet, at the same time, it works very efficiently. Humans can perform it subconsciously, without a recognizable effort, and without taking a remarkable amount of time lag between hearing the language and understanding what has been said. What seems even more remarkable that this efficiency is not lost when the environment gets noticeably noisy or when only parts of the acoustic information are transmitted to listeners, as is the case in conversations on the telephone, or when people talk to each other in a pub or at a cocktail party. The ease with which a rather perfect performance is reached by listeners even leads to the fact that in general, linguistically naive adults would not see speech perception as a challenge (cf. Juszyk, 1997). Several factors influence auditory speech recognition. However, before speech is recognized, it first has to be uttered. What seems to be a tautology is often regarded rather dilatory in research on speech perception.

The events leading from speech production to a successful perception can be briefly summarized as follows.[7] Firstly, some word or words is/are produced by the speaker, then the resulting physical signal is transmitted (mostly via the air) into the ear of the listener, where it is afterwards encoded into neural responses that are mapped subsequently onto a lexical representation in the brain, activating words leading finally to the recognition of the word that has been uttered by the speaker. Even this short summary of processes, however, is not free of theoretical bias(es) and implicit assumptions, e.g. 'there is something like a lexical representation in the brain'. At the same time, it is only a very crude and vague description of the events. For example, the process of neural encoding is not elaborated

[7] This summary starts at the point of time after the conceptualization of what the speaker wants to say, as well as after the point in time when planning and sending the motor-commands for the articulators are already over (for the processes prior to this point in time, see, e.g. Levelt, 1989).

any further. Possible biases in the description that also set apart different models of speech recognition are assumptions concerning the architecture of the lexicon, or what exactly is stored in the mental lexicon, including assumptions about the size of the units that are stored: are there words, segments, features, or are there any smaller (or larger) units which make up the representations in the lexicon?

Theories and models of lexical access, lexical representation, speech perception and phonology in general have to account for observations based on natural human languages. The chain of processes leading to successful speech recognition as sketched above is what models of speech production and perception should strive to explain. The short summary above started with the natural origin of speech, – the speaker. For successful modeling of speech perception, processes connected to speech production have also to be incorporated into linguistic theories, since there is no perception without production and *vice versa*. If incorrect assumptions are made about what is being produced by speakers, theories of perception (i.e. how do humans perceive what other humans produce) might be flawed to a severe degree.[8] The relationship between production and perception has been found to be very systematic in many cases (e.g. Pierrehumbert, 2003b; Kingston, 2006).

One of the reasons that make speech perception an extremely interesting case to be explained by linguistic theory is variation, as has been mentioned superficially in the introduction. If every word was uttered identically by every speaker of a given language, speech perception would indeed be a trivial task, both for the listener as for the linguist. For the listener, the mapping from speech input to representation would not require much effort as the speech input would be rather invariant; for the linguist the quest for explaining speech perception would not be as thrilling a quest as it is in reality, the "lack of invariance" problem would no longer exist (cf. Perkell & Klatt, 1986). Thus, variation in speech production makes speech perception both complex as well as interesting. There is little if no debate that there is a huge amount of variation occurring in everyday conversation (e.g. Flege, 1988; Kohler, 1990, 1995; Lindblom, 1990; Johnson, 1997, 2004a; Kirchner, 1998, 2001; Greenberg, 1999; Greenberg & Fosler-Lussier, 2000; Lahiri & Reetz, 2002, accepted; Tucker, 2007). A common tendency for speakers is to reduce what they say. Complete words are targeted by reduction, segments or phonemes are not produced perfectly, they are lenited, or sometimes omitted completely. Thus, variation in general and reduction in particular can be seen as two of the original factors that drive the quest for a theoretical modeling of speech perceptions (see, Goldinger, 1998, for a similar argumentation). The enormous amount of variation is one of the few observations most linguists in general and phonologists and phoneticians in particular would agree on. However, the debate starts as soon as one asks about the amount or regularity of such variation, and how listeners deal with reductions occurring in conversational speech. Is reduction a process that is predictable?

---

[8] There might be one exception, though, because when children acquire language, they arguably perceive first, before they begin to produce speech.

Do listeners make use of reductional regularities, supposed they exist? How much variation is tolerated by listeners? This dissertation focuses on questions like these.

Variation can be seen not only as the driving force for the existence of theories of speech perception, but the way variation is handled is also the characteristic that sets apart different models of speech perception. There are two possible ways how theoretical approaches treat variation.[9] Firstly, it can be seen as a problem that aggravates successful recognition; consequently, some processes have to compensate for variation before a successful recognition is possible. Or, in the worst case, if there is too much variation, successful perception will be harder if not impossible. A second view sees variation as informative for its own sake and as a building force for lexical representations. These two views and their ability to explain and predict human behavior will be juxtaposed in this dissertation, exemplified by two models of speech perception, the FUL model, elaborated by Lahiri & Reetz (2002) and an exemplar model, X-MOD as proposed by Johnson (1997).[10]

One other point that drives the quest for phonological theories is the observation that there exist both language specific as well as universal processes that have to be accounted for. Listeners and speakers of different languages use very particular phonetic instantiations for phonological features (cf. Kingston & Diehl, 1994; Bradlow, 1995; Pierrehumbert 2000, 2001a; Kingston, 2007 and references therein). This observation holds for language perception as well as for production; both are always adhering to the phonological system of the particular language and of its (phonological) contrasts (e.g. Lahiri & Marslen-Wilson, 1991, 1992; Ghini, 2001a,b, 2003; Pierrehumbert, 2003b). Other features and processes, however, are connected to the human ability to speak in general, and thus have to be universal. Hence, theories and models of speech perception and production have to include both language particular and language universal assumptions. Similarly, they have to differentiate between language specific and universal rules and processes in their assumptions.

A third requirement for linguistic models is that they are able to predict the amount of variation that occurs and why this is so. They have to explain what kinds of variations are to be expected, and make assumptions why other processes will not occur in spontaneous speech. Or else, they have to explain why "anything" goes. The better the models and theories are able to explain and predict the actual data, i.e. natural speech, the more desirable is their use in linguistic theory.[11]

There is a vast body of literature on language production and perception. Similarly, there are many different theories and models that aim at explaining these two human abilities. It is impossible to present them exhaustively in a dissertation. Therefore, an exemplary overview for each of the two opposing views regarding variation will be provided. Models which see variation as a possible "challenge" for perception will be set apart

[9] Reductions are treated as special instance of variation in this dissertation. Whenever special assumptions have to be made that are different from the more general theme of variation, this will be discussed in more detail.
[10] Variation in itself is nothing special to language, of course, (cf. Lieberman, 1986).
[11] In the remainder of this dissertation, the terms theory and model will often be used interchangeably. The (philosophic) question of what differentiates a model from a theory is set aside, this is clearly outside the scope of this dissertation, even more so since, within different fields of research such as, for example in Political Science and in Linguistics, the meaning of the two terms is used differently.

from models regarding variation as source for additional information. To contrast the two opposing views as sharply as possible, two of the most promising approaches representing the two opposing views will be introduced and discussed. As a consequence, they come from very different theoretical frameworks. Some of their assumptions are compatible, whereas others are not.

On the basis of the distinction drawn above, one can lend two labels to the ways variation is handled. On the one side, views are "abstractionist", on the other end they are "episodic" or "exemplar-based".[12] What does that mean? Models that adhere to the former view assume one abstract representation that is deprived of both a lot of redundant and indexical information, as well as of other variation that is characteristic for natural speech. Indexical information encompasses many details of the speech signal, such as information about the identity of a speaker, gender, or dialectal origin of the speaker. The "episodic" view sees indexical information as crucial part of the lexicon and hypothesizes many lexical entries for the same item. As its central assumption, this view treats lexical representation as very concrete. In its most basic instantiation, this view assumes that every time a word is heard, an exemplar of this word is stored along with a lot of indexical and in fact redundant information. Note that there exist many similar models on both sides of the borderline between abstractionist end exemplar-based models, of which only these two are examined here. The models that have been chosen as examples for the dichotomous typology of "abstraction" versus "exemplar" are the Featurally Underspecified Lexicon (FUL) model, as proposed by Lahiri and Reetz (2002, accepted) and the exemplar model as proposed by Johnson (1997), with some refinements of approaches by Goldinger, (1998) and Pierrehumbert (2001a, 2003a, b). What makes a comparison of the two models possible despite their apparent differences is that both have their main focus on speech perception.

Note that this dissertation does not aim at finding the one and only, the "best" and "true" theory. It is impossible to give evidence that allows for such far-reaching conclusions. However, this dissertation seeks to present evidence from natural speech that is able to point to strengths and weaknesses of the frameworks in question. Two kinds of reduction processes, regressive place assimilation and "massive reduction (cf. Johnson, 2007) are examined and help to point to the successes of the different models to predict and explain data from natural speech. This method is a crucial process for the advancement in theoretic development (see Brown, 1990 for a similar point).

Recently, "mixed" models, combining abstract and episodic representations have been suggested (e.g. Goldinger, 1998, 2007; Pierrehumbert, 2002, 2006b). The emergence of mixed models is primarily attributable to considerations that neither a "pure" abstractionist model, nor a "basic" episodic approach will very likely be able to explain everything. Thus, there might be a need to extend them or bring together different frameworks at a certain

---

[12] The question of abstractness is also a very basic question that is asked in phonetics, for example, concerning motor commands for different speaking rates and their representation in the lexicon (cf. Reetz & Jongman, 2009: 89 and references therein).

point in time. However, before this can be done, the power of the pure models should be examined in more detail. It is important to test the models and how they are able to explain data from natural speech, and how far the data can be accommodated with the existing assumptions. Only when strengths and weaknesses of pure models are well understood and studied, "synergetic" effects of two different views can be expected. If one comes to assume a hybrid model where an intervening phonological coding level is added to the exemplar storage level, it is also important to know what level accounts for what effects.

In the remainder of this chapter, the exemplary models will be presented, their main assumptions will be explained, and crucial points will be highlighted. These predictions will then be compared to data of corpora analyses and of five experiments in Chapters 3 and 4.

2.2 The Featurally Underspecified Lexicon Model

The first model that is examined in this dissertation, as representative for an abstractionist framework, is the Featurally Underspecified Lexicon (FUL) model, as proposed by Lahiri and colleagues (e.g. Lahiri & Evers, 1991; Lahiri & Marslen-Wilson, 1991, 1992; Reetz, 1998; 1999a; 2000; Lahiri, 2000a,b; Ghini, 2001a,b; Lahiri & Reetz, 2002). The model has evolved as psycholinguistic advancement of traditional phonological underspecification theories (cf. Kiparsky, 1982; Archangeli, 1988; Pulleybank, 1988; Avery & Rice, 1989; for an overview and critique, see Steriade, 1995) and has an instantiation as automatic speech recognition (ASR) model where the model's basic assumptions can be tested.[13] All of these underspecification theories – as the name already suggests – assume that lexical representations are not completely specified, i.e. that not all possible features are part of an abstract underlying representation in the lexicon. They differ, however, as for the extent to which underspecification is assumed and about the actual architecture of the mental lexicon. In this dissertation, FUL will be treated as representative for abstractionist models in general and for models assuming underspecification in particular. FUL is one of the most prominent current models assuming underspecification, and has proven a successful approach for many different phonological and morphological phenomena (e.g. Ghini, 2001a; Obleser et al. 2003a,b, 2004; Scharinger, 2006; Kabak, 2007; Wetterlin, 2007), most notably for assimilation processes, that will also be discussed in the next section of this dissertation (cf. Lahiri & Reetz, 2002; Wheeldon & Waksler, 2004, Lahiri et al., 2006; Zimmerer et al., 2009; and references therein). In the upcoming paragraphs, the basic assumptions of FUL will be depicted. Particular assumptions and predictions will be also elaborated before each analysis in the upcoming chapters.

[13] Other linguistic areas also assume underspecification, however, the term is used differently in syntax, or semantics, for example. Underspecification theory in those areas is not (necessarily) related to phonological underspecification.

2.2.1 Basic Assumptions

2.2.1.1 Representation: The Mental Lexicon

FUL posits that each morpheme has one single, abstract representation in the mental lexicon. Whether this strict assumption of only one representation per morpheme can be upheld for all function words, such as *und* 'and' which can be produced with very unpredictable reductions, is not clear (cf. Jones, 1972; Kaisse, 1985; Hall, 1999). However, for lexical words, this assumption is crucial, especially for the results reported in Chapter 4.

Representations of morphemes are built up of matrices of monovalent universal phonological features, except for the two pairs [CONSONANTAL]/[VOCALIC] and [SONORANT]/[OBSTRUENT] which are binary and opposing, and each segment has to be specified for one of each of the pairs. For the other features, they are assumed to be either absent or present, but they are not marked with [+] or [-] in the representations. The features themselves and their hierarchical organization as assumed in the model are based both on universal principles of phonological alternations as well as on perceptual mechanisms (Lahiri & Reetz, accepted). The matrices have a language universal basic set-up being hierarchically organized (cf. Clements, 1985, 2003; Clements & Hume, 1995; see also Figure 2 below, or Halle, 1995; Halle et al., 2000). One characteristic that sets apart FUL from many other abstractionist frameworks and even from many underspecification approaches is the assumption that vowels and consonants share the same features (cf. Lahiri & Evers, 1991; Lahiri & Reetz, 2002, accepted).

In the model, there is no intermediate level of representation such as segments or syllables in the lexicon. This point is crucial especially for reductions and deletions, as will be elaborated in the sections below. Figure 2 depicts the complete set of features that are building up the lexicon in FUL. It also gives an overview over their hierarchical organization. This hierarchical organization is based on theoretical phonological considerations as well as data from many different languages (e.g. Clements, 1985, 2001, 2003; Lahiri & Evers, 1991; Halle, 1995; Ghini, 2001a; Lahiri & Reetz, accepted). Many phonological processes that occur in languages of the world show that some features can group together in phonological processes such as assimilations, whereas others cannot. There are also universal implications that lend support to a tree structure of phonological features.

One of the most important basic assumptions of FUL is that only features that are contrastive and unpredictable are part of the lexicon, leading to lexical representations of morphemes that are possibly underspecified. Predictability also implies that underspecification is strongly dependent on the phonological system of a particular language; thus, which and how many features are underspecified varies from language to language. This also means that segments can have different specifications in different languages. For example, in

German, as in most other languages of the world, the PoA feature [CORONAL] is assumed to be unspecified. Another example for underspecification in German is exemplified by the specification of vowels. They are not specified for the feature [NASAL] in German, because it is not a phonological contrast in that language. In other languages, as Bengali for example, where nasality of vowels is contrastive and not predictable, vowels are specified for this feature (Lahiri & Marslen Wilson, 1991).

This underspecification method meets one of the requirements for phonological theories and models as discussed above to be both language specific and language universal. The method is able to accommodate language specific representations, such as the differentiation what features are underspecified in which language. Further language particular elements are phonological (rewrite) rules (cf. SPE, Chomsky & Halle, 1969) that allow for language particular rules as well as for universal rules. At the same time, the model incorporates as well universal characteristics of languages of the world and their systems, such as one basic feature tree for all these languages. Coming back to Figure 2, depicting the complete set of phonological features that are assumed in FUL; if a segment is underspecified, its corresponding place or tongue root node is used as "place holder" in the feature matrix, but otherwise empty. For production, default rules fill in the necessary features from which the final articulatory score is derived. This will also be explained in more detail further below.

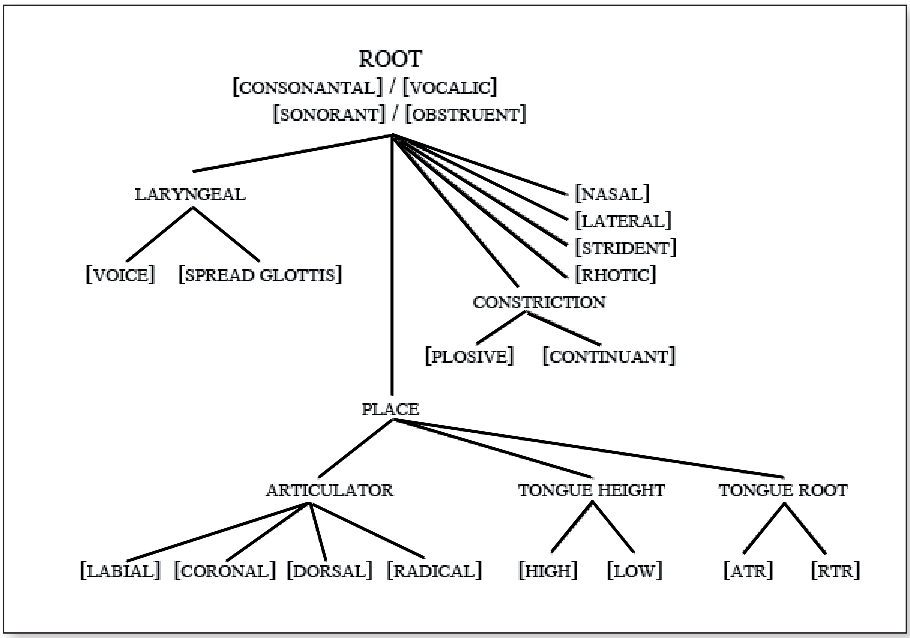


Figure 2:
Feature tree as assumed in the FUL model, taken from Lahiri & Reetz, 2009

2.2.1.2 Speech Perception

FUL makes very explicit assumptions about speech perception as it evolved as a speech perception model, subsuming a theory of lexical representations as well as of lexical access. Many assumptions concerning speech perception are not unique to FUL. Rather, FUL also includes assumptions of different psycholinguistic models of speech perception that have found solid support in experimental research, most notably the Cohort Theory developed by Marslen-Wilson and colleagues (Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Zwitserlood, 1989; Marslen-Wilson, 1990). In the Cohort Theory, a search for the correct word candidates begins as soon as information from the acoustic signal enters the perceptual system. This search initiates the activation and then selection of word candidates that match the acoustic input best (Marslen-Wilson & Zwitserlood, 1989; Marslen-Wilson, 1990). Such a metric capturing the amount of fit between inputs and representation is a part of many current models of spoken word recognition (e.g. McClelland & Ellman, 1986; Marslen-Wilson, 1990; Boersma, 1997, 1998; Goldinger, 1998). However, the matching mechanism for FUL has some peculiarities setting apart the framework from other models of speech perception. One important characteristic of FUL concerns its evaluation metric. Here, the assumptions are made very explicitly. They are a very central part of FUL and are also implemented in the ASR system that is built on the FUL structure and architecture. At the heart of lexical activation is a three-way matching algorithm that distinguishes between a "match", "no-mismatch" and "mismatch" condition (e.g. Lahiri & Reetz, 2002, 2009; Scharinger, 2006). Prior to the actual matching process, the speech signal is analyzed into perceptual phonological features, which are directly mapped onto the representations in the mental lexicon. Possible word candidates are activated or rejected depending on the information that has been extracted from the signal. The extraction of features is an automatic process, based on the speech signal, without an intervening level of representation. The matching algorithm functions as follows.

Firstly, a *match* occurs if the information from the signal is the same as in the lexicon. This match will increase the activation score of a morpheme. An example for a match is the extraction of the PoA feature [DORSAL] from the speech signal. Every candidate that has this PoA feature at their respective position will have an increased matching score.

The second possibility is exactly the opposite condition, the *mismatch*. A *mismatch* occurs, whenever the featural information from the signal is not compatible with the lexical representation, i.e. when the features are mutually exclusive. This is the case when the feature [LOW] is extracted from the signal and applies to every corresponding candidate in the lexicon that has the featural specification [HIGH]. If [LOW] is extracted, this information cannot activate candidates that are specified as [HIGH] in the lexicon. All possible morpheme candidates with high are therefore rejected. Another example of mutually exclusive features can be exemplified for consonants when the PoA feature [CORONAL] is extracted from the signal. This information mismatches with all candidates having a [LABIAL] or [DORSAL] specification in the lexicon.

The third matching condition is the so-called *no-mismatch*. This condition is crucial and sets apart underspecification from many other phonological frameworks. It occurs whenever the information the listener extracts from the signal is not conflicting with the features in the lexicon, or if no feature is extracted from the signal and is matched to any corresponding feature in the lexicon. The following examples illustrate this condition. For consonants, the extraction of [LABIAL] and the representation of [CORONAL] segments in the lexicon creates a *no-mismatch*. Since [CORONAL] is not specified, the labiality from the signal is matched against "nothing". This will keep the candidate activated, but not increase its score. This is true for all incoming PoA features in combination with all [CORONAL] consonants. Even when [CORONAL] is extracted from the signal, there will be no perfect match because in the lexicon, there is nothing the matching algorithm can match the information to. One further example illustrates the assumption of FUL that some features are not extracted at all from the signal. For instance, for vowel height, [MID] is not defined. Only [HIGH] and [LOW] are acoustically defined and there exists additionally an undefined acoustic space in between the values of the two features. This is crucial in many ways. Firstly, for [MID] vowels, there will be no perfect match (as for any underspecified segment). Secondly, and more generally, although the physical signal, i.e. the information carried by the sound waves, is "completely" specified, some features will not be extracted. The lack of overlap between the two categories [HIGH] and [LOW], for example also allows for natural variation in the signal that will not disrupt successful speech perception. From this architecture, one further crucial assumption can be deduced, [HIGH] vowels are assumed to be never realized as [LOW] and *vice versa*. If they are, the model predicts that the respective candidates will no longer be activated.

As can be seen, this matching mechanism is in line with the Cohort-Model advocated by Marslen-Wilson and others (e.g. Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980). This architecture predicts that as soon as acoustic information is available, this information is analyzed into features and the lexicon is directly accessed. At first, many different candidates are activated. As the amount of information from the signal increases, the number of activated morphemes decreases, since more and more mismatching information occurs and more and more morphemes cease to be viable candidates. Only those that are either matching or having no-mismatch scores will stay activated. In the end, the morpheme with the highest score will be the one that is selected amongst its competitors. The matching score can also be expressed by a scoring formula (Lahiri & Reetz, 2002: 641, Figure 3).

$$\text{SCORE} = \frac{(\text{NR. OF MATCHING FEATURES})^2}{(\text{NR. OF FEATURES FROM SIGNAL}) \times (\text{NR. OF FEATURES IN LEXICON})}$$

Figure 3: Scoring Formula in FUL, as in Lahiri & Reetz, 2002

Since competing word hypotheses will be evaluated after the signal enters the perceptual system, the size of the cohort that is activated is crucial for word recognition as well. This competition can for example be influenced by frequency effects, both of the candidates in question as well as frequency of the cohort members. “Dense” neighborhoods, i.e. neighborhoods with many different similar word candidates, decrease the amount of activation for a given candidate, in that other competitors are also activated (e.g. Goldinger et al., 1989; Luce & Pisoni, 1998; Vitevitch & Luce, 1998; Vitevitch et al., 1999; Dell & Gordon, 2003; Ziegler, 2003). At the same time, this inhibition is additionally influenced by the frequency of the neighbors. *Ceteris paribus*, dense neighborhoods with high frequency neighbors will lead to lesser activation than sparse neighborhoods with low frequency competitors (e.g. Luce & McLennan, 2005). Generally, frequency effects can be handled by FUL. A small change in resting activation depending on frequency, will allow for the modeling of these effects.

In its current version, FUL explicitly only takes into account the levels of phonetics, phonology, and morphology. However, natural perception also includes matching in syntax, semantics and pragmatics. These levels can be added to the basic structure of the model. So far, this inclusion has not been modeled explicitly, however, it is obvious that during the process of perception, different candidates will have to take into account also semantic and syntactic information at a later stage, especially when acoustic information is not enough for finding a matching candidate.

One advantage of this architecture compared to many other approaches with abstract lexical representations is that FUL allows for incomplete and variable information in the speech signal that does not disrupt successful language recognition. At the same time, however, this also means that variation, at least to some extent, has to be regular and not coincidental. If features are varied by chance, not following (phonological or morphological) rules, recognition can be severely reduced if not rendered completely impossible. The amount and predictability of variation will be examined in the following chapters of this dissertation. Note that for the assumption of FUL, phonetic variation is different from rule based phonological or morphological variation. Such variation in FUL is not part of the grammar, nor is it part of the lexicon. It is assumed to impede successful recognition, depending on the scale of deviation it creates from the abstract representation or the canonical pronunciation (for a convincing argumentation, why phonetic processes should not be part of grammar, see e.g. Kingston, 2006). If variation is lawful, FUL does not only allow for it, the model even predicts that variation will be produced by speakers.

2.1.1.3 Speech Production

In the timeline of speech perception, production precedes perception. Phonological processes occurring during speech production are explicitly treated in the basic outline of the model. As in traditional generative frameworks, phonological rewrite rules are applied to the underlying forms in the lexicon if they meet the conditions for rule application (cf. SPE, Chomsky & Halle, 1963). Additionally, there are rules specific to underspecification models. Ultimately, empty (i.e. underspecified) features have to be produced at some point in time. Therefore, for production, unspecified features are executed by phonological (default) rules. For a canonical pronunciation, the speech signal is assumed to be fully specified. To turn back to the example of underspecified [CORONAL] segments in German: if the speaker is to utter a word containing for example an /n/, the PoA specification in the lexicon is empty. For production, the default rule inserts the feature [CORONAL]. This rule is exemplified in (1).

(1) ARTICULATOR [] -> [CORONAL]

FUL thereby is able to predict assimilation patterns that are observable in many languages of the world (cf. Jun, 1995, 2004). At the same time, the model also predicts and explains what kinds of assimilations are very unlikely, and why this is the case. Assimilation is one of the reasons, why underspecification has evolved in phonological theory. Additionally, since phonological rules are part of the grammar, any (morpho-)phonological rules can be incorporated. The model also allows for phonetic variation, to a certain extent, although this kind of variation has not been the primary concern of FUL. This is even true for variable rules, which do not apply in an "either or" fashion but in a probabilistic manner, depending on many, partly non-linguistic factors and are more phonetic than phonological (e.g. Labov, 1967; 1969; Cedergen & Sankoff, 1974; Raymond et al., 2006). The crucial difference between purely phonetic and phonological variation in a framework of classical generative phonology is that phonetic variation is a question of performance, not competence. One reason for this is that phonetic variation is not predictable in the same way as phonological variation. Many factors, as already discussed in Chapter 1, have an impact on the acoustic properties of speech, factors that cannot be included in the grammar of an abstract model. Therefore, there is an upper limit to the variation an abstract model as FUL is able to deal with. If variation occurs in a larger scale than this limit, it is impeding successful word recognition. Depending on the amount of deviation of the signal compared to the abstract representation, this deviation is even expected to be fatal for speech recognition.

Care has to be taken what variation occurring in the speech signal is phonetic and what is phonological. This is especially true for the regularities that are variable (c.f. Labov, 1969; Cedergen & Sankoff, 1974; Raymond et al., 2006). Pure tendencies are very likely not incorporated as rules into the grammar (see also Kingston, 2006). The point here is quite obvious, since FUL claims that the phonological variation can be dealt with in speech perception, whereas phonetic, or random, not-rule based variation is impeding speech perception of making it impossible if deviations from the abstract representation are too large. The differentiation of what is phonological and what is not should not be made ad-hoc; clear definitions have to be established on theoretical points. A rather conservative approach should be made concerning the addition of variable rules into the grammar (see also Kingston, 2006 for an argumentation to keep phonetics and phonology separated, or Arvaniti, 2007; Lahiri, 2007).

Both for production and for perception, FUL assumes feature matrices as basic unit for representation and lexical access, i.e. matching. So far, time constraints have not been included in the assumption of FUL, neither does the model explicitly assume strict time alignment, nor is the possibility of overlapping of feature matrices an explicit assumption. Therefore, a promising extension of the model would be to include explicitly time constraints in the model. Assuming looser temporal feature representations and allowing for overlapping phonological features due to coarticulatory processes would render the model even more explanatory. Consider the possible pronunciation of the German word *haben* ('have') that is produced by many speakers as [ham], whereas the canonical Duden-like pronunciation would be [haːbn̩]. A complete syllable has been deleted, and one segment is missing. Regarding the lexical representation of the word, it becomes apparent that almost all features of the lexical representation are uttered by the speaker, only the alignment of the segmental realization is a little bit blurred, the labiality of the /b/ is sustained on the [m], as is the nasality of the /n/. The fact that a syllabic nasal becomes a plain nasal is no grave deviation from a featural point of view. This example illustrates how a featural representation is superior in dealing with natural variation compared to a segmental representation. The possibility that the actual time ordering of features can be loose, and the repercussions for the general built up of underspecified segments, has been also discussed and subsequently been incorporated in a different phonological underspecification theory (e.g. Lodge, 1992; 1995).[14] Yet another approach where temporal constraints are also loosened and where an automatic feature extraction mechanism is a central part of the speech perception process has been elaborated by Carsen-Berndsen (1998). Assumptions from both accounts could be incorporated into FUL, to explicitly dealing with the temporal aspects of the lexical representations as well as with speech production and the perception.

Frequency effects that occur in language production can also be included in FUL. There is evidence for the faster production of high frequency words compared to words with lower frequency (Caramazza et al., 2001). This frequency effect does not follow directly from the models' architecture, but it is also possible to explain it straightforwardly, paralleling the assumptions for speech perception in the section above.

What is crucial about the FUL model regarding the upcoming chapters is its separation of phonetics and phonology. While phonological (rule-based) variation is assumed to be dealt with by listeners without problems, phonetic variation (in a rather random fashion) can result in problems during the process of speech recognition.

2.3 Exemplar Models

In the previous section, the FUL model as example of a single representation abstractionist model has been presented. If the two possibilities of representational assumptions are thought to be located on a continuum, such an abstractionist model with a single representation for each word is on one extreme end. Pure exemplar models, on the other hand, are located at the opposing extreme of this continuum. Such models maintain drastically different assumptions, not only with respect to lexical representations (e.g. Johnson, 1997; Lacerda, 1997; Goldinger, 1998; Pierrehumbert, 2001a, Goldinger & Azuma, 2003; Hawkins, 2003) but also with respect to lexical access and speech production.[15] Exemplar models are usage based models, in the sense that usage of language – both perception and production – are crucial for shaping a speaker's grammar (see also Langacker, 1987, 2000 for a definition of the term "usage-based approaches"). The label "exemplar models" encompasses many different approaches (cf. Johnson, 2007). What they all have in common is that they are assuming lexical representations with many exemplars for each lexical item. They differ, for example, in the assumption of how many exemplars are stored in memory, or how speech is produced (see for example Hintzman, 1986; Johnson, 1997; Goldinger, 1998; 2007; Pierrehumbert, 2003; 2006b). Exemplars,

[14] There are also aspects in this kind of underspecification that set apart the two models drastically. At this point, however, the focus has been laid on the possibility to include loose temporal order into the FUL model.
[15] Exemplar-based models are not unique to phonetics and phonology. They have been proposed also for other linguistic areas, such as syntax and morphology (cf. Gahl & Yu, 2006).

that is the storage of many detailed instances of encountered words with ample phonetic detail, is a feature of all these models.

The model that is presented in this section is one single example for all of the approaches of the episodic design. Again, it is a rather basic model. This basic model is the X-MOD model proposed by Johnson (1997), a model that many subsequent approaches used as their respective point of departure (cf. Pierrehumbert, 2006b). X-MOD was proposed as extension to the LAFS model introduced by Klatt (1979), which was a single entry model (cf. Johnson, 2004a). In X-MOD, Johnson applied basic assumptions from exemplar theories to speech perception (e.g. Hintzman, 1986; Nosofsky, 1988; Kruschke, 1992). The following summary depicts this model as proposed by Johnson (1997) with some ramifications – for example for production, as advocated, by Goldinger (1998) or Pierrehumbert (2001a).

2.3.1 Basic Assumptions

2.3.1.1 Representation: Multiple Exemplars with Fine Phonetic Detail

Exemplar-based models do not see variation as a problem for speech recognition in the same sense as models assuming abstract representations. Due to their architecture, variation is directly built into the lexical representations. Whereas the lexical representation of FUL is rather "poor" in that only one very abstract featural entry is assumed for each word, exemplar models have very rich and multiple entries. Indexical information such as speaker identity is stored along with many other properties of words in exemplar models. Normalization processes that reduce the variation and a matching to single abstract entries therefore become superfluous (e.g. Jacoby, 1983; Johnson, 1997, Johnson & Mullenix, 1997). Very basic exemplar models assume that an exemplar is created every single time a token is encountered in speech, no matter how similar it is to already existing memory traces (e.g. Hintzman, 1986). However, X-MOD – as most models do – attenuates this assumption. There are different reasons plausibly explaining why not every single time a token is heard, a new exemplar is created in memory. First, one can doubt whether there actually would be enough processing power and space in the brain for every single word heard from birth to death (viz. the "head-filling-up" problem, cf. Johnson, 1997). Besides, and even more important, it is not only resource-demanding to store every exemplar, but also searching through them whenever a new exemplar is heard during speech perception. Another possible reason that explains why listeners do not create an exemplar representation upon every word encounter is that it is not learned or ignored (cf. Hintzman, 1986). Forgetting is also a factor that diminishes the number of episodes that are stored in long term memory. Memories are known to decay over time (e.g. Squire, 1986; Sloman et al., 1988; Pierrehumbert, 2001a). Forgetting can be modeled as probabilistic function, possibly affecting all exemplars alike, or as a process that is mainly targeted at older exemplars, since these are remembered less vividly for independent reasons, or exemplars that are rarely used. Finally, after having stored many exemplars, a new token may be extremely close to another one that is already stored in memory. The granularization (or resolution capacity) of the memory does not realize the difference if it is sufficiently small. Thus, tokens that differ with less than one just noticeable difference (JND) from each other will not be stored separately. It is plausible to assume that in such cases, listeners will not create a new exemplar, but strengthen the representation of an already existing exemplar (Pierrehumbert, 2001a). Due to the high amount of acoustic properties that are stored alongside the exemplars, they really have to be close in order to be stored as one exemplar. The question of how small or large the JND's are, clearly depends on the size of the granularization of the memory and on the level where the memory trace is created, that is, more closely to a segment, or a word, or even a phrase. "As a result, an individual exemplar – which is a detailed perceptual memory – does not correspond to a single perceptual experience, but rather to an equivalence class of perceptual experiences" (Pierrehumbert, 2001a). Although X-MOD does not assume that every event creates a trace, an exact description of how many exemplars are actually stored in the lexicon is not provided either.[16] Arguably, each time a new item is heard an exemplar is created. Newness can arise for two possible reasons. Either it is a word that has never been heard before. The listener does not find any corresponding similar entry and adds this episode to the lexicon. If possible, also semantic information and morphological features are stored along with the exemplar. The second possibility for a new entry is that the word itself has been heard before, however, this instantiation is noticeably different from any prior exposure, consequently leading to a new episode in memory. This can occur if a talker with a remarkable voice, or a special kind of pronunciation is uttering the word, or when the word has particular features, such as rare reduction processes that set apart this utterance from any other utterance of this word.

Generally, categorization in X-MOD occurs through association between a set of auditory properties and a set of category labels (Johnson, 1997: 147).[17] These associations include various properties of the acoustic signal that are stored along with the exemplar, such as information about the speaker gender, or even the speaker identity along with linguistic information (semantic content). Exemplars can be considered as link between the acoustic input and category labels. Figure 4 illustrates these assumptions.

[16] Actually, most exemplar models do not explicitly state a clear definition of what creates an exemplar that is completely new or adds to the strength of an already existing episode.

[17] The exact nature of acoustic properties is not defined in Johnson's original proposal (cf. Johnson, 1997). For the sake of this dissertation, it does not matter whether formant values are stored, or whether acoustic properties are represented by other measures. It suffices to assume that there is little loss of information between an actually encountered episode and the resulting trace in memory.

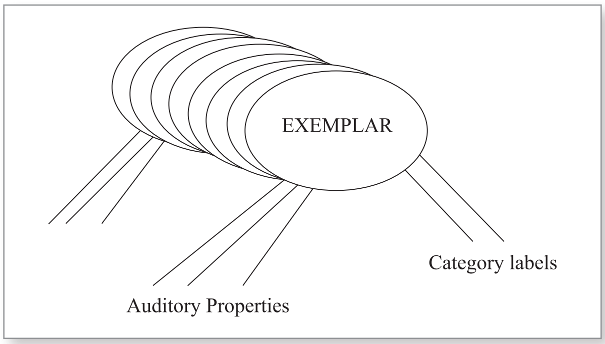


Figure 4: Categorization in the X-MOD model, as proposed in Johnson, 1997

Categories are represented by clouds of episodic memory traces that have been labeled as member of the respective category. When an exemplar is to be categorized anew, its acoustic properties are compared to every stored exemplar. The similarity of the incoming exemplar to other already stored episodes is calculated determining the activation level of the incoming exemplar. If the exemplar creates a good match, its level of activation gets high (cf. Johnson, 1997). Subsequently, the sum of activation from a complete category determines into which category the new exemplar is categorized. Exemplars are stored on a cognitive map (Johnson, 1997; Pierrehumbert, 2001a). Acoustic similarity thus also determines the location of representation: if two tokens are similar, they are represented more closely together than dissimilar tokens. Representations keep their auditory properties, as for example certain formant values, F0 values, or information about the speaker, even after categorization and representation. This procedure is able to create categorical decisions in continuous representations. The following example is an adoption of an example from Pierrehumbert for a new vowel token (2001a) that has to be categorized as either [iː] or [eː]. It is described only for one acoustic property (i.e. F1) but can be also applied also to a multidimensional categorization with many acoustical properties. For the sake of simplicity the assumption is that [iː] and [eː] are assumed to differ only in F1. Figure 5 depicts the situation. The x-axis symbolizes F1, whereas the y-axis shows the activation level of each exemplar that has been already stored in memory. Exemplars for [iː] have dashed lines, whereas exemplars of [eː] are drawn in solid lines. The new exemplar's [V?] acoustic location is symbolized by '*', the window that is taken for comparison is indicated by the black frame. As can be seen, there is a remarkable amount of variation stored in the lexicon. The new vowel falls into a region where there are exemplars of both vowels stored. Thus, a situation of ambiguity arises. Below the axis there is a window over which similarity is computed. Within this comparison window, there are more exemplars of [eː] that also have a higher level of activation. Therefore, the new exemplar will also be categorized as [eː] (cf. Pierrehumbert, 2001a). As can be seen from the example, in contexts with a possible uncertainty, exemplar models predict a tendency towards the more frequent categories, which has also been shown experimentally (cf. Pierrehumbert, 2001a).

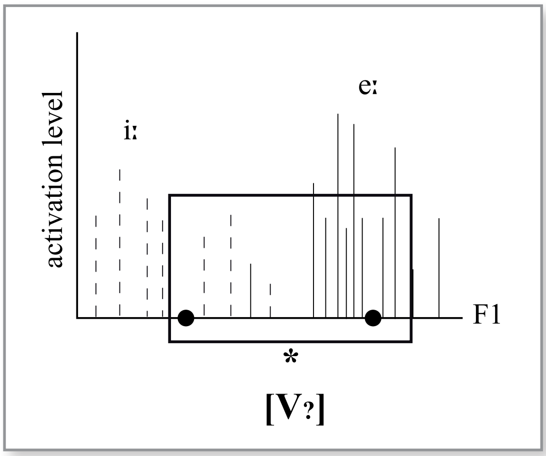


Figure 5: Categorization of a new vowel exemplar as in Pierrehumbert (2001a)

Although there are no prototypes stored in the lexicon, X-MOD can account for prototype effects, depending on the distribution of experienced episodes. Generally, prototype accounts are crucially different from exemplar models in at least one respect. In models assuming such prototype storage, what is stored are not necessarily single episodes, rather, the most important unit is some average that is built from encountered exemplars (in Figure 5, hypothetical prototypes are indicated by small circles). In the above example, a prototypical [iː] could be one with a F1 of the average F1 of all observed [iː] and for [eː] accordingly. This means that there is not necessarily an existing episode with this characteristic. On the other hand, exemplar models only assume actually encountered items as represented in memory, no prototypes are created. Nonetheless, they can handle prototype effects because a new exemplar in the center of a category, surrounded by exemplars with a high activation strength will be a perfect representative of that category, irrespective of whether it has been encountered before or not and appears therefore as prototypical. And X-MOD is also able to explain why extreme exemplars are judged to be the best ones, because there are fewer competitors in the extreme regions of an exemplar. This is true even if they are usually not encountered in real life situations (cf. Johnson, 2007).

Since similarity to existing items is crucial, and auditory properties such as speaker identity are retained in representations, words uttered by the same speaker or a speaker with a similar voice are most likely to be categorized closely together. Additionally, because speaker identity is a stored acoustic property, it is possible that exemplars are grouped together according to these properties. For instance, all words uttered by 'speaker X' would cluster together. Therefore, X-MOD indirectly compensates for speaker variability (cf. Johnson, 1997: 149). Of course, in addition to phonetic information, non-phonetic information (e.g. semantics) is stored as well (cf. Johnson, 2007).

One of the obvious strengths of exemplar theory is the ability to model frequency effects in an elegant way. Frequency effects are pervasive in speech in general and also in language processing (e.g. Bybee, 2000a, b, 2001, 2002, 2007; Bybee & Hopper, 2001; Pierrehumbert, 2001b; 2006a). Accounting for frequency effects comes at no cost for episodic models, since the storage of encountered and produced exemplars will always have an influence on subsequent speech processing, and thereby frequency effects are handled intuitively. When categories are encountered more frequently in speech, they are represented by more exemplars and become stronger (e.g. Johnson, 1997; Pierrehumbert, 2001a). Remember that each exemplar has an associated strength, or resting activation level, that is dependent on frequency. High-frequent and more recent experiences have a higher resting activation level than low-frequent or remotely encountered exemplars (e.g. Pierrehumbert, 2001a). The storage of exemplars has another elegant characteristic, indirectly connected to frequency. Since words are stored with a lot of exemplars and many acoustic properties, word-specific variation patterns are part of the lexical representation. Word-specific variation is also known to be affected by frequency, in that high frequency words tend to be lenited more often and to a higher degree than infrequent words (e.g. Bybee, 2000b, 2001). However, even when frequency does not play a role for word-specific variation patterns in that the unreduced variant is more frequent, it still is inherently part of the lexical representation. Effects of frequency can also lead to the assumption that infrequent phonemes can only survive when they are sufficiently distinct from other phonemes (e.g. Pierrehumbert, 2006b: 524). More frequent phonemes have an influence on the categorization of less frequent phonemes, since they are represented by more exemplars, thereby more easily attracting phonemes that fall in an area of uncertainty between two categories, as in the example of categorization discussed above.

An important characteristic of X-MOD is that it is not a segmental model (cf. Johnson, 2004a). By virtue of its exemplar storage with acoustic properties, and the possibility of multiple categorization and labeling, the model allows for both very fine grained basic units, such as acoustic formant values or single sounds, and for the representation of larger chunks, such as syllables, words or even phrases (cf. the different assumptions of e.g. Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001a; Local, 2003; Wedel, 2006; Wade & Möbius, 2008). X-MOD is not necessarily restricted to one basic unit, but as a basic assumption, words are stored as exemplars (Johnson, 1997; 2007).[18]

[18] It is always desirable if a model or a theory makes explicit claims and assumptions and thereby becomes falsifiable. X-MOD can be described by mathematical formulas, making it very explicit. In these formulas, parameters such as attention weights and sensitivity variables are represented in the equations (cf. Johnson, 1997, based on Nosofsky, 1988).

One critical feature of X-MOD and this is true for all pure exemplar models, is that if words are regularly encountered in a reduced manner, there should be an advantage for the reduced variant over the more canonically produced item. This assumption will also be tested in Chapter 4 of this dissertation.

2.3.1.2 Speech Perception

Speech perception for exemplar models reflects the architecture of the lexicon. Since every time speech is recognized, an exemplar is stored, the processes of perception and the creation of lexical exemplars are also very similar. As in FUL, there is no intervening level of representation between the acoustic input and the lexical representation in memory. The acoustic input (the speech signal) is transformed into an auditory neural spectrogram, which in turn is compared to all stored exemplars (Johnson, 2004a). Consequently, the exemplars that created the highest amount of activation compared to the incoming candidate will succeed in speech perception.

Prior to X-MOD, exemplar models have been assumed and tested mainly in the visual domain and with non-speech simulations where stimuli were not time-varying (cf. Johnson, 1997). However, an important feature of speech is its inherent time variability. Therefore, a time parameter has also been included in X-MOD. The inclusion of the time factor increases the complexity of the approach. At the same time, if speech is to be accounted for, it is an absolutely necessary assumption. After a short period of time (about 10ms) auditory properties are evaluated from the speech signal. A similarity match between these properties and exemplars is carried out subsequently. However, the assumption is that not at every time frame the matching is carried out anew, only when changes in the signal are recognized by a surprise detector (Johnson, 1997: 152). This surprise is possible due to syllables, words, etc.

Since X-MOD allows for fine phonetic detail (and thereby variation) to be stored in the representation of a lexical item, due to its very architecture, variation is seen as informative and "extensively" used during speech perception. If a word is encountered that has been uttered already by the same speaker or one with a similar voice, the activation of that exemplar is higher as for words uttered by dissimilar speakers and words that have not been encountered before. Exemplars that have been encountered frequently and recent episodes are recognized faster and better all else being equal. This assumption about speech perception is also crucial for the experimental evidence presented in Chapter 4 of this dissertation.

Additionally, also non-phonetic information that is stored with exemplars is activated through a resonance mechanism that is targeted at non-phonetic properties of exemplars (cf. Johnson, 2007). Resting activation of exemplars that fit the topic of sentences get increased and thereby make it easier to be recognized if top-down information (i.e. fitting to the topic) and bottom-up information (i.e. matching to an exemplar, or cloud of exemplars) are converging (Johnson, 2007).

One of the original goals for X-MOD was to model talker normalization in speech perception (Johnson, 2007). Johnson indicated that speaker normalization is possible without recursion to an actual process of normalization in an exemplar based approach (Johnson, 1997). Since variation is stored in the lexicon, as are exemplars of many speakers, acoustic properties of new speakers can be perceived as close to or identical to episodes that have been already stored in memory.

2.3.1.3 Speech Production

In its most basic outline, X-MOD is focused mainly on speech perception (e.g. Pierrehumbert, 2001a). This is true for FUL as well. Both models, however, can easily be used for the modeling of speech production. Additionally, Pierrehumbert (2001a) demonstrated how exemplar models such as X-MOD can be extended to account also for speech production phenomena. In X-MOD, speech perception and production are closely linked. The decision to produce a certain category leads to the activation of this label. Subsequently, an exemplar from that category can be randomly chosen and then be produced (Pierrehumbert, 2001a).[19] The stronger the exemplars are, the more likely they are produced. However, the production of the chosen exemplar will not usually be "perfect", resulting in additional variation (Pierrehumbert, 2001a). This variation can be modeled differently, resulting in random variation, systematic bias and entrenchment. Even neutralization processes can be captured by this kind of modeling (Pierrehumbert, 2001a).

A further assumption in the X-MOD framework is that the production-perception link is based on one's own speech. A crucial result of this basis is that incoming speech from other speakers will be assumed to be produced according to one's own gestural knowledge, so called ego-exemplars (cf. Johnson, 1997). This gestural own-interpretation can be assumed to be parallel to the difference in personal semantic and pragmatic meanings two different speakers have about identical words and phrases. The result is in the case of semantics a close approximation of what the partner in a conversation conveyed, however the approximation is not necessarily perfect (Johnson, 1997: 154). Additional assumptions concerning speech production are elaborated by Pierrehumbert (2001a).

---

[19] The assumptions depicted here are really basic. More elaborate choice rules and procedures can be imagined. However, for the purpose of this dissertation, the basic assumptions will be sufficient. Alternative models also exist (e.g. the PEBLS model by Kirchner & Moore, 2008) but will not be regarded any further.

The most important feature of speech production in X-MOD is that variation is also inherently rooted in the model. Since encountered episodes are the basis for production, even the same speaker will exhibit a huge amount of variation. The differentiation between phonetics and phonology in this kind of model is not that important. Rather, phonological systematicity rather is the co-product of phonetic exemplar storage and labeling, both for production and perception.

After having laid out the basic assumptions and the characteristics of the models that are juxtaposed in this dissertation, they will be evaluated in their ability to predict and explain data from conversational German in the upcoming chapters. Chapter 3 concentrates on regressive assimilation of PoA across word boundaries, whereas Chapter 4 investigates massive reduction phenomena in natural German. The crucial predictions of the models will be repeated and made more explicit for each of the processes that are examined.

«…The little things are infinitely the most important»
Sherlock Holmes *(Arthur Conan Doyle, A Case of Identity)*

Chapter 3 – **A Case of Phonologically Based Reduction? Regressive Assimilation of Place of Articulation**

3.1 Introduction

In the previous chapter, FUL (cf. Lahiri & Reetz, 2002) and X-MOD (cf. Johnson, 1997) have been presented, their basic assumptions have been laid out, and important characteristics have been highlighted. In the third chapter of this dissertation, the models and their predictions will be compared with respect to natural speech data. Assimilations constitute one of the reduction processes that are distinctive for variation in natural speech. They are phonological and phonetic processes occurring in many languages of the world. The results of these processes may be possible neutralizations of featural contrasts.[20]

The most common and basic explanation for assimilation is that two neighboring sounds become more similar to each other in the process of speech production. Not surprisingly, assimilation is not only occurring in most of languages of the world, it is also one of the most thoroughly investigated phonological and phonetic processes starting from more than a hundred years ago and attracting research interests until today (e.g. Sweet, 1877; Jespersen, 1922; Nolan, 1992; Kenstowicz, 1994; Jun, 1995; Wood, 1996; Snoeren et al., 2006; Dilley & Pitt, 2007; Hayes, 2009; to name but a few). There are phonological assimilation processes that occur independently of speech rate (from slow to fast) and independently of speech register (from formal to informal). For example, in Russian, obstruents assimilate progressively in voicing (cf. Hayes, 2009). However, there are also well known assimilation processes that do not always occur, even if the context would trigger it. Such an assimilation process exists in French, where word final obstruents may or may not assimilate in voicing to initial segments of upcoming words (cf. Snoeren

[20] This chapter, especially concerning the data, is based on Zimmerer et al., (2009).

et al., 2006). It is this latter kind of assimilation process that poses a possible challenge to traditional linguistic theories, both for production as well as for perception. This is due to the fact that on the one hand it is hard to predict when it occurs and when it does not, and on the other hand, listeners' expectations have to fit the speech input. In addition to the optional nature of assimilations, there is also a debate about whether several of these assimilation processes are only incomplete assimilations, that is, coarticulation processes where acoustic information from the target and the triggering segment are still present. It has been claimed that many assimilation processes, such as English assimilation of PoA do not result in a complete neutralization (cf. "near-neutralization", Hayes, 2009, e.g. Nolan, 1992; Gow & Hussami, 1999; Gow, 2001, 2002; Gow & Im, 2004; for recent results, see Snoeren et al., 2006; Dilley & Pitt, 2007). Despite the optional and gradient nature of fast speech processes, neutralization due to assimilation can be perceived as complete. One piece of evidence comes from historical developments of English, where assimilations may even lead to orthographic changes. Orthography tends to be very conservative and even if a pronunciation change has occurred, the spelling often remains unaltered. However, when the orthography changes (without formal institutional intervention such as the German *Rechtschreibreform*), one can be relatively sure that some change has in fact taken place. For instance, words with the negative prefix {in-} have been borrowed into English from Romance at different times. A word like impossible could be spelt earlier as <inpossible>: *It es bot foli al þi talking, And als an inpossibile thing* **1300** Cursor M. 14761 (OED, 1989: 732). The <n> is now always pronounced as a [LABIAL] nasal. Hence this place assimilation which changed the [n] of {in-} to [m] when [LABIAL] consonants [p, b, m] followed is now always reflected in spelling. Listeners must have perceived the assimilation which then has led to a change in the (conservative) orthography. A new formation like *input*, which does not consist of the negative prefix, preserves the <n> in spelling, although it is also pronounced with a labial nasal [m].

This chapter investigates regressive assimilation of PoA across word boundaries in natural German. The remainder of this chapter is organized as follows. First, a corpus analysis is provided, which explores the actual amount of regressive assimilation in natural German. In the second part of this chapter, results from two experiments (a forced-choice phoneme identification task and a free transcription experiment) are reported that shed light on the question whether assimilated segments can be differentiated from underlying segments by native German listeners.

3.2 Corpus Analysis of Regressive Place Assimilation Across Word Boundaries

The process that is investigated in this part of the dissertation is a prime example of an assimilation process that is both optional and gradient in nature: regressive assimilation of PoA in German (cf. Kohler, 1995a; Wiese, 1996). Generally speaking, as already briefly addressed in Chapter 1, speakers seem to be rather careless and inconstant in their speech production. Evidence for this carelessness can be found easily in the Kiel corpus (IPDS, 1994). There is an enormous amount of several variation processes that can be observed even in the pronunciation of single words. Consider for instance the German word *einverstanden* ('agree-PAST PARTICIPLE') occurring 47 times in the Kiel corpus (IPDS, 1994). This word has 23 different variants in the database (for a complete list see Appendix A).[21] Actually, as for the example of *irgendwie* ('somehow') presented in Chapter 1, there is no single utterance in the corpus that is exactly matching the canonical pronunciation [ˈʔaɪnfɐˌʃtandən]. In most of the cases, the pronunciation departs from the perfect canonical pronunciation by more than one deviation. Not only are there many types of variation, but again, they are optional and need not be complete, hence, they may still be perceptible because of this remaining acoustic information. Remnants of a deleted sound may still be present as in [ˈa̰ɪ̰nfɐˌʃtanʔn], or [ˈa̰ɪ̰nfɐˌʃtan̰]), where apart from the seemingly complete deletion of some segments, glottalization indicates that a [CORONAL] stop (i.e. [d]) has been drastically reduced.[22] The corpus' transcriptions of place assimilations suggest that there exist complete neutralizations of a featural contrast, exemplified by a variant of *einverstanden*, which is produced as [ˈaɪmfɐˌʃta̰nn]. In this variant, among other variation processes, the [n] is assimilated to the labiality of [f].[23] However, this complete assimilation only reflects what transcribers decided when faced with a binary choice for [n] and [m] during their (rather broad) phonetic transcription. As in the example of /d/-deletion, there may still be some traces of the original [n] in the signal, despite the categorical transcription.

In this vein, Nolan (1992) argued that assimilations were more likely to be gradient than complete, with remnants from underlying segments still present in the articulatory gestures of a segment. This view is also taken by Gow (2001, 2002, 2003). Indeed, some researchers express doubt concerning the very existence of complete assimilation (Gow, 2002; see also Snoeren et al., 2006). For Bulgarian, a study by Wood, (1996) using X-ray data, investigated assimilation (palatalization) of alveolar stops. His findings suggest that assimilation (palatalization) is neither pure coarticulation nor complete neutralization: It

[21] The "-h" symbol in the Kiel transcriptions has been ignored, because it has many correlates in reality (e.g. aspiration, release). These are not relevant here. The corpus transcription (SAMPA) has been converted into common IPA transcription.
[22] Glottalization was not treated as an instance of complete deletion, rather as some remnant of a severely reduced segment in order to keep the two processes apart.
[23] Neutralizations occur when speakers eliminate contrastive featural contrasts of segments in speech production. For instance, when they produce a segment such as /n/ – underlyingly [CORONAL] – as a [LABIAL] [m] due to a complete assimilation to the place of articulation of an upcoming [LABIAL] segment, such as [b]. "Complete" means that the resulting [m] (underlyingly [n]) is not different from an underlying /m/ being produced as [m].

appears to be a preplanned process with the purpose of reducing conflicting movements for the articulator, while at the same time, the resulting gesture is different from both target places (Woods, 1996). For French, Duez (1995) has shown that in spontaneous speech, place of articulation contrasts are not neutralized completely.

The view that assimilations are never complete is also not unchallenged. In a recent extensive coverage of regressive assimilation of naturally spoken American English (Buckeye Corpus of Conversational Speech, Pitt et al., 2007), Dilley & Pitt (2007) found that 9% of CORONAL (alveolar) word final stops and nasals were transcribed as assimilated to the place of articulation of the following consonant (LABIALs and VELARs).[24] Acoustic measurements consisting of the change in the second formant (F2) and amplitude of the preceding vowel showed that these frequently did not differ between the assimilated consonants and the canonical labials and velars. They conclude that "assimilation is often complete or nearly complete in spontaneous speech" (Dilley & Pitt, 2007: 2350). One must note, however, that the F2 values were gradient for both consonants labeled as assimilated, as well as for those in an assimilatory context (i.e. followed by LABIALs or VELARs) as compared to alveolars in a non-assimilatory context (i.e. followed by other alveolars). As the authors report, a possibility exists that the real number of assimilations is underestimated, since even some instances of those that were labeled as unassimilated could be actually assimilated, because the labelers are always reasonably conservative (Dilley & Pitt, 2007). Ellis and Hardcastle (2002) provide further important findings concerning assimilation from an articulatory basis. Regarding the completeness of assimilations, their results were somewhat inconclusive, however. They found evidence both for gradient assimilation as well as for complete neutralization. What made their results even more interesting is that their study revealed speaker-dependent strategies: whereas some talkers seem to produce gradient, incomplete assimilations, there were also speakers showing complete assimilations with no residue of the underlying place of articulation left (Ellis & Hardcastle, 2002). Kim and Jongman (1996) demonstrated another case of complete neutralization. In Korean, word-final [CORONAL] obstruents completely neutralize to [t]. This finding was supported by a perception experiment where listeners could not reliably tell the (underlying) final segment of the words they heard (Kim & Jongman, 1996). Another area of research where there is a debate about the (in)completeness of the neutralization of a (featural) contrast is the process of final obstruent devoicing, as it occurs in German and other languages (e.g. Port et al, 1981; Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985; Piroth & Janker, 2004; Warner et al., 2004; 2006). Some researchers have found a complete neutralization of the featural contrast (e.g. Piroth & Janker, 2004), whereas others suggest that there are still cues of the underlying segments that listeners possibly can use (e.g. Port & O'Dell, 1985). Evidence from Dutch suggests that orthography plays also a role in neutralization

[24] Regressive assimilation occurs when in a sequence of two segments S1 and S2, S1 assimilates in some feature(s) to S2. Progressive assimilation occurs when S2 assimilates to S1.

processes (Warner et al., 2006). For German, there are still too many questions that are not answered satisfactorily. Although progressive place assimilations are reported to be frequent and complete within a word in German, cf. *geben* [gebən] > [gebm̩], regressive assimilation across words is more controversial (Wurzel, 1970; Dressler et al., 1972; Vater, 1979; Benware, 1986; Hall, 1992; Wiese, 1996; but see Kohler, 1995a). As Wiese (1996) states, when it is possible to pronounce two words as a single unit, regressive assimilation is more likely (cf. *man kommt* 'one comes' pronounced as [maŋ kɔmt] versus no assimilation in *der Mann kommt* 'the man comes' [man kɔmt]). More definite conclusions regarding regressive assimilations in German is difficult since in his words, "... first, there is little systematic study of such differences, and, second, at the tempo of fast speech, assimilation is certainly possible in the latter example" (Wiese, 1996: 221). Nonetheless, regressive assimilations across words are not unknown and the possibility of this process is at least mentioned by most of these authors.[25]

Kohler, however, explicitly claims that regressive place assimilation takes place across word boundaries (Kohler, 1995a: 206; see also Kohler, 1990) and cites several examples where such assimilations occur. One such example is *bunt machen* 'to make colorful' [bʊnt maxŋ] being pronounced as [bʊmp maxŋ]. A study on the Viennese variety of German by Dressler and his colleagues also mentions the possibility of regressive place assimilation in fast speech (Dressler et al., 1972). Thus, despite the increasing number of spoken language corpora which are used in recent publications, such as in Snoeren and colleagues (2006) and Dilley and Pitt (2007), there is still a dearth of statistically reliable data as to what extent connected speech phenomena like assimilations actually occur in other languages. Moreover, even less is known about how they are perceived by normal listeners and trained phoneticians.

Assimilation predictions

Before the results of the corpus analysis are reported, the predictions for the two models, FUL and X-MOD are discussed. The repercussions of reduction processes due to assimilation are modeled differently in FUL and X-MOD, their prediction also sets them apart. For X-MOD, assimilation processes are not problematic, neither for production nor for perception. Concerning production, it does not make a real difference whether they are complete or not. The only thing that matters is frequency of occurrence and acoustic similarity to existing exemplars in the lexicon (cf. Johnson, 1997; Pierrehumbert, 2001a). X-MOD does not make *a priori* predictions as to the completeness of neutralization. Everything depends on the amount of variation that speakers of a given language (i.e. German)

[25] Along with Wiese (1996), Benware (1986) sees the 'phonological word' as the only domain where regressive place assimilation can occur. He cites Kallmeyer (1981) for a case of regressive place assimilation in *kaputt gegangen* 'has broken down', where the final /t/ of *kaputt* 'ruined' is pronounced with a [k]. The phrase *kaputt gehen* consisting of two words is interpreted as a single 'phonological word' in the sense that they form a very close unit, different from usual words in a phrase (Benware, 1986: 129).

produce. The more variation to be observed, the more likely are speakers to also vary in their pronunciation. For X-MOD, frequency of occurrence and acoustic properties are decisive. Since X-MOD, and virtually all exemplar-based approaches have incorporated phonetic processes into the lexical representation, the need to discriminate between complete and incomplete neutralization becomes superfluous. Articulatory explanations such as salience of PoA cues explain why [CORONAL] segments assimilate to neighboring segments but not *vice versa* (e.g. Ohala, 1990; Kohler, 1991). The degree of assimilation is not of primary interest for X-MOD. What is crucial for X-MOD is the frequency of occurrence of assimilations. Depending on how often assimilated segments are encountered, or on how coarticulation is an articulatory necessity, and depending on the speech register, representations of assimilated sounds are created in the lexicon, hence, they will be also produced.

Concerning FUL, the situation is somewhat different. In FUL, there is a clear division between phonological and phonetic processes, their effects may seem to be identical, but this is not necessarily true. For the model, there are actually two different assimilation processes, one phonological and one phonetic. A common misperception of the model in the literature is that there exists the claim that assimilations always have to be complete (cf. Gow 2002, 2003). This is an oversimplification of the model's predictions. The model predicts that there will be cases of complete neutralization. These cases occur through feature spreading, when the PoA feature of a neighboring segment spreads and fills the empty PoA feature slot for [CORONAL] segments. However, there is still the possibility of graded phonetic coarticulation. In these cases, FUL does not predict that the neutralization is complete. Another prediction of FUL is that assimilations occur asymmetrically, spreading is only possible from [LABIAL] or [DORSAL] to [CORONAL], but never vice versa. Therefore, FUL would predict asymmetric assimilation patterns which are not not due to differences in salience and the amount of effort that is necessary for producing featural contrasts, but rather emerge as a phonological process due to the representation of features.

After having laid out the expectations of both models, the results of the corpus analysis are reported. This section is first divided into a separate analysis for function words and a separate analysis for lexical words. There is ample evidence that the phonological and phonetic behavior of these two word categories is different (e.g. Selkirk, 1984; Kaisse, 1985; Nespor & Vogel, 1986; Hall, 1999; Ogden, 1999; Philipps, 2001; Local, 2003; Kabak & Schiering, 2006; Bybee, 2007). For instance, function words in German are often drastically reduced (Hall, 1999). For English, it has been reported that /m/ in the function words *I'm* may assimilate to neighboring segments, whereas /m/ in content words such as *time* do not (Ogden, 1999; Local, 2003). Therefore, it is important to also examine whether function words behave differently when it comes to regressive place assimilation in a comparison of the findings of the two separate sections.

In the analysis, the sequence of consonants across word boundaries is referred to as C_1 and C_2. The word final segment (C_1), which is a possible candidate for assimilation, is also called TARGET and the word initial segment of the following word (C_2) is the TRIGGER. In (2) a schematic example of regressive place assimilation is depicted. Note that C_1 could be any stop, fricative or nasal in German, whereas C_2 could be any obstruent or nasal which may occur in that position. Word finally (i.e. in TARGET position), as indicated already in the discussion of prior research results, voiced stops and fricatives are regularly devoiced in German (*Auslautverhärtung* – 'final devoicing' – see Kohler, 1995a; Wiese, 1996; Hall, 2000, and references therein). Consequently, in production, there are no word-final voiced obstruents.[26]

(2) TARGET (C_1) and TRIGGER (C_2) in word sequences

$[\ldots C_1]_{\omega 1}$ $[C_2 \ldots]_{\omega 2}$

e.g. *machen* *müssen* ('make must')

The analyses of the speech data provide answers to the following questions: how often do German speakers assimilate regressively across word boundaries? Is there a particular place of articulation for word final consonants favoring assimilation? For instance, are [CORONAL] sounds more likely to assimilate than [LABIAL] ones? Does the manner of articulation of the word final segment matter for regressive assimilation such that, for example, nasals assimilate more often than plosives in running speech? Does the place and manner of articulation of the word initial consonant correlate with regressive assimilation? Does the lexical status of the first word (function words vs. lexical words) increase the probability of assimilation, since function words are supposed to be less stable and more vulnerable to alterations?

For the analysis, all possible contexts of regressive place assimilations of nasals and obstruents were counted, and all cases where assimilation actually occurred were summed up. Homorganic C_1-TARGET and C_2-TRIGGER combinations were thus ignored. Additionally, utterances where speakers produced false starts or where technical problems led to incomplete speech signals were excluded from further analysis. If C_1-TARGET or C_2-TRIGGER consonants were parts of hesitational markers such as *ähm* or *m(hm)* ('ahem, hm') like in *machen äh(m) wir*, they were also not included in the analyses. Furthermore, utterances where a possibility of progressive place assimilation existed and thus, target and trigger could not be identified unambiguously were not included. For example, the assimilated [m] in a phrase like *haben wir* ('have we') [haːbən viːɐ] spoken as [haːbm viːɐ] has two potential triggers, the preceding [LABIAL] [b] or the following [LABIAL] [v]. Consequently, such cases were not considered in

[26] Since final devoicing affects all places of articulation alike, there was no differentiation between voiced and voiceless segments. For the analyses reported in this dissertation, it does also not matter, whether *Auslautverhärtung* is a case of complete neutralization or not. It is assumed to be complete, but the results would be still valid if it was not.

the analysis reported here. Finally, cases where the underlying final segment (C_1) was deleted were also not taken into account. This was done to rule out possible confounds connected to deletions. Thus, phrases like *und Mittwoch* ('and Wednesday') [ʊnt ˈmɪtvɔx] pronounced without word final [t] as [ʊn ˈmɪtvɔx] were not included despite that a possible context for assimilations existed.[27] All obstruents and nasals were treated as possible triggers (C_2). The phonological features of the consonants that were taken into account, both as target and trigger, are given in Table 2.[28]

Table 2:
Obstruents and nasals in German and their phonological PLACE *features*

LABIAL	bilabial, labiodental	[m, p, b, f, v, pf]
CORONAL	alveolar, palatoalveolar, palata	[n, t, d, s, ʃ, z, ç, ts, tʃ]
DORSAL	velar	[ŋ, k, g, x]

3.2.1 Regressive Place Assimilation for Function Words

First, the behavior of function words concerning regressive assimilation across word boundaries is presented. Since there is considerable controversy concerning the question of which words count as function words, the classification in the Kiel corpus was followed (in the corpus, function words are marked with a final "+" in their transcription). An overview of the different kinds of function words occurring in the database is given in (3). The function words could be either trigger or target. Note that the target word's syntactic category is ignored.

(3) Examples for different function word categories in the Kiel corpus

a) Auxiliaries: *bin, hatte, gewesen, möchte* ('am, had, been, would like')
b) Determiners: *der, die, das, ein, eine* ('the-MASC,-FEM,-NEUT, a-MASC,-FEM')
c) Pronouns: *ich, wir, Sie, Ihre,* ('I, we, you-HON, you-HON.GEN')
d) Prepositions: *in, am, bis,* ('in, at-DAT, until/to')
e) Demonstratives: *diesen, dieser, diesem* ('this-CASE')
f) Conjunctions: *und, aber, zwar* ('and, but, but/namely')

[27] If there is a deletion and no assimilation on the preceding segment, it is not clear whether the deleted segment itself was assimilated. If the preceding segment assimilates, it is not clear whether the deleted segment triggered the assimilation, or the first segment of the upcoming word.
[28] The features are based on Lahiri & Reetz (2002). Palatals are assumed to be [CORONAL], as in many phonological accounts (e.g. Lahiri & Evers, 1991; Clements & Hume, 1995; Kenstowicz, 1994; for a different view, see for example Hall, 2000). The segments [x, ç] are assumed to be underlyingly placeless since the place of articulation of the preceding vowel determines the place of articulation of the fricative – [CORONAL] after front vowels, [DORSAL] after back vowels. For sake of simplicity, we refer to the underlying fricatives as /x/ or /ç/. Note that the segments [ŋ, tʃ, s, x] do not occur in initial position in German, except in a few loanwords.

Overall, 4144 function words qualified as target (C1) in a sequence of two consonants at word boundaries. Out of these, in 266 (6.4%) instances the target C1 was transcribed in the corpus as having been pronounced with a different place of articulation compared to the canonical form, e.g. *ein Montag* [aɪm ˈmoːntax] instead of [aɪn ˈmoːntax]. Tables 3 i-iii show the data for all occurrences of TARGETS and the corresponding TRIGGERS, with the numbers and percentages of assimilated segments.[29]

Table 3:
C_1-TARGETS and C_2-TRIGGERS for all assimilated function words.
The lightly shaded cells highlight assimilations.

(i) Function words ending in a [LABIAL]

Assimilation	LABIAL > CORONAL						LABIAL > DORSAL	
C_2 Triggers / C_1 Target PLACE [LABIAL]		n	t, d	ts	z	ʃ		k,g
/m/ 27/583 4.6%	m>n 4.3%	1/76	13/204	4/79	5/177	0/4	m>ŋ 4.3%	4/43
/p, b, f, v/ 0/141	p>t	0/18	0/81	0/10	0/32	0/0	p>k	0/0
Sum 27/724 3.7%	23/681 3.4%						4/43 9.3%	

(ii) Function words ending in a [CORONAL]

Assimilation	CORONAL > LABIAL					CORONAL > DORSAL	
C_2 Triggers / C_1 Target PLACE [CORONAL]		m	p, b	pf	f, v		k,g
/n/ 225/1230 18.3%	n>m 16.2%	44/187	33/142	3/4	88/703	n>ŋ 29.4%	57/194
/t, d/ 4/200 2.0%	t>p 2.3%	4/43	0/21	0/2	0/107	t>k	0/27
/s/ 1/1021 0.1%	s>f 0.1%	0/138	0/161	0/1	1/534	s>x	0/187
/ç/ 2/510 0.4%	ç>f 0.6%	0/91	0/56	0/0	2/186	ç>x	0/177
Sum 232/2961 7.8%	175/2376 7.4%					57/585 9.7%	

[29] The fricative [x] is the only [DORSAL] consonant function words end with. Due to final devoicing, only voiceless obstruents occur in C_1-TARGET position.

(iii) Function words ending in a [DORSAL]

Assimilation	DORSAL > LABIAL					DORSAL > CORONAL					
C₂ Triggers / C₁ Target PLACE [DORSAL]		m	p, b	pf	f, v		n	t, d	ts	z	ʃ
/x/ 7/459 1.5%	x>f 2.8%	0/70	0/49	0/5	6/94	x>s 0.4%	0/51	0/98	0/24	1/37	0/31
Sum 7/459 1.5%	6/218 2.8%					1/241 0.4%					

The results clearly indicate that although regressive place assimilation is not an obligatory process, there are final segments of function words assimilating across word boundaries. Table 3 ii indicates that 232 out of 2961 [CORONAL] sounds assimilate in place to the following segment, most of them /n/. Out of a total of 1230 /n/ final function words, 225 or 18.3%, were labeled as assimilated; 168 out of 1036 (16.2%) words ending in /n/ assimilated to [m] and 57 out of 194 (29.4%) changed to [ŋ], when followed by [LABIAL] or [DORSAL] consonants, respectively (e.g. *man* 'you/one' canonically /man/, produced as [mam] or [maŋ]). Out of a total of 200 function words ending in /t/, only 4 assimilated to [p] when a [LABIAL] followed, and none assimilated before [DORSAL] segments. Overall, 1021 function words ended in /s/, one of which was assimilated to a [LABIAL] [f]. Finally, out of 510 /ç/ final function words, 2 assimilated to [f].

Concerning functions words ending in a [LABIAL] sound (Table 3 i), there were a total of 724 of which 27 assimilated, all of which were /m/. There were 583 instances of /m/ final function words and 27 were labeled as having changed its place of articulation, 23 (i.e. 4.3%) to [n] when followed by a [CORONAL], 4 (9.3%) to /ŋ/ when followed by a [DORSAL]. None of the 82 /p/ or 59 /f/ final function words assimilated. As for the [DORSAL] final function words (Table 3 iii), they all ended in /x/, and 7 out of 459 instances (1.5%) showed assimilation – 6 times to [f] when a [LABIAL] followed, and one to [s] when a [CORONAL] consonant followed.

From the data, it also becomes evident that there are clear asymmetries in the patterns of assimilation. [CORONAL] sounds assimilate more frequently (7.8%) than other places of articulation; cf. [DORSAL] (1.5%) and [LABIAL] (3.7%).[30] Another asymmetry concerns the manner of articulation of the targets that undergo assimilation. Nasal sounds are more prone to assimilation than stops, and fricatives assimilate the least. Are these results special to function words or do content words behave similarly? The next section examines these questions.

[30] Almost all the cases of /m/ assimilating to /n/ could also be analyzed as being a wrong case-marking, a phenomenon that is well known for many German speakers (Bayer & Brandner, 2004; Schiering, 2005); *den* 'the-ACCUSATIVE' instead of *dem* 'the-DATIVE' etc. However, here we treated them as any other case of assimilation.

3.2.2 Lexical Words

For lexical words, the speakers of the corpus produced 2916 possible environments for regressive place assimilation. A first striking observation compared to the results for function words is that there were more C_1 [DORSAL] segments. However, this is not surprising since the number of possible word final [DORSAL] segments is higher for content than for function words. Out of all possible environments, 127 (4.4%) assimilations were actually realized. An overview of the different Targets and Triggers is presented in Tables 4 i-iii.

Table 4:
C_1-TARGETS and C_2-TRIGGERS for all assimilated lexical words.
The lightly shaded cells highlight assimilations.

(i) Lexical words ending in a [LABIAL]

Assimilation	LABIAL > CORONAL						LABIAL > DORSAL	
C_2 Triggers / C_1 Target PLACE [LABIAL]		n	t, d	ts	z	ʃ		k,g
/m/ 1/34 2.9%	m>n 3.3%	0/5	1/20	0/2	0/2	0/1		0/4
/p, b, f, v/ 0/42	p>t	0/5	0/26	0/0	0/3	0/2		0/6
Sum 1/76 1.3%	1/66 1.5%						0/10 0%	

(ii) Lexical words ending in a [CORONAL]

Assimilation	CORONAL > LABIAL					CORONAL > DORSAL	
C_1 Triggers / C_1 Target PLACE [CORONAL]		m	p, b	pf	f, v		k,g
/n/ 97/1050 9.2%	n>m 9.4%	14/160	29/285	0/0	46/497	n>ŋ 7.4%	8/108
/t, d/ 24/531 4.5%	t>p 4.9%	18/146	3/143	0/1	2/184	t>k 1.8%	1/57
/s, ç, ʃ/ 0/386		0/68	0/59	0/0	0/165		0/97
Sum 121/1970 6.1%	112/1708 6.6%					9/262 3.4%	

(iii) Lexical words ending in a [DORSAL]

Assimilation	DORSAL > LABIAL				DORSAL > CORONAL					
C_2 Triggers / C_1 Target PLACE [DORSAL]		m	p, b	f, v		n	t, d	ts	z	–
/ŋ, k, g/ 0/342		0/25	0/12	0/78		0/27	0/158	0/13	0/22	0/7
/x/ 5/528 0.9%	x>f 2.5%	0/32	0/37	3/53	x>s 0.5%	0/15	0/343	0/18	2/26	0/4
Sum 5/870 0.6%	2/237 1.3%				2/633 0.4%					

The data for lexical words shows a very similar assimilation pattern to that of the function words. [CORONAL] segments undergo regressive place assimilation in 121 cases, of which 97 were nasals (Table 4 ii). Among the nasals, 8 /n/ (7.4%) were realized as [ŋ]. The rest, i.e. 89 /n/ (9.4%) were produced as [m]. For lexical words, final [t]s accounted for 24 cases (4.5%) of regressive assimilations. Of the 24 instances where [t] was assimilated, there was one utterance where [t] became [k] (1.8%), 23 cases showed assimilation to [p] (4.9%). No [CORONAL] fricative changed its place of articulation. As for [LABIAL] target segments, there occurred one assimilation: A word final [m] assimilated to [n] preceding a coronal stop (Table 4 i). No other [LABIAL] segment assimilated. [DORSAL] segments assimilated 5 times, all of them were [x]; 3 of them assimilated to [LABIAL], 2 to [CORONAL] (Table 4 iii).

Overall, the data of the lexical words also revealed two kinds of asymmetries. First, the nasal consonants assimilated more often than stops or fricatives. The second asymmetry concerns again the place of articulation of the target segment; [CORONAL] sounds undergo regressive place assimilation much more frequently (6.1%) than [LABIAL] (1.3%) or [DORSAL] (0.6%) segments. Before the results of the corpus study are evaluated with respect to the predictions of FUL and X-MOD, an analysis is carried out to compare the behavior of function words and lexical words.

3.2.3 Comparison of Function and Lexical Words' Behavior

Generally, the pattern of assimilation was the same for function words and lexical words, although the former underwent assimilation more frequently. Overall, the Kiel corpus offered 7060 possible C_1-C_2 sequences for regressive place assimilation. Of these, 393 cases of assimilation could be observed (see Table 5); i.e. 5.6% of the possible sequences were actually assimilated. Function words summed up to 266 assimilations, whereas lexical words accounted for 127 instances of regressive place assimilation. However, the sheer number of assimilations is somewhat misleading because function words also occurred more often as targets for assimilation than lexical words. Overall, there were 4144 function words (58.7%) and 2916 lexical words (41.3%) as targets. The percentage of assimilations that actually occurred is therefore drastically less different between the two categories: 6.4% of the function words and 4.4% of the lexical words assimilated regressively. Nonetheless, function words were assimilated significantly more often regressively than lexical words, as a *Chi-Square* test revealed (χ^2=13.9, p<0.001).[31]

Table 5: Assimilation of function and lexical words combined.

C_1 Target				C_2 Triggers					
PLACE	Total	Assimilated		[LABIAL]		[CORONAL]		[DORSAL]	
[LABIAL]	800	28	3.5%	–		24/747	7.0%	4/53	7.5%
[CORONAL]	4931	353	7.2%	287/4084	7.0%	–		66/847	7.8%
[DORSAL]	1329	12	0.9%	9/455	2.0%	3/874	0.3%	–	
Sum	7060	393	5.6%	296/4539	6.6%	27/1621	1.7%	70/900	7.8%

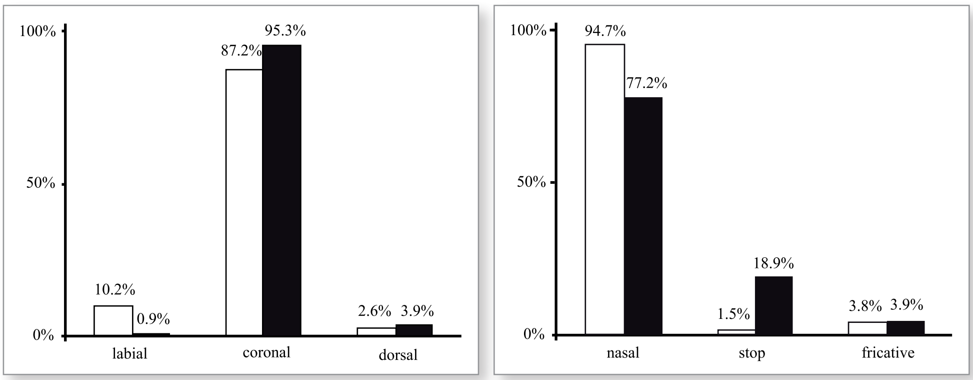


Figure 6:
Relative percentages of regressive place assimilations (based on the total number of assimilated sequences) across different PLACE (left) and MANNER of articulation (right). Function words are represented by light bars, lexical words by dark bars.

[31] If not indicated otherwise, the statistical analysis was calculated using the SAS software JMP (Version 5.0.1.2., SAS, 2002).

Figure 6 depicts the relative percentage of assimilations depending on manner and PoA of C_1 in lexical and function words. For both function and lexical words, nasals are the most frequent to assimilate (350 out of 393 – 89.1%). Stops assimilate in 28 (7.1%) cases and fricatives in 15 cases (3.8%). Concerning PoA, out of a total of 393 assimilated TARGETS, overwhelmingly the [CORONAL] sounds (353 out of 393 – 89.8%) assimilate to the place of a following segment across word boundaries, whereas [LABIAL] (7.1%) and [DORSAL] (3.1%) segments usually do not. In general, [CORONAL] TARGETS (C_1) by far outnumber the other PoA (4931 or 69.8%). The fewest number of TARGETS are [LABIAL] sounds (800 or 11.3%). The only [DORSAL] segment that assimilates is [x] – both in function words as well as lexical words.

At this point, a parenthesis is necessary, because the analysis did not differentiate between C_1 and C_2 sequences that were within one phrase or sequences that crossed phrase boundaries. Out of the 7060 items analyzed in the data, there were 1174 (16.6%), either crossing a period, a question mark, or a comma in the transcription. Of all 18 cases where C_1 and C_2 were separated by a question boundary, none showed assimilation. Concerning periods, a total of 310 sequences occurred in this category. There was one (out of 188) assimilation occurring in a [CORONAL] – [LABIAL] context. Overall a comma separated 848 of the 1174 sequences. In this category, there were 13 assimilations. 10 (out of 319 – 3.1%) occurred in a [CORONAL] – [LABIAL] context, 2 (out of 42 – 2.4%) showed an assimilation of [x] to [f] in front of [f], and 2 cases (out of 441 possible sequences – 0.5%) had an assimilation of [x] to [s] in front of [z]. Thus, although phrase boundaries do impede assimilation, at least regarding commas, there are cases where assimilation even occurs across those boundaries.

3.2.4 Discussion

To summarize, place assimilations across words in German is controversial. Some authors claim that such assimilations do not occur (cf. Wurzel, 1970; Vater, 1979; Wiese, 1996), while others assert the opposite (cf. Kohler, 1995a). This controversy was one of the reasons for a systematic analysis of assimilations across word boundaries in conversational German. The Kiel corpus data suggests that although such assimilations are not frequent, they do occur – overall, approximately 6% of possible assimilatory sequences did undergo a change in place of articulation. Function and lexical words were analyzed separately since they are claimed to be different, and indeed, there is a significant difference in the number of assimilations between the two categories although the assimilation patterns were the same. Function words are more likely to assimilate than lexical words, hence, the former are more prone to reduction than lexical words. Moreover, the data revealed clear asymmetries in the pattern of assimilations that actually occurred. One asymmetry is that the PoA of the targets (C_1) undergoing assimilation: [CORONAL] sounds are more frequently assimilated than [DORSAL] and [LABIAL] consonants. A second asymmetry concerns the manner of articulation: nasals assimilate more often than stops or fricatives. This result replicates also earlier findings (cf. Mohanan, 1993; Jun, 1995).

Coming back to the predictions for assimilations the two models (i.e. FUL and X-MOD) make, the results of the corpus analysis are not very indicative for the performance of either of them. This might be not that surprising, because, as already discussed in Chapter 2, both of the models have their main focus on speech perception rather than on speech production. For production, the predictions have not been far apart, albeit the explanations are rather different. The results show a clear asymmetry in the PoA assimilation patterns in that overwhelmingly, [CORONAL] segments undergo this kind of reduction, whereas [LABIAL], and [DORSAL] segments do not. Both models can explain these results, however the underlying reasons for doing so are rather different. The result found above was in fact predicted and explained by the FUL model. In FUL, only [CORONAL] obstruents are expected to assimilate regressively in PoA, because in their representation, there is no PoA feature specified. The prediction of the model is that there are cases of complete assimilation and that these occur in an asymmetric fashion. The results support these predictions. Almost all of the cases where assimilations of [DORSAL] or [LABIAL] can be observed are attributable to an alternative explanation. Furthermore, their number is clearly much smaller than the amount of regressive assimilation for [CORONAL] segments. Crucially, though, FUL assumes that there are cases of complete neutralization of a contrast. If the transcription of the phoneticians is correct, this prediction of FUL has been found to be correct in natural speech data, even in regressive assimilation that occurs across word boundaries.

For the exemplar-based model, the production results are not that crucial, since variation is assumed to be observable in any case, be it symmetrical or asymmetrical. However, X-MOD is also able to explain the asymmetrical results via the salience hypothesis for PoA cues, assuming that the PoA cues are weaker for [CORONAL] segments than for other PoA features. This salience hypothesis could also be applied for manner asymmetries in that PoA features are assumed to be weakest for nasal sounds. Additionally, the results of the corpus study have also repercussions for the perception in the exemplar-based model. For X-MOD, actually occurring word variants are stored in the lexicon. When new episodes find their way to the mental representation, they are labeled for subsequent retrieval. When, for example, an assimilated segment is encountered during speech perception, the listener is able to infer the underlying PoA information due to additional cues from sentence context (e.g. syntax, semantics, phonology). In such a case even completely assimilated segments would be labeled with their underlying PoA feature. This labeling would subsequently create a more wide spread representation for [CORONAL] segments, because they occur with more variation in natural speech.

For X-MOD, thus, the assimilation patterns are completely phonetic in nature, whereas for FUL, the main focus lies on the role of phonology in regressive place assimilation. One piece of evidence that it is actually phonology playing a crucial role in determining assimilation comes from cross-linguistic observations for assimilations. These observations suggest that assimilation processes are not only determined by phonetics because there exist various such processes that are language specific. Languages also differ with respect to what is allowed and what is not (cf. Jun, 1995). This is true despite the fact that there are cross-linguistic assimilation patterns that can be observed (cf. Mohanan, 1993; Jun, 1995). In a purely phonetic framework, all languages should behave alike.

The analysis that has been presented so far crucially hinges on the transcriptions of trained phoneticians. Despite the fact that orthographic as well as canonical transcriptions could have biased their decisions toward unassimilated segments, they opted for assimilation transcriptions in about 6% of the possible cases. This does not rule out, however, that there were still traces from underlying segments left that could not be transcribed correctly. Transcribers had to decide at a certain point, whether what they heard as segment “X” or “Y”. Thus, this transcription decision maybe underestimates what acoustic remnants from underlying segments could still be present. At least, this is what speech perception research suggests: although some sounds might seem to be assimilated, there may still be residual cues for listeners to identify the underlying segments (cf. Gow, 2002). Therefore, the perception of naive listeners has to be tested and the results have then to be compared to the transcriptions of the trained phoneticians, in order to gain a deeper understanding concerning the completeness of transcribed assimilations. Two perception studies were conducted using selected material from the Kiel Corpus (IPDS, 1994). The first experiment used a forced choice phoneme identification task on word fragments from selected dialogues. The second experiment was based on a free transcription task where subjects were asked to transcribe what they heard. The goal of both experiments was to observe how listeners perceived segments that had been labeled as assimilated in the corpus. If there were any remnants of the original segment, listeners should be able to use that information affecting the speed and accuracy of identification as compared to unchanged segments.

3.3 Perception of Regressive Place Assimilation in German

In the previous section, the amount of regressive assimilation of PoA features has been the object of investigation. Seemingly, there are cases of complete neutralization in conversational German. However, so far, one cannot tell whether they are real or whether transcribing conventions have omitted information that is still present in the speech signal and subsequently used by listeners when exposed to cases of what seems to be complete neutralization of a PoA feature contrast.

Nolan (1992) showed that assimilation processes were gradient and that target information was available in assimilated sequences (see also Gow, 2002; for voicing assimilation see Snoeren et al., 2006). Listeners have been shown to be sensitive to these gradient assimilations in production. In Nolan's study, for example, listeners could identify residual alveolar gestures in 40% of the assimilated tokens (Nolan, 1992: 271). Results by Gow also indicate that listeners use the information of the underlying place of articulation even in segments that auditorily sounded as if they were completely assimilated (Gow, 2002). There are also studies examining different assimilation processes that have shown that assimilations can be only partial and that listeners are sensitive to residual cues left (e.g. Manuel, 1995). Manuel (1995), for example, found that in a sequence [n ð], as in *win those*, where the /ð/ became a nasal, the PoA was not that of a "real" [n], suggesting that some featural information of the original nasal was still available to the listener.

Therefore, the question that is investigated in the following task is: do naive listeners (naive both with respect to the goal of the experiment as well as not having additional information from the context) perceive the assimilated and unassimilated segments from the Kiel dialogues in the same way as trained phoneticians who used speech analysis tools? Therefore, two experiments are carried out to investigate in how far listeners actually perceive assimilated segments as neutralized with respect to PoA. For the experiments, the focus was on nasals (/n/ and /m/) since the choice of assimilated segments was larger than for oral stops and it was possible to choose stimuli from several speakers, thereby lessening speaker dependence (for details see **Materials** below).

3.3.1 Experiment 1: Phoneme Identification

A timed forced-choice identification task was chosen for the first experiment. Subjects had to decide whether the auditory stimuli included either a [LABIAL] [m] or a [CORONAL] [n]. This method was chosen to determine the speed as well as the accuracy of the subjects' decision. The focus of the experiments was not just on the assimilated stimuli, but also on stimuli that had been labeled as "unchanged from the canonical" by the transcribers of the corpus – that is, underlying /n/ or /m/ which were spoken and heard as [n] and [m]. The issue was whether the responses to the unchanged stimuli differed across varying contexts – VOWEL, [LABIAL], [DORSAL], [CORONAL]. The crucial conditions with a set of examples are listed in Table 6. The segmental context from which the stimuli were extracted is underlined twice. Since the CORONAL nasal assimilated most frequently, conditions where /n/ was assimilated to [m] were used.

Table 6: Examples of stimuli with the vowel [eː] for Experiments 1 and 2. Column 1 gives the Kiel transcription. Column 2 provides the orthographic contexts from which the stimuli were extracted and column 3 gives the three conditions – unchanged UNASSIMILATED-/m/ and UNASSIMILATED-/n/, and ASSIMILATED.

Kiel corpus transcription	Example stimuli in orthography	Condition	
		UNASSIMILATED-/m/	
[eːm]	... von d**em** **a**chtzehnten Juni?	VOWEL CONTEXT	/m/-VOWEL
[eːm]	... mit d**em** **B**ericht	LABIAL CONTEXT	/m/-LABIAL
[eːm]	... dann d**em** **D**ienstag …	CORONAL CONTEXT	/m/-CORONAL
[eːm]	… und d**em** **g**anzen Kram.	DORSAL CONTEXT	/m/-DORSAL
		UNASSIMILATED-/n/	
[eːn]	... Freitag, d**en** **e**rsten …	VOWEL CONTEXT	/n/-VOWEL
[eːn]	… für d**en** **B**ericht …	LABIAL CONTEXT	/n/-LABIAL
[eːn]	… in d**en** **d**eutschen …	CORONAL CONTEXT	/n/-CORONAL
[eːn]	... d**en** **g**anzen Tag ...	DORSAL CONTEXT	/n/-DORSAL
		ASSIMILATED	
[eːm]	… über d**en** **B**ericht …	LABIAL CONTEXT	

Materials

The stimuli for the perception task consisted of a vowel-nasal (VN) sequence extracted from real words (CVN or VN), and were taken from 27 different speakers (13 female, 14 male) of the Kiel corpus (IPDS, 1994). At most 5 items were taken from any given speaker. The segmental context was thereby as similar as possible and at the same time it was possible to make the perception task speaker-independent. The two vowels in the VN sequences that were chosen for the experiment had been transcribed as either a mid [eː] or a low [ɑ] vowel. The extracted sequences with [ɑ] form possible words: *an* [ɑn] 'on, at.ACC' and *am* [ɑm] 'at.DAT', whereas the [eːn] and [eːm] sequences do not. A set of sentences from which the [eː] sequences were extracted is given in Table 6 and corresponding [ɑm/ɑn] sequences are given in Appendix B.

Subsequently, VN-items were cut at zero-crossings in order to avoid clicks at item boundaries using both visual as well as auditory information. The first identifiable glottal period was taken as the beginning of the vowel. However, when there was an extensive amount of coarticulation from the preceding segment (i.e. at the word onset), up to four glottal periods were cut off to ensure that the consonantal onset could no longer be perceived. The end of the nasal in the VN-items was determined at points when the amplitude of the waveform dropped markedly or at the beginning of the closure of the following consonant. Thus, the nasal itself was left untouched, but any contextual information in the following closure would have been removed. For each vowel (i.e. [eː]/[ɑ]), 10 [CORONAL]#[LABIAL] assimilated sequences (ASSIMILATED category), and 10 each of unassimilated [CORONAL] (UNASSIMILATED-/n/) and [LABIAL] (UNASSIMILATED-/m/) items were chosen. This added up to 60 different stimuli. The unassimilated items were cut out of different contexts (see Table 6 and Appendix B); three preceded a [LABIAL] consonant, three a [CORONAL] consonant, two a [DORSAL] consonant, and two were originally followed by a vowel. The amplitude of the items was equalized.

Identification predictions

After having presented the way the stimuli have been created for the experiments, it is necessary to make explicit the predictions the two models make. Whereas the predictions of the two models were quite similar for the patterns of assimilation that can be observed in natural speech, they diverge for the effects of assimilation on perception. FUL assumes that there are cases of complete assimilation. For those, the prediction is that there is no difference in reaction time and accuracy for completely ASSIMILATED-/n/ tokens compared to underlying, UNASSIMILATED-/m/ tokens. If there will be a difference in accuracy and reaction time at all, [CORONAL] segments are expected to be different from the other two kinds of segments. This might be a surprising prediction since the PoA feature [CORONAL] could be taken as most informative when encountered in speech perception. However, the prediction follows the feature extraction logic of the basic assumptions of the model, as laid out in Chapter 2. Most descriptions of assimilations suggest that coronal consonants are more vulnerable to variation in the context of consonants with other places of articulation (c.f. Paradis & Prunet, 1991). Consequently, one could expect that [LABIAL] and [DORSAL] C_2 contexts would leave more acoustic traces in UNASSIMILATED-/n/ stimuli than [CORONAL] and [DORSAL] C_2 segments influence UNASSIMILATED-/m/ stimuli. This would make it more difficult for listeners to come to a definite decision for the UNASSIMILATED-/n/ stimuli. Therefore, FUL expects slower reaction times (RTs) for UNASSIMILATED-/n/ in [LABIAL] and [DORSAL] contexts. For items in the VOWEL or homorganic (i.e. [CORONAL]) consonantal context, there is not necessarily a RT difference. Based on the basic matching assumptions, one could speculate whether [CORONAL] items are generally recognized somewhat slower than [LABIAL] items, because even if [CORONAL] can be extracted from the signal, there is no PoA feature this information can be matched to in the lexicon. In so far as the difference between ASSIMILATED-/n/s and UNASSIMILATED-/m/s are concerned, no difference in the speed of reaction is expected, assuming that the ASSIMILATED-/n/s exhibit complete neutralization. However, whether the ASSIMILATED-/n/ items were equally well heard as [m] as the UNASSIMILATED-/m/s depends on whether the assimilation as perceived by the transcribers was reasonably complete. Thus, both the reaction time measures as well as percentage of [m] and [n] responses are important for deciding whether assimilations were really realized as complete or not.

The case is different for X-MOD. For X-MOD, it is rather unimportant whether there are cases of complete assimilation or not. Whatever kind of variation occurs in spoken language will be stored as episodic trace in memory. Therefore, X-MOD predicts that [CORONAL] segments should be recognized faster than segments with other PoA and with higher accuracy even despite the higher amount of variation they occur in. There are several reasons for this. First of all, if there is any information suggesting PoA for [CORONAL] in the signal, the listener can be rather sure that the speaker actually produced [CORONAL], despite the lack of salience in context. Additionally, [CORONAL] segments are the ones that occur most often word finally. Therefore, there are many more exemplars close to the ones the speakers produced, leading to a higher resting activation level. In turn, decision processes are faster. What is not completely clear is the predicted behavior for assimilated tokens and [LABIAL] nasals. There are two possibilities. As indicated in the discussion of the prior section, listeners use context information for labeling segments in memory. Thus, in the acoustic map, there are exemplars both for [n] as well as for [m] close to [LABIAL]. This makes the decision harder in that area for speech perception, especially when items are heard without context. The [LABIAL] segments are rarer than their coronal counterparts word finally. Therefore, subjects should react slower to them compared to [CORONAL] items. Assimilated tokens may fall together with [m] in that they are slower than [n] because they are located within the [LABIAL] area acoustically, which leads to further uncertainty. Or alternatively, they could fall together with [n], since [CORONAL] can be also extracted from the signal if there is no complete assimilation, making decisions easy for [n]. It is not completely clear, however, how exactly listeners will react to assimilated tokens from the corpus. For the experiment, assuming that they are completely neutralized, listeners should treat them like underlying [m], since they have been deprived of further context if X-MOD is correct.

Subjects and procedure.

Overall, 18 undergraduates from the University of Konstanz with no reported hearing disorders participated in the experiment and were paid for their participation. They were tested in groups of five or less and were given oral as well as written instructions. A push-button box with two buttons labeled [m] and [n] was placed in front of each subject. They were instructed to listen to the syllables presented over headphones and decide as quickly as possible whether the consonant was [m] or [n] and press the appropriate button with the index finger of their dominant hand. Before the test began, the subjects familiarized themselves to the task with practice items, but were given no feedback about the 'congruency' of their decisions.[32] Each item occurred 5 times during the experiment, adding up to 300 items presented in a randomized order. The sequence of presentation was as follows. Each item was preceded by a warning tone of 300 ms followed by 200 ms of silence. After each test stimulus, there was a pause of 1500 ms where subjects had time to push the button before the next sequence began. Reaction time measurements began at the onset of the nasal segment. The stimuli were played from a SONY DAT recorder and presented over headphones (Sennheiser HD520II). In the set up, a central experimental hardware box connected the DAT recorder, the response boxes and a Macintosh computer, where the reaction times were recorded (Reetz & Kleinmann, 2003). A single experimental session lasted approximately 18 minutes excluding the practice items.

Results

The responses of all 18 subjects entered the reaction time analysis.[33] Responses faster than 200 ms and slower than 1000 ms were excluded leading to the exclusion of 133 responses (2.5% of the data). None of the subjects showed an exceptionally high number of responses which were too slow or too fast. A subsequent ANOVA with reaction times as a dependent variable and the factors SUBJECT (as random variable), RESPONSE ([m] or [n]),[34] UNDERLYING (UNASSIMILATED-/m/, UNASSIMILATED-/n/, ASSIMILATED), CONTEXT (nested under UNDERLYING) (/n/-CORONAL, /n/-LABIAL, /n/-DORSAL, /n/-V, /m/-CORONAL, /m/-LABIAL, /m/-DORSAL, /m/-V, ASSIMILATED), ITEM (nested under UNDERLYING and CONTEXT), RESPONSE X CONTEXT (nested under UNDERLYING) and UNDERLYING X RESPONSE was calculated using the REML estimation.[35] There was a main effect of CONTEXT ($F(6,5181)=9.03$, $p<0.001$) and

[32] The term congruent is used for responses where the transcription of the corpus was the same as the subjects' decision and incongruent for the opposite case.
[33] The analysis was carried out using SAS statistic software JMP, version 5.0.1.2. (SAS, 2002). Since the interest of the analysis is on the influence of the response on the reaction time, the responses are treated as a factor.
[34] Since we are interested in the influence of the response on RT, response is treated as a factor.
[35] The REML (Residual Maximum Likelihood) estimation does not substitute missing values with estimated means and does not need synthetic denominators; rather the individual factors are tested against the whole model. This method is more conservative than the traditional EMS (Expected Mean Squares) estimation. Results that were not significant did not reach the 5% level.

RESPONSE (F(1,5181)=15.37, p<0.001), and the interaction of CONTEXT X RESPONSE was also significant (F(6,5181)=4.70, p<0.001). SPEAKER and REPETITION were not significant factors in the ANOVA. They are therefore not reported. The Least Square Means of the RT measures for both [m] and [n] responses for each CONTEXT are given in Table 7.

Table 7:
Least Square Means of reaction times for three main categories in all contexts for both [m] and [n] responses [% values are computed for each row by $100 * Nx/(N_{Response[m]}+N_{Response[n]})$]

CONTEXT	**Response [m]**			**Response [n]**		
	N	%	RT ms	N	%	RT ms
UNASSIMILATED-/m/	1643	93.2	536.3	120	6.8	580.4
/m/-LABIAL	467	89.1	535.0	57	10.9	518.3
/m/-CORONAL	523	97.9	531.0	11	2.1	573.2
/m/-DORSAL	310	88.6	547.9	40	11.4	647.8
/m/-VOWEL	343	96.6	531.3	12	3.4	582.1
UNASSIMILATED-/n/	405	23.1	547.1	1346	76.9	547.2
/n/-LABIAL	141	26.8	592.9	385	73.2	553.4
/n/-CORONAL	92	17.6	520.6	432	82.4	528.4
/n/-DORSAL	95	27.2	536.2	254	72.8	570.1
/n/-VOWEL	77	21.9	538.6	275	78.1	536.8
ASSIMILATED – (LABIAL CONTEXT)	1534	87.5	545.8	219	12.5	580.0

Several pair-wise post-hoc comparisons were made for the critical conditions, the interpretations of which are summarized below with individual figures.

(i) Recall that based on the analysis of the Kiel corpora transcriptions, the expected congruent responses are [m] for the UNASSIMILATED-/m/ category and [n] for the UNASSIMILATED-/n/ category. The percentage of congruent responses is revealing. For UNASSIMILATED-/m/ stimuli, 93% of the responses were [m], and only 7% were [n]. In contrast, for UNASSIMILATED-/n/ items, almost a quarter of the stimuli were identified as the opposite [m] – 23% [m] vs 77% [n]. Obviously, listeners had more difficulty with the UNASSIMILATED-/n/ stimuli than with UNASSIMILATED-/m/ items. A *Chi-Square* analysis revealed a significant difference (χ^2=1773.63, p<0.001). The reaction times also reflect the same pattern. For the congruent responses, [m] for UNASSIMILATED-/m/ and [n] for the UNASSIMILATED-/n/, the reaction times across these categories (536 ms and 547 ms respectively) are statistically different (t=2.15, p<0.05). There is a much larger difference between the reaction times for [m]- and [n]-responses to the UNASSIMILATED-/m/ stimuli (536 ms vs. 580 ms, t=2.97, p<0.005). Likewise, there is a significant difference between the incongruent [m]-responses of UNASSIMILATED-/n/ and the [n]-responses of UNASSIMILATED-/m/ (547 ms vs. 580 ms, t=2.04, p<0.05). The RT of [m] or [n] responses to the UNASSIMILATED-/n/ category are essentially identical. This suggests that it was more difficult for the listeners, and hence they were slower, to give [n] responses to UNASSIMILATED-/m/ stimuli when they were uncertain.

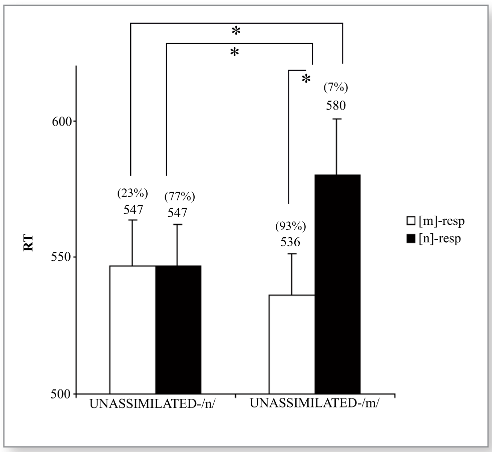


Figure 7a:
[m] and [n] responses to UNASSIMILATED-/n/ and UNASSIMILATED-/m/ stimuli in percent and with their reaction times as bars. Asterisks indicate significant differences in reaction times. White bars represent [m]-responses and grey bars show [n]-responses.

(ii) Since there were four contexts, the next point to address is if any particular context is responsible for the worse identification of UNASSIMILATED-/n/ than UNASSIMILATED-/m/ (see Figure 7b). With respect to percentage of congruent responses, in all contexts more than 89% of the UNASSIMILATED-/m/ stimuli were congruently responded to as [m]. This was not the case for the UNASSIMILATED-/n/ stimuli, where 27% of the responses were [m] in the [LABIAL] and DORSAL contexts. When an UNASSIMILATED-/n/ item was preceding another CORONAL or a vowel, the responses were better comparable to the UNASSIMILATED-/m/ stimuli, viz. around 80% [n] responses. To test whether parallel results are reflected in the reaction times, pair-wise comparisons across all 4 contexts – VOWEL, [CORONAL], [DORSAL], [LABIAL] – were calculated (see Figure 7b). For the [m] responses to UNASSIMILATED-/m/, there were no significant differences in reaction across any of the contexts.

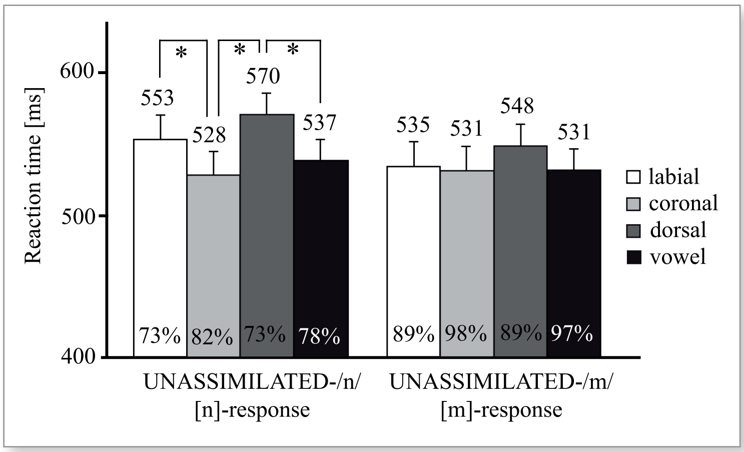


Figure 7b:
[n] responses to UNASSIMILATED*-/n/ and [m] responses to* UNASSIMILATED*-/m/ stimuli differentiated by context. Percent of responses are given in numbers and the bars represent the reaction times with significant differences indicated by asterisks.*

Thus, UNASSIMILATED-/m/ (extracted from *dem, am* etc.) stimuli were heard and reacted to as [m] equally fast regardless of which context they had been extracted from. Would the same pattern for [n] responses to the UNASSIMILATED-/n/ category stimuli occur? Based on the corpus analysis, we know that /n/ is more vulnerable to coarticulation from following consonants with different places of articulation. There could therefore be a difference between the contexts [DORSAL], [LABIAL] on the one hand versus [CORONAL] and VOWEL on the other. In the former contexts, the /n/ may have more coarticulation cues of the place of articulation of the following [DORSAL] or [LABIAL] consonant, making it more difficult to label the UNASSIMILATED-/n/ as [n] in a reaction time task, whereas in the [DORSAL] context, the /n/ is in its "ideal" environment. The pair-wise comparisons confirmed this prediction. The [n] responses to UNASSIMILATED-/n/ in [CORONAL] context differed significantly from the responses to UNASSIMILATED-/n/ in [LABIAL] context ($t=-2.82$, $p<0.005$) as well as from the [DORSAL] contexts ($t=-3.99$, $p<0.001$). Another significant difference emerged in the comparison of the [n]-responses to UNASSIMILATED-/n/ in the [DORSAL] and the VOWEL contexts ($t=2.91$, $p<0.005$). There were no further significant differences between any other contexts for the [n]-responses. Thus, the [n]-responses to UNASSIMILATED-/n/ in the coronal and vowel contexts, which are the most neutral contexts in terms of coarticulation, are significantly different from the [LABIAL] and [DORSAL] contexts. It is therefore safe to conclude that the coarticulation cues from the (deleted) following [LABIAL] and [DORSAL] consonants were strong enough to slow down the subjects' [n] responses to these stimuli. Recall that these consonants had been labeled as [n] by phoneticians who had recourse to both visual and auditory cues and were under no time pressure.

In sum, the [LABIAL] and [DORSAL] contexts had a deceleration effect on the [n] responses for UNASSIMILATED-/n/ stimuli as compared to its homorganic CORONAL context. This effect was not observed for the UNASSIMILATED-/m/ stimuli in the [CORONAL] and [DORSAL] contexts in comparison to its homorganic [LABIAL] context. For the UNASSIMILATED-/m/ stimuli, the subjects' speed and their response were unaffected by the context of other places of articulation, from which one may deduce that there were less coarticulation cues which could confuse them. Thus, there was an asymmetry in the stimuli even where trained phoneticians had transcribed the sounds carefully.

(iii) Remember that for ASSIMILATED stimuli there was only one context, because they were always (by definition) extracted from a [LABIAL] context. The crucial question to gain further insight in how far the assimilations were produced completely is whether these stimuli differ from the UNASSIMILATED-/m/ stimuli in the same context. The UNASSIMILATED-/m/ stimuli in [LABIAL] context can be seen as the most prototypically produced [LABIAL] features without coarticulation and they are taken as clear examples of [m]. Since an effect of coarticulation of the [LABIAL] context in the UNASSIMILATED-/n/ stimuli was found, these were also taken for comparison. With respect to the percentage of congruent responses, the [m] responses to the ASSIMILATED stimuli and the UNASSIMILATED-/m/-LABIAL were almost identical – 88% vs. 89%. Further, there were no significant differences in reaction times in the [m] or [n] responses to these categories. From this one can conclude that subjects were equally fast in responding to the assimilated [m] and the canonical, UNASSIMILATED-/m/ stimuli (e.g. [eːm] from über *de**n** **B**ericht* vs. [eːm] from mit *de**m** **B**ericht*).

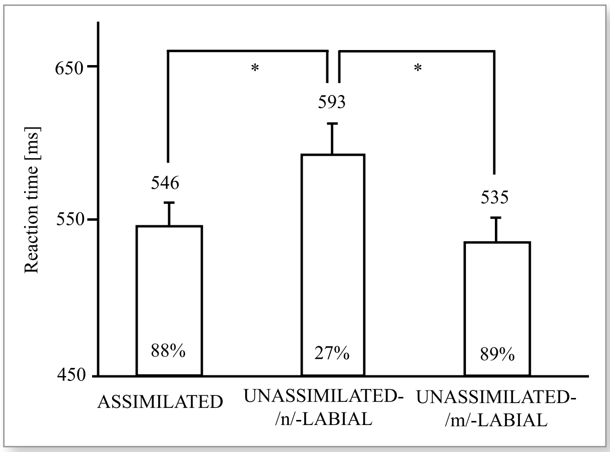


Figure 7c: Percentages and reaction times for [m] responses to ASSIMILATED, UNASSIMILATED-/n/-LABIAL and UNASSIMILATED-/m/-LABIAL stimuli. Asterisks indicate significant reaction times differences.

As for the RT of [m] responses to UNASSIMILATED-/n/-LABIAL stimuli, these were different from the [m] responses to the other two categories (ASSIMILATED vs. UNASSIMILATED-/n/-LABIAL $t=4.08$, $p<0.001$; UNASSIMILATED-/n/-LABIAL vs. UNASSIMILATED- /m/-LABIAL $t=-4.64$, $p<0.001$), indicating that although there was sufficient coarticulation, these stimuli were different from those that were considered by the transcribers as real ASSIMILATED or canonical UNASSIMILATED-/m/ items. Crucially there was no difference between the [m] responses in the ASSIMILATED and the UNASSIMILATED-/m/-LABIAL categories ($t=-1.65$, $p<0.1$). Thus for listeners, the ASSIMILATED stimuli were similar to the UNASSIMILATED-/m/-LABIAL but not to the UNASSIMILATED-/n/-LABIAL.

Recall that the task in Experiment 1 was a forced choice task where subjects had to choose between [m] or [n] as possible responses. To determine in how far the forced choice task of Experiment 1 created a possible bias in the subjects' responses, a second experiment was run where the listeners were free to choose and write down what they heard. If Experiment 2 shows the same pattern of results, then one can conclude that the context-dependent responses of UNASSIMILATED-/n/ stimuli were not caused by the fact that the listeners were forced to choose between [m] or [n]. Further, Experiment 2 allows for a closer analysis of unassimilated stimuli in [DORSAL] context since in Experiment 1 the listeners had no option of providing [DORSAL] responses. A discussion of the results in the light of the two models is given after the data of Experiment 2 and the acoustic analysis are presented.

3.3.2 Experiment 2: Phoneme Transcription Task

Experiment 1 allowed for a first investigation of the acoustic nature of assimilated stimuli in the Kiel Corpus (IPDS, 1994) and the repercussion for perception. However, subjects had only two possible response buttons, i.e. [n] or [m] to choose from. As has been shown in the analysis, especially UNASSIMILATED-/n/ items in [LABIAL] context produced a high amount of incongruent responses. This is arguably due to coarticulatory cues. For items in [DORSAL] context, one could also expect coarticulatory cues influencing subjects' responses. However, it is not clear how subjects would react in this situation, since there was no possibility to indicate "something else". In order to examine the nature of incongruent responses further, a free transcription task was chosen, where subjects could write what they heard without being restricted to two responses, in fact without being restricted to a NASAL response at all.

Materials and design

The stimuli were identical to Experiment 1 except that there was a longer pause between two items (2500 ms instead of 1500 ms), sufficient for writing the syllables but not too much time to think about the stimuli. Each page in the booklet had space for ten items. Warning tones were added after every 10 items, prompting subjects to turn to the next page of the booklet. This was done to ensure that if a subject missed an item, it was possible to correctly resume at the beginning of the next page. Thus, as in Experiment 1, subjects listened to 300 stimuli.

Subjects and procedure

Ten students from the University of Konstanz served as subjects, and none had taken part in the earlier experiment. They were tested individually and were paid for their participation. The set up and equipment was the same as in Experiment 1. Written instructions were given to the subjects prior to the experiment and they received the same practice items as before. They were asked to write down what they heard as quickly and accurately as possible. No instruction was given with reference to nasals, syllables or the "wordness" of the items. Given German orthography, if subjects heard nasals, subjects were expected to transcribe them using one of the three possible responses <m>, <n> or <ng>.

Results

In all, there was only 1 missing response and three were not a nasal. These 4 items were discarded (0.13%). The nasal responses were split up into the three main CATEGORIES as above (LABIAL, ASSIMILATED and CORONAL), based on the original labeling in the Kiel corpus. A total of 2996 transcribed items went into the analysis. Across all categories subjects heard 2032 <m> (67.8%), 890 <n> (29.7%) and only 74 <ng> (2.5%), of which 41 (i.e. 55.4%) came from UNASSIMILATED-/n/ in a DORSAL context.

Within the individual categories, the nasal segments were transcribed as follows (see Figure 8). UNASSIMILATED-/m/ segments were transcribed as <m> in 959 cases (96.2%), <n> in 33 instances (3.3%), and <ng> in 5 (0.5%) cases. ASSIMILATED tokens were transcribed as <m> in 926 cases (92.6%), as <n> in 70 (7.0%) cases, as <ng> in 4 cases (0.4%). UNASSIMILATED-/n/ were transcribed as <n> in 787 (78.8%) cases, <m> 147 times (14.7%), and <ng> in 65 (6.5%) instances.

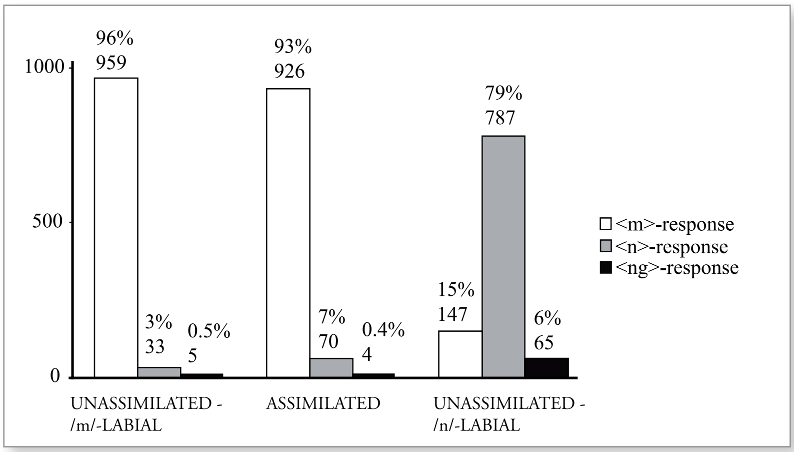


Figure 8: Total number of responses and percentages within the three main categories.

Insofar as congruent responses are concerned, the percentage of [m] responses to ASSIMILATED and UNASSIMILATED-/m/ categories is far higher than the corresponding [n] responses to the UNASSIMILATED-/n/ category (93%, 96% vs. 79%).

Discussion

The free choice task was taken on to ensure that the incongruent responses in Experiment 1 were not due to the fact that subjects were forced to choose between only two nasals (i.e. [n] and [m]). A particular concern was that the large number of <m>-responses to UNASSIMILATED-/n/ stimuli had been biased by the forced choice task. However, Experiment 2 shows that this was not the case. First, there were only 3 non-nasal responses, and second, 97.6% of the entire responses were transcribed as <m> or <n>. In fact, the pattern of results was the same as in Experiment 1. On the whole, the UNASSIMILATED-/n/ stimuli were more difficult to identify congruently as <n> (79%) and were subject to context dependent responses, as compared to the UNASSIMILATED-/m/ or ASSIMILATED items, both of which were congruently identified as <m>, 96% and 93% respectively. As in Experiment 1, the UNASSIMILATED-/n/ stimuli in the context of [LABIAL] consonants were identified as <m> 15% of the time (Experiment 1: 27%). In contrast, there were only 3% <n> responses to UNASSIMILATED-/m/ items. Overall, the accuracy of Experiment 1 for [LABIAL] and assimilated tokens was even higher in Experiment 2, possibly due to the longer time subjects had for their decisions. The results for the ASSIMILATED category are very much the same as in Experiment 1. They were largely perceived as [LABIAL], indicating the completeness of assimilation. In general, this experiment replicates the same asymmetry as was observable already in the identification task and the corpus analysis.

One remaining issue is the acoustic differences between the different conditions, in particular between the ASSIMILATED-LABIAL, the canonical UNASSIMILATED-/m/ against the UNASSIMILATED-/n/-CORONAL. Since the ASSIMILATED nasals did not differ in perception from the canonical UNASSIMILATED-/m/, one would conjecture that the acoustic differences would also be minimal.

3.3.3 Acoustic Measurements

As indicated in section 3.1, there is one important issue that has also been reflected in the literature on place assimilation. It is the question whether acoustic cues can be found that relate to listeners' decisions for [n] or [m] (e.g. Nolan, 1992; Gow, 2002; Dilley & Pitt, 2007). Following Dilley & Pitt's (2007) approach, the stimuli from the experiments were subject to an acoustic analysis. In their study, they compared assimilated segments with their underlying counterparts. Since their results are based on the variation in the F2 of the preceding vowel, the same measure was taken and applied to items from the perception test. In this dissertation, the most prototypical items were analyzed, namely the ASSIMILATED items were compared acoustically with UNASSIMILATED-/m/ stimuli in [LABIAL] context and UNASSIMILATED-/n/ stimuli in [CORONAL] context. Since the number of items from the experiments was too small for calculating an ANOVA, additional items from the Kiel corpus were randomly selected. There is one important difference between the stimuli from the experiments reported here compared to Dilley and Pitt's: in the former stimuli, the final consonant (i.e. the nasal), was not deleted and acoustic information with respect to PoA could also be extracted from the nasal segment. Thus F2 measurements were also taken at the midpoint of the nasal segment itself.

Method

The difference in the F2 frequency values in Hertz were measured between the middle of the vowel and immediately before the beginning of the nasal murmur of all 20 ASSIMILATED items, 6 UNASSIMILATED-/m/-LABIAL and 6 UNASSIMILATED-/n/-CORONAL items being used in the perception studies as an indication for the amount of possible assimilation. In order to base a statistical analysis on a more thorough database, the measurements of 4 ASSIMILATED and 18 UNASSIMILATED-/m/-LABIAL and 18 UNASSIMILATED-/n/-CORONAL items were randomly added with the respective vowels. Overall, the measurements of 72 items were analyzed – 36 for each vowel (i.e. [eː]/[ɑ]), 24 for each condition (i.e. ASSIMILATED, UNASSIMILATED-/m/-LABIAL and UNASSIMILATED-/n/-CORONAL). As in Dilley and Pitt, a mixture of automatic and hand taken measurements was performed (Dilley & Pitt, 2007). Formant values were taken from the estimation provided by PRAAT (Boersma & Weenink, 2007) as well as wide-band spectrograms. In case that the estimated formant values differed from the spectrograms, the spectrogram readings were followed. Dilley and Pitt could measure only the difference between midpoint and endpoint of vowels to gain information about the place of articulation of the upcoming segments, since their items included cases where the consonant in question had been deleted. Since the nasal consonant was never deleted in this case, it was also possible to measure the F2 frequency in the midpoint of the nasal segments (F2 measurements on the nasal differ for [LABIAL] and [CORONAL] nasal consonants, cf. Stevens, 1998: 487-507). The F2 values at the midpoint of the nasals were measured the same way as in the vowels.

Results

F2 differences in the midpoint and endpoint of preceding vowels were subject to an ANOVA with CONDITION (ASSIMILATED, UNASSIMILATED-/m/-LABIAL and UNASSIMILATED-/n/-CORONAL) and VOWEL as independent variables, as well as the interaction of the two factors (VOWEL X CONDITION). Post-hoc tests were performed for the contrasts between the conditions. Figure 9a summarizes the results for the F2 differences.

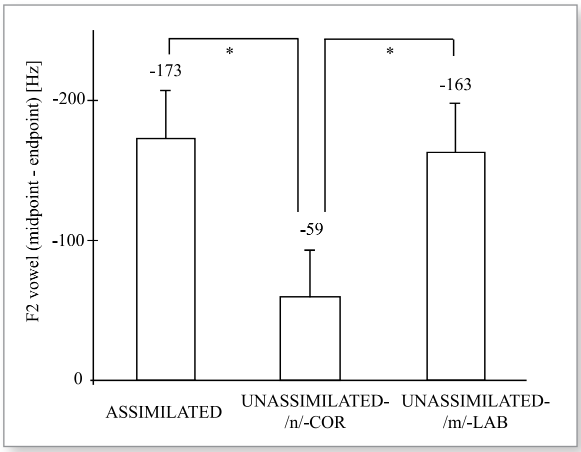


Figure 9a:
Differences between F2-frequency measures at the middle and at the end of the vowel in Hz. Significant differences between these difference values are marked by asterisks.

Concerning the F2 difference analysis, there was a main effect of both condition (F(2,66)= 10.7106, p<0.002) and vowel (F(1,66)= 3.3052, p<0.05), but no significant interaction. A post-hoc test revealed that UNASSIMILATED-/n/-CORONAL items were significantly different from ASSIMILATED (t=-2.317, p<0.05) and UNASSIMILATED-/m/-LABIAL (t=2.1242, p<0.05) items, but the latter two were not significantly different from each other.

For the F2 measurements taken at the midpoint of the nasal consonants (see Figure 9b) the same ANOVA design was used. The following results emerged: there was a main effect of CONDITION (F(2,66)=5.1775, p<0.01), but no effect of VOWEL, neither was the interaction (CONDITION X VOWEL). A post-hoc test showed that UNASSIMILATED-/n/-CORONAL items were significantly different from ASSIMILATED (t=-2.605, p<0.02) and UNASSIMILATED-/m/-LABIAL (t=2.9385, p<0.005) items, but the latter two were not significantly different from each other. Figure 9b depicts the least square means of the nasal F2 measurements.

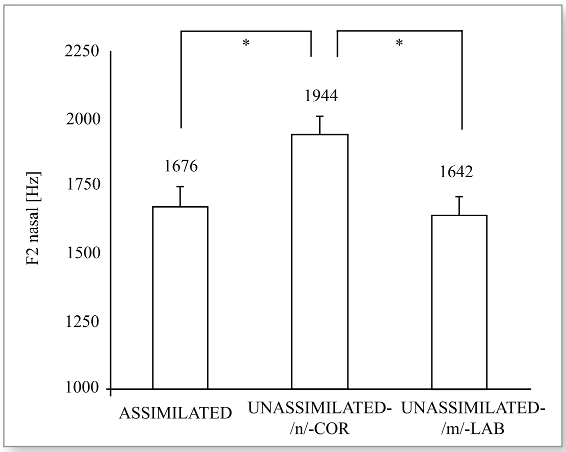


Figure 9b: Least square means of F2-frequencies at nasal midpoints for the investigated conditions, significant differences are indicated with asterisks.

Discussion

The F2 differences between vowel and nasals and the F2 measurements of the nasals parallel the perception results. There is no significant difference between the F2 difference of the vowel preceding the ASSIMILATED coronals and the canonical UNASSIMILATED-/m/. Neither does the nasal F2 differ in these two categories. Corresponding to the perception results, there is a significant difference both in the F2 of the nasal and the F2 difference for the ASSIMILATED and UNASSIMILATED-/m/-LABIAL nasals on the one hand and the UNASSIMILATED-/n/-CORONAL on the other. The results indicate that subjects take these acoustic cues as basis for their decision when deciding on whether they heard [m] or [n].

After the results of the two experiments and the acoustic analysis have been presented, a closer evaluation of the successes of X-MOD and FUL in explaining the data is warranted. The perception experiments indicated that there are cases of complete assimilation. Listeners reacted to these items as if they were underlying UNASSIMILATED-/m/ items in a [LABIAL] context. The first conclusion drawn from the results is that the transcribers of the Kiel corpus were very accurate in deciding whether segments have been assimilated. If at all, their transcription has been conservative, somewhat underestimating the amount of assimilation that occurs in natural speech. This can be seen in the high amount of errors for segments that had been labeled as [CORONAL] in [LABIAL] and [DORSAL] contexts. These findings have been supported also by an acoustic analysis of assimilated items compared to unassimilated, underlying segments.

Coming back to the predictions of the two models that have been outlined in Chapter 2 and before the experiments, a first evaluation of the success of the respective models is possible. The results of the perception experiments are more in line with FUL than they are with X-MOD. Clearly, subjects had most difficulties with [CORONAL] nasal segments in contexts where assimilation would be possible. Clearly, subjects were not fastest for [CORONAL] segments. Thus, they could not use the acoustic information "not quite labial" for deciding fast and accurately for [n]. This would be the prediction of X-MOD, though. For FUL, VN sequences that sound like [m] are expected to be fastest and most accurate, since there is [LABIAL] in the signal and at the same time [LABIAL] exists in the lexicon. When the results for the [CORONAL] nasal is split up into different contexts, the RT for [CORONAL] became as fast as [LABIAL] in contexts where no additional PoA information occurs on the nasal. While there is no significant RT difference in those cases, accuracy is still worse for [CORONAL] segments. In how far this part of the results is decisive for any of the two models is not clear. Further studies that control for effects of context and lexical access are needed to tear apart the effect of contexts. This could shed further light on the question of how the matching mechanism in FUL can be observed in natural speech.

The fact that ASSIMILATED-/n/ were treated as if they were underlying UNASSIMILATED-/m/ in [LABIAL] context is not decisive for the success of either of the two models. Since the experimental stimuli had been deprived of context, subjects could not rely on further information that would be present in natural speech situations. The results indicate that the assimilation is perceived as complete, a result both models could explain. However, the two models differ with respect to the status of this result. For FUL, it is crucial that there are cases of complete assimilations in natural speech. Remember that due to the spreading of

PoA features, FUL predicts that these instances of phonological reduction result in complete neutralization of a contrast. The additional variation that occurs in natural speech is also expected by FUL. On the other hand, for X-MOD, there is no differentiation of the two assimilatory processes. Neither is it crucial whether there are cases of complete assimilation. For both models, though, one would expect that cases of complete assimilation would be perceived as complete, especially when they are presented out of context.

3.4 General Discussion

The focus of this chapter was an investigation of the extent to which regressive place assimilations across words exist in conversational German and how listeners perceive them. This enterprise served the purpose of comparing the success of X-MOD and FUL in predicting and explaining the experimental findings.

Overall, FUL fares somewhat better than X-MOD. But the data of naturally occurring assimilation is, for itself, not enough to decide for one model or the other. While X-MOD's strength is that generally, variation is not seen as problematic and even predicted, FUL assumes two different kinds of variation processes, phonetic and phonological in nature. The data reported for speech perception is not very telling. Both models can account for the patterns observed in natural speech. One piece of evidence that suggests that a differentiation between phonological and phonetic processes is crucial comes from cross-linguistic observations. It has been shown that different assimilation patterns exist (e.g. Jun, 1995). This differentiation into phonological versus phonetic processes is crucial for FUL, whereas X-MOD does not differentiate the two assimilation processes. On the other hand, X-MOD's general predictions based on phonetic variation are more explicit than how FUL predicts phonetic variation.

Concerning perception of assimilated tokens, the results suggest that there are cases of complete neutralization where subjects cannot differentiate between "real" and "assimilated" [m]. The results from the two perception experiments are handled more easily by the FUL model, making this model slightly mode successful than the episodic framework. Yet, the findings are insufficient for far reaching conclusions about the adequacy of either model. Assimilations, while an important testing ground for linguistic theories, are relatively "harmless" in their effect, because they target only one segment (or to be more precise, the value of a single feature). Thus, a more informative way to evaluate the success of the two approaches comes from reduction processes that target more than one segment. These reduction processes create more drastic deviations from canonical, underlying pronunciations. Such reductions have been coined "massive reductions". They are discussed in the next chapter and the amount of occurrence and the effects they have on perception shed further light on the strengths and weaknesses of the two models.

«Though unmusical, German is the most expressive of all languages.»
Sherlock Holmes *(Arthur Conan Doyle, His Last Bow)*

Chapter 4 – **Deletions and Massive Reductions**

4.1 Introduction

The previous chapter examined reduction caused by regressive assimilation of PoA in German. This kind of reduction targets (the value of) a single feature, the overall structure of words is thereby not changed. Thus, the most dramatic outcome of regressive place assimilation – from a point of view of speech perception – is a word differing in one single feature from the canonical pronunciation, which it also does in a rather predictable way. However, there are also reduction processes that have a more profound impact on the structure of words: deletions are reduction processes that can lead to the loss of a segment, a syllable, or even larger chunks of a word. These will be examined in this chapter of the dissertation.

Reductions in general and deletions or lenitions in special are characteristic processes for natural speech (e.g. Kohler, 1990, 1995; Johnson, 2004a; Mitterer & Ernestus, 2006). Several factors have been proposed why speakers seemingly are so careless in the production of their speech. Most often, the "sloppiness" of the speakers has been connected with their alleged intrinsic want to minimize effort of speech production (e.g. Flege, 1988; Lindblom, 1990; Kohler, 1990; Boersma, 1998; Kirchner, 1998; 2004, Ernestus et al., 2006; Mitterer et al., 2008). If this assumption is correct, there is, at the same time, a conflict of interests, since the goal of the speaker to minimize effort, is constrained by the needs of the listener who likes to understand, also with as little effort as possible, what has been said. Speakers usually do not utter speech only for the sake of uttering it, but they want to be understood by listeners making them to react to what they said, either by replying or by physically acting upon the speech act (cf. Jakobson et al., 1963, Byrd & Tan, 1996; or Boersma, 1998 – citing Passy, 1891). For the sake of "perfect", i.e. successful speech comprehension without additional effort, listeners would prefer canonically produced words without variation over massively reduced ones. This idea of conflicting interests is elaborated and formalized in the H&H Theory by Lindblom (1990), (see also Flege, 1988; Kohler, 1990; Byrd & Tan, 1996; Mitterer et al., 2008).

However, also a rather different view of lenition has been proposed, in which lenition (reduction or deletion) is considered informative rather than a loss of information for the listener (Kingston, 2006). The kind of lenition that has been examined by Kingston is targeting single consonants and is pursued by speakers to not interrupt the stream of intensity within a prosodic constituent, indicating constituent boundaries where there is no lenition. Kingston showed convincingly, that for the lenition of consonants within a prosodic constituent, minimization of effort is not an adequate explanation (Kingston, 2006).

In conversational speech, though, there can also occur more severe reductions targeting more than one segment. When the amount of reduction in a naturally produced word compared to its canonical pronunciation is significant, the term "massive reduction" seems a very apt one (viz. Johnson, 2004a).

Whatever the exact reason for reductions, as a first step, it is necessary to examine the exact amount of reduction actually occurring in spontaneous speech. As for Chapter 3, this dissertation concentrates exclusively on German conversational speech. After having reliable analyses of the extent of deletions, a discussion is possible in how far it is correct to label these reductions as "massive". After establishing an overview over the amount of reduction that occurs in conversational German, a second investigation allows for a closer inspection of underlying sources for the occurring reductions exemplified by the case of final /t/. It seems that there is no monocausal explanation, but a better understanding of the following questions can be reached: are these reductions phonetically based or rather due to phonological processes? Phonetically based reductions, which have been shown to depend on speech rate or speech register, pose some problems for many theories of speech perception, notably abstractionist models, since most of them have been built on assumptions of clear speech (cf. Keating, 1998) or do not see phonetic variation as part of the grammar (cf. Lahiri & Reetz, 2002; Johnson, 2004a; Kingston, 2006). This point is also crucial for the two models that are compared in this dissertation; especially when in the second part of this chapter, repercussions of massive reduction for speech perception is examined. At the same time, there are also phonological or rule-based reductions occurring in natural speech being caused by (optional) phonological rules. The differentiation has to be made on theoretical grounds and may not be made *ad-hoc*. But first, it is necessary to see what amount of reduction actually occurs in conversational German.

Coming back to the most crucial characteristic of massive reduction processes – i.e. more than a single segment is reduced – it is important to bring to mind that such changes pose different challenges to word recognition models than assimilation processes, especially if they assume only one abstract representation for each word (or morpheme). Consider the following hypothetical German example. If a German speaker intended to say *Kranz* [krants] ('wreath') and deleted the nasal [n] completely, the outcome would be a word with a completely different meaning: [krats] *kratz* ('itch-IMPERATIVE), and this deletion is not even massive. If such deletions do not occur based on phonological rules that can be "undone" by the listeners, arguably, it will be hard for them to get to the meaning of the intended word, no matter whether you assume an abstractionist or an episodic model of speech perception. However, if this process is occurring from time to time, listeners will encounter this variation more often. This example sets apart the two models FUL and X-MOD. While for FUL the variant without [n] will still lead to a rejection of the intended lexical entry, X-MOD would predict that listeners could, depending on the frequency, still activate the correct entry. This is not to say that FUL does predict that the listener ultimately never understands what has been intended. For FUL, however, other levels of word recognition would also have to come to the aid of the listener. In natural speech, words usually are not uttered in isolation, but in context. Thus, if the sentential and semantic context clearly point to *Kranz*, the listener could finally achieve a correct perception despite the incorrect pronunciation.

Crucially, irregular or massive reduction possibly sets apart the two models. While X-MOD assumes that if reductions occur naturally (i.e. with some frequency larger than zero), exemplars are stored, and subsequently also reduced variants are perceived correctly. For FUL, if there is no phonological reduction, depending on the exact kind and amount of reduction, listeners will have problems in understanding the intended word, unless sentential context with additional information (i.e. morphology, syntax, semantics) is able to correctly perceive what has been said.

The example of *Kranz* just mentioned assumes that the [n] is completely absent. Another possibility, however, is that such complete reductions are only rarely encountered in natural speech, and that speakers do not delete segments completely. In the example, it would also be possible to retain nasalization on the vowel, which would (in German or English, or for that matter any language that does not have nasal vowels phonemically) be enough for the listener to induce the presence of a nasal segment (cf. Lahiri & Marslen-Wilson, 1991).

To summarize, the list of questions that will be examined in this chapter of the dissertation are the following: firstly, how large is the amount of deletions that occur in spontaneous speech and subsequently the listeners have to deal with. Secondly, are there phonological rules leading to massive reductions, or are they unpredictable and mainly phonetically based? Thirdly, what factors have an impact on the amount of deletion that is encountered in natural speech? When those questions have been answered, the next step is an examination in how far these reductions in natural speech have an impact on speech perception. This last question is very important for testing the success of the theoretical frameworks that are examined and evaluated in this dissertation.

This chapter is organized as follows. After giving an overview over existing literature on deletions in conversational speech, the expectations for FUL and X-MOD are discussed. Thereafter, the amount of reductions and deletions occurring in natural speech is analyzed and the data is scanned for regularities of those processes. Aspects that influence the amount of deletion are also examined. Next, a study on /t/ deletion in German is reported using a verb-production paradigm. In the second part of this chapter, the repercussions these findings have on speech perception are examined. Transcription studies and priming experiments are reported that shed some light on these questions and allow for an evaluation of FUL and X-MOD.

4.2 Production Data

4.2.1 Massive Reductions and Deletions in the Literature

Before we turn to the corpus analysis, results from prior research are discussed. For a long time, linguistic research has focused on perfect speech exclusively (cf. Johnson, 2004a; Cutler, 1998). However, there was always a smaller group of linguists, mainly (socio-)phoneticians, who also investigated more casual speech (e.g. Picket & Pollack, 1963; Pollack & Picket, 1963; Labov, 1966; Dressler, 1972; Stampe, 1979; Guy, 1980; Neu, 1980; Dalby, 1986; Kohler, 1990). More recently, casual speech has generally received more attention in linguistic research (e.g. Ernestus, 2000; Shockey, 2003; Johnson, 2004a; Mitterer & Ernestus, 2006; Sumner & Samuel, 2005; Snoeren et al., 2006; Dilley & Pitt, 2007 Ernestus & Baayen, 2007; Tucker, 2007; Tucker & Warner, 2007).

Most studies that examined the amount of variation and reduction occurring in conversational speech have been conducted with (American) English data (e.g. Lieberman, 1963; Manuel, 1991, 1995; Johnson, 2004a; Sumner & Samuel, 2005; Dilley & Pitt, 2007; Tucker, 2007). This is foremost due to the fact that for other languages there exist fewer phonetically transcribed corpora of conversational speech. Only lately and with a rather small number, have researchers created corpora in languages other than English. For Dutch, there is the Corpus of spoken Dutch (Oostdijk, 2000), and one other partially transcribed corpus created by Ernestus (Ernestus, 2000). For Japanese, there also exists a corpus of spontaneous speech (Maekawa et al., 2000). Concerning German, the Kiel Corpus is the only corpus of conversational German that is completely phonetically transcribed (IPDS, 1994).[36]

[36] The so-called "Lindenstrassencorpus" is treated also as part of the Kiel Corpus (Peters, 2001, IPDS, 2007).

The majority of the studies – this will also become clear in the following paragraphs – concentrated on the deletion of a single segment (i.e. /t/) or a natural class of segments (e.g. [CORONAL] consonants) and their reduction behavior in natural speech (e.g. Wright, 1994; Raymond et al., 2006). Much rarer are studies that tried to examine reduction with a more overview-like character (such as, e.g. Johnson, 2004a). In turn, these latter studies usually refrained from a systematic treatment of reductions, because there are too many different factors facilitating reduction that have been suggested and examined. Having a close examination of all would lead to an unfeasible amount of control over speech and too many variables that could not be included in one single analysis. At the same time, controlling for the *explicandum* makes it easier to test different *explanans*. Thus, usually researchers opted for either an investigation of deletion of a single selected segment, or they presented data in an overview-like fashion. The approach in this dissertation is a mixed one. First, an overview over the deletions that occur in conversational German is given. Some factors that have been established as affecting deletion in the literature (e.g. gender differences, cf. Byrd, 1994) are tested. However, a complete list of factors that influence reductions and their impact on conversational German is beyond the scope of this dissertation, and probably any corpus of naturally spoken language that is available to date. Following this rather overview-like treatment of deletion processes in German, in a next step, one special kind of deletion process in German is investigated (i.e. final /t/ deletion in {-st} morphemes). In concentrating on a single segment in a single morpheme, control over many factors (e.g. stress, frequency) is possible and consequently, a more detailed analysis of other factors (e.g. CONTEXT, GENDER) becomes viable.

As for assimilation of the PoA feature, [CORONAL] segments figure most prominently in linguistic research on reduction processes. The process of reduction having drawn most attention in linguistic research and being examined quite extensively is flapping. Most studies have focused on American English flap, where flapping is a regular process (e.g. Connine, 2004; Fukaya & Byrd, 2005; Ranbom & Connine, 2007; Tucker, 2007 and references therein). While flapping is a common reduction process in American English, (possibly neutralizing contrasts as in *writer/rider* minimal pairs) it is (almost) absent in German conversational speech and has no regular occurrence. Therefore, flapping is not examined in this dissertation.

However, flapping is not the only process of reduction that is common in natural speech. Another kind of lenition that is prominent both in recent linguistic research as well as in spontaneous (American) English is the deletion of alveolar stops (e.g. Guy, 1980; Neu, 1980; Sumner & Samuel, 2005; Mitterer & Ernestus, 2006, Raymond et al., 2006; Mitterer et al., 2008). Again most studies dealing this kind of deletion have focused on (American) English. Different than flapping, this process also occurs though in conversational Dutch and German. Therefore, /t/-deletion in German will be examined more closely in this dissertation as well in section 4.2.3. One question concerning /t/ deletion that is still not examined satisfactorily is the role of morphological and phonological influences on this kind of reduction.

Traditionally, one can distinguish between extra-linguistic and linguistic factors for segment deletion in general and /t/-deletion in particular (cf. Raymond et al., 2006). Extra-linguistic factors comprise speaker characteristics such as gender, age, and social class. For instance, Wolfram (1969) showed higher deletion rates for men than for women (see also Byrd, 1994; Neu, 1980, for similar results),[37] while Guy (1992) found that older speakers deleted /t/s and /d/s less often than younger speakers. Differences in social class and dialect have also been identified as influencing parameters of /t, d/-deletion (e.g. Labov, 1967; Wolfram, 1969). Another factor for segment deletion is speaking rate. A number of researchers found higher deletion rates for fast speech than for slow speech (cf. Guy, 1980; Byrd & Tan, 1996; Fosler-Lussier & Morgan, 1999).[38] Byrd and Tan (1996) investigated in an electropalatographic study reduction in consonant clusters [d#g], [g#d], [s#g] and [g#s] across (non) word boundaries and how speakers achieve faster production (# indicates a word boundary). They found that speakers use both segment reduction, as well as overlap of segments (or gestures, or features) if they want to speed up talking. However, manner, place of articulation and syllabic position was found to also play a role for reduction. They additionally demonstrated that not all speakers use the same strategies. A crucial result of their study was that [CORONAL] /d/'s were reduced more than any other segment.

On the other hand, linguistic factors involve word category, frequency, phonological context and morphological structure. Regarding word category, Neu (1980) found higher deletion rates for the function word *and* than for other words in the same context. Function words notoriously differ in their behavior from content words (see also Chapter 3, or, e.g. Selkirk, 1984; Kaisse, 1985; Nespor & Vogel, 1986; Hall, 1999; Ogden, 1999; Phillips, 2001; Kabak & Schiering, 2006). It has been shown independently in psycholinguistic research that frequency of occurrence is a determining factor for speech processing (cf. Frauenfelder et al., 1982; Forster, 1990; Marslen-Wilson, 1990; Meunier & Segui, 1999; Segui & Meunier, 1999). Regarding deletions, Jurafsky et al. (2001) provided evidence that for content words, frequency positively correlated with deletion rate (see also Pluymakers et al., 2005). These findings contrasted with the results of Raymond et al., (2006), who found only marginal frequency effects. These differences could be explained also by the fact that besides frequency also predictability plays a role in determining deletions. Predictability is a factor that has been examined already in earlier studies such as Lieberman (1963) who demonstrated that words in a highly predictable semantic context were reduced compared to words that were less predictable. There seems to be also a correlation between

[37] However, Raymond et al., (2006) did not find a gender difference in the rate of medial /t, d/-deletion.
[38] Byrd & Tan (1996) investigated segment reduction, not deletion *per se*. However, since we assume deletion to be at the endpoint of reduction processes, their results are applicable to deletion data as well.

the two factors (i.e. frequency and predictability). It appears that the findings of Jurafsky and colleagues reflected the different behavior of word category as well, in that content words are less predictable from sentence context than are function words, and that thereby, content words are more prone to frequency effects. Note, however, that Raymond et al. (2006) investigated word medial /t, d/-deletion, whereas the majority of studies focused on word final deletion. This could presumably explain the differing findings. Positional effects have been also found to have an impact on deletion of segments. Greenberg (1999) as well as Raymond et al. (2006) showed that, overall, deletion is less likely in the syllable onset than in syllable coda position. Furthermore, deletion rates differ according to whether syllables are stressed or not. For stressed syllables, deletion rates are generally lower than for unstressed syllables (Zue & Laferriere, 1979; Turk & White, 1999; Greenberg et al., 2002; Turk & Shattuck-Hufnagel, 2007). Note that these findings are in line with the observation that function words are more prone to deletion than content words, since the former are less likely to be in a prominent (stressed) position. Moreover, segments flanking alveolar stops decisively influence their deletion rate, depending on syllable position. For instance, Mitterer & Ernestus (2006) showed that in Dutch, the likelihood of /t/-deletion in word final position is highest if preceded by /s/ or followed by bilabials. Other studies demonstrated higher deletion rates in positions followed by consonants than in positions followed by vowels (Guy, 1980; Labov, 1967; Neu, 1980; Wolfram, 1969). The same studies indicated that similarly, the preceding context caused more deletions if it was consonantal than if it was vocalic. Besides the manner of articulation of the following segments, it has been shown that the place of articulation of these segments also influences alveolar stop deletion rates. Fasold (1972) found more deletions of alveolar stops if these were followed by consonants with the same place of articulation than if they were followed by consonants with a different place of articulation.

A corpus study on spoken Dutch by Janse and colleagues also investigated variation in deletion of word final /t/ (Janse et al., 2007). They found, analyzing the corpus of spoken Dutch CGN (Oostdijk, 2000) that /t/ is deleted or reduced frequently in a /st#b/ context. The results were supported also by an analysis of /t/ reduction in another corpus of Dutch, the IFA corpus (van Son et al., 2001). In about 85% of the time, /t/ is not canonically produced.

Examining Schwa deletion in French, Steriade and Fougeron (1996, Fougeron & Steriade, 1997) showed that what could be considered a complete neutralization due to Schwa deletion, on a more closer analysis actually is not a case of complete neutralization. In French, a Schwa is optional, for example, in an utterance like *de rôle* [dərol] 'some role' which can be produced without Schwa as *d'rôle* [drol]. If the deletion of the Schwa were complete, the resulting utterance would be homophonic with the word *drôle* [drol] 'funny'.

However, a close acoustical analysis showed that the [d] is different in the two cases at least for some speakers, making the neutralization only incomplete. Thus, there are French speakers maintaining differences in production when Schwa is deleted. These acoustic cues that are left by some speakers can possibly distinguish the two words. Fougeron and Steriade found that they are in fact used by listeners when they discriminate the two words correctly. However, the cues are not extremely reliable; their results showed that discrimination was not always successful and also dependent on the speaker (Fougeron & Steriade, 1997).

Similar results were obtained by Spinelli and colleagues (2003), showing that French liaison of final /r/ in a case like *dernier oignon* 'last onion' where liaison creates an utterance almost identical to *dernier rognon* 'last kidney', does not create absolute neutralization. Listeners are able to extract the fine differences for successful perception, however, as in the Fougeron and Steriade's study, and also the competitor gets activated. Recasens (2004) investigated consonant reduction in Catalan and found a syllable position effect for reductions in heterosyllabic consonant clusters.

For German, there is still a dearth of studies examining deletions and the parameters that determine it. Not surprisingly, phoneticians from the Kiel University, who were responsible for creating the Kiel corpus, are also leaders in examining variation and reductions in conversational German (e.g. Kohler, 1990, 1995a,b, 1996; Rehor, 1996; Rehor & Pätzold, 1996; Simpson, 1998; Rodgers, 1999; Wesener, 1999; Kohler & Rodgers, 2001). Kohler (1990) identified four general processes that occur in conversational German leading to variation in pronunciation: (a) r-vocalization, (b) weak forms, (c) elisions, (d) assimilations. After giving examples and restrictions to the processes, he translates the observations into a set of 19 ordered rules, adhering to the rule formulation in generative tradition (Kohler, 1990: 77-82). He thereby incorporates phonological <u>and</u> phonetic rules into a possible grammar of German. The rules have to be also refined in order to take into account restrictions based on syntax, semantics, context, and so on (Kohler, 1990: 77). For example, Kohler introduces a rule that deletes /t/ in *und* 'and'. As Kohler also notes, the process of rule application can be stopped at different points leading to different results in pronunciation (Kohler, 1990: 82). Further restrictions may be introduced to account for various degrees of rule application. This dissertation does not aim at correcting these rules. However, a more quantitative approach will be taken, that also tries to establish the amount of rule(s) application that actually are produced by German speakers. Thereby, it will be also possible to gain more insight into the amount of variation that is produced when Germans talk to each other. Additionally, the repercussions for speech perception were not dealt with at all. Another objective of this dissertation may be seen also in abstracting away from rules that are targeted at single words. The question that is examined in the second part of this chapter can be exemplified with the following word from Kohler's observations: is a massively (or "extreme" as Kohler calls it) reduced *dem* 'the-DATIVE' which would canonically be pronounced as [de:m] and actually can be uttered as a mere [m], possibly recognized as a variant of dem by German listeners?[39] Further studies have been provided by researchers in Kiel investigating the amount of reductions in the Kiel Corpus (e.g. Rehor, 1996; Rehor & Pätzold, 1996). Even if all the results are combined, a more thorough investigation of the reductions that occur in German conversational speech is still missing.

A factor that also possibly affects deletion rate is what is assumed to be stored in the mental representation. An interesting case illustrating the importance of what is stored is Schwa deletion, which occurs numerously in German. Especially in {-en} at the end of a word, Schwa is deleted more often than not (see the section below and Kohler, 1996, for (Amercian) English, see Patterson et al., 2003). For Schwa deletion it can be argued, that underlyingly, there is no Schwa present, however, partly due to orthographic conventions, speakers insert it in some cases in natural speech. The clearer (the more hyperspeech), the more probable is the insertion of Schwa. An influence of orthography on speech production is not unheard of and seems plausible for Schwa, as has been shown by Warner and colleagues, for example. They demonstrated that orthographical, but not morpho(phono)logical status can lead to incomplete neutralization in speech (Warner et al., 2004; Warner et al., 2006). Also, Ventura and colleagues (2001) investigated the role of orthography and found a profound influence on how Portuguese subjects reacted to experimental stimuli. An investigation of the amount of Schwa reduction in German is also conducted in the upcoming section. The question of what is stored is also crucial for the analysis of /t/ deletion. For this phenomenon, also morphological factors of storage are investigated (see 4.2.3).

Deletion predictions

Both models have their main focus on perception of speech. However, one can deduce from their basic assumptions what they predict for language production as well. X-MOD has as point of departure the observation that naturally spoken language is highly variable (Johnson, 1997). Whether linguistically or extra-linguistically, variation is always expected. Reductions and deletions are such kinds of variations. For X-MOD, the differentiation between these underlying factors that determine deletion are not important. For later perception, frequency of occurrence is crucial in that deleted or massively reduced variants of words are still recognizable if they occur frequently in natural speech (cf. Johnson, 2004a). Thus, X-MOD explicitly assumes a great deal of reduction in natural speech.

[39] This example is reduction of a function word. They are known to behave differently and be subject to more extreme reduction processes than lexical words (see also Chapter 3 and references therein). However, the same question can be, and actually is asked in this dissertation for massive reduction of content words as well.

The predictions for FUL are also deduced. However, for FUL, there is a crucial difference between phonetic and phonological factors influencing deletion processes (e.g. Lahiri, 2007). If such processes are regular, such as assimilation of PoA, FUL can account for such variation via inclusion of phonological rules. Phonetic variation is not treated via phonological rules. For phonetic variation, FUL does not make explicit claims. It is expected to occur, but FUL assumes that the deviation from underlying abstract representation is in the majority of the cases not too drastic. Otherwise perception will be distinctly worse.

To summarize, as for assimilation, both models expect variation. What sets them apart is that X-MOD makes these expectations explicit and includes variation in the representation, whereas for FUL, only phonological processes are made explicit. For perception (section 4.3) the predictions of the two models are diverging. Thus, the results of the reduction data have to be also kept in mind when the results for speech perception are presented. The production data taken alone do not allow for a deeper understanding of the nature of representation in the mental lexicon.

4.2.2 Corpus Analysis

4.2.2.1 Amount and Nature of Deletions in Conversational German

If possible repercussion of processes occurring in conversational speech for speech perception can be discussed, it is crucial to know what exactly these processes are that occur in naturally spoken German. Therefore, a corpus study will be carried out. As for the production data on regressive place assimilation, the data basis is the Kiel Corpus (IPDS, 1994). As in Chapter 3, the complete corpus was taken as basis for the amount and regularities of massive reductions that were produced in conversational German. For the analysis, cases of deletions that were transcribed with uncertainty ("%") in the Corpus, as well as cases, where despite a deletion, transcribers indicated that some minimal remnant of the canonical segment was still produced were treated as completely deleted for the analysis (cf. Kohler et al., 1995). This was done to have an estimation of a "worst" case scenario of deletion and reduction from the point of view of the listener. Only complete words were taken into account for the analysis, words that were uttered partly as false starts were ignored.

Again, as in the previous chapter, it is examined whether function words and content words behave differently concerning (massive) reductions and deletions. In this chapter, however, the results for function words should be treated more cautiously than the results for lexical items. This is linked directly to the way the corpus is created: While for consonants, transcriptions did not differ between function words and content words, transcribers made a difference in their exactness of transcriptions concerning the vowels of function words and content words.[40] They expected *a priori* that vowels in function words exhibited a certain amount or articulatory reduction. Hence, only when the amount of reduction seemed disproportionate, that is, if there occurred "syncope, monophthongization of diphthongs and Schwa reductions", did transcribers actually mark the reduction in the corpus (Kohler et al., 1995: 40).[41] The consequence of this transcription policy for function words, is for example, a reduction of a canonically /eː/ to [ɛ] or [e] is not transcribed consistently, whereas a reduction of [ɛ] to [ə] should be transcribed regularly. For lexical words, such variation would be transcribed in any case (Kohler et al., 1995: 39). Since reduction is a priori expected, it is also plausible to assume that less care has been taken for the transcription of vowels in function words. Evidence for this thesis comes from Experiment 2 in section 3.3.2, where subjects transcribed vowel-nasal stimuli. Remember, that 60 items (30 [ɑ] and 30 [eː]) were repeated 5 times each and transcribed by 10 subjects (60*5*10), resulting in the transcription of 3000 items. Overwhelmingly, the experimental stimuli were cut from function words (see Table 6 and Appendix B). Responses from all subjects went into the analysis. Nine responses were discarded (0.3% of the data): twice, there was no vowel transcribed, once the there was no transcription given, two responses were disyllabic, and thus it was not clear which vowel to take into account, three responses included vocalic <r>, one response had a diphthong. That left 2991 responses for the analysis. The overall pattern of responses is given as a confusion matrix in Table 8. The columns show the different vowel-responses given by the subjects, the rows split up the data by the stimuli-vowel, row percentages are given in italics.

Table 8: Overall confusion matrix of the vowels transcribed by subjects in Experiment 2

	<a>	<e>	<i>	<o>	<u>	<y>	<ä>	<ö>	<ü>	
[e]	2 *0.1*	5 *0.3*	1485 *99.2*	0 *0*	0 *0*	0 *0*	0 *0*	0 *0*	4 *0.3*	1497
[ɑ]	885 *59.1*	200 *13.4*	30 *2.0*	214 *14.3*	26 *1.7*	1 *0.1*	46 *3.1*	79 *5.3*	13 *0.9*	1494
	887	205	1515	214	27	1	46	79	17	2991

For the <e>-stimuli, the transcription results are as follows. They were almost exclusively (99.2%) judged to be <i> i.e. a [HIGH] vowel. Five items were transcribed as <e>, four responses were given as <ü>, two as <a>. For the <a>-items, the results are clearly different. For this vowel, 59.08% of the items were "correctly" transcribed as <a>. 70 vowels (4.7%) were transcribed as [HIGH] vowels. Overall, 990 times (29.8%) did subjects

[40] For regressive assimilation of place of articulation, thus, there is no reason to assume a difference in transcription accuracy.
[41] Schwa reduction means reduction of a vowel to Schwa.

transcribe the correct vowel. These results suggest that there is an enormous amount of variation for the vowels that is not captured by the corpus transcription of function words. This impression is affirmed by an acoustic analysis of the stimuli, where F1 and F2 of the formants of the stimuli were measured. This was done in medial vowel position of each item. Figure 10 displays the F1/F2 values of the stimuli. What becomes apparent is the amount of acoustic variation, especially for the <a>-stimuli that is reflected also in the transcription of the subjects.

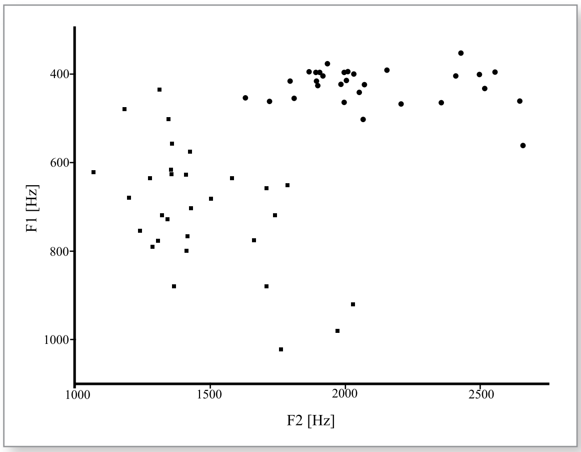


Figure 10:
F1/F2 values of the vowel of the experimental stimuli used in Experiment 1 and 2, F1 is depicted on the y-axis, whereas F2 values lie on the x-axis. Circles indicate <e>-stimuli, squares are for <a>-stimuli.

For deletions, this transcription caveat should not play a role, though, it is reasonable to keep in mind this peculiarity of the Kiel corpus. The effects of the transcription bias can also be observed in the first analysis presented below.

Overall, 37470 words are analyzed in the data set; they were produced by all of the 42 speakers that comprise the corpus. The speakers produced between 35 and 2715 (Mean: 892.1, SD: 640.4) words. Of the overall amount of words, 16409 (44%) were produced canonically. On the whole, speakers uttered 16681 content words and 20789 function words. Content words were uttered canonically 41.6% of the time, whereas function words had no change in 45.5% of the time. A *Chi-Square* test revealed that this difference is significant (χ^2=57.84, p<0.0001), although, this factor taken alone is not a good predictor (R^2=0.0011).[43] As already discussed in the preceding paragraph, the relatively small number of deviations from canonical pronunciations for function words arguably reflects the fact

[43] If not indicated otherwise, χ^2 is calculated using the Pearson test.

that the transcribers paid less attention to the reduction of vowels for function words than for content words (cf. Kohler et al., 1995). To further underpin this argument, the masculine definite article *den* 'the-ACC.' is a prime example. It occurs 673 times in the corpus. Canonically, the determiner is produced as [deːn]. It is transcribed as being produced canonically in 640 cases (95.1%), 26 transcriptions show a deletion (3.9%) and 7 cases (1%) other reductions (5 reductions to [ə] and 2 changes to [ɪ]). This small number of unreduced determiners seems to underestimate the pronunciation variation, especially of the vowel.

Male speakers produced words in 42.8% canonically, whereas female speakers had no deviation from the canonical pronunciation in 45.2% of the time. This difference was also significant (χ^2=21.4, p<0.0001), but again, taken alone the explanatory power is rather small (R^2=0.0004). Canonically produced words were produced by the different speakers in an overall percentage range from 52.6% to 31.4%. Figure 11 illustrates the amount of canonically produced words: overall, split up into gender and word category (function word vs. content word).

A nominal logistic model was calculated with the factors GENDER, WORD CATEGORY (function word and content word) SPEAKER (nested under GENDER) as random factor and GENDER x WORD CATEGORY. In this analysis, gender was no longer a significant factor (Wald-χ^2=1.02, p=0.1772). SPEAKER (Wald-χ^2=226.36, p<0.0001) and word category (Wald-χ^2=46.68, p=0.0001) were significant main factors, the interaction of word category and gender was significant as well (Wald-χ^2=13.38, p<0.0005).

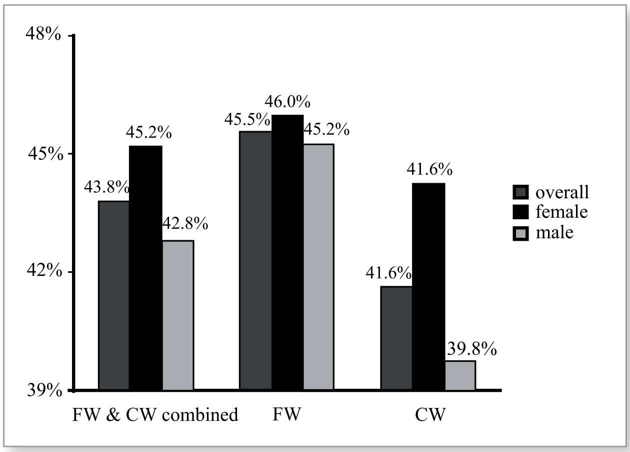


Figure 11: Percentage of canonically produced words, split up by gender, and overall

In the next analytical step, a deletion rate was calculated for the words by dividing the number of deletions by the number of segments in a word. The mean deletion rate for all words was 0.16 (SD 0.18). Function words had a rate of 0.17 (SD 0.21) whereas content words exhibited a deletion rate of 0.11 (SD 0.14). For individual speakers, deletion rates ranged between 0.11 and 0.21. This means that between 10% and 20% of the segments in all words were deleted by the speakers. For individual words, this rate had a value between 0 (i.e. no reduction) for 19171 words, and at the opposite end, 28 cases had a deletion rate of 1 (every segment of the word was deleted).[44] 27 words that were reduced completely were function words, only one of these "phantom words" was a content word (i.e. jetzt). Overall, 13 different words were affected by complete omission (a complete list is given in Appendix C).

An ANOVA was calculated to test the effect of the different factors in a combined way. Thus, GENDER, WORD CATEGORY and SPEAKER (nested under GENDER as random variable), as well as GENDER x WORD CATEGORY were taken as factors in the analysis. In this analysis, GENDER was not significant (F(1,37426)=2.3972, p<.129), whereas the factor WORD CATEGORY is significant (F(1,37426)=10.311, p<.0001). The interaction gender x word category is not significant (F(1,37426)=10.311, p=0.2824). However, since R^2 equals 0.034, this model itself is not explaining the amount of variation that can be observed in the data. A more fine-grained analysis on the basis of segments seems thus more promising.

Therefore in the upcoming paragraphs, a closer examination of the individual segments is reported. As in the analysis of assimilation, the "-h" transcription was ignored, since there is not a one-to-one correspondence of the symbol to an actual physical event and seemingly, there does not exist a complete consistency in transcription. For instance, "-h" could be interpreted as a transcription of aspiration, because it denotes some significant burst (Kohler et al., 1995). However, for 14658 /t/'s there are 5248 (36%) where "-h" is transcribed. On the other hand aspiration in German is arguably an indication for voiceless stops. Hence an investigation for /d/ gives further evidence for "-h" as being hard to interpret. There would be many cases of devoicing, even intervocalically. Of the 7906 /d/ in the database, there are also 5914 (75%) cases where "-h" is transcribed. Therefore, for the deletion analysis, "-h" is ignored. Furthermore, some relabeling of the corpus had to be performed for the deletion analysis. The transcription in the Kiel corpus is left-to-right. Therefore, when segments were deleted, the question for the transcribers was always which segment to mark as deleted. For example, the word guten ('good-CASE'), canonically [guːtən], was produced by a speaker as [guːn]. In the corpus transcription, the apical stop /t/

[44] At this point the question could be raised to what extent a word can be deleted completely. Is it really true that the speaker actually intended to utter this word? There are two reasons why arguably, speakers wanted to utter the word. One reason to assume that intentionally this word was uttered, is that also words where some residue was left (indicated by "-MA" in the transcription) as well as deletions where transcribers were unsure (indicated by "%") were counted as deletion. This concerned 21 of the 28 cases of complete deletion. However, a second point is due to the fact that for the corpus transcription, first, the complete utterances were orthographically transcribed, and only later came the phonetic transcription. In one case, for example, the complete sentence is transcribed: *Nein, richtig, das wäre mir jetzt zur Not auch noch recht.* 'No, right, that would be OK for me in case of need'. The word where the deletion rate is 1 is *jetzt* 'now'. A non-systematic study of how listeners would transcribe the sentence showed that *jetzt* is heard invariably by listeners. However, when only the part of speech where the word is supposed to be is analyzed, *jetzt* really seems completely deleted.

is transcribed as being changed to [n], the /ə/ as well as the /n/ as deleted. While the number of deletions is correct, the number of segments that were changed is rather questionable. In the actual pronunciation, the [n] is still present. In such cases, the transcription was changed manually for the analysis into a [t] that got deleted, and an [n] that was produced canonically, whereas the [ə] labeling was not affected by the relabeling.

Table 9:
The 10 segments that are deleted most often in the Kiel corpus, how often they occurred and the respective deletion rate

Segment	Number of occurrences	Deletion rate
ã	1	1
ʔ	10749	0.814
ə	10184	0.644
t	14658	0.214
l	3795	0.192
d	7906	0.184
g	2724	0.181
b	3235	0.152
ʊ	2485	0.117
n	18860	0.106

Overall, 167229 segments were analyzed. Out of those, 26998 segments (16.1%) were deleted. The 10 segments that were deleted most often are listed in Table 9. The segment that got always deleted (deletion rate of 100%) is nasalized /ã/, occurring once in the corpus in the word *arrangiert* ('arranged'), a loanword from French that is pronounced with nasalized [a] as [arãˈʒiːɐt] canonically in German. However, [ã] is not a phoneme of German and only occurring once in the corpus, therefore, no further conclusion can be reached from this single deletion, consequently, this segment is excluded from further analysis. The two segments that are deleted second and third most often illustrate another idiosyncrasy of the Kiel corpus: /ʔ/ and /ə/.

For /ʔ/, the phoneme status in German is not clear, and most phonologists do not treat it as phoneme of German (cf. Hall, 2000: 65). It can be produced syllable-initially preceding vowels if the syllable does not have an onset otherwise. The transcribers of the Kiel corpus assumed the glottal stop to be a phoneme of German that should be produced canonically. Following the transcription, there should be 10749 [ʔ] in the corpus of which only 1995 are actually produced, hence 8754 (81.4%) are deleted. In many cases, the deletion is followed by a transcription of glottalization or creaky voice (i.e. “-q”) in the corpus. The insertion of glottalization can be observed for 5436 cases (62.1%) of the deleted cases. However, the cases where [ʔ] is produced show a glottalization in 78.9%, therefore, insertion of glottalization cannot be considered remnant of [ʔ] in case of deletion.

The segment that is ranked next on the table is [ə], which should be produced 10184 times according to the transcription of the Kiel corpus, if all words were always produced canonically. Out of these 10184 cases, /ə/ was deleted 6559 times (64.4%). Most of the instances of [ə] deletion occur when Schwa is followed by a [n] word finally. As already indicated in the introductory part of this chapter, the word final {-en} can be assumed to be either a [ən] underlyingly, this is what the Kiel corpus transcription opted for, or, one can assume a syllabic nasal only [n̩]. Of the 10184 [ə] in the corpus, 6204 occur in front of [n], of which 91.8% (5696) are deleted. Out of the 6204 [ən] occurrences in the corpus, 5522 are word final, with a deletion of [ə] in 5226 cases (94.6%). The 682 cases where [ən] is not word finally, have a deletion rate of 68.9% corresponding to a deletion in 470 cases. 3980 instances of underlying [ə] are not preceding [n]. In these cases, the vowel gets deleted in 863 occurrences (21.7%). Combined with non-final [ən] occurrences, this results in 4662 cases, of which 1333 (28.6%) are deleted. This is still a high amount of deletion, but one that is comparable to the next segment (i.e. /t/) in Table 9, for example. The main conclusion that can be reached for word final [ən] syllables is that it would be more sensible to assume only a syllabic nasal underlyingly and that instances where [ə] is produced constitute a case of strengthening. This strengthening could be used by talkers to indicate word boundaries, parallel to the production data for [ʔ], and to Kingston’s (2006) argumentation, where lenition indicates the absence of a boundary and perfectly produced (or hyperarticulated) segments strengthen upcoming constituents. Consequently, the cases of Schwa in word final [ən], as well as all [ã]’s and [ʔ]’s, were excluded from further analysis to have a more realistic database that is not influenced by outliers with questionable status. After the exclusion of these segments, 150957 segments were further analyzed. As can be seen in Table 9, the segment that is deleted with the next highest probability is /t/. For /t/, there does not exist a question of the phoneme status or its underlying presence or absence in the representation. If the Kiel speakers were perfect, they should have produced 14658 /t/’s. However, of these, 3141 are deleted (21.4%). Within the Top10 of deleted segments, another vowel occurs with a deletion rate of 11.75%, namely [ʊ]. Remember that the task for the subjects was appointment making, therefore, a lot of numerical words occur for the dates, and in many of them *und* (‘and’) is part of the number (e.g. *fünf**und**zwanzigster* ‘twenty-fifth’), where the /ʊ/ gets reduced regularly, paralleling English ‘n’ constructions such as ‘Rock’n’Roll’.

The inevitable question than is in how far the difference in "proneness" of segments to be deleted is a factor that explains the variation in deletion frequency. A *Chi-Square* test was calculated for SEGMENT as single factor. The result showed that the deletion patterns for different segments are significantly discriminative (χ^2=11629.55, p<0.0001). The nature of the SEGMENT alone accounts for a notable share of variation (R^2=0.1307). Related to the segment itself as a factor are the phonological contexts of a segment (PRECEDING and FOLLOWING) which were analyzed in turn. The *Chi-Square* test for FOLLOWING CONTEXT showed a significant effect (χ^2=1858.59, p<0.0001), but R^2 dropped markedly (R^2=0.0234). On the other hand, preceding context was analyzed. Here, the analysis produced again a significant effect (χ^2=4522.78, p<0.0001; R^2=0.0529). It is obvious, of course, that phonotactic constraints restrict a free combination of segments, and thus this is a very crude way for an analysis. But for a first estimation, the data is analyzed as if every factor is independent of any other factor.

The following comparison contrasts the deletion behavior of vowels and consonants. Overall, including the remaining cases of [ə], vowels account for 54735 canonical data points in the analysis (36.3%). Of these, 2820 (5.2%) are actually deleted. The remaining consonants (without the glottal stop) sum up to 96222 segments (63.7%) of which 10197 (10.6%) are transcribed as being deleted. A *Chi-Square* test corroborate that this difference is significant (χ^2=1312.9, p<0.0001). This factor SEGMENT TYPE (vowel vs. consonant) taken alone, however, is not a good predictor (R^2=0.019). Thus, vowels are more stable than consonants. This is rather crucial for the syllable structure of words, which seem rather unaltered because except for Schwa, vowels and consequently the syllabic structure of words are rather stable.

As for the word-based analysis, the percentage of deleted segments for the factor GENDER was calculated. Male speakers should have uttered 88909 segments, of which they deleted 8158 (9.2%). Assuming canonical pronunciation, female speakers should have produced 62048 segments, but they deleted 4859 of them (7.8%). Again, a *Chi-Square* test was calculated. This test shows that the difference in deletion behavior between male and female speakers is significant (χ^2=83.86, p<0.0001). The factor GENDER on its own does also not account for a large fraction of the variation (R^2=0.001).

In a next step, the difference of deletion patterns for segments that are part of function words and those that are members of content words is examined. Underlyingly, the segments of function words sum up to 56636 segments, where 6080 (10.7%) of them are deleted. Content words have 6937 deleted segments of 94321 underlying ones (7.4%). The effect of WORD CATEGORY was emphasized by a *Chi-Square* test (χ^2=513.25, p<0.0001, R^2=0.0057).

The next factor that is examined on its own is FREQUENCY of occurrence. Frequent units have been found to delete with higher probability than infrequent units. Frequency was determined by the number of canonical occurrences of a segment in the dataset divided by the overall number of canonical segments. The *Chi-Square* test showed a significant effect for frequency (χ^2=2311.99, p<0.0001) with a moderate explanatory power (R^2=0.0261).

Subsequently, a nominal logistic model was calculated with the factors SEGMENT TYPE (consonant/vowel), SEGMENT (nested under SEGMENT TYPE), PRECEDING CONTEXT, FOLLOWING CONTEXT, GENDER (f/m), WORD CATEGORY (part of function word, or content word), and SPEAKER (nested under GENDER).[45] In this analysis, GENDER was no longer a significant factor, nor was SEGMENT TYPE. SEGMENT (Wald-χ^2=6012.1, p<0.0001), SPEAKER (Wald-χ^2=584.68, p<0.0001) PRECEDING CONTEXT (Wald-χ^2=3369, p<0.0001), FOLLOWING CONTEXT (Wald-χ^2=1347.17, p<0.0001) and WORD CATEGORY (Wald-χ^2=1213.82, p<0.0001) were all significant main factors (R^2=0.2075).

In a next analysis, the deletion rate for vowels and consonants is examined separately. First, the 54735 underlying vowels are examined. The first analysis concerns the SEGMENT itself. This factor is significant in a *Chi-Square* test (χ^2=1213.82, p<0.0001) and responsible for a considerable amount of variation (R^2=0.2093). For vowels, different than for consonants, the PoA and height features were not analyzed. This was done because first of all, vowels can have two different PoA features at the same time, for instance, [u] is both [DORSAL] and [LABIAL]. This difficulty is aggravated by the occurrence of diphthongs such as [ai], where they have a specification with a combination of [DORSAL] [LOW] and [CORONAL] [HIGH]. Thus, categorization into single PoA or height features does not work as well as for consonants (see below).

Again, the impact of the factor GENDER is evaluated. Female talkers account for 22604 canonical vowels (41.3%). Out of those, they deleted 1029 (4.6%). Male talkers should have produced 32131 vowels (58.7%), but they deleted 1791 (5.6%) of them. The subsequent *Chi-Square* test reveals the factor gender to be significant (χ^2=28.71, p<0.0001; R^2=0.0013).

Then, as already discussed in section 4.2.1 another factor that possibly has an impact on deletion probability for vowels is whether they are stressed or not. Vowels in the Kiel corpus are either marked as stressed, unstressed or with secondary stress.[46] Out of the 54735 vowels in the analysis, 35632 are unstressed, of which 2768 are deleted (7.8%) – this number includes Schwa, that, by definition, is unstressed, thus not all vowels can occur in all conditions. Primary-stressed vowels occur canonically in 16721 cases, 46 of them (0.3%) are deleted. Secondary-stressed vowels (2382 overall) are deleted in 6 cases (0.3%). The factor STRESS (unstressed, primary, secondary) was found to fairly contribute to the variation (R^2=0.0918; χ^2=1430, p<0.0001).

[45] The factor FREQUENCY was not taken into account because the model's power did not improve at all when it was included (R^2=0.2075).

[46] The transcription convention for vowels in function words is that they are always labeled unstressed.

Now, the impact of PRECEDING and FOLLOWING CONTEXT is reported. For FOLLOWING CONTEXT, the analysis revealed an indicative (R^2=0.1073) and significant effect (χ^2=3281, p<0.0001). Also, preceding context was significant (χ^2=2112.75, p<0.0001; R^2=0.0801). This result may have been biased because some cells had a count of less than 5, however.

Furthermore, the role of WORD CATEGORY was probed for the deletion probability of vowels. Vowels in function words summed up to 22726 canonical segments. Of these, 1471 were deleted (6.5%). Vowels in content words could have been uttered in 32009 cases, of which 1349 got deleted (4.2%). This difference was found to be significant, but not very revealing (χ^2=138.7, p<0.0001; R^2=0.0061).

Finally, a nominal logistic model was computed with the factors SEGMENT, PRECEDING CONTEXT, FOLLOWING CONTEXT, GENDER (f/m), WORD CATEGORY (part of function word, or content word), STRESS (unstressed, primary, secondary), and SPEAKER (nested under GENDER). Only GENDER was no longer a significant factor. SEGMENT (Wald-χ^2=1667.63, p<0.0001), SPEAKER (Wald-χ^2=364.41, p<0.0001) PRECEDING CONTEXT (Wald-χ^2=744.91, p<0.0001), FOLLOWING CONTEXT (Wald-χ^2=1082.08, p<0.0001), stress (Wald-χ^2=364.41, p<0.0001) and WORD CATEGORY (Wald-χ^2=11.19, p<0.0005) were all significant main factors (R^2=0.2075).

Now the pronunciation of the 96222 underlying consonants is examined closer (10.6% deletion). However, in view of the analysis for the effect of place of articulation, /h/ was excluded from further analysis, leaving 95120 segments (10.6% deleted). Paralleling the analysis for vowels, the role of the segment itself is calculated. The factor SEGMENT is significant (χ^2=5154.71, p<0.0001; R^2=0.086). Thus this factor is less explanatory for consonants than for vowels. For consonants, it is possible to investigate the role of PoA features in a meaningful way. It becomes emergent that [CORONAL] consonants are deleted more often than [LABIAL] or [DORSAL] consonants. Out of 64471 [CORONAL] consonants, 7494 are deleted (11.6%), [LABIAL] segments occur 18461 times underlyingly, out of which 939 (5.1%) are deleted, [DORSAL] segments account for 11469 underlying consonants, of which 932 are deleted (8.1%). The variable PLACE shows a significant effect (χ^2=733.4, p<0.0001; R^2=0.0134). Next, deletion rates for stops, fricatives and nasals are reported. Stops accounting for 32133 underlying segments are deleted in 5385 cases (16.8%). Nasal segments are deleted in 2122 out of 24978 (8.5%) cases, whereas fricatives are the most stable segments (932 deleted of 25213 underlying fricatives – 3.7%). The *Chi-Square* test revealed the difference for MANNER of articulation to be significant (χ^2=2740.1, p<0.0001; R^2=0.0529).

Next, the context effects were investigated for consonants, being highly significant for following context (χ^2=2015.82, p<0.0001; R^2=0.031) as well as for preceding context (χ^2=4962.87, p<0.0001; R^2=0.0736).

As for the vowel analysis, the gender effect was examined. Female speakers, accounting for 38982 consonants deleted 3791 (9.7%) of them, whereas male talkers who accounted for 56138 consonants deleted 6293 (11.2%). The subsequent *Chi-Square* test revealed the significance of this factor, albeit some very small explanatory power (χ^2=53.52, p<0.0001; R^2=0.0008).

Concerning the effect of whether consonants were parts of function words or content words, the following results were found: for function words, 4334 out of 33403 consonants were deleted (13.6%), whereas in content words, deleted segments (5530 out of 61717) accounted for 9%, a difference that was significant (χ^2=499.42, p<0.0001; R^2=0.0075).

After analyzing the factors distinctly, a nominal logistic model was calculated, where they all were analyzed in a single model. This model included the factors SEGMENT, PLACE, MANNER PRECEDING CONTEXT, FOLLOWING CONTEXT, GENDER (f/m), WORD CATEGORY (part of function word, or content word), and SPEAKER (nested under GENDER). In this analysis GENDER, MANNER and PLACE were no significant factors. SEGMENT (Wald-χ^2=970.23, p<0.0001), SPEAKER (Wald-χ^2=515.76, p<0.0001) PRECEDING CONTEXT (Wald-χ^2=3887, p<0.0001), FOLLOWING CONTEXT (Wald-χ^2=1702.56, p<0.0001), and WORD CATEGORY (Wald-χ^2=737.65, p<0.0001) were all significant main factors (R^2=0.2170).

4.2.2.2 Discussion

The data analysis revealed several factors that have an impact on the deletion of segments. However, despite the large number of data points in the analysis, an interpretation of the results becomes very hard. This is partly due to the uneven distribution of segments and some co-occurrence restrictions.

Furthermore, the explanatory power of the statistical analysis in not very high. This shows that there are many different effects that contribute to variation in language production. The results are also indicative that language is dependent on multiple factors, with a large amount of enlaced factors that cannot easily be separately analyzed. It also illustrates a dilemma linguists are faced with. On the one hand, natural speech data is important for theoretical modeling. On the other hand, when there are too many data points, it is not very easy to perform an adequate statistical analysis any more. A more sensible way to analyze data thus is to concentrate on a single segment, control some of the factors, and intentionally vary other factors. The repercussions these results have for X-MOD or FUL will be discussed after the case study of a single segment in a fixed position which will be reported in the next section.

4.2.3 Case Study: Final /t/ Deletion in Verbal Paradigms

4.2.3.1 Introduction

The statistical analyses conducted in the preceding part of this dissertation as well as the overview over the findings of previous works have shown quite plainly that many factors influence the deletion behavior in natural speech (viz. the rules established by Kohler, 1990). However, due to the intertwining nature of many of these factors, the analyses in the previous section so far failed to produce a clear pattern of deletion variation. Since it is difficult (if not impossible) to investigate all factors simultaneously and independently, it is better to concentrate on a few and examine more closely their interaction. In section 4.2.2, the single factor with the highest impact on deletion rate was the segment itself. Deletion percentages differed considerably between segments (see Table 9). By keeping the segment in question constant, a more detailed analysis is possible. Therefore, in this section, the objective is to concentrate on a single segment (i.e. /t/) and to control for some other of the factors that have been identified in section 4.2 so far. In particular, an interesting question is whether /t/ deletion rates differs depending on the regularity or irregularity of a given complex verb form or depending more on the phonological context. Furthermore, the construction of the speech corpus controlled for the word category that only content words, or to be more specific, only verbs were produced. The analysis further concentrated on verb forms for the 2nd person singular in the present tense in German. All the verb forms end in the morpheme {-st} (for more detail of how the corpus was constructed, see section 4.2.3.2 below). This course of action allows for an examination of /t/-deletion in German and its possible interaction with morphology.

Before this analysis is embarked, a short summary of the factors that have been found to have an impact on /t/-deletion and that are examined in this section is given (for a more general overview with more details see section 4.2.1). Then, corpus data from a production study aimed at the separation of phonetic, phonological and morphological factors for /t/-deletion in German is presented.

As the results of the previous section underpin, /t/-deletion is not completely regular in a phonological sense, (compare it to Final Devoicing, cf. Brockhaus, 1995; Kohler, 1995a; Wiese, 1996; Piroth & Janker, 2004), however, it is not completely random, either (cf. Chapter 3, Section 4.2.1, or Raymond et al., 2006 and references therein). Kohler describes the /t/-elision as a "possible" process in German when the /t/ is in-between two apical (i.e. [CORONAL]) fricatives (Kohler, 1995: 209), Raymond et al., 2006 model /t,d/-deletion in American English with variable rules (Cedergen & Sankoff, 1974).

The effects of the following factors will be highlighted by the production study that is reported below. First, it will be investigated, whether female speakers are more accurate in their production as male speakers. A GENDER effect was found by Wolfram (1969): men deleted more frequently than women (see also Neu, 1980; Byrd, 1994). Raymond et al. (2006) however, did not find a consistent effect of gender on deletion. This finding was also obtained in the previous section of the dissertation when other factors were included in the statistical analysis.

A second factor that will be examined is the effect of hesitational pauses on final /t/ deletion. It has been found that DYSFLUENT PRODUCTIONS (e.g. characterized by hesitational pauses) may have an impact on the probability of deletions, in that segments are strengthened when occurring in dysfluent contexts (e.g. Fougeron & Keating, 1997; Fox Tree & Clark, 1997; Shriberg, 1999; Kingston, 2006).[47]

The third factor investigated more thoroughly is the following (phonological) CONTEXT. Due to the corpus construction (see below) the preceding context was held constant. Context as factor for deletion probability has been identified by many linguists investigating [CORONAL] stop deletion (cf. section 4.2.2.1, or, e.g. Labov, 1967; Wolfram, 1969; Fasold, 1972; Guy, 1980; Neu, 1980; Mitterer & Ernestus, 2006).

Finally, it has been shown that morphology also affects deletion of /t/. If it carries morphological function as in the case of the English past tense marker /t/, deletion is less likely than if it does not carry such a function (Guy, 1980; 1992; Neu, 1980). The way the corpus was constructed controlled for this factor, in that the /t/ was always part of the same suffix. Another morphological factor for /t/-deletion has been identified by Hay (2003). She observed that stem final /t/s are deleted to a different degree depending on the transparency of relationship between stem and word form. In cases where the stem was more transparently related to the word form (e.g. *swift-ly*), she found fewer deletions than in cases where the stem was less transparently related to the word form (e.g. *list-less*). Transparency of stem-to-word relation has been expressed by the so-called relative frequency (Hay, 2001; Hay & Baayen, 2005). If a stem is transparently related to a morphologically complex word form containing this stem, the stem frequency tends to be higher than the surface frequency of the complex word form. Conversely, if there is no transparent relation between stem and complex word form, the complex word form tends to have a higher frequency than its stem. RELATIVE FREQUENCY is thus a measure of how likely it is that a particular complex word form is decomposable into its constituent morphemes. Returning to deletion rates, the likelihood of /t/ deletion in morphologically complex forms is negatively correlated with the likelihood of their decompositionality. RELATIVE FREQUENCY and its role in determining /t/ deletion will also be investigated in the upcoming section. Furthermore, some studies found effects of (ir)regularity in morphological processes (Pinker & Prince, 1988, 1994; Pinker & Prasada, 1993; Marcus et al., 1995; Pinker, 1998; Clahsen, 2006a, b). So far, it is not clear, whether irregular inflected verbs behave differently when it comes to the reduction

[47] Jurafsky et al., (1998), however, found that this effect may interact with word function and other factors.

of parts of inflectional phonemes (i.e. /t/) than regular inflected verbs. For X-MOD, the most crucial effect (which is correlated with regularity) is based on frequency of occurrence, whereas for FUL, all the verbs should behave alike.

Especially concerning morphology, German verb inflection provides an ideal testing ground; this is true for several reasons. Most importantly, the distinction between regular and irregular word forms does not necessarily align with a distinction between "decomposable" and "not decomposable". In the German verb system, the 2nd person singular suffix {*–st*} is added to verb stems regardless of whether the verbs are regular or irregular. Irregularity is defined in terms of past tense formation: the verb *graben* ('to dig') is irregular since its past tense is not formed with the regular past tense suffix {*–te*}, but expressed by a stem vowel change (**grab-te-st* vs. *grub-st* 'you dug'). On the other hand, the past tense of the regular verb *baden* ('to bath') involves the regular past tense suffix without a vowel change (*bade-te-st* 'you bathed'). Irregular verbs may also show a stem vowel alternation in the present tense. For graben, the 2nd person singular is *gräbst*, not **grabst*, while the present tense of regular verbs never shows such alternations (*bad(e)-st*, not **bäd(e)-st*).[48]

Apart from the fact that the 2nd person singular suffix consistently attaches to regular and irregular stems, there are further reasons why these forms are ideal for the investigation of word final /t/ deletion. First, the *{–st}* suffix is unique in German inflection. It only expresses the 2nd person singular for verbs. Besides, within the verbs' paradigm, omitted final /t/s do not cause ambiguities, while ambiguities outside the paradigm are still possible (e.g. /t/ deletion in *hau-st* 'you beat' results in *haus* 'house' or *hau's*, a reduced form of *hau es* 'beat-IMP it'). The alveolar fricative is sufficient to distinguish the 2nd person from all other person/number combinations. This is important regarding the findings of Guy (1980, 1992) and Neu (1980), who found differences in /t/ deletion depending on the morphological function of the alveolar plosive. Note that for some irregular verbs, the 2nd person is additionally marked by a stem vowel change in the present tense, making the final /t/ even more redundant.

Next, the 2nd person forms provide a consistent preceding context (/s/), in which /t/ deletions are to be expected (cf. Mitterer & Ernestus, 2006, for Dutch). Furthermore, the preceding /s/ can be either part of the stem or part of the suffix. In the verb form *hau-st* from the infinitive *hauen* ('to beat'), [s] surfaces as part of the suffix, while in the form *haus-(s)t* from hausen ('to house, to dwell'), [s] surfaces as part of the stem or is possibly "ambimorphemic".

Regarding a controlled data set with regular and irregular verbs, comprising /s/- and other stems as inflected 2nd singular forms, we are faced with the problem that no existing corpus of spoken language could provide us with the necessary stimuli. The use of the corpus that has been used as database in the preceding sections, the Kiel corpus (IPDS,

---

[48] There are dialects that allow for the form *grab-st* in the present tense. However, in Standard German, this form is not grammatical.

1994) would have been ideal. However, there are hardly any 2nd person singular forms in this corpus, since the spontaneous conversations were based on the usage of honorable 2nd person forms which are equivalent to the 3rd person plural and have a {*-en*} suffix. Secondly, the corpus is based on a restricted vocabulary, since the conversations are all about appointment making. Another natural result of the corpus structure is that any control over the following context of the forms of interest is hard to achieve, if not impossible. Finally, the rather random conversational samples make it very difficult to control for extra-linguistic variables such as gender, age, dialect region and so forth. For these reasons, an own corpus had to be created, which will be described in the next section.

4.2.3.2 Corpus Construction

The rate of /t/ deletion crucially depends on the task subjects have to perform, or more precisely, the speech register they use (e.g. Wolfram, 1969; Guy, 1980; Fosler-Lussier & Morgan, 1999; Jurafsky et al., 2001, 2002; Mitterer & Ernestus, 2006; Raymond et al., 2006). In read speech, subjects reduce words less drastically und delete segments less often compared to (fast) natural, conversational speech. In order to achieve a natural way of speech production while simultaneously being able to control for specific verbs and the context in which they occur, we opted for a verb paradigm production task. In such a task, subjects have to produce inflected forms of a verb's paradigm. Subjects are given the infinitive of the respective verb as well as the personal pronouns for each inflected form, but not the form itself. Thus, subjects have to provide the correct word forms by themselves, whereby the task is not a simple reading task. This increases the probability of a natural way of speaking. Furthermore, it is obvious that producing verb paradigms in a fast way is not a task that occurs in natural speech situations, but for native speakers, the task itself is not very complex or complicated, therefore, subjects do not have to concentrate too much on their production, but rather can produce the verbs fluently.

For the production task, 50 verbs (25 irregular, 25 regular) were chosen. A complete list of the verbs is given in Appendix D. All verbs were disyllabic. The lemma frequency of the verbs as provided by the CELEX database for German (Baayen et al., 1995) ranged between 1 and 426 (Mean=89.2; SD=79.8). Care was taken to match the Mean Lemma Frequency for the regular and irregular verbs (Mean(IRREGULAR)=107.7 Mean(REGULAR), 70.6; StdErr=15.7, F(1,48)=2.7947, n.s.). Furthermore, irregular verbs comprised of verbs that have a change of the stem vowel in the 2nd and 3rd person singular in the present tense (11 irregular verbs) and verbs that did not alternate in their stem vowel (14 irregular verbs). Yet another factor that determined the choice for respective verbs was whether the stem final segment was a /s/ or not. 16 words with stem final /s/ (8 irregular, 8 regular) were chosen for the corpus construction. Additionally, verbs were chosen that are homophonic with stem final /s/ verbs in the 2nd person singular. For example, both the verb *hauen* 'to beat' and *hausen* 'to dwell' have as the second person singular verb form *du haust* 'you beat/dwell', but underlyingly, *du hau-st* 'you beat' and *du haus-st* ('you dwell') are differently. A question that could be also investigated due to this choice of verbs is whether phonetic detail could possibly differentiate the two verb forms. If, for instance the /ss/ is produced differently than the /s/ or if there is a difference concerning /t/ deletion, listeners could differentiate the /s/ stem verbs from the non-/s/-stem counterparts.

Table 10:
Paradigm of the verb ***hauen*** *'to beat' and the cells of the paradigm that had to be produced for the verbs in the respective conditions, is indicated by ×. Column 1 indicates the pronouns for the respective verb form.*

Pronoun Person/Number			Verb stem + Suffix e.g. hau-en 'to beat'	CONDITION I	CONDITION II	CONDITION III
Ich	'I	– 1st sg'	*hau-e*	×	×	
Du	'you	– 2nd sg'	*hau-st*	×	×	×
Er/Sie	's/he	– 3rd sg'	*hau-t*	× (he)		× (she)
Wir	'we	– 1st pl.'	*hau-en*	×	×	
Ihr	'you	– 2nd pl.'	*hau-t*	×	×	×
Sie	'they	– 3rd pl.'	*hau-en*	×		×

Subjects were asked to produce the verbs and (parts) of their paradigm. Every subject had to produce the verb in three different conditions. Table 10 gives an overview of the complete paradigm for the verb *hauen* (to beat) and the three production conditions. In CONDITIONI, the complete paradigm of the present tense had to be produced (e.g. "*ich haue, du haust, er haut, wir hauen, ihr haut, sie hauen*"). In this condition, the 2nd person singular form of the verb is followed by a vowel [e:] of the pronoun *er* (he). In CONDITIONII (cf. Column 4 in Table 10), four inflected verb forms were required. In this condition, the crucial verb form was preceding the pronoun *wir* (we) with a voiced labiodental fricative [v] as initial segment (i.e. "… *du haust, wir hauen* …"). Finally, in CONDITIONIII, the verb form of interest was followed by the singular feminine pronoun *sie* (she) (i.e. "… *du haust, sie haut, wir hauen* …"). Canonically, this pronoun would be produced with an initial [z]. However, in the Southern German dialects of our speakers as well as in fast and conversational speech, this fricative is realized as a voiceless [s].

Thus, each subject had to produce the 2nd person singular in three different phonological contexts (i.e. [eː], [v], [s]). This allows for a close control over the contexts in which the final /t/ occurred. Each verb had to be produced in every condition creating 150 verb production conditions, which were pseudo-randomized in one list.

Method

Overall, 10 subjects from the Universities of Frankfurt and Konstanz (6 female, 4 male) participated in the production task. They received monetary compensation for their participation and were not told the purpose of the study beforehand. Subjects were given the infinitive of each verb in the center of a power point slide (e.g. *hassen* 'to hate'). Underneath each form, the relevant personal pronouns according to the conditions illustrated in Table 10 indicated which forms had to be produced (e.g. *ich, du, wir, sie*-PL. 'I, you, we, they'). Hence, subjects had to create the paradigm forms for themselves and did not perform a pure reading task. They were unaware of the purpose of the study (i.e. an investigation of the reduction of final /t/ in 2nd person singular verb forms). Since every verb occurred in each condition, subjects had to perform the production of 150 paradigms, including 150 times the verb in the 2nd person singular. Subjects could determine the speed of presentation for themselves. When they pressed the mouse button, the next trial was presented on the screen. During the session no feedback was given as to the accuracy of their production. Additionally, emphatic orders as "do not slow down" or to "speed up a little" were presented on the screen. These orders were given randomly and did not correlate with the subjects' performances. The purpose of these instructions was to keep the speaking rate at a high level.

Subjects were asked to produce the verb forms as quickly as possible. At the same time, we wanted to ensure that each form was produced correctly at least once. Therefore, we instructed subjects not to worry about mistakes. If they realized they made a mistake, they just should repeat the word in question. Therefore, in order to correct a speech error, they only had to repeat the wrong item correctly. This was done to make subjects care less about mistakes and concentrate less on a perfect pronunciation. A pretest had shown that asking subjects also to avoid mistakes resulted in productions that were much less natural and slower.

Subjects received written instructions before the task. First, a training session with different verbs than in the test session ensured that subjects could familiarize with the task. Then, the test session began. Overall, the production task lasted approximately 25 minutes including instructions and the training session.

4.2.3.3 Results

In German, a canonically produced final /t/ consists of three physical events: a complete (alveolar) closure, followed by a release and considerable aspiration. The process of final /t/ deletion does not occur in an "either-or" fashion, rather there are gradual differences in the lenition of /t/. The canonically produced /t/ is on the one extreme end of this gradual pronunciation, whereas the completely deleted /t/ lies on the other extreme. A dichotomous decision of either "/t/ produced" or "/t/ deleted" is not always easy to perform (cf. Mitterer & Ernestus, 2006). For the analysis of /t/ deletion, however, such a dichotomy is crucial. The following criteria were used to decide whether a /t/ in question was deleted or not:

Final /t/ was labeled as "deleted", when there was none of the three characteristic events in the speech signal. That is, neither a closure, nor a release nor aspiration could be found in the signal, as in the example of Figure 12(a).

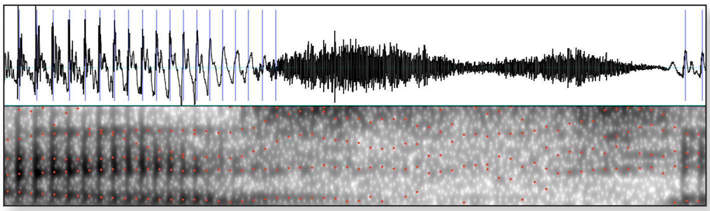


Figure 12(a):
Example for deleted /t/. There is no indication for any of the three physical events in the speech wave form and in the spectrogram. Deletion of /t/ , due to the context of another preceding /s/ leads to a sequence of two alveolar fricatives.

If all three physical events were present in the signal, /t/ was assumed to be present and produced canonically (cf. Figure 12(b)).

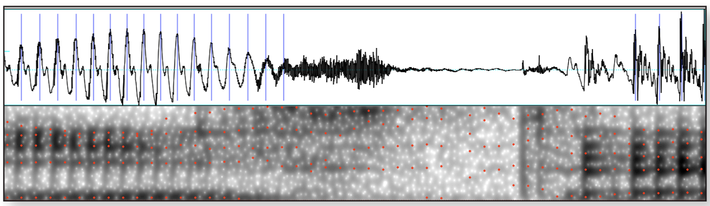


Figure 12(b):
Example for canonical /t/. Closure, release and aspiration phases are clearly visible in the signal. The closure of the alveolar stop was labeled with the corresponding IPA symbol and the release and aspiration phase with /h/.

Besides a complete omission of the final /t/, there was another pattern that occurred regularly in the data. Subjects produced an audible and visible closure and release, but abstained from producing aspiration. This was treated as /t/ reduction (cf. Figure 12(c)). There was no instance, where there was only aspiration but no closure.

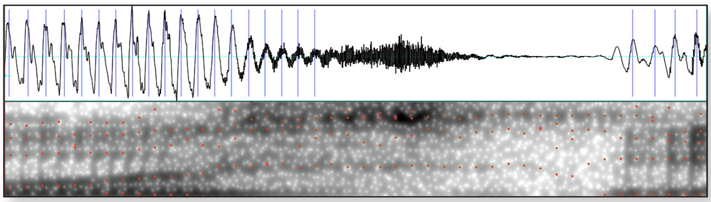


Figure 12(c):
Example for reduced /t/. There is a visible closure period labeled by the IPA symbol for the alveolar stop, but no clear aspiration.

The labeling of the corpus was carried out by a phonetically trained graduate student unaware of the purpose of the study. The program used for this task was PRAAT (Boersma & Weenink, 2007). If the /t/ deletion occurred in the /s/ context, the result of the deletion was a /ss/ segment. The length of the two segments was determined by halving its complete length if there were no cues for segment boundaries present as, for example, indicated by a drop in the signal's amplitude different to what can be seen in Figure 12(a).

For the analysis, 54 cases had to be excluded (3.6% of the overall data). In these cases subjects did not produce the correct verb form, or a wrong verb, or they did not produce the verb form at all.

Separate analyses for the deletion and reduction rates as defined above were calculated. We also analyzed the duration of the preceding /s/, depending on several factors. For the dichotomous deletion and reduction variables, a separate multiple logistic analysis with the factors GENDER, SUBJECT (nested under GENDER), PAUSE (period of silence after the 2nd person singular forms), FREQUENCY (log values of the relative frequency per million, the relative frequency, as discussed above, was based on the ratio of surface frequency and infinitive frequency, according to CELEX, Baayen et al., 1995), VERB CLASS (regular, irregular), STEM (s-stem, other stem) and CONTEXT (following /s/, /v/, or /eː/) was performed.

There was a main effect of GENDER (Wald-χ^2=49.77, p<0.001), reflecting the fact that males deleted /t/s more often than females (30.4% vs. 13.1%), as well as a main effect of PAUSE (Wald-χ^2=27.92, p<0.001), showing that the deletion rate dropped if subjects made pauses after the relevant word forms (from 24.3% to 3.7%). Furthermore, there was a strong effect of CONTEXT (Wald-χ^2=85.32, p<0.001). Most deletions occurred before /s/ (45.5%), while fewest deletions were found before /eː/ (3.3%). The deletion rate before /v/ was intermediate (11.5%). Finally, there was a significant interaction of STEM X CONTEXT (Wald-χ^2=6.51, p<0.05), which was driven by a higher deletion rate for non-/s/-stems than for /s/-stems (14.0% vs. 6.4%). None of the other factors or interactions were significant (all χ^2<4.00). In all, the model turned out to be quite explanatory (R^2=0.4327).

The reduction analysis yielded partially similar results. There were fewer reductions if the word forms were followed by a PAUSE (19.3% vs. 29.7%, Wald-χ^2=18.65, p<0.001). However, in contrast to the deletion data, females reduced more often than males (30.0% vs. 23.8%, Wald-χ^2=13.44, p<0.001). There was a significant interaction of STEM and VERB CLASS (Wald-χ^2=6.98, p<0.01), driven by higher reduction rates for irregular than for regular forms with /s/-stems (30.3% vs. 21.6%). There was no CONTEXT effect, and all other factors and interactions were not significant (all χ^2<3.00).

Finally, we calculated a mixed-model ANOVA (Baayen, Davidson, & Bates, to appear) for the duration of the preceding /s/ with SUBJECT and ITEM as random variables (using the REML).[49] Since we were interested in possible compensatory lengthening processes of the preceding /s/ depending on the realization of the /t/, we introduced another variable REALIZATION with the values "deleted" (if /t/ was deleted), "reduced" (if /t/ was reduced) and "canonical" (if /t/ was neither deleted nor reduced). Apart from this variable and GENDER, we used the same factors as for the logistic analyses. In order to avoid a quantitative variable, we transformed the relative log frequencies into a dichotomous variable. We considered the relative log frequency to be high if the value was above the median of the total distribution, and low, if the value was below the median of the total distribution.

The ANOVA showed a main effect of PAUSE (F(1,1343)=98.81, p<0.001) and STEM (F(1,86)=9.86, p<0.003). The fricative was produced longer if there was a pause after the word form (119 ms vs. 95 ms) and if the stem ended in /s/ (107 ms vs. 96 ms). Furthermore, /s/ was significantly longer if /t/ was deleted compared to reduced or canonical /t/ realizations (112 ms vs. 98 ms/96 ms; F(2,1360)=48.51, p<0.001). There was also a main effect of CONTEXT (F(2,1349)=8.92, p<0.001) and a significant interaction with REALIZATION (F(4,1351)=4.94, p<0.002). Generally, the fricative was longer if there was a following /s/ (113 ms) than if there was a following /eː/ (91 ms) or /v/ (96 ms). This effect depended on REALIZATION and only held if /t/ was not deleted. If /t/ was deleted, the duration of /s/ did not differ between /s/ and /eː/ (t=1.14, p<0.15) and between /s/ and /v/ (t=0.26, p<0.81). Besides, STEM interacted both with CONTEXT (F(2,1348)=8.41, p<0.001) and REALIZATION (F(2,1356)=6.68, p<0.002). The duration of /s/ in /s/-stems differed from the duration of /s/ in non-/s/-stems only if the following context was also /s/ (126 ms vs. 107 ms, t=5.76, p<0.001). Furthermore, the same difference was found only if /t/ was

[49] This statistical analysis was calculated with SPSS (Version 15).

reduced (108 ms vs. 93 ms, t=4.21, p<0.001). Within the /s/-stems, there was no duration difference between a canonical and a reduced /t/ realization (98 ms vs. 109 ms, t=2.00, p<0.08). Similarly, there was also no duration difference in these pairs of comparisons in the non-/s/-stems (96 ms vs. 93 ms, t=0.31, p<0.77).

4.2.3.4 Discussion and Conclusions

The objectives of the corpus construction were to investigate phonological and morphological factors on word final alveolar /t/ deletion in German. Partially, the findings on word final /t/ deletion in other Germanic languages such as Dutch and English were replicated. However, the results presented here also show contradictory tendencies, especially with respect to expected morphological effects. Overall, final /t/ was deleted in 289 cases (of 1446 possible /t/ realizations, 20%). This deletion rate is almost identical to the overall /t/ deletion rate of 21.4% in the Kiel Corpus. The fact that the overall deletion rate of final /t/ as part of a suffix was almost identical to the Kiel corpus, where neither context nor position nor morphological status are taken into account, is indicative for the adequacy of the corpus construction as a method to resemble natural speech data.

First, concerning extra-linguistic factors, the analyses showed a stable gender effect on /t/-deletion. Male speakers deleted /t/ more often than female speakers. This is in line with the findings of Byrd (1994), Neu (1980) and Wolfram (1969). Next, it was found that hesitational pauses decreased the amount of deletions. If one parallels hesitational pauses with fluency, the results again conform to previous studies showing fewer deletions in dysfluent speech or at prosodic breaks also indicated by pauses (e.g. Fougeron & Keating, 1997; Fox Tree & Clark, 1997; Shriberg, 1999; Kingston, 2006).

In the Introduction part, several linguistic factors were discussed that have been found to influence the amount of /t/ deletion. For phonological factors, the corpus was constructed in a way that allowed only for an examination of the following context. Our results confirmed once more previous investigations showing a strong influence of the following context. In particular, the vocalic context demoted the deletion rate, while a following coronal fricative led to a deletion rate of almost 50%. The labio-dental fricative produced an intermediate amount of deletion. The latter two findings are somewhat different from Mitterer & Ernestus (2006), who found the highest amount of deletion in front of a (bi)labial consonant. However, this difference is likely to result from differences in the data sets. In this corpus, the /s/ context allowed for cluster simplification since the preceding context was consistently the coronal fricative /s/ (cf. Kohler, 1995a: 209), and the /v/ fricative is labiodental, where the result of /t/ deletion is not a cluster of two identical segments, and /v/ is not bilabial, which arguably could also have an influence on deletion of preceding word final /t/. Other phonological factors were kept constant across the different conditions and are not investigated further.

When it comes to the morphological factors, the morphological status itself was irrelevant in this analysis, since every instance of final /t/ arose in the same suffix (i.e. {–st}). An objective was to investigate the role of relative frequency on the amount of /t/ deletion which has been suggested as crucial in determining the amount of /t/ deletion (cf. Hay, 2003). These findings on final /t/ deletion, however, do not lend support to the effect of relative frequency on final /t/ deletion. This result is also not in line with accounts that propose a dual mechanism for irregular and regular verbs. No difference was found between these two verb classes.

As indicated above, complete /t/ deletion can be considered as extreme end on a reduction scale. We therefore also investigated the amount of final /t/'s that were reduced. The analysis of the reduction of /t/ shows that the results are not identical to the ones obtained for /t/ deletion. Overall, 398 words had the final /t/ reduced (out of 1446, i.e. 27.5%).

The extra-linguistic variables showed a strong gender effect. However, this time it was female speakers to reduce more often than the male speakers. This could be interpreted as showing that male speakers reduce more drastically, but not more often than female speakers. The gender effect has been found in some studies, and was absent in others (cf. Byrd, 1994; Raymond et al., 2006, or Section 4.2.2.). Thus, GENDER as factor might be influenced also by other factors such as speaking rate, or grade of reduction. As in the deletion analysis, pauses led to fewer final /t/ reductions. This is in line with lenition accounts that show that (prosodic) boundaries are indicated by more canonical productions of segments (e.g. Kingston, 2006). For linguistic factors, the emerging reduction patterns are rather different than for deletion patterns. Context was no significant determiner for /t/ reduction. The role of morphological factors on final /t/ reduction is also somewhat different from the ones for /t/ deletion. Whereas neither verb class nor relative frequency contributed as predicting main factors, there was an interaction of Verb class and /s/-stem. This interaction was driven by higher deletion rates for irregular forms with /s/-stem. At this point, no meaningful explanation can be given.

Finally, the duration of /s/ was analyzed to see whether deletion of /t/ resulted in different /s/ realizations. Previous research suggested that in a final /st/ cluster, the /s/ is shorter than a single final /s/. This difference was held constant even after /t/ deletion and interpreted as cue for listeners for an underlying /t/ (Mitterer & Ernestus, 2006). In this verb corpus, subjects seem to compensate for /t/ deletion in that /s/ were produced longer in these cases, however. This result is opposite to findings by Mitterer and Ernestus (2006) where subjects did not compensate for /t/ deletion. They showed also that in perception studies the short [s] was taken as cue for an /st/ cluster, regardless of the presence of the plosive in the signal. The data presented in this dissertation suggests that, this strategy is not viable cross-linguistically, since German speakers lengthen the /s/ when the /t/ gets deleted.

4.2.4 Discussion of the Production Data

In a summary of the results, it is safe to conclude that deletions do occur regularly, though not rule based, in conversational German. Less than half of the words are produced canonically. Deviations from the underlying representation are the norm in language production rather than the exception. The huge amount of variation becomes also evident by the analysis of deletions in the Kiel corpus. When questionable phonemes are set aside, 8.9% of the underlying segments are deleted. If the segments with questionable phoneme status are included, 14.1% of the segments are deleted. A considerable amount of segments is missing in natural speech. In some cases, the deletions and reductions are even massive. Especially such words pose a possible challenge to word recognition.

More specifically, the combined results from the Kiel corpus analysis and the study on final /t/ reduction show that /t/ reduction is a general process in German. The process seems to be regular but not based on a traditional phonological rule, such as final devoicing and occurs across different phonological contexts alike. However, in its extreme form, i.e. complete deletion, phonological context is crucial. For the prime objective of this dissertation, the evaluation of X-MOD and FUL, the data reported here does not add evidence for or against one of the models. Many factors have to be included when analyzing reductions in general. The analysis showed that it is almost impossible to account for all the reductions that occur in natural speech by rules. This finding is water on the mills for episodic approaches such as X-MOD. They allow and expect random variation in natural language in a large amount as a basic assumption (e.g. Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001a). This is also what we find in naturally spoken German. However, the findings are not very decisive. Exemplar-based approaches have been suggested as way to include huge amounts of variation in the representation, the fact that variation occurs, thus, is not very telling, and rather trivial for episodic approaches. Only if no variation were found, an argument against episodic storage could be made. What makes episodic models more attractive if only the production results are taken into account, is that variation is explicitly treated in these models. Reduction processes are included in representations because they occur in naturally produced language. FUL, on the other hand, does not deny the existence of such reduction processes. For FUL, only phonological rules are included in the grammar of a speaker. Other variation is based on phonetic variables, but not part of the mental representations. Depending on the extent of variation, speech perception is affected. The prediction of FUL for perception is that especially those reduction processes that are not predictable and those that change words severely have deteriorating effects for speech perception. These different predictions will be tested in the next section of this dissertation.

Before we turn to the presentation of the perception side of massive reductions, a more general observation has to be made. An important result of this section of the dissertation so far is that despite the usefulness of corpus-studies for a more realistic modeling of linguistic theories, a concentration on only one source of information is dangerous. If only a general overview over processes that occur in a corpus is given, the factors that influence these processes cannot be analyzed completely. Many factors are dependent on other variables, despite the large number of data points, the distribution is not normal, and not all combinations of every variable can be tested consequently. Therefore a combination of the investigation of the behavior of a single segment, and on all the segments that are produced in a corpus is the most promising strategy. While naturally produced corpora allow for the discovery of general trends and processes, the actual individual contribution of these are best studied in a more controlled way. Furthermore, a mixed strategy for the creation of speech corpora (i.e. mixed between controlled for factors or free speech) is also a very promising approach. This method allows for the close control over many variables, allowing for the examination of other variables without confounding influences.

4.3 Perception Data

Within a model assuming discrete, abstract phonological entities as basic units in the mental lexicon, such as the Featurally Underspecified Lexicon (FUL) model of speech perception (Lahiri & Reetz, 2002, accepted), it is assumed that reductions, deletions, insertions, and assimilations are processes that modify canonical pronunciations in natural speech. The FUL model is able to explain, and predict some of this variation, such as assimilations (e.g. Zimmerer et al, 2009; for other phenomena, see Ghini, 2001a,b; Scharinger, 2006; Lahiri et al., 2006; Wetterlin, 2007). Yet, reductions and deletions pose possible problems for a model with abstract representations, as has been pointed out in the literature, especially for perception of altered word variants (cf. Johnson, 2004a). On the other hand, accounting for the handling of reductions and deletions is one of the strengths of exemplar-based models. Due to their architecture, variation is part of the lexical representation (e.g. Johnson, 1997, 2004a; Goldinger, 1998, Pierrehumbert, 2001a). Traces of heard episodes are retained with ample phonetic detail. These episodes allow for recognition of severely reduced words if they have been heard before. Recognition of lenited variants even improves if they are frequent and common, because similar episodes will ultimately lead to a larger exemplar cloud with a higher resting activation. This in turn predicts more activation when a similar, reduced probe is encountered during auditory word recognition (e.g. Johnson, 2004a; Goldinger, 1998).

Ever since data from natural speech corpora became more prominent in linguistic research and the enormous variability in general and reductions and deletions in particular have become tangible, purely abstract models have been increasingly called into question. The argumentation for such a critique can be summarized as follows: since there is only one (featural) representation in the lexicon, (irregularly)[50] reduced words will not activate the (perfectly) stored entries in the lexicon. Also, many deletion processes are not predictably linked to certain features or segments, rather to individual words or words in certain contexts. Therefore, there are no rules that can predict the resulting variation, and there is no mechanism in "undoing" the massive reductions before mapping them onto lexical representations. Furthermore, if the exact matching requirement is slackened, word recognition will also fail, since too many word candidates will be activated, and there would be no possibility to decide on the correct candidate. In natural speech, there occur many, sometimes "massive" reductions, therefore abstract models would predict that auditory word recognition would fail. However, data from natural speech tells us otherwise: listeners do not have difficulty in understanding words, even when they are massively reduced. Therefore, other models, such as exemplar-based ones, should be preferred (Johnson, 2004a).[51]

While this argumentation is coherent and logically true, it is also inherently fraught with a problem. It is entirely based on data from production and on impressionistic data on sentence perception (i.e. there is massive reduction and listeners still understand what has been said). So far, it has not been established whether massively reduced word forms actually do or do not activate the correct word in the lexicon, or whether it is due to other processes that allow for (near) perfect recognition of reduced words in context.[52] The question of whether considerably reduced words activate the intended lexical entries when they are encountered out of context by listeners is investigated in this chapter.

What do we know about the effects of reductions and deletions on speech perception? Some studies have examined some aspects of these effects. An important finding that has been reported by several studies using different methods is that regularly reduced words are able to activate the correct lexical entries (e.g. Lahiri & Reetz, 2002; Gow, 2003; Connine, 2004). However, there is also evidence that even if in natural speech they occur more often than unreduced versions of the word, they are less successful than (less frequent but) more canonically produced words (e.g. Ernestus & Baayen, 2007; Tucker, 2007; Tucker & Warner, 2007).

---

[50] Irregular is used here in the sense 'not predictable, and not rule based'.

[51] There are also other points of critique. However, the main point is ultimately that abstract models would fail in recognizing massively reduced words.

[52] Because the argumentation is largely based on production data, without actual data from natural speech perception, it can be flawed to some degree: It is also forbidden to do something like "crossing streets when the light is red", however, as real life tells us, there are many people who do this nonetheless, sometimes with severe consequences for themselves and others.

As already discussed in the first part of this chapter, alveolar stops are very prone to reduction processes. Not surprisingly, they have been very prominent not only in research on the reduction processes themselves, but also on the effects these reductions have for speech perception (e.g. Sumner & Samuel, 2005; Tucker, 2007; Mitterer et al., 2008 and references therein).

An exemplary series of experiments on the effects of reduction for perception was carried out by Tucker and Warner (Tucker, 2007; Tucker & Warner, 2007). They compared words with "regular" flaps and words that had the flap reduced. For this investigation, they regarded the flap version of the word as the canonical one and examined the effect of further reduction of that flap, a process that is very common in conversational American English (e.g. Kiparsky, 1979; Zue & Lafierre, 1979; Selkirk, 1982). They investigated the effect of reduction using cross-modal identity repetition priming with lexical decision. Both variants were able to activate the intended word. But more important, their results also show that words with a reduced flap were worse primes than words with canonical flap (Tucker & Warner, 2007; Tucker 2007). This effect was obtained irrespective of frequency of occurrence (Warner & Tucker, 2007: 1952).

Flapped und unflapped variants of medial /t/ and /d/ and their effect on lexical activation was investigated by McLennan and colleagues (McLennan et al., 2003). In a series of six long term repetition priming experiments using shadowing, lexical decision and a mixture of these tasks, they examined the potential difference in lexical activation between the two pronunciation variants for *atom*, /ætəm/, and /æɾəm/, respectively. Note that the flapped variant is inherently ambiguous, since the word *Adam* /ædəm/ can also be produced as /æɾəm/. McLennan and colleagues found that, for these alveolar stimuli, both flap and unflapped versions in almost all experimental setups were able to activate the intended lexical entry. Furthermore, they found a specificity effect only in a lexical decision task with nonwords that were hard to detect (e.g. *bacov* – very close to *bacon*) where only identical variants lead to a priming effect in a second block. Interestingly, when the nonwords were easy to detect due to their similarity to existing English words (e.g. *thushshug*), the specificity effect was lost. These results suggest that both naturally occurring variants are able to activate an abstract underlying representation. McLennan and colleagues interpreted their results also as indication for an additional exemplar-based representation. However, it is not completely clear whether their experimental setup really tapped into the actual long-term lexical representations (cf. Sumner & Samuel, 2005). One important difference setting apart these results from the ones by Tucker and Warner (2007) is how flaps are treated. Whereas McLennan and colleagues treat the flap as a reduced item, Tucker and Warner assume the flap as the unreduced variant, and only regard the reduced flap as a lenited segment.

Sumner and Samuel (2005) examined the effect of reduction of word final /t/ in English on speech perception. The /t/ has several possible realizations in natural speech (cf. Raymond et al., 2006; Dilley & Pitt, 2007; or, for medial /t/, cf. Patterson & Connine, 2001; Tucker, 2007 and references therein). For instance, they can occur as glottal stop or glottalized version of /t/. Actually, the glottalized version is the most frequent type of word final /t/ occurring phrase finally in the dialect that was examined by Sumner and Samuel. However, in a short-term semantic priming experiment, all three variants were able to prime semantically related targets. There was no benefit for the most frequent of these variants, i.e. the glottalized /t/. Arbitrary variation, however, did not activate the correct lexical entry. In a long-term repetition priming experiment, though, variants were not equally effective as primes: only canonically produced items were successful. These results suggest that variants are not (necessarily) stored in long-term memory, even if they are the most frequent variants, and that one abstract "canonical" representation can account for the results obtained by these authors. The results also indicate that minimal variation, if lawful and naturally occurring, is tolerated by listeners, but arbitrary changes are not accepted.

Janse and colleagues (2007) also focused on the question how Dutch listeners cope with variation in final /t/. In a corpus study of naturally spoken Dutch they found highest deletion or reduction rates of /t/ in /st#b/ contexts. In a series of experiments they found that although lexical activation is possible when the /t/ is deleted, more canonically produced words fare better. Not all kinds of reductions are detrimental to speech recognition. Mitterer and Ernestus (2006) have shown for example that listeners are able to restore reduced /t/'s in running speech. Other works focusing on phoneme restoration effects have found similar effects (e.g. Warren & Obusek, 1971; Samuel, 1996).

Similarly, Deelman and Connine (2001) examined the effect of unreleased /d/ and /t/ word finally on the perception of such variant productions. For both variants, released stops are less frequent than their unreleased counterparts. In a cross-modal semantic priming experiment, they found that both variants showed comparable priming for a target that was semantically related, but no effect of variant frequency emerged. In a phoneme monitoring experiment, however, results were different. For /t/ stimuli, released variants had a much faster detection rate than their unreleased counterpart. For /d/ words, the advantage for released variants was still observable, but much smaller. These results also suggested that the amount of deviation from the canonically produced word influences recognition. The more words deviated from their canonical form the worse was the subjects' performance for those stimuli (Deelman & Connine, 2001).

Another kind of reduction was examined by Ernestus and Baayen (2007), who showed in an auditory lexical decision task that Dutch prefixed words such as *bestraten* ('to pave'), where the Schwa of the prefix <be-> was deleted, were recognized slower than "perfectly" produced words where the Schwa was present. This result was independent of the form frequency or the frequency of its stem form (Ernestus & Baayen, 2007: 776).

LoCasto & Connine (2002) investigated vowel reduction and deletion in words such as *elephant* or *police*, where the unstressed vowel can be reduced to Schwa or is even deleted completely (but see, for example, Manuel, 1991, arguing that there is only an incomplete neutralization due to deletion, differentiating a seemingly deleted variant of *police* and *please*). The result of an acceptability rating experiment conducted by LoCasto and Connine revealed that items that had the vowel deleted were rated less acceptable than counterparts with the vowels reduced (i.e. with a Schwa). These results suggest that the representation includes a vowel (or Schwa) (LoCasto & Connine, 2002: 216). The length of the word, that is the number of matching segments, attenuated the effect, however. Two subsequent form repetition priming experiments further investigated the effect of reduction and deletion on perception. LoCasto and Connine (2002) found an overall advantage for vowel reduced variants, which were produced with Schwa. The results also suggest that under some circumstances reductions and even deletions are tolerated. LoCasto and Connine (2002) argue that if the remaining matching segments are enough, no further process is necessary to activate the correct lexical entries. However, when deletions create words that do not have enough redundant information and activate too many competitors, phonological knowledge must kick in to undo a rule based deletion. In their account, the higher the amount of redundant matching information, the more tolerable are deviations to the canonical pronunciation.

Examining massive reductions that were not always rule based but still natural, Ernestus et al. (2002) found in a transcription task, that reduced words were transcribed significantly worse when presented without context. This was a replication of results obtained for English by Picket and Pollack in the 1960's (Picket & Pollack, 1963; Pollack & Picket, 1963).

Furthermore, many researchers have examined the effect of unnatural feature mismatches, either explicitly or in order to compare the results of natural arising mismatches (e.g. Connine et al., 1993; Coenen et al., 2001; Deelman & Connine, 2001; Bölte, 2001; Bölte & Coenen, 2002; Lahiri & Reetz, 2002). Those unnatural mismatches are usually created by changing segments or features of segments, although there are no phonological rules or even phonetic patterns that allow for such changes. Results of these studies show that mismatching features can be tolerated in some cases (e.g. Connine et al., 1993, Lahiri & Reetz, 2002), whereas other mismatching conditions block lexical access (e.g. Lahiri & Reetz, 2002). Another result of this body of research is that even if mismatching items could access the lexical entries, there was a cost associated to the arbitrary deviation. Priming Studies have shown that some kinds of reductions and delineations from perfect speech are "allowed" in speech perception, i.e. they are able to activate lexical items despite the fact that they are not perfect. Similarly, in a study that did not concentrate on reductions, Smolka and colleagues (2007) investigated the priming of German participles. They showed that participles that were built illegally, such as *gekäuft* (something like 'bough-ed') or *geworft* (something like 'throw-ed') instead of *gekauft*, or *geworfen* ('bought' or 'thrown') respectively, were able to activate their stem verb. Thus, even illegal, and "unnatural" – in the sense: not occurring in natural speech – variation was able to activate the correct entry in the lexicon, since the illegality did include the important information about the verb stem after decomposition.

The results consistently illustrate that even lawful, predictable variation that can be tolerated in word recognition processes most often causes a deterioration of speech perception. However, most of the studies have used stimuli that were produced intentionally for the purpose of the experiment (e.g. Lahiri & Reetz, 2002; LoCasto & Connine, 2002; Sumner & Samuel, 2005; Ernestus & Baayen, 2007; Tucker, 2007; Janse et al., 2007; Ranbom & Connine, 2007) or manipulated, for example by cutting out parts of words (e.g. Deelman & Connine, 2001). This sets apart this dissertation from most prior research. For the experiments reported in this dissertation, there will be only words from natural speech data, that is, words that have been produced in the course of corpus construction. One reason for choosing only items that were unintentionally produced is that intentionally produced items might lack important acoustic features and characteristics that occur in natural speech which can be used by listeners during word recognition (see, e.g., Manuel, 1991 for deletions; or Nolan, 1992 for assimilations). In the perception literature, naturally reduced items have been used both in identification tasks (e.g. Ernestus et al., 2002) as well as in lexical decision experiments (e.g. Ernestus & Baayen, 2007) and even in a semantic priming experiment (Snoeren et al., 2008). Up to date, however, no study has been reported using reduced words from natural speech in a series of experiments in a combination of transcription and lexical decision. The study reported by Snoeren and her colleagues (Snoeren et al., 2008) is a prime example underlining the importance of naturally produced experimental items. In their study, they used items that were produced for the experiment, but in a way that ensured for a rather natural production. Snoeren and her colleagues found evidence that French listeners were able to make use of minimal differences in cases of what seemed complete voice assimilations, and were able to differentiate between the underlying segments (Snoeren et al., 2008). This result clearly shows that items that are produced intentionally for the experiment or manipulated and controlled, might actually be different from natural speech and miss important information listeners can rely on in natural settings, casting doubts about the reliability of the results that are obtained with such "unnatural" stimuli.

The crucial results of previous research examining the effect of reduction and deletions on auditory word recognition are the following: studies clearly indicate that a) some reductions are tolerated; b) perfectly produced speech seems to be favorable in speech perception (faster RT's in some cases); c) some deviations are not tolerated in speech perception; d) for successful recognition of reduced words, context helps (see also, e.g. Sheldon et al., 2008).

In the upcoming sections (4.3.1 – 4.3.3) a series of three experiments is reported examining the effects of massive reductions on speech perception. The experimental paradigms, that is, transcription of reduced words, identity repetition priming of reduced words and a combination of these methods will shed light on the question how well listeners are able to understand reduced words when they are presented out of context.

4.3.1 Experiment 3: Transcription of Words out of Context [53]

Listeners can deal with naturally occurring regular reductions if they are the results of rule based patterns, and if they are limited in scale. However, how do listeners react when they are faced with more severe, yet natural deviations from a canonical form? This experiment was designed to find out how subjects would react to reduced words without context. Thus, sentential context including semantic, syntactic, morphological, phonological, and phonetic (i.e. acoustic) information, could not be used by the subjects to resolve any ambiguities or problems in speech perception. Subjects had to perform a single word transcription task, where they were asked to write down the word they heard and indicate their confidence of the transcription they provided.

Transcription expectations

Based on evidence from earlier studies the expectation was that indeed listeners would have more problems in transcribing reduced items correctly (cf. Picket & Pollack, 1963; Pollack & Picket, 1963, Ernestus et al. 2002). The transcription experiment was not designed to set apart the two theoretical frameworks. Actually, both theoretical frameworks predict that subjects will have problems in transcribing massively reduced words out of context. The difference between the frameworks concerns the explanation why this is the case. This experiment also allows for an estimation whether there are differences between the reduction patterns and differences in the kinds of mistakes subjects will make in their transcriptions.

[53] Preliminary results of this research have been presented in Zimmerer et al., 2008

The FUL model, assuming one abstract lexical entry for each word, upholds that severely reduced words are not able to activate the correct lexical entry, since when too many features are absent (due to the deletion of a segment, for example) there is mismatching evidence between the acoustic signal and the lexical representation. The listener's lexical entries may sometimes be correctly activated if the deviation from the canonical representation is not too large, or, given some time to think, the correct entry will be activated, since it is still the best match. However, there is also a good chance, that there will be no successful recognition of severely reduced items, as illustrated by the *Kranz/kratz* example in the introductory part of this chapter.

For the exemplar model, correct activation is also not always to be expected. Although the reductions occur naturally in spoken German, (and hence, listeners will have encountered and stored very similar instances of the words,) the problem might be that too many word candidates are activated, and subjects cannot always decide for the correct entry. However, correct transcriptions are also expected. Since there is no way in a simple transcription experiment to control for exact subjects' prior encounter of reduced words, this experiment, although being a first important examination of the effect of reductions on speech perception, will be followed by other experiments in order to gain better understanding on the kind of model that can explain best the data. The results of this experiment will also serve as an aid in choosing the experimental items for the repetition priming experiments reported in Section 4.3.2 and 4.3.3.

Materials

Overall, 92 word pairs were selected from the Kiel Corpus (IPDS, 1994). Each word pair consisted of an unreduced and a reduced instance of that word. For the experimental items, an item was labeled reduced when, according to the transcription provided with the corpus, it was produced with more reduction than the unreduced counterpart. In general, unreduced words showed reduction of two or less segments, whereas reduced words were transcribed as having at least two segment reductions. Additionally, 16 items were added for control between the two groups. The experimental items were taken from utterances of 35 different speakers (15 female, 20 male). Of the 200 items (92 reduced, 92 unreduced, 16 fillers), 125 were uttered by male speakers, 75 by females. No speaker uttered more than 15 words that were used for the experiment. The items were cut out of their phrase context using PRAAT (Boersma & Weenink, 2007). The number of syllables reached from 1 to 4 syllables, (Mean: overall=2.3 (SD=0.78); word pairs=2.3 (SD=0.76); control=2.3 (SD=1.01)). Two lists were constructed with each 46 reduced, 46 unreduced words and

16 fillers. The resulting experimental lists and the experimental items presented a huge amount of variation, one that probably even exceeds natural conversations, since we usually do not interact with such an amount of different speakers and in natural speech situations words are not heard without their phrase context. The complete set of words as used in Experiment 3 is given in Appendix E. The experimental lists were recorded on a CD and presented over headphones (Sennheiser HD520II).

Subjects and procedure

In all, 22 students from the University of Konstanz participated in the transcription experiment (12 female, 10 male). All were native Germans, and did not report on any hearing problem. They were tested in groups of four or less and paid for their participation. They received written instructions, which, if necessary, were additionally explained orally. A booklet for the transcription was placed in front of them on a table. They were instructed to listen to words and subsequently write down what they thought the word was. Additionally, they were asked to write down how certain they were about their transcription. This confidence rating ranged from 0 (absolutely uncertain) to 10 (absolutely certain). Before the test phase, subjects were familiarized with the task by help of 25 practice items. They were not given any feedback about the 'correctness' of their transcriptions.

The sequence of presentation was as follows. Each test item was preceded by a warning tone of 300 ms and a 200 ms pause. After each test stimulus, there was a pause of 4 s for the transcription and the confidence rating. A single experimental session lasted approximately 15 minutes including the practice items.

Results

The subjects provided transcriptions and confidence ratings, which will be reported in turn. First, the correctness of the transcriptions was analyzed. If the word that had been transcribed and the word that had been uttered by the speaker of the corpus were completely identical, the transcription was labeled as correct. The only deviations that were still labeled as correct were case insensitivities. This was done, since words like *M/mittag* '(after) noon', as a noun or adverb, or *S/schaden* '(to) damage' as verb or noun can be written both with capitals and without. All other deviations were treated as incorrect transcriptions, i.e. as a different word to the one that the speaker uttered. That means that various inflectional or other affixes setting apart the transcription from the experimental items were treated as wrong transcriptions. Words where no transcription was given were also treated as incorrect.

Overall, 446 (out of 2376, = 18.8%) of the experimental items were not transcribed correctly. Out of those, 76 items, or 3.2% of the overall transcriptions, were not transcribed at all (34 of those received a confidence rating by the subjects of '0', the rest, i.e. 42 words, none). Control items were transcribed correctly in 97.4% of the time. Accuracy rates for individual items reached from 81.8% to 100% correct for control items. There was no significant difference of ACCURACY across the two experimental groups. They were transcribed correctly 96.6% in one group, the second group performed the transcription correctly 98.2% of the time. For further analysis of ACCURACY, control items were not taken into account.

Individual analyses showed that subjects reached a rate of accuracy between 75.9% and 85.2%. Unreduced items were written correctly in 94.6% of the cases, whereas reduced items in only 62.3%. Figure 13 depicts the amount of correct transcriptions. Accuracy for individual items ranged from 0% to 100% for reduced items, and from 27.3% to 100% for unreduced items.

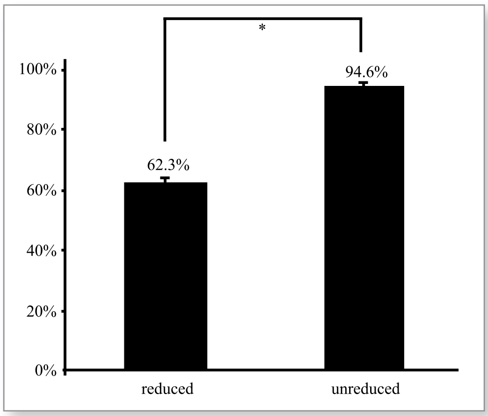


Figure 13: Correctness of transcriptions, comparing unreduced and reduced items, with standard error. Significant differences are indicated with an asterisk.

To further investigate the differences between the categories, a statistical analysis of variance (ANOVA) was carried out. For the analysis, ACCURACY was examined as dependent variable. The factors SUBJECT (as random variable), WORD and CONDITION (reduced, unreduced) were used as independent variables. As in Chapter 3, the REML estimation was chosen for the analysis. The analysis revealed that both CONDITION ($F(1,1910)= 500.0524$, $p<0.0001$) and WORD ($F(91,1910)= 8.9132$, $p<0.0001$) were significant main effects. To summarize, reduced items were transcribed significantly worse than their unreduced counterparts ($t=-22.36$, $p<0.0001$).

Since subjects not only transcribed what they thought they heard, but also indicated how confident they were about their transcription, the second analysis is based on the confidence ratings as provided by the subjects. In 54 cases (2.3%), no confidence rating was given. Filler items were not rated in 2 out of 352 cases (0.6%), unreduced words received no confidence rating in 10 out of 1012 cases (1.0%) and reduced items had no confidence rating in 42 out of 1012 instances (4.2%). These cases were not taken into account for further analyses of confidence rating. Note that out of the 54 items without confidence rating, 32 items also had no transcription given.

On an overall average, subjects rated their transcriptions with 8.23. The confidence ratings for control items reached 9.53 on an overall average, for both groups transcription was identically rated with a mean of 9.53. Confidence ratings for individual control words were given on average between 8.18 and 10. Unreduced words were rated with slightly less confidence than control items, 9.21. The lowest confidence ratings (with 6.75 on average – paralleling the less correct transcriptions) were given to reduced words. For further analysis, the control items were excluded. Figure 14 illustrates the differences between reduced and unreduced items.

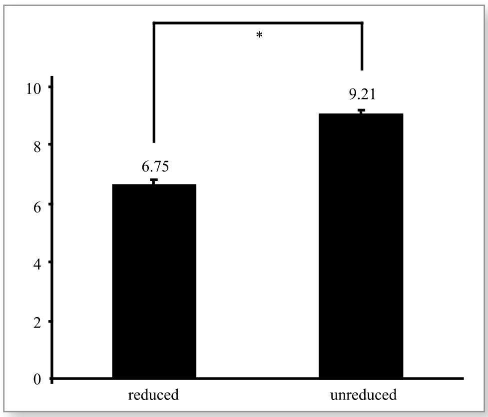


Figure 14: Confidence rating for transcriptions, comparing unreduced and reduced items, with standard error. Significant differences are indicated with an asterisk.

The average confidence ratings for individual stimuli ranged from 4.1 to 10 for unreduced items, and were between 1.56 and 10 on average for reduced items. To further investigate the differences between the categories, the same ANOVA as for the accuracy rates was calculated for the confidence rating as dependent variable. (Again, the factors SUBJECT (as random variable), WORD and CONDITION (reduced, unreduced) as independent variables were also entered into the ANOVA, using REML estimation). The analysis revealed that both

CONDITION (F(1,1858)=606.1834, p<0.0001) and WORD (F(91,1858)=10.311, p<0.0001) were significant main effects. The results for the confidence ratings clearly reveal the same kind of asymmetry as did the accuracy rates: reduced words were transcribed with less confidence than their unreduced counterparts (t=-24.62, p<0.0001).

As explained above, only completely identical transcriptions were treated as correct in the analysis. However, the transcriptions provided by the subjects also allowed for a more detailed analysis of the mistakes that differentiated between the word that was intended by the speaker of the corpus and the word that has been transcribed. The transcription mistakes were split up into five different categories. Table 11 displays the amount of mistakes in each category.

Table 11: Different kinds of mistakes, split up by category, percentages in parentheses

	unrelated	1st segment correct	wordform incorrect	rhyming response	1st feature correct	no response	overall
unreduced	11 (1.1)	13 (1.3)	8 (0.8)	11 (1.1)	7 (0.7)	5 (0.5)	55 (5.4)
reduced	115 (11.4)	91 (9)	48 (4.7)	18 (1.8)	40 (4)	70 (6.9)	382 (37.8)
Sum	126 (6.2)	104 (5.1)	56 (2.8)	29 (1.4)	47 (2.3)	75 (3.7)	437 (21.6)

A closer analysis of the mistakes as depicted in Table 11 indicates that subjects made very different transcription errors. The categorization into different error groups followed independent linguistic factors that have been identified as playing a role in speech perception and production. In 126 cases, the transcription was unrelated to the intended word, for example, *Feuerwehr* 'firefighters' was transcribed instead of *Karneval* 'carnival'. 115 such cases occurred for reduced words, 11 transcriptions were completely unrelated unreduced items. In 104 cases, the transcription of the first segment was correct, albeit not the complete word. This category of mistakes was chosen, since researchers have identified the first segment of a word as crucial for subsequent speech perception (cf. Connine et al., 1993 and references therein). Such a case occurred, for example, when the intended *dauern* 'to last' was transcribed as *dort* 'there', both beginning in [d]. Reduced items were correct in the transcription of the first segment in 91 cases, but only in 13 instances for unreduced words. Word form mistakes occurred if subjects transcribed, for example, *bekommt* '(he) gets, (you-pl.) get' instead of *bekommen* 'to get'. This kind of error is relatively "harmless".

Because there was no phrase context, subjects could not rely on additional information which word form would be correct. In a sentential context, for example, the personal pronoun helps to identify the correct word form. Such mistakes could be observed in 56 cases (48 for reduced items, 8 for unreduced ones). Rhyming responses occurred, when subjects transcribed a word that rhymed with the intended word, for example *fassen*, 'to catch' instead of *passen* 'to fit'. Rhyming responses were observed overall in 29 cases (18 for reduced items, 11 for unreduced words).[54] One possible mistake was characterized by an incorrect transcription of the first segment, while at least one feature of the first segment was (e.g. PoA) correct. Such mistakes are different than mistakes where the first segment was correct. It is indicative, however, that listeners extracted at least parts of the features of the first segment. This kind of mistake occurred in 47 instances (40 for reduced items, 7 for unreduced ones). An example for this kind of mistake is a subject transcribing *Woche* 'week' instead of *brauchen*, 'need' where there is a match of the place of articulation feature [LABIAL] of the first segment. In 75 cases, (70 reduced, 5 unreduced) no transcription was given at all. As can be seen, there is no case of transcription mistakes, where reduced words were transcribed more correctly. The results and analyses clearly indicate that, for words from conversational speech, reduction impedes correct transcription, if they are presented without context.

Discussion

Both accounts, FUL and X-MOD, correctly predicted the results obtained by this experiment, namely, that reduced words are harder to transcribe than unreduced words. This experiment also replicated results from previous research (e.g. Picket & Pollack, 1963; Pollack & Picket, 1963; Ernestus et al., 2002). However, if listeners encounter reduced words in conversational speech, they seemingly are still able to understand what the speaker just said. One explanation for this discrepancy between the results of Experiment 3 and natural speech lies in the nature of the stimuli themselves. In natural speech, words are not uttered in isolation, and if so, there is less reduction. Listeners are usually able to extract more information from context. Context is a crucial factor for understanding (e.g. Picket & Pollack, 1963; Pollack & Picket, 1963, Ernestus et al., 2002, also some assimilation literature). In this experiment, listeners could not rely on information from context at all. Besides, the transcription task is no online task. While this experiment allows for certain insights about the easiness of perception and allows for important conclusions concerning the kind of mistakes subjects made, the exact reasons for the problematic perception of reduced words cannot be studied in more detail. Since it is an offline task, subjects had

[54] Note that in the case of rhyming, the first segment was not considered. So, *lassen* 'to let' would be categorized equally, despite the fact that *fassen* 'to catch' also matches in the feature [LABIAL] with the first segment.

(limited) time to think about their responses. It is unclear whether the increased problems of transcribing words correctly arose since subjects had many exemplars activated in the lexicon that all could fit with the incoming speech signal or whether the signal was so different from an abstract representation that a correct transcription was not possible. Therefore, a second experiment was planned that could tap more into online speech perception processes. The results of Experiment 3 provided also a possibility to select the experimental items that were transcribed rather poorly for Experiments 4 and 5.

4.3.2 Experiment 4: Identity Repetition Priming

Experiment 3 replicated earlier results indicating that massively reduced words are hard to identify for listeners without having access to phrasal contexts (cf. Picket & Pollack, 1963; Pollack & Picket, 1963; Ernestus et al., 2002). However, single word transcriptions are not an online task, hence there was no control over the processes in the subjects' minds with the final result of a transcription. The subjects' problems to correctly transcribe the experimental items could be due to several factors that cannot be controlled by this experimental setup. To tease these factors apart, Experiment 4 was designed to tap into online processing during speech perception. For this experiment, the method of cross-modal identity repetition priming was used, as this method taps into online processing.

Cross-modal identity repetition priming with lexical decision is a paradigm that is well suited to answer the questions that have been raised by the results from the corpus analysis and Experiment 3. In such an experiment, subjects hear a word (prime) over headphones. If the same word (target) appears immediately thereafter on a screen, subjects react faster in deciding whether what they see is a word than if the same word appears on the screen after having heard a unrelated word before. If the reaction times are faster in the case of having heard the same word compared to an unrelated word, a priming effect is observable, if the reaction times are slower compared to the control item, inhibition is the result. Cross-modal priming has one important advantage over unimodal priming setups. Since the prime is presented acoustically over headphones, the timing of the target presentation can be manipulated very accurately and reaction time measurements are very reliable (e.g. Tabossi, 1996). Additionally, using this method precludes a pure acoustic pattern matching that would be possible in a unimodal auditory presentation. This technique has been used in a variety of studies to tap into processes of lexical access. As discussed above, the priming paradigm has been used, for instance, to investigate the effects of reductions on perception. However, mostly, even for research on the effects of reduction on perception, stimuli have been used that were produced just for the sake of the experiment (e.g. Lahiri & Reetz, 2002; Sumner & Samuel, 2005; Tucker, 2007; Tucker & Warner, 2007). While this procedure ensures for a better control over the kinds and amount of variation that the subjects are exposed to in the experiment, it still may miss some important factors that are given in natural speech. So far, for priming experiments, especially in a combination of different methods, no items taken from casual speech were used before.

Priming predictions

For this experiment the predictions of the two competing models discussed in Chapter 2 are not equal any longer. The very idea of exemplar models is to build variation – such as reduction – directly into the lexical representation. This enables the listener to correctly activate words that are not canonically produced. A normalization process is therefore not necessary assuming exemplar representations (cf. Johnson, 1997). Since the reductions that are tested here do naturally occur in speech, it is very plausible that the listener has encountered similar exemplars of the words before. Therefore, there should be exemplars of these words in the subjects' mental lexica that enable them to correctly activate the lexical entry. Since the priming method taps directly into lexical representations, the exemplar model predicts two possible outcomes: 1) Despite the imperfect nature of the input, listeners will be able to correctly activate the lexical entry upon hearing such a word. Therefore, a subsequent exposure to the same word in written modus will show a priming effect, as is also expected for the unreduced items. 2) If, on the other hand, the reductions are still not able to activate the correct entry alone, but due to a possible similarity to more than one lexical entry many entries will be activated, there should not be any facilitation but rather inhibition since there is too much competition between different items.

On the other hand, the FUL model predicts that reduced items, albeit being naturally occurring, will not be able to activate the correct lexical entry. Despite the plausible prior exposure to very similar words, the abstract representation will not be found during lexical access. Because there are no phonological rules that account for the reductions and deletions, too many deviations will make it impossible to match the features extracted from the incoming signal to the lexical representation, the correct entry will not be activated, and hence there will be no priming observable. Instead, the results will be the same as if the subjects heard an unrelated word before.

Materials

From Experiment 3, 30 word pairs where the reduced item was transcribed rather poorly, were chosen as experimental stimuli for Experiment 4. The 30 reduced items were transcribed correctly only in 11.8% of the time. Their confidence rating was on average 3.8. The unreduced counterparts had a correct transcription rate of 95.2% and a confidence rating of 9.1 on average. Taken together, the pairs were transcribed correctly at almost chance level (53.5% correct) and had a rating of 6.5. Additionally, a Phonetics graduate student of the University of Frankfurt transcribed the experimental items, to have independent evidence for the amount of reduction that occurred in the experimental word pairs. The transcriptions were subsequently compared to the canonical pronunciation of the Kiel corpus. This additional transcription was also a way to control for the amount and quality of information subjects in the priming experiment would be exposed to. Results clearly showed that reduced items were in fact different compared to the more canonically uttered words. For the 30 reduced items, words had at least two segmental deviations compared to a canonical transcription, at most 7 segments were changed, the mean number of changed segments was 4.1. Unreduced items had a mean change rate of 1.6, ranging from 0 to 5 changed segments. A change parameter that estimated the amount of changes per word was calculated by dividing the number of changes by the number of segments of each word. Reduced items had a change rate of 0.71, whereas unreduced items had a change rate of 0.27. If only reductions and deletions were counted, reduced items had an average rate of 2.73 (range 2 to 5) segments, whereas unreduced items had an average of less than one segment deleted or reduced (0.93, range from 0 to 3).[55] Again, a reduction and deletion rate was calculated by dividing the amount of reductions and deletions by the overall word length. Reduced items had a rate of 0.48, whereas unreduced items' rate was 0.16. Thus, reduced items had a rate that was three times as high as unreduced words. The results clearly indicate that the items that were used in the priming experiment as reduced variants of the word actually were in fact massively reduced.

30 words which were unrelated to the targets were also chosen from the corpus. Care was taken to ensure that overall, there was no statistically significant difference in word class and no difference in frequency between these primes and the reduced/unreduced word pairs. These items were unreduced utterances of words. Additionally, 10 filler items which were also unrelated to the respective target were added to the word triplets. Since subjects had to perform a lexical decision task, 40 pseudo word targets were added, preceded by another 40 words from the corpus. Three experimental lists were created using a Latin square design ensuring that every subject was exposed to each target only once in either of the three priming conditions (reduced, unreduced, control). Altogether, there were as many word as nonword decisions to be made, and within the word condition, 20 targets were unrelated to their prime, and 20 items were identical. Overall, subjects responded to 80 words. The filler items and the pseudo word items were identical across lists. The lists had the same randomization. Only primes/controls differed across the lists. The 140

[55] Assimilations were not treated as reductions in this estimation, but as a separate change category. Furthermore, insertions and segmental changes that did not belong to any of the 4 categories were included in the change count.

experimental stimuli (30 reduced, 30 unreduced, 30 control, 10 filler, 40 words preceding the nonword) were produced by 34 different speakers (14 female, 20 male). No speaker uttered more than 10 items that were used in either of the lists. The experimental set up of the crucial items is depicted in (4), a complete list of all prime/target pairs is given in Appendix G.

(4) Experimental setup for one target word

Gloss	***Prime (auditory)***	**Target (visual)**
'doctor'	Doktor REDUCED	
'broadcast'	Sendung CONTROL	DOKTOR
'doctor'	Doktor UNREDUCED	

Subjects and procedure

A total of 62 students (37 female, 25 male) from the University of Konstanz participated in this experiment for monetary compensation or course credit. None of them had participated in Experiment 3. All were native Germans and did not report on any hearing disorder. They were tested in groups of four or less.

Each list comprised of 80 items subjects had to react to. Each item was preceded by a warning tone and a 250 ms pause. Directly after the presentation of the prime, the target was displayed for 500 ms on a computer screen, which was placed at a distance of 50 cm in front of the subjects (font size: 36, upper-case). Reaction times were recorded from the point in time where the visual stimulus was presented. In the setup, a central experimental hardware box connected the DAT recorder, the response boxes and a Macintosh computer, where the reaction times were recorded (Reetz & Kleinmann, 2003). Subjects had 2500 ms to respond to each item, before a new trial began. Subjects were familiarized with the task in a short training session with items that were not part of the experiment. The experiment including instructions and training phase lasted about 10 minutes. In order to have subjects tested for a longer period of time, two different experiments that were completely unrelated to this experiment were also conducted. No interference was possible between these experiments. Experiment 3 was always the second experiment of the series. Overall, the experimental session lasted about 40 minutes.

Reaction time measurement began with the presentation of the visual targets, i.e. directly after the auditory presentation of the prime.

Results

A total of 4960 responses were given. Errors occurred when subjects gave wrong responses (n=230) or no response at all (n=6). Additionally, responses that were too fast (below 300 ms; n=4) or too slow (above 1500 ms; n=7) were treated as error. Of the total data, 247 responses (4.98%) were errors. Subjects with an error rate of 15% or more were excluded from further analysis. This led to the exclusion of 5 subjects.

For the subsequent statistical analysis, responses to fillers and pseudo words were not taken into account; only reduced/unreduced word pairs and the respective control items were analyzed. Both accuracy and reaction times were statistically evaluated. First, ACCURACY was examined. Responses were given a value of "1" if they were correct, and "0" if they were incorrect, in order to analyze the data statistically. The closer to 1 for the average of the correct responses, the more correct were the subjects' lexical decisions. An ANOVA was calculated for ACCURACY as dependent variable. Independent variables were SUBJECT (as random variable), CONDITION (unreduced, reduced, control) and TARGET. The number of erroneous responses differed between conditions ($F(2,1622)=5.8507$, $p<0.0029$). A post-hoc test revealed that targets preceded by unreduced primes were significantly different from control primes ($t=-3.365$, $p<0.0008$) and also significantly different from reduced primes ($t=-2.215$, $p<0.0269$), but reduced primes and control primes did not affect the accuracy significantly different ($t=-1.151$, $p<0.2501$).

Secondly, an ANOVA was calculated for the Reaction Time (RT) data. For this analysis, only correct responses were taken into account. Again, SUBJECT (as random variable), CONDITION (unreduced, reduced, control) and TAGET were entered into the calculation as independent factors. Reaction times differed between conditions ($F(2,1560)=108.5383$, $p<0.0001$). Post-hoc tests showed that, as for accuracy rates, the unreduced condition differed significantly from both the reduced ($t=13.482$, $p=0$) and the control condition ($t=11.819$, $p<0.0001$), but that there was no significant difference between control and reduced items ($t=-1.614$, $p=0.1068$). Figure 15 shows the amount of facilitation (control – unreduced/reduced) depending on the two conditions. The unreduced, i.e. more canonically articulated items produced a significant priming of about 74 milliseconds. The reduced items produced 10 milliseconds of inhibition; however this inhibition was not significant.

Table 12:
Reaction times for lexical decision on targets in three conditions in milliseconds and standard error

Condition	**Reaction Time LSM [ms]**	**Standard Error**
Control	602.06	10.26
Reduced	612.15	10.25
Unreduced	528.49	10.22

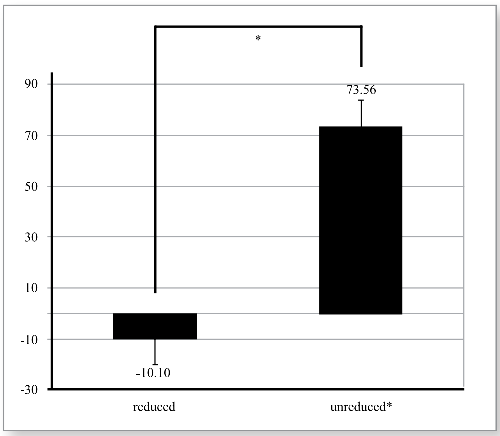


Figure 15:
Facilitation compared to control condition for reduced and unreduced primes in milliseconds. Significant differences are indicated by an asterisk (Unreduced means that this condition is significantly different from the control condition).*

Discussion

The results of this priming experiment indicate that massively reduced words were not able to activate the respective correct entry in the mental lexicon. Subsequent lexical decision times were not affected by the prior exposure to the identical word if it was reduced. The information provided by the acoustic signal or reduced items was also not able to activate several different entries that compete for recognition, because there was not significant inhibition. This inhibition would be expected if more than one lexical candidate was activated by the input and due to the incompleteness of information, there would be no winning candidate when the visual target is produced.

The results (no priming effect for reduced items but robust priming effects for canonically produced words) were as predicted by the FUL model but were not in accordance with the predictions made by exemplar models.

In the light of the results of Experiment 4 which was an online task, the findings of Experiment 3 where words had to be transcribed in isolation can also be interpreted with more confidence. In Experiment 3 reduced stimuli were transcribed less correctly than their unreduced counterparts. The results of Experiment 4 suggest that this was due to the fact that the speech input did not activate the correct lexical entry rather than it was due to the fact that the stimuli activated too many lexical entries, respectively.

One important caveat that could potentially reduce the explanatory power of Experiment 4 has to be mentioned at this point. One assumption for this experiment was that the reductions occurring in the Kiel corpus in particular are representative of in natural speech in general. The assumption therefore was that it is plausible that listeners have encountered such reductions before and have been able to create exemplar representation similar to the ones encountered in the priming experiment. However, a real and exact exemplar with more indexical information about the words could not have been built, since subjects never encountered the exact same exemplar before, arguably none of them was ever exposed to the dialogues of the Kiel Corpus (IPDS, 1994). Although the exemplar model is constructed as to deal with variation via similarity (e.g. Johnson, 1997; Pierrehumbert, 2001a), the fact that subjects never encountered the same episode before could lead to the non-activation of the imperfect (reduced) prime. To test this hypothesis, a third experiment was carried out.

4.3.3 Experiment 5: Transcription and Priming Combined

Experiment 4 showed that massively reduced words were not able to activate the correct lexical entry. While this result indicates that theories assuming abstract representations model lexical activation more correctly than episodic accounts, theories assuming exemplar representation could also explain the results. They would have to assume that the stimuli were rather different from reduced words occurring regularly in normal speech. Consequently, there would be no similar exemplars in the lexicon, since listeners were never exposed to the exact same or even similar exemplar before. To test the effect of prior exposure to items for recognition, a third experiment was created. This experiment was designed to allow for a testing of the effect of prior exposure to reduced words for speech recognition.

The challenge that the architecture of such an experiment had to meet was to ensure that listeners could build exemplar representation in a first phase of the experiment that could be activated at a later stage in a priming paradigm. Results from prior studies on the effects of reductions for recognition using transcription tasks suggested that context would enable subjects to transcribe reduced words better. Assuming exemplar storage of episodes, the line of argumentation is as follows: if a subject listens to a phrase and transcribes the word in question correctly, one can assume that the subject perceived the correct word and that this item created a trace in the lexicon. Subsequent exposure to that same item trace should lead to a successful activation of this stored item and lead to an observable priming effect.

Priming predictions

Experiment 5 combines the methods of Experiment 3 and 4. The first part of Experiment 5 consists of a transcription task. In this transcription task, however, subjects were exposed also to the phrase context of the original utterance. The second part of Experiment 5 is identical to Experiment 4. Due to results of previous studies, the expectations for the first transcription part of the experiment, taken alone, are that transcription accuracy will improve for reduced items. Subjects have now ample information from context on which they can rely their transcription. Contextual information will clearly help to narrow down the possible choices of word candidates. Both accounts would predict this result. For abstract models, correct recognition is possible due to increased information from other levels in recognition. Exemplar models would also assume that this increase of information makes it easier for subjects to choose the correct word from possibly too many that have been activated in the first transcription experiment and transcribe these items more correctly.

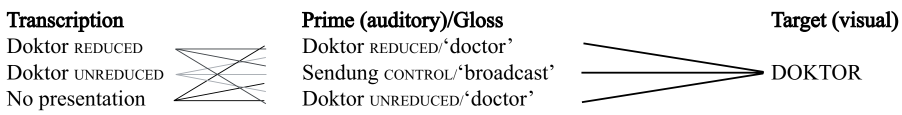
The predictions for the priming part of this experiment combined with the transcription task are different for the two approaches. While the FUL model assumes no activation if the prime is a reduced item, irrespective of whether the same item has been heard before or not, exemplar models predict that there will be an effect of prior exposure. If subjects could create a trace in the lexicon, the subsequent exposure to the same word, presented as prime, should activate this trace. This activation of the same exemplar should be visible in reaction time facilitation, since especially recently experienced episodes are assumed to be generally easier to access than older episodes (cf. Tenpenny, 1995 and references therein). If, however, subjects were not exposed to the same item before, no trace could be created, thus, no priming should arise for reduced words in those cases. Both models predict that unreduced items are able to produce a stable priming effect.

Material

In the priming part of this experiment, the same experimental items as in Experiment 4 were used. For the transcription task, the reduced and unreduced word pairs were given in their original sentential context.

Overall, nine different experimental lists were constructed. In (5) the experimental architecture of the experiment is illustrated. In every condition, there were 20 sentences presented for transcription. These were in three different relations to the items of the priming experiment: a) subjects heard the same word, either reduced or unreduced, for transcription that was later also prime in the second part of the experiment, b) subjects heard the opposite realization of the word in the word pair for the transcription experiment and the priming experiment respectively, i.e. they transcribed the reduced word and in the priming experiment, the unreduced word was presented, or *vice versa*, c) subjects were not exposed at all to an item that occurred in the priming experiment. These three options were crossed with the three priming conditions (reduced, unreduced, control). The materials and lists for the priming experiment were identical to Experiment 4. A list of the sentences that were used in the experiment is given in Appendix H. The experimental lists were recorded on CD.

(5) Experimental setup for one item for Experiment 5



The length of the sentences ranged from 5 words up to 16 words (Mean=9.8, StdDev=3.3). An ANOVA showed that concerning mean sentence length, there was no significant difference between the nine lists, with average sentence length ranging from 9.25 to 10.9 ($F(8,171)=0.64$, $p=0.742$). Another factor that differentiated the sentences that had to be transcribed was the actual position of the stimuli in question. The crucial word could occur on any position from the very first word in the sentence up to the last one. The actual position ranged from 1 to 16 (concerning percentage of words of the sentences, the crucial items were on average presented after 61% of the complete sentences, ranging between 9.1% and 100%); the overall mean position of the crucial word was 6.2 (StdDev=3.65). Again, there was no significant difference between the absolute word positions for the nine groups, ranging from an average of 5.3 to 7.1 (F(8,171=0.3685, p=0.9316). As indicated by the range of possible positions and the sentences differing in length, subjects had no clues which word was of main interest and therefore no special treatment of the word was to be expected.

Subjects and procedure

90 students from the University of Konstanz participated in Experiment 5 (61 female, 29 male). They were paid for their participation or received course credits. All were native German speakers and did not report a hearing impairment. Subjects were tested in groups of four or less. First, the transcription part of the experiment had to be performed. Their task was to transcribe the sentences they heard. For their transcription, subjects received numbered and lined sheets of paper, where enough space for the transcription was provided. After half of the transcriptions (10 sentences), the page had to be turned over. Subjects were familiarized with the task in a training session of four practice sentences, which were taken from a different corpus and whose content words did not occur in a later stage of the experiment. They did not receive any feedback about their transcription performance. After the transcription experiment, and a short pause of less than 3 minutes, the identity repetition priming part of the experiment began. The procedure was identical to Experiment 4. Overall, the experimental session lasted less than 20 minutes, including all pauses and the practice items.

The presentation sequence for the transcription task was as follows. Each sentence was preceded by a warning tone and a 200 ms pause. After the sentences, subjects had 16 seconds for transcribing what they had heard. The presentation equipment was the same as in Experiments 3 and 4. For the lexical decision part, the presentation was identical as in Eexperiment 4.

Results

The sentence transcription task of Experiment 5 was analyzed first. Due to recording problems, one group had to transcribe only 19 test sentences. The correctness of the transcription was evaluated as in Experiment 3. For this analysis, only the crucial items were examined. There was no evaluation overall correctness of sentences' transcriptions. Since the crucial items were at different positions in the sentences and subjects did not have any clue which item was crucial, this method was chosen. Only transcriptions that were identical with the intended word were treated as correct, as in Experiment 3. Note that mistakes where no transcription was provided could arise in three different ways in this experiment. Sometimes, subjects simply did not have enough time for a transcription of the complete sentence. If the crucial item was in the missing part of the sentence, no transcription was given. However, sometimes, subjects also omitted words in the mid-part of their transcription, either because they "forgot" to transcribe the word, or because they did not recognize it. One example exemplifying this evaluation dilemma is that in a number of cases, where subjects omitted several words of the sentences in the middle of the utterance, they wrote "…" in their transcriptions. Thus, the exact source for this omission was not clear: was it because they did not understand stretches of speech or because they did not have enough time for the transcription and decided to transcribe only the initial and final parts of the sentences, or was it that they knew that there were words which they did not understand? Since this differentiation is hard to make and there is no objective way to decide for one reason or another, all omissions were treated the same. Figure 16 depicts the differences in the accuracy rate of the transcriptions for reduced and unreduced words in context. Overall, 266 of 1790 words were not transcribed correctly. This is an accuracy rate of 85.1%. Reduced items were not transcribed correctly in 180 cases; unreduced words had an erroneous transcription in 86 instances. As can be seen, reduced items were still transcribed worse than unreduced items. However, there was enormous improvement in accuracy for the reduced items. They were transcribed correctly in 80% of the cases. Recall from Experiment 3 that the reduced items that were used in Experiment 4 and in this transcription task, were transcribed correctly in Experiment 3 only 11.8% of the time, compared to 80% in this experiment, when presented in context. Interestingly, the accuracy rate for unreduced items deteriorated from 95.2% in Experiment 3 to 90% in this experiment. There was also variation across the nine experimental groups. Accuracy ranged between 80% and 93% for the nine groups. Accuracy rates for individual items ranged from 23% to 100% for reduced words and fell inbetween a range from 40% to 100% for unreduced words. The individual analysis of accuracy also revealed that 11 out of 20 reduced words were always transcribed correctly, for unreduced words 16 items were always transcribed correctly.

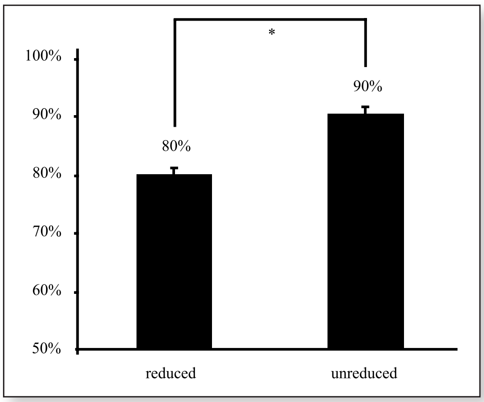


Figure 16:
Overall correctness of transcriptions for the crucial words, comparing unreduced and reduced items, with standard error. Significant differences are indicated with an asterisk.

To further investigate the differences between the categories, an ANOVA was calculated with ACCURACY rate as dependent variable. The factors SUBJECT (as random variable), WORD, GROUP (i.e. experimental list), RELATIVE ITEM POSITION (ITEM POSITION/ SENTENCE LENGTH) and CONDITION (reduced, unreduced) were used as independent variables and were also entered into the ANOVA. Again the REML estimation was used. The results showed that both CONDITION ($F(1,1910)= 500.0524$, $p<.0001$) and WORD ($F(91,1910)= 8.9132$, $p<.0001$) were significant main effects. To summarize, reduced items were transcribed significantly worse than their unreduced counterparts ($t=-10.3$, $p<0.0001$).

As for Experiment 3, a more detailed analysis of mistakes was performed. The errors were labeled into the same 5 categories as in the first transcription experiment. Table 13 exemplifies the different mistakes.

Table 13: Different kinds of mistakes, split up by category, percentages in parentheses

	unrelated	1st segment correct	wordform incorrect	rhyming response	1st feature correct	no response	overall
unreduced	5 (0.6)	7 (0.8)	3 (0.3)	0 (0)	0 (0)	71 (8)	86 (9.7)
reduced	14 (1.6)	12 (1.3)	41 (4.6)	4 (0.4)	17 (1.9)	92 (10.2)	180 (20)
Sum	19 (1.1)	19 (1.1)	44 (2.5)	4 (0.2)	17 (0.9)	163 (9.1)	266 (14.9)

Out of the 266 incorrect transcriptions, 163 words (9.1% of the overall items, 61.3% of the mistakes) were not transcribed at all by subjects. Reduced words were omitted 92 times, constituting 51% of the mistakes for reduced words, whereas unreduced words were not transcribed in 71 instances, which equals 83% of the entire amount of transcription errors for unreduced words. In this experiment, subjects had to transcribe more than one word under considerable time pressure listening to a stretch of a conversation only once. Remember, that giving no response to an item does not mean that the sentence was not transcribed, but that the crucial word was omitted. This is different from Experiment 3 where subjects listened to only one word. Therefore, it is also indicative to analyze the transcription results and exclude responses where the crucial word was missing. When the items where no transcription was given were excluded from analysis, overall transcription accuracy was 93.7%. Reduced items were transcribed correctly 89.1% of the time, unreduced items in 98.2%.

The second largest error subjects made were word form errors, and these mistakes occurred particularly often for reduced words. A closer inspection of this error revealed that word form mistakes were also regularly accompanied by a person/number pronoun error in transcription for verbs. For one item, the reduced word *Sinn* ('sense'), however, there existed the possibility of producing the word in question grammatically correct both with and without the case marker <-e>. Transcribers opted more often (in 21 from 30 cases) for the transcription of *Sinne* 'sense-DAT' with <e> (which would be produced as [ə]). This variant would be a more formal way of expressing the phrase. The 9 remaining cases had the word transcribed closer to the actual, less formal, pronunciation without <-e>. However, the 21 transcriptions with the case marker were still treated as false transcription.

Only reduced words were transcribed with errors where the first segment had features of the first segment correct or where there was a rhyming word given. In 19 cases subjects provided a transcription where the first segment was correct, but not the rest of the word. Out of a total of 1790 transcriptions, only 19 or 7.1% (14 reduced, 5 unreduced) of the transcriptions were completely unrelated to the intended word. This is a clear improvement compared to Experiment 3, where 29% of the mistakes were transcriptions that were completely unrelated.

Next, the results of the identity repetition priming part of Experiment 5 were analyzed without inclusion of the transcription task. Again, the error analysis is reported first. The definition of errors is as in Experiment 4. When a wrong response was given (n=347), or no response at all (n=12), they were counted as error. Equally, responses faster than 300 ms (n=8) and responses slower than 1500 ms (n=26) were not treated as correct. Of the complete data set, 393 responses (5.5%) were errors. Subjects with an overall error rate of 15% or more were excluded from further analysis. Based on these criteria, 4 subjects were excluded from further analysis.

Again, for all ANOVAs in the upcoming section, responses to fillers and pseudo words were not included in the analysis. The ACCURACY of responses was evaluated identically as in the analysis of Experiment 4: responses were given a value of "1" when they were correct and a "0" when they were incorrect. The closer the mean was to "1" the more accurate were the responses. For the first ANOVA ACCURACY was dependent variable. Independent variables were SUBJECT (as random factor), CONDITION (unreduced, reduced, control), and TARGET. Accuracy rate was significantly different across conditions, ($F(2,25679=9.2553)$, $p<0.0001$). Post-hoc tests showed that all three conditions were significantly different from each other in accuracy rate.

The next step was the calculation of ANOVA for the RT data. As for experiment 4, SUBJECT (as random factor), CONDITION (unreduced, reduced, control), and TARGET were included as factors. Reaction times differed significantly across conditions ($F(2,2443)=125.5289$; $p<0.0001$). Post-hoc tests were performed and revealed that responses to targets that were primed by reduced items did not differ from the control condition ($t=1.045$, $p=0.2961$), but the unreduced condition was significantly different from both the reduced ($t=13.181$, $p=0$), and the control condition ($t=14.161$, $p=0$). Table 14 shows reaction times for the different conditions, whereas Figure 17 depicts facilitation (CONTROL – REDUCED/UNREDUCED) for the two categories, i.e. REDUCED/UNREDUCED. In the unreduced priming condition, a robust effect of 73 ms facilitation could be observed, whereas the reduced items resulted in a mere 5 ms facilitation that was not significant.

Table 14:
Reaction times for lexical decision on targets in three conditions in milliseconds and standard error.

Condition	Reaction Time LSM [ms]	Standard Error
Control	577.64	9.64
Reduced	572.33	9.62
Unreduced	505.73	9.62

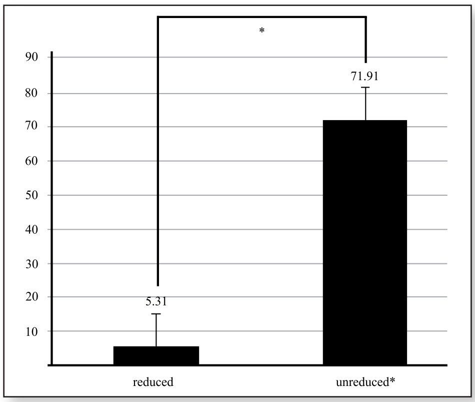


Figure 17:
Facilitation in ms compared to control condition for reduced and unreduced primes in milliseconds, significant facilitation is indicated by asterisk.

In the next step, the results from both experimental parts were combined. Another statistical analysis was carried out to examine whether the transcription experiment, and the chance to create exemplar representation, had an effect on the results in the lexical decision task. Again, responses to fillers and pseudo words were not included in the analysis. Also all cases where subjects did not transcribe the crucial word correctly, were excluded from further analysis. This was done to ensure that either subjects had heard the word correctly or the word had not been heard before. The ANOVA for ACCURACY rate as dependent variable had SUBJECT (as random factor), CONDITION (unreduced, reduced, control), and TARGET, TRANSCRIPTION CONDITION (reduced, unreduced, not presented), and CONDITION x TRANSCRIPTION CONDITION as factors. The results showed that Target ($F(29,2201)=1.9579; p<0.0017$) and Condition ($F(2,2201)=9.0516, p<0.0001$) were the only main effects; TRANSCRIPTION CONDITION ($p=0.21$) was not significant and the interaction was also not significant ($p=0.37$). Post-hoc tests revealed that all conditions were significantly different from each other.

The ANOVA for RT as dependent variable had the same factors as the accuracy analysis: SUBJECT (as random factor), CONDITION (unreduced, reduced, control), and TARGET, TRANSCRIPTION CONDITION (reduced, unreduced, not heard), and CONDITION x TRANSCRIPTION as independent factors. Note that only the items where subjects provided a correct transcription and correct responses were analyzed. The results showed that CONDITION ($F(2,2095)=108.8875; p<0.0001$) and TARGET ($F(29,2095)=7.1374; p<0.0001$) were significant main factors, but TRANSCRIPTION CONDITION ($p=0.1061$) was not. There was also no significant interaction ($p=0.5265$). Post-hoc tests revealed again, that reduced and control items were different from unreduced words, but not from each other.

Discussion

Discussion of the transcription task

The transcription of words in context showed that context is important for disambiguating reduced words for a successful recognition. There was an enormous improvement in transcription accuracy for the items compared to Experiment 3 where subjects had to transcribe words without their sentential context. The importance of context is also revealed by the fact that the number of mistakes that were completely off was reduced drastically compared to Experiment 3 where subjects could not use any context for their transcriptions. A seemingly unexpected result was that overall, unreduced words were transcribed less correct than in Experiment 3. However, a closer inspection revealed that this was not caused by the transcription of "wrong" words, but rather due to the fact that no transcription of the words was provided at all (71 instances out of 86 errors). There are three possible sources for missing transcriptions: either the word was not heard, or it was forgotten, or subjects, due to time restrictions, did not have enough time to transcribe the word. More thorough examination of these mistakes also revealed that out of the 71 missing transcriptions, 53 occurred with a word at position 9 or later, a position where time constraints are increasingly important. This was clearly different for reduced items, where out of 180 errors, 92 were omissions, and 28 of them occurred at position 9 or later. If the omission mistakes were not counted, results for both reduced and unreduced words would be much better as in Experiment 3.

Anecdotic evidence also underlines the importance of context. One subject explicitly reported after the experimental session that some words were only possible to be transcribed with the knowledge of the rest of the sentence. While the subject was unaware of the purpose of the experiment and where the critical stimuli were located, this is additional evidence that subjects make use of the information that the context provides.

The results for the transcription part of the experiment are clearly a replication of earlier findings, in that context is important when words are reduced (e.g. Picket & Pollack, 1963; Pollack & Picket, 1963; Ernestus et al., 2002). The results of the transcription task also allow for the planned examination of the effect of prior exposure to exemplars of words.

Both accounts have correctly predicted the results of the transcription part, as was the case for Experiment 3. Note that this task is more similar to everyday natural conversations than Experiment 3. When listeners are faced with natural speech, they usually understand very well what has been said. Actually, without correct word recognition, no sensible conversation is possible. If some of the words are reduced, there is usually ample context that helps to allow for correct perception nonetheless. In everyday conversations, recognition "mistakes" as hearing *Sinn-e* 'sense-DAT' when *Sinn-Ø* 'sense-UNMARKED' was said, does not impede mutual understanding, it is a mere difference in speaking style.

Again, the two models differ in their explanations for this excellent performance. In exemplar theoretic frameworks, words, whether reduced or unreduced, are mapped to prior encountered traces in memory. Since in these memory traces, there are many different variable exemplars, recognition is successful. Besides this main way of recognition, exemplar models also have a role for other sources of information that allow for recognition. Syntactic, semantic, morphological and pragmatic information is also assumed as important factors for recognition. However, for abstract models, such as FUL, this extra information is more crucial for word recognition. The reduced word by itself is not able to activate the correct lexical entry. When more context is available, the information that did not allow for correct lexical activation can be used in combination with the information of the context and finally lead to the activation of the intended words, since the search for possible candidates can be narrowed down by additional information. The results of the transcription task thus are in accordance with both frameworks, and the transcription experiment is not able to tear apart the more correct explanations as for why subjects improved in the transcription accuracy. Given the results of the first two experiments, however, the explanation of the abstract model seems more plausible. From an exemplar-based model, this transcription part of the experiment might be crucial when the second part, the lexical decision task with identity priming is analyzed. When subjects transcribed a word correctly, one can assume that they created an exemplar trace in their memories, and that subsequent exposure to that or similar exemplars would lead to genuine lexical activation and a speeding up in lexical access.

4.3.4 Discussion of the Priming Experiment, and Both Tasks Combined

The pure results of the identity priming experiment task without an inclusion of the transcription task replicated the results of Experiment 4. Massively reduced words were not able to activate the correct lexical entry, whereas there was a robust activation effect for words that were produced rather canonically. The results are very similar to the ones obtained in Experiment 4. However, there was an overall decrease in reaction time compared to Experiment 5. Subjects reacted faster in all three conditions. The reason for this decrease in reaction time is not clear, since this effect cannot be attributed to primary exposure to the same exemplar, because also control items and nonwords and fillers that were not presented before were faster in the lexical decision task of Experiment 5. An one-way ANOVA with RT as dependent variable, and CONDITION, EXPERIMENT AND CONDITION x EXPERIMENT as factors revealed that both CONDITION and EXPERIMENT were significant factors, but there was no interaction. Every single condition was significantly faster in Experiment 5 than in Experiment 4.

When the correct transcriptions of the first part were included, the results for the priming experiment did not differ. The results indicate that prior exposition to the same exemplar for reduced words did not lead to any improvement of activation. This was also true for the unreduced words, where there was no difference in activation independent of whether the transcription experiment had the reduced or unreduced word presented in context. Thus, despite the possibility for subjects to create lexical entries (episodes) after transcribing words correctly, no effect of prior exposure was found in the priming paradigm. However, as for Experiment 4, the results are in line with the FUL model, assuming single, abstract, featural representations. A prior encountered reduced or unreduced variant of a word is not assumed to have an effect on the underlying abstract representation. Therefore, even a subsequent exposure to the same variant is not different compared to an exposure to the same word but different exemplar or no prior exposure at all.

The results suggest that, while the overall assumption that listeners understand correctly what other people said is valid, the conclusion that this is the case for every single word at every point in time is not warranted. Transcription experiments and repetition priming experiments both provided evidence that for the almost perfect performance of listeners in natural conversations, context is a crucial factor. Deprived of this information, listeners are not able to perceive massively reduced words correctly. Abstract models such as FUL have a straightforward explanation for these results, namely, that the featural information provided by the reduced acoustic signal is insufficient for activating the correct lexical entry. On the other hand, however, exemplar models have some difficulty to explain the results reported here. This is not to say that there is absolutely no possibility for an exemplar based explanation. However, additional assumptions and stipulations have to be made, and applying Ockham's razor then clearly favors the abstractionist view.

For exemplar models, there are several ways to account for the data with different further assumptions delineating from a rather basic model as X-MOD:

1) Listeners do not store "faithful" exemplars, but echoes of actually experienced exemplars (e.g. Goldinger, 1998 testing the MIVERVA2 model proposed by Hintzman, 1986, 1988 and references therein). An echo is an activation aggregate of all traces that occurs when a word probe is matched to exemplars stored in the lexicon. Echoes are stored in memory, and may contain information that has not been part of the exemplar itself (for a summary, see Goldinger, 1998). Therefore, even when the same word is heard twice, there will never be an exact match. This architecture might explain why even after having the possibility to create an exemplar representation, no activation was found. The subsequent exposure to the reduced items was just ineffective for these echoes since too little traces have

been similar to the repetition of the word, hence, no activation occurred. However, this explanation is not very plausible. Since the first exposure created a trace or an echo in the lexicon, even if it is an imperfect echo with much noise added to it, a second exposure to the same word should at least be able to activate this echo to some extent, irrespective of noise and the impossibility of an exact match. Therefore, some effect should be observable, but this was not the case. Additionally, exemplar models have been assumed for the very reason to allow for recognition despite variation and reductions. If the assumption of exemplar-based lexica is correct, listeners should be able to handle this kind of naturally occurring variation without problems.

2) Another way of explaining the data from an exemplar-based approach is the assumption that listeners do not have word-based exemplar storage, but larger chunks of representation, such that traces are sensitive to (sentential) context (e.g. Bybee, 2002; Wade, 2007; Wade & Möbius, 2008). Therefore, words presented without context cannot create a robust activation, even if they are identical to parts of chunks that are part of episodic representations. The exact size and number of episodes that are stored is actually one of the assumptions, where different episodic models diverge considerably and where it is not clear how many levels of representation are needed (e.g. Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001a, 2006a; Bybee, 2002, Wade & Möbius, 2008). However, one of the most important strengths of exemplar-based representations is to be able to deal with variation in general and with naturally occurring reductions in particular. The bigger the sizes of the chunks that are stored holistically as exemplars, the smaller the chance for a match between an incoming signal and the episodic representation. Therefore, this kind of purely large scale representation deprives the model of parts of its most pronounced strength. Proponents of storage of larger chunks usually assume multiple representational possibilities. The exact size of these chunks is not easy to be determined, as frequency of collocations plays a seemingly decisive role for storage of larger chunks, as well as phonological and morpho(phono)logical considerations (see, for example, Bybee, 2002 and references therein). Additionally, for the exemplars that could have been built in this experiment, there is no obvious pressure for subjects to store complete sentences. These sentences are not comprised of high frequency collocations, and arguably for most of the sentences, subjects do have heard the same phrases before nor is it probable that they will encounter them again. Thus, storage of the complete sentences will not help in subsequent speech perception processes.

3) Another possibility that renders the results of the experiments reported somewhat disputable is that despite the correct transcriptions, subjects did not create an exemplar in the mental lexicon at all and the transcription part of the experiment did not reach the lexicon. This argumentation is not directly linked to assumptions of exemplar models.

The creation of exemplars is supposedly an automatic process that takes place whenever listeners are exposed to speech and listen to it (arguably even when they overhear speech unintentionally). When listeners try to understand what has been said, these exemplars are created and remembered. When a word is transcribed, listeners have perceived it correctly, otherwise the transcription would not be possible. Thus, listeners can also be expected to create a lexical trace in memory. Therefore, even if an exemplar is not a perfectly produced instance of a word, such an exemplar should still be recognized subsequently.

4) The time course of activation may be different for exemplars depending on the time that is needed for processing (McLennan & Luce, 2005; McLennan, 2006). If processing (lexical decision) is easy, then underlying, abstract representation influences word perception (McLennan & Luce, 2005: 317). Many exemplar models are not pure exemplar models, but mixed models in the sense that they assume both exemplar representation and abstraction (e.g. Goldinger, 1998, 2007; McLennan & Luce, 2005). The time course hypothesis posits that indexical exemplar information becomes available only later in the speech recognition process. McLennan and Luce found specificity effects when the nonwords were harder to be recognized (e.g. *bacov*) than when this decision was easier to make (e.g. *thushshug*) (McLennan & Luce, 2005). And although in the experiments reported in this dissertation, there occurred activation for perfectly produced instances that were able to match the abstract representation, the effect of exemplar activation was not possible at that point in time, but would have been observed later. Connected to this possible caveat is the question whether the experiments really tapped into the underlying lexical representation (see, also McLennan et al., 2003 for a discussion of this topic and 3), above). However, the cross-modal repetition priming paradigm has been shown in many instances to be able to reliably tap into underlying representations (cf. Tabossi, 1996; Scharinger, 2006). Although indirect semantic priming would be even more apt to do so, this caveat does not seem to hold. There were no prior results using one of the paradigms, therefore the repetition priming paradigm was a promising first step. Additionally, the nonwords used in the experiments (see Appendix F for a complete list of the nonwords from the Experiments 4 and 5) were rather hard to be discriminated from real words. Therefore, we would expect also results comparable to those reported by McLennan & Luce, 2005; namely, that specificity effects should occur. However, this is not what the results show.

In sum, the assumptions of a model like FUL assuming single abstract representations are more straightforward and do not need any stipulation to deal with the results. In combination with the transcription experiments, the results indicate that context is crucial when single words do not match sufficiently to the abstract representation. The featural information that is gathered during the first stage of recognition can be used at a later stage when syntactic plausibility or pragmatic information is evaluated in speech perception and helps the listener to further narrow down the choice among candidates for recognition. Without context, however, massively reduced words do not contain enough matching information to activate correct lexical entries. Take, for example the word *natürlich* 'natural(ly)', the Duden-like pronunciation would be /naˈtyːɐlɪç/. The graduate student's transcription of the unreduced variant was /naˈtyɐlɪç/, differing only in the length of the /y/ vowel, being not contrastive in German. Compare this to the transcription of the reduced variant, /dəˈdʊç/. For the listener, recognition is quasi impossible despite some non-mismatching features on the first positions, at the latest when the word final fricative is encountered, since either a vocalized <r> or a <l> is expected for the intended lexical *natürlich* item to be activated. Thus, compared to this lexical item, in the FUL model matching mechanism, a mismatch occurs, and the lexical entry for *natürlich* is no longer activated.[56] Therefore, for the subsequent visual presentation, no priming is expected for this item. This is exactly what we find in the experiments reported here.

[56] This is not to say that no candidate is activated at all. Maybe, in this example, the word *dadurch* 'through there' is activated. It would be canonically produced as /ˈdadʊɐç/. Indeed, in Experiment 3 the transcription was *dadurch* by three subjects, whereas *natürlich* was transcribed only in two instances.

«… the skillful workman is very careful indeed as to what he takes into his brain-attic.»
***Sherlock Holmes** (Arthur Conan Doyle, , A Study in Scarlet)*

Chapter 5 – **Summary and Conclusions**

"Speech is variable". What seems like a tautological statement about spoken language has important repercussions for linguistic theories but has also been neglected for a long time for the modeling in phonological theory. This statement, or to be more precise, the variability of natural speech, can be incorporated very differently into phonological theories, the two extreme ways to handle variation are a) to see it as informative and store variation directly in the mental representation, or, b) not to store it and to have routines that lead to an abstraction away from variation. Following this distinction, at the outset of this dissertation the questions that were raised concerned the adequacy of two different phonological models, one representative for each way to handle variation, (i.e. X-MOD, Johnson, 1997; and FUL, Lahiri & Reetz, 2002) and their ability to predict and explain reduction processes in naturally spoken German with their effects for speech perception. Both models have their main focus on speech perception, but also make (partly implicit) assumptions for speech production.

What became very apparent in the analysis of spoken German is that Germans, as speakers of other languages as well, seem to be rather sloppy in their speech: they reduce segments regularly. There appear to be not exclusively phonological rules in a traditional generative sense accounting for the reductions. Rather, Germans reduce segments more randomly. From a point of view of speech perception, the scenario is not as bad as depicted vividly by Hocket (1955) in Chapter 1 of this dissertation, though. The random reductions do not affect all the segments alike (cf. Chapter 3 & 4). For instance, vowels are very stable in speech production (cf. Chapter 4). However, for an evaluation of the successes and failures of FUL and X-MOD, a more detailed summary is needed. In the following paragraphs, the findings of this dissertation are summarized step by step, before general conclusions are depicted.

The first area of investigation in this dissertation was regressive assimilation of place of articulation (PoA) across word boundaries in conversational German (Chapter 3). There does not exist agreement among phonologists with respect to the status of this process. First of all, linguists debate whether this reduction process does occur in German at all (cf. Kohler, 1995a; Wiese, 1996), and secondly, they disagree whether regressive assimilation results in a complete neutralization of featural contrasts. These questions have repercussions for the correctness of the assumptions of linguistic models in general, but as well for the two models that had been introduced in Chapter 2 (i.e. X-MOD and FUL) in particular. The Kiel corpus of spontaneous speech (IPDS, 1994) served as database for the first test of the successes or failures of the two models.

Concerning the question of the amount of occurrence of regressive place assimilation, the corpus study revealed that it actually occurred across words in German in approximately 6% of possible sequences of consonants. This is comparable but slightly less often than in American English, as reported by Dilley & Pitt (2007). Another finding of the analysis of the corpus was that function words were more likely to assimilate than lexical words. Furthermore, there was an asymmetry concerning the direction of assimilation. As the analysis revealed, [CORONAL] sounds ([t, s, ʃ, ç, n]) assimilated more often than [LABIAL] segments ([p, f, m]), or [DORSAL] consonants ([k, x, ŋ]). Moreover, the manner of articulation of consonants also mattered for regressive assimilation; nasal consonants were far more likely to assimilate than obstruents. These results (Section 3.2) reflected what trained phoneticians transcribed who had additional information such as the speech signal, spectrograms and the context. When it comes to an evaluation of the adequacy of the two models, the production data alone did not provide decisive evidence for either of the two models. Both models are able to explain the corpus data. For FUL, regressive assimilation due to phonological rule application is expected. The model also predicts that only [CORONAL] segments undergo phonological assimilation of PoA. These predictions have been corroborated by the corpus analysis. X-MOD, on the other hand, can also explain the asymmetric behavior. If a salience-effect is posited, such that when two competing features are produced, [CORONAL], as the less salient feature will be repressed, the results are explained straightforwardly. Based on the corpus study alone, no further conclusions should be made. Only when the perception side of speech is taken into account, more reliable conclusions are possible.

Subsequently, two perception experiments were conducted (forced choice and free transcription) to test how fast and accurately naive listeners' responses would support the analysis of the production data based transcriptions of trained phoneticians. In these experiments, the production asymmetries were exploited by using experimental stimuli from different speakers of the corpus. In using stimuli that had been labeled as either ASSIMILATED (/n/ > [m]) or UNASSIMILATED (UNASSIMILATED-/n/, UNASSIMILATED-/m/).

First, the results indicated that the transcription of the Kiel corpus is very accurate concerning regressive place assimilations and that the transcribers were rather conservative in deciding whether a sound was completely assimilated or not. Subjects' responses and the transcription correlated very highly. When they had to decide on the ASSIMILATED stimuli, they reacted as if they heard UNASSIMILATED-/m/'s. This is an indication as to the completeness of the assimilation. A difference in responses did only occur for UNASSIMILATED-/n/ stimuli in the context of [LABIAL] or [DORSAL] segments, reflected in the high amount of response variation in the /n/-category. The RT of Experiment 1 lend further support to the accuracy data. The differences in the RT of congruent and incongruent responses of UNASSIMILATED-/m/ and UNASSIMILATED-/n/ responses are especially informative. First, the incongruent [n] responses to UNASSIMILATED-/m/ stimuli were significantly slower than the corresponding incongruent [m] responses to UNASSIMILATED-/n/ stimuli. Second, there is a stronger context effect in the reaction times for the UNASSIMILATED-/n/ stimuli than for the UNASSIMILATED-/m/ stimuli split up into different context. Whereas the reaction times for the congruent [n] responses to the UNASSIMILATED-/n/ stimuli differed by context, no such difference was found for unassimilated-/m/ stimuli. This is an additional hint that the transcribers of the Kiel corpus were 'conservative' and labeled the UNASSIMILATED-/n/-LABIAL as [n] rather than [m].

Furthermore, an acoustic analysis was performed with the experimental stimuli of Experiment 1 and 2. As in Dilley & Pitt (2007), the F2 measures of the middle and end of the vowel were taken; additionally, F2 at the nasal mid point was analyzed. Corresponding to the perception results, it was found that the change in the F2 from the middle to the end of the vowel did not significantly differ between the UNASSIMILATED-/m/ and ASSIMILATED consonants. Similarly, the nasal formant measure did not differ between these categories indicating that the ASSIMILATED tokens shared these acoustic categories with the canonical, UNASSIMILATED-/m/. Both the perception results and acoustic analysis of the stimuli suggest that complete assimilations do occur in running speech (/n/ > [m] in a [LABIAL] context). This conclusion is supported by the responses to some assimilated tokens which were judged by subjects in the experiments to be [m] 100% of the time. Clearly, however, assimilations are gradient as can be seen in the responses to the UNASSIMILATED-/n/-LABIAL stimuli. Although transcribers labeled them as [n], they were often perceived as [m]. Gradualness of assimilation is most distinct for the [CORONAL]-category where we see the greatest amount of (response) variation.

In combination with the results of the production study, the findings on regressive place assimilation allow for a first evaluation of the two models. The asymmetry between [CORONAL] versus [DORSAL] and [LABIAL] both in production analysis ([CORONAL] consonants assimilate more than the others) and in perception ([CORONAL] segments vary most in perception) has been frequently noted in the literature (cf. Lahiri & Evers, 1991; Paradis & Prunet, 1991; Ghini, 2001a; Gumnior et al., 2005), but which of the models fares better?

Assimilation was one of the very reasons to promote underspecification theory. Not surprisingly, FUL is able to handle the observations from perception as well as production. The "unmarkedness" and asymmetry of [CORONAL] segments have been built into FUL directly via underspecification (Lahiri & Reetz, 2002). Crucially, the results indicate that there exist cases, where the PoA contrast is completely neutralized. This is, what the model expects in when the unspecified feature is not produced as [CORONAL] via a default production rule. FUL predicts in such cases when assimilation takes place before the default production rule applies, a complete neutralization of the PoA contrast, as a consequence of spreading of the PoA feature of neighboring segments. Based on underspecification, not only complete assimilations are expected but also an asymmetrical direction of assimilation, in that only [CORONAL] segments are expected to assimilate and they are not expected to trigger assimilation, because only specified segments can spread their PoA features. This is consistently confirmed by the production data of the Kiel corpus. Assimilations almost exclusively occur with [CORONAL] segments that assimilate to either to [LABIAL] or [DORSAL] PoA, but almost never vice versa.

More generally, there are two kinds of assimilation processes for FUL. Besides the complete neutralization process, there also exists the possibility of phonetic variation. The latter is due to coarticulation or the overlapping of different gestures. This possibility is not modeled explicitly in FUL, as is true for phonetic variation in general, but it exists nonetheless. From the point of view of theory evaluation, positing two different processes (i.e. phonological and phonetic) might seem not as elegant as positing only one, which is what X-MOD does.

Concerning the success of X-MOD, the model does not differentiate between phonetic and phonological variation. Every kind of variation that is encountered by listeners is represented in the mental lexicon. X-MOD predicts that there is gradual variation, and does not make specific claims as to the completeness of assimilations. Salience of perceptual cues can be taken to account for the asymmetric assimilation behavior of [CORONAL] segments, since the PoA feature cue of [CORONAL] can be assumed to be less salient than for example [LABIAL]. Thus, on first sight, X-MOD is more attractive in explaining the assimilation data. However, when the results of the perception are taken into account, X-MOD gets less attractive.

Overall, FUL is more successful and correct in predicting the findings of Chapter 3 than is the episodic model. Interestingly, FUL can also explain the findings of the RT data. FUL predicts that [CORONAL] segments should not be faster than [labial]s, despite the fact that listeners should know that when [CORONAL] is heard, the segment has to be [CORONAL] and not [LABIAL] (or [DORSAL] for that matter). Since [CORONAL] sounds do not have a PoA feature in the lexical representation that can be matched, a non-mismatch condition arises. On the other hand, [LABIAL] features can be matched onto a [LABIAL] feature in a lexicon. Although non-mismatch conditions need not to be always slower than matching conditions, they are expected not to be faster, as is exactly the result that was obtained for Experiment 1.

In X-MOD, encountered speech is labeled and stored in memory. Because [CORONAL] segments are encountered often, there can be expected a cloud of many exemplars in the lexical representation. This cloud should also be expanded in the acoustical space because of the variation in PoA acoustic cues. Crucially, due to assimilation, there will be also instances of [CORONAL] segments that are stored within the area where [LABIAL] segments are represented. Furthermore, [LABIAL] segments occur less often in natural speech than [CORONAL] segments (see Tables 3, or 4, and section 4.2). Thus, their resting activation level should be lower. When listeners encounter to a speech item and have to decide whether what they heard was a [LABIAL] or a [CORONAL] segment, they should be fastest, when they heard a [CORONAL] segment, because due to the asymmetry in assimilation direction, there is no doubt that [CORONAL] in the speech signal can only occur for [CORONAL] segments. For [LABIAL] items, on the other hand, there is the possibility that what has been encountered in speech is either a true, underlying [LABIAL], or it could have been also an underlying [CORONAL] that had been assimilated. Consequently, listeners should react fastest and most accurate to [CORONAL] segments, even for those that have been partially assimilated, because of the asymmetry in assimilation direction. However, this is not how the listeners in Experiment 1 and 2 behaved. They did neither react faster to [CORONAL] items, nor did they identify them more accurately. Thus, while X-MOD's explanation of the production findings seems more elegant and needs fewer assumptions, the model fails on the perception side.

Deletions occurring in conversational German were the second kind of reduction processes investigated in this dissertation (Chapter 4). They can lead to "massive reductions" (cf. Johnson, 2004a) and have more severe repercussions for speech perception than assimilations. While assimilation processes change the value of a single feature, deletions result in the complete loss of the segment and consequently alter the pronunciations of words more drastically. This is even more so in cases where speakers delete more than one segment of a word. In a first step, the amount of deletion in conversational German was examined. While about 60% of the words were not produced canonically, the amount of deletions that occurred was less dramatic. Adhering to the transcription in the Kiel corpus, 16.1% of the underlying segments were deleted. However, this number included cases of segments where the underlying phonemic status is at least questionable, such as [ʔ]. After excluding these from further analysis, speakers deleted 8.6% of all canonical segments. Several statistical analyses indicated that many factors contributed to the deletion patterns that were observable in naturally spoken German. These factors were the segment itself, the preceding and the following context, whether the segment was a consonant or a vowel, whether the segment was part of a function word or a lexical word. Finally, a factor was also the gender of the speaker. When all of these were included in a single analysis, the difference between male and female speakers as well as vowels and consonants did no longer reach significance, though. At the same time, this statistical model accounted only for about 20% of the variation. The results were similar when vowels and consonants were analyzed separately and when additional consonant- or vowel-dependent variables were added in these separate models.

A first conclusion of the deletion analysis was that there exists an enormous amount of variation in the deletion patterns. Many different, not always independent factors have an influence on deletion of segments. Deletions occur regularly, but they are not rule based in natural speech; yet, the amount of deletion is not extraordinarily high. This corpus study was not able to provide unequivocal and consistent results for an evaluation of the two models, though. Too many factors have been shown to have a possible influence on segment deletion. Despite the size of the corpus, no sensible way of controlling for all of these was possible. Furthermore, because of the lack of control over many factors, the data was not distributed equally. Thus, the statistical models alone are not very indicative for an evaluation of the two models.

Therefore, in a next step, deletion of a single segment was investigated. The segment of choice was /t/. On the one hand, /t/ got deleted quite often in the corpus (21.4% of the time) and on the other hand, there is no doubt as to the phonemic status of /t/ in German. Since the objective was to investigate a single segment in a controlled condition, a /t/ with a constant preceding consonant (/s/) was chosen. The /t/ in question was part of the German suffix for 2nd person singular in the present tense (i.e. {*-st*}). Since these forms did not occur in the Kiel corpus, a new corpus was constructed where subjects produced verbal paradigms with this suffix in three different contexts (preceding /v/, /s/, or /e:/). The results of the corpus study showed that context is crucial in determining the amount of /t/ deletion. However, also other factors (such as hesitational pauses or gender differences) could be identified. The deletion rate of 20% was very similar to the overall /t/ deletion rate in the Kiel corpus (21.4%), an indication as to the adequacy of this newly constructed corpus.

The main conclusion that can be drawn of the two corpus studies is that speakers of German show a huge amount of variation in their deletion behavior. Many factors have an influence on the probability of a segment to get deleted. Furthermore, only when several methods are applied, a meaningful result can be reached. Both an analysis of the overall deletion rate in a corpus, as well as a more controlled examination are crucial for obtaining interpretable results. In an overall examination of deletion in a corpus, there is no possibility to control which words are uttered in which contexts. At the same time, the number of data points gets incredibly large, and yet, phonotactic constraints or general frequency distributions lead to skewed distributions. Therefore, the concentration on a single segment in a newly created corpus was crucial. In this corpus subjects produced words in a very controlled and yet natural way, allowing for a better estimation of the effects of factors such as following context or pauses. Subsequently, the results were more stable and could be interpreted better.

Concerning the adequacy of the models, if the production data is considered alone, again, X-MOD, seems to explain the data more elegantly than FUL. In any case, however, both models are able to explain the results. For X-MOD, deletions do not pose a huge challenge. Variation is one of the most basic assumptions of the model. Any variation that occurs in spoken language is also reproduced in the lexical representation via episodic storage. Subsequently, it is also possible for speakers to produce episodes of reduced variants that have been encountered before. Additionally, if one assumes an intrinsic desire of speakers to minimize effort in speech production, even more variation and deletion is expected. The results of the production studies corroborated these expectations. On the other hand, for FUL, there are no explicit claims concerning the kind of reduction process that was investigated. Since there appear to be no phonological rules that explain the deletion patterns, these patterns are assumed to be part of phonetic instantiation of speech rather than processes that are part of a speaker's grammar. For FUL, such reduction processes are possible, but the prediction is that on the perception side of the conversation, there will be problems for successful speech recognition.

Again, X-MOD needs fewer assumptions than FUL to explain the observed data. However, this seeming advantage of X-MOD should be seen a little bit critical as well. Exemplar-based models such as X-MOD have been proposed exactly for the reason to deal with variation in natural speech and have been based on the assumption that this variation exists (e.g. Johnson, 1997, 2004a; Goldinger, 1998; Pierrehumbert, 2001a). Therefore, the fact alone that there exists variation in natural speech is not evidence for the correctness of the model. Rather if there were no variation, the model's basic assumptions would have to be refuted. Thus, as for the assimilation part, the production data alone does not add very decisive results. A more telling test is provided from examining the effect of deletions on speech perception, because the two models make rather different assumptions how listeners deal with massive reductions.

In a series of three experiments, the effects of massive reductions on speech perception were examined. In Experiment 3, subjects transcribed reduced and unreduced words out of their sentence context. The transcription accuracy for reduced words was significantly worse than for unreduced words. However, the reason for this rather poor transcription accuracy could not be analyzed, because transcription does not allow for an investigation of online processing of lexical access. Thus, both could explain the results, however, the explanations were very different. For X-MOD, the poor results reflected the uncertainty due to many activated entries. For FUL, the results were due to the uncertainty because the reduced items had too little information in the speech signal. Therefore, a method that tapped into online lexical access processes was chosen for the next Experiment. Experiment 4 examined lexical activation of reduced words out of context with a cross-modal repetition priming paradigm. The results indicated that massively reduced words were unable to activate the correct lexical entry. Thus, the poor transcription performance of Experiment 3 was rather caused by too little information in the speech signal than by the problem of choosing the correct word candidate for transcription. In a next step, Experiment 5 tried to estimate how the effect of prior exposure to the identical experimental stimuli could have an impact on a later priming study. Here, the reduced items were transcribed better when they were presented in their sentence context compared to Experiment 3, but the priming results did not change, irrespective of whether subjects heard the same word before or not.

Taken together, these results indicate that FUL is more successful to explain the behavior of listeners than X-MOD. X-MOD assumes that naturally occurring reduction should be stored in the mental representation. Thus, the stimuli that had been chosen for the experiments should have been encountered in similar variants before. Therefore, listeners could be expected to either activate the correct lexical entry, or if the information of the speech signal was too little, many more items should have been activated, making a decision harder. However, the priming results showed that the reduced words were not able to activate the correct entries at all. This lack of activation could have still been based on the lack of prior exposure to a similar or identical variant of the word. But even when subjects heard identical variants before and were able to create an episodic trace in memory (as verified by a correct transcription of the word), subsequent exposure did not activate the correct lexical entry. These results call into question the exemplar representation as a mean to deal with variation in natural speech.

FUL, on the other hand, correctly predicted the results of all the experiments. If the reductions are too massive, the model assumes that the correct lexical entry cannot be activated. As has been shown already in the production analysis, there are no phonological rules in the grammar of the listener that can undo the changes that can subsequently be matched successfully to the underlying form. For the listener, the information in the speech signal is too little – there might also be too much mismatching information in some positions – to activate the correct lexical entry. FUL also predicted that this lack of activation is independent of prior exposure. Experiment 5 provided evidence that this assumption is also correct. Another assumption of FUL was indirectly supported by the data in that subjects were able to correctly transcribe massively reduced words when they were presented within their sentence context. Thus, information from other levels than phonology alone is crucially needed for successful speech perception of massively reduced words. The results are clearly advocating for an abstract representation, as proposed in FUL (Lahiri & Reetz, 2002, accepted).

These results shed a different light onto the findings of Section 4.2, where deletions were analyzed. While X-MOD's explanation seemed more elegant and needed fewer assumptions than FUL, there is an overall advantage for FUL in assuming abstract representations and a division of labor for phonetics and phonology, especially when the results for production and perception are combined (cf. Kingston, 2006; Arvaniti, 2007; Lahiri, 2007). What makes FUL an even more successful model compared to other models with an abstract representation and a division between phonology and phonetics is that its representational structure is not based on segments but on features (cf. Lahiri & Reetz, 2002), making perception more flexible than phoneme-based approaches (cf. Johnson, 2004a).

A potential weakness of FUL is that for processes outside the phonological level, further assumptions need to be made more explicit. This clearly is a very promising field for future research. The results for the corpus studies indicate that at least an important part of the reduction processes are not phonological in nature. In its current state, FUL allows for phonetic variation, and is able to handle the data from natural speech. However, a definitive elaboration of where the processes of phonetics reside, what exactly they are, how they are exactly handled needs to be provided; the borderline between phonetics and phonology has to be defined more exactly. Clearly, the role of phonetics is better defined for the perceptual mechanisms in FUL; an explicit extension of the model for the phonetics of production seems desirable.

The results of this dissertation showed that for the interplay between phonology and phonetics 'anything goes' is not the correct assumption to make. This is true both for production as well as for perception. First of all, deletions, albeit being more random than rule-governed are not as drastic as one could imagine. There are also clear tendencies, for instance vowels are deleted rarely. Secondly, listeners have limits as to which deviations from an abstract lexical representation are tolerated. As this dissertation shows, massive reductions are not recognized without information from the sentence context. Here also lies a challenge for future research; namely to identify how this additional information is used by listeners during speech perception, and to identify the boundary between which deviations are tolerated by listeners and which are not.

A further generalization of the results presented here is that linguistic models that do not assume a clear division between phonetics and phonology will probably fail to explain the data from natural speech (for a similar argumentation, see e.g. Kingston, 2006). Most of such models build variation directly into the lexicon or the (phonological) grammar (e.g. Boersma, 1997; 1998; Kirchner, 1998; 2001; 2004). Therefore, one can label these approaches as phonetically based. These models regularly have analyses of speech production data as starting point. Therefore, they include phonetic variation directly in their basic setup. Their success for handling production data subsequently is not surprising; their success for modeling speech production is undisputed. Problematic for these models is the perception side, however. Because when phonetic rules become phonologized such that for instance massively reduced words are the outcome of these rules, perception of these variants should not be problematic. However, the results presented here tell us otherwise. Concentrating on production of natural speech, therefore, can also result in flawed assumptions.

Similarly, reduction and deletion explanations that focus solely on effort minimization are not realistic either. The reduction processes have been shown to be far more complex than can be explained by an effort-based approach alone (cf. Kirchner, 1998, 2004). Independent evidence also calls into question a purely effort-based explanation (e.g. Kingston, 2006; Cho et al., 2007; Tabain & Perrier, 2007; Kraehenmann & Lahiri, 2008). For instance Kraehenmann and Lahiri (2008) showed that Swiss speakers increased their effort and expanded a length contrast in utterance initial position, a position where this effort is completely in vain, since listeners are not able to use the cue of the closure duration at all (for a general treatment of this phenomenon in Thurgovian, see Lahiri & Kraehenmann, 2004).

Furthermore, on the other extreme side are models that do not allow for variation at all. These models are also doomed to fail. A classic Optimality Theoretic approach, for instance, does not allow for the amount of variation that can be observed in natural speech, because there is only one single absolute constraint ranking per (speaker) grammar (cf. Prince & Smolensky, 1993; McCarthy, 2002). Adaptations to this basic theory have been made to handle variation (e.g. Anttila & Cho, 1998; Anttila, 2002; Anttila & Fong, 2004; Coetzee, 2006). Further research has to show, in how far they can account for all the data from natural speech.

Another recent development has been the emergence of models and theories with assumptions that have been treated in this dissertation as complimentary; these models are both abstract <u>and</u> episodic. Rather than opposing the two assumptions, they combine the two views (e.g. Goldinger, 1998, 2007; Luce et al., 2003; Luce & McLennan, 2005; Cutler et al., 2006; Ranbom & Connine, 2007; Pierrehumbert, 2006a). While there is some attractiveness in adding different assumptions from different approaches, at some point these models run the risk of being unfalsifiable. Unless it is made absolutely clear what roles the two representations have, or why this is the case, there is no possibility for clear-cut predictions. Otherwise, such models are in danger of being adjusted *ad-hoc* to results from research. There is no question that after knowing in more detail the strengths and weaknesses of both "pure" accounts, the absolute need for a mixed model could arise. However, before that point is reached, it is absolutely necessary that the pure accounts are examined to their maxima. Only after knowing more about the perception of natural speech, the need for a combined model should be given in.

A further crucial finding of this dissertation is that only when several experimental methods are combined, and when results of the investigation of different processes are taken into account, meaningful conclusions about the success or failure of theoretical models are possible (see also Kessinger & Blumstein, 1998). The results of Chapter 3 demonstrated that featural representations with underspecification are successful in explaining and predicting the results, whereas the results of Chapter 4 indicated more generally, that the abstractionist approach is crucial in explaining the findings. A combination of these results showed FUL as model that is able to explain all the findings. Furthermore, this dissertation provides a comparison of labeled corpus data and real-life performance, bridging corpus studies for speech production and online measures for speech perception. The combination of these methods is a mean to strengthen the ultimate explanatory power of the results. While, for example, production results indicated that exemplar-based models, such as X-MOD (cf. Johnson, 1997) could model the reduction and variation patterns occurring in natural speech more successfully (see also Johnson, 2004a), the overall picture changed when results from perception studies were combined with the findings for production. The lack of experiments that investigate the effect of massive reductions on perception render abstractionist models in general and FUL in particular appear as not adequately dealing with natural language. However, this dissertation provided evidence that the contrary is true by combining corpus analyses and perception experiments. The data analysis revealed several factors that have an impact on the deletion of segments. However, despite the large number of data points in the analysis, an interpretation of the results becomes very hard. Almost every factor is significant in a single analysis. Furthermore, the explanatory power of the statistical analysis is not very high. This is additional evidence that there are many different effects that contribute to variation in language production. The results are also indicative that language is dependent on multiple factors, with a large amount of enlaced factors that cannot easily be analyzed separately. Thus, a more sensible way to analyze data is to concentrate on a single segment after having an overview of the processes that occur in natural langue. This allows for a control of some of the factors and enables an intentional variation of others. While natural speech stimuli do not allow for a complete control over what information is actually included in the speech signal, they are closer to the reality for listeners than any stimuli that have been consciously created for the sake of experiments. The results that have been presented in this dissertation should be seen as a starting point for further more interdisciplinary studies on natural speech.

An interesting direction for future research concerns the investigation of the lexical activation produced by reduced variants. Moreover, a thorough examination of the role of sentence context and further studies on the exact extent of reductions and deletions that still permit lexical access are necessary. Furthermore, additional research based on conversational speech has to be conducted to investigate how listeners determine word boundaries in connected speech, especially when they are reduced since natural speech usually does not have clear demarcations of word boundaries. This is an extension to research investigating the units that are crucial for word recognition (e.g. Mehler et al., 1981; Cutler et al., 1986; Cutler, 1997; Christophe et al, 2003; Christophe et al, 2004). In addition, more languages from different language families have to be investigated. The majority of the languages that have been investigated are Germanic languages (e.g. Kohler, 1990; Ernestus, 2002; Johnson, 2004a; Dilley & Pitt, 2007). Of crucial importance for this enterprise is the creation of more corpora of natural speech in many different languages with a reliable phonetic transcription.

Furthermore, a promising field of research concerns an investigation of the information that helps to identify speakers and adjust recognition expectation to different listeners (see also Schweinberger, 2001; Newman & Evers, 2007). If listeners "know" that a special speaker does not produce a certain segment, for example, it is effective to adjust the expectations accordingly. From a point of view of an abstract model, the lexicon will not be adjusted, only some expectations for perception. Thus, some kind of "speaker filters" could be created by listeners. Exemplars may be used for creating such filters; but as the evidence of this dissertation suggests, they are not part of the lexicon.

One important note at this point is that the disagreement between the two theoretical frameworks that have been examined in this dissertation concentrates on lexical representation of language. The more general question of whether exemplars are stored in the brain is not touched upon at all. Clearly, humans store episodic memory traces in their brains. For instance, we can remember with great detail the latest movie we have seen in cinema; we can also remember the best pizza we have eaten in our life, its toppings as well as quite exactly how it tasted (we might also forget parts of these memory traces, though). These are all episodes that every human has stored in the brain. However, the crucial question that sets apart the two models is whether humans store these episodic traces in what linguists call the “lexicon”, or whether this language component has rather different, abstract representations which are independent of encountered exemplars. The evidence provided in this dissertation suggests that exemplars are not stored in the mental lexicon and that representations are rather abstract.

The results of Chapters 3 and 4 taken together suggest that a model assuming abstract representations with single lexical entries based on features for each morpheme (i.e. FUL, Lahiri & Reetz, 2002; accepted) is able to account for the data from natural speech. Thus, not every instance of a heard word is stored with ample phonetic detail. On the contrary, “poor” and simplistic abstract representations best explain the human behavior. This technique (i.e. not to store everything) was also one of Sherlock Holmes’ maxims that led to his legendary successes. Rephrased to fit the linguistic needs, one could quote him like this: “listeners are very careful as to what they store in their brain-attic.”

«'Data! Data! Data!' he cried impatiently.' I can't make bricks without clay.'»
Sherlock Holmes *(Arthur Conan Doyle, The Adventure of the Copper Beeches)*

Appendices

Appendix A:
All pronunciation variants of einverstanden ('agree-PAST PART.) in the Kiel corpus transcribed according to IPA conventions.

	phonetic transcription	**deviations from canonical transcription**
	[ˈʔaɪnfɐˌʃtandən]	canonical transcription, no deviations
i	[ˈa̰ɪ̰nfɐˌʃtan̰n]	1 segment deletions, 2 glottalization
ii	[ˈa̰ɪ̰nfɐˌʃtanʔn]	1 segment deletions, 1 glottalization, 1 weakening
iii	[ˈa̰ɪ̰nfɐˌʃtan̰]	2 segment deletions, 2 glottalization
iv	[ˈa̰ɪ̰nfɐˌʃtan]	3 segment deletions, 1 glottalization
v	[ˈa̰ɪ̰nfɐˌʃtandn]	1 segment deletions, 1 glottalization
vi	[ˈa̰ɪ̰nfɐˌʃtanhn]	1 segment deletions, 1 glottalization, 1 weakening
vii	[ˈaɪmfɐˌʃta̰n]	3 segment deletions, 1 glottalization, 1 assimilation
viii	[ˈʔa̰ɪ̰nfɐˌʃta̰n]	2 segment deletions, 2 glottalization
ix	[ˈʔa̰ɪ̰nfɐˌʃtanʔn]	1 segment deletion, 1 glottalization, 1 weakening
x	[ˈʔa̰ɪ̰nfɐˌʃtan̰]	2 segment deletions, 2 glottalizations
xi	[ˈʔa̰ɪ̰nfɐˌʃtan̰n]	1 segment deletions, 2 glottalizations
xii	[ˈʔa̰ɪ̰nfɐˌʃtandn]	1 segment deletion, 1 glottalization
xiii	[ˈʔa̰ɪ̰ntfɐˌʃtanhn]	1 segment deletion, 1 glottalization, 1 weakening
xiv	[ˈʔaɪnfɐˌʃtanhn]	1 segment deletion, 1 weakening

	phonetic transcription	deviations from canonical transcription
xv	[ˈa̰ɪ̰nfɐˌʃtann]	2 segment deletions, 1 glottalization
xvi	[ˈa̰ɪ̰ntfɐˌʃtan̰]	2 segment deletions, 2 glottalizations, 1 insertion
xvii	[ˈa̰ɪ̰ntfɐˌʃtan̰n]	1 segment deletions, 2 glottalizations, 1 insertion
xviii	[ˈa̰ɪ̰nvɐˌʃtann]	2 segment deletions, 1 glottalization, 1 voicing
xix	[ˈʔaɪnfɐˌʃtanhn]	1 segment deletion, 1 weakening
xx	[ˈaɪnfɐˌʃanʔn]	2 segment deletions, 1 weakening
xxi	[ˈaɪnfɐˌʃta̰n]	3 segment deletions, 1 glottalization
xxii	[ˈaɪnfɐˌʃtan̰]	3 segment deletions, 1 glottalization
xxiii	[nfɐʃtan̰]	4 segment deletions, 1 glottalization

Appendix B:
Examples of contexts from which [am] and [an] stimuli were extracted for Experiments 1 & 2. For details of <e>-stimuli, see Table 6.

Kiel corpus transcription	Example stimuli in orthography	Condition	
		UNASSIMILATED-/m/	
[ɑm]	.. daran **am** **e**rsten, ...	VOWEL CONTEXT	/m/-VOWEL
[ɑm]	… das denn **am** **B**esten	LABIAL CONTEXT	/m/-LABIAL
[ɑm]	...wir **am** **s**echsten	CORONAL CONTEXT	/m/-CORONAL
[ɑm]	... wir **am** **g**ünstigsten ...	DORSAL CONTEXT	/m/-DORSAL
		UNASSIMILATED-/n/	
[ɑn]	… sieht das d**ann** **a**us …	VOWEL CONTEXT	/n/-VOWEL
[ɑn]	D**ann** **b**rauchen wir ...	LABIAL CONTEXT	/n/-LABIAL
[ɑn]	Ist d**ann** **d**er…	CORONAL CONTEXT	/n/-CORONAL
[ɑn]	… aber m**an** **k**ann …	DORSAL CONTEXT	/n/-DORSAL
		ASSIMILATED	
[ɑm]	Und d**ann** **b**rauchen wir ...	LABIAL CONTEXT	

Appendix C:
List of words that were deleted completely and how often this was the case.

Word	Number	Gloss.
auch	1	too
da	1	there
dann	1	then
ein	3	a/an/one
einen	1	a/an/one-CASE
es	2	it
ich	3	I
ist	7	is
ja	3	yes
jetzt	1	now
mit	1	with
sie	1	she/they
und	3	and
Total	**28**	

Appendix D:
*List of verbs used for the corpus production (*regular, ** irregular, gloss. in parentheses).*

baden*	(bath)	hausen*	(house)	reiben**	(rub)
bannen*	(ban)	kauen*	(chew)	reihen*	(rank)
beissen**	(bite)	kleben*	(glue)	reisen*	(travel)
bergen**	(recover)	knallen*	(bang)	reissen**	(rip)
blaehen*	(billow)	kneifen**	(pinch)	ruhen*	(rest)
blasen**	(blow)	kochen*	(cook)	russen*	(grime)
braten**	(fry)	kriechen**	(crawl)	speien**	(spit)
buchen*	(book)	loben*	(praise)	speisen*	(dine)
buessen*	(purge)	mieten*	(rent)	sperren*	(bar)
fliehen**	(flee)	missen*	(miss)	spinnen**	(spin)
fliessen**	(flow)	pfeifen**	(whistle)	stechen**	(stab)
fressen**	(feed)	preisen**	(glorify)	streifen*	(swipe)
frieren**	(freeze)	pressen*	(press)	tilgen*	(amortize)
giessen**	(water)	quellen**	(swell)	wachsen**	(grow)
graben**	(dig)	raffen*	(gather)	waschen**	(wash)
hassen*	(hate)	raten**	(guess)	werben**	(advertize)
hauen**	(beat)	rauben*	(rob)		

Appendix E: Word list of the first transcription Experiment 3.

Filler:

Alternative	(alternative)
Angriff	(Attack)
anrufen	(Call)
Bier	(Beer)
Familie	(Family)
gelten	(apply)
Gesellschaft	(Society)
Glück	(Luck)
Gruppe	(Group)
lächerlich	(ridiculous)
maximal	(Max.)
Norden	(North)
nützlich	(useful)
Raum	(Room)
reden	(Speak)
Sendung	(broadcast)

Word Pairs:

Abend	(evening)
Abstand	(distance)
allerdings	(however)
Anschluss	(affiliation)
Arbeit	(work)
aufsuchen	(call)
Auge	(eye)
bedeutendsten	(most significant)
beginnen	(begin)
Beispiel	(example)
bekommen	(get)
Bericht	(report)
Besuch	(visit)
bisschen	(a little)
brauchen	(need)
Büro	(office)
dauern	(take)
Doktor	(doctor)
ehrlich	(regular)
eigentlich	(regular)
einfach	(simple)
einverstanden	(agreed)
empfangen	(receive)
Entschuldigung	(sorry)
Erinnerung	(memory)
erledigen	(handle)
eventuell	(maybe)
Fach	(compartment)
fahren	(drive)
finden	(find)
gehen	(go)
genau	(exact)
Geschichte	(history)
Glas	(glass)
günstig	(cheap)
halten	(hold)
hervorragend	(excellent)
Hotel	(hotel)
ideal	(ideal)
Kaffee	(coffee)
Karneval	(carneval)
kriegen	(become)
Kuchen	(cake)
Mittag	(noon)
möchten	(like)
Moment	(moment)
natürlich	(natural)
nehmen	(take)
nennen	(call)
notieren	(note)
passen	(fit)
Pause	(pause)
Problem	(problem)
Rahmen	(frame)
Reise	(journey)
rufen	(call)
ruhig	(calm)
Sache	(matter)
schade	(too bad)
schaffen	(create)
schauen	(look)
schlagen	(hit)
schlecht	(bad)
schreiben	(write)
Seminar	(seminar)
Sinn	(sense)
Sitzung	(meeting)
Spitze	(peak)
stark	(strong)
stehen	(stand)
Tasse	(cup)
Termin	(date)
Tisch	(table)
tragen	(wear)
über	(about)
üblich	(usual)
übrig	(remaining)
unterbringen	(accommodate)
unternehmen	(venture)
vereinbaren	(arrange)
Verfügung	(disposal)
verstehen	(understand)
vielleicht	(perhaps)
Vorschlag	(proposition)
wahrscheinlich	(probably)
wissen	(know)
Woche	(week)
wollen	(want)
wunderbar	(wonderful)
Wunsch	(wish)
ziemlich	(fairly)
zusammen	(together)

Appendix F: List of nonwords for the lexical decision tasks in experiment 4 and 5.

barne	grinze	regul
begrunnen	hils	sannen
benusten	kefen	schan
biesen	kichen	schlumm
breisen	kons	schnaupen
brulde	kontilt	tappich
bunen	lofen	ufen
bust	mot	urbe
delmen	mönde	wides
dreblich	nitz	wuhl
driben	olke	zauger
flagen	ollend	zilgen
fost	pucken	
gelugen	püst	

Appendix G:
Prime/target pairs and the condition of Experiment 4 and 5.
(U/R = Unreduced, Reduced; CT = Control)

beispiel (U/R)	beispiel	abstand (CT)	moment
kuchen (CT)	beispiel	moment (U/R)	moment
bisschen (U/R)	bisschen	genau (CT)	natürlich
eventuell (CT)	bisschen	natürlich (U/R)	natürlich
brauchen (U/R)	brauchen	nehmen (U/R)	nehmen
gelten (CT)	brauchen	rufen (CT)	nehmen
dauern (U/R)	dauern	nennen (U/R)	nennen
stehen (CT)	dauern	schaffen (CT)	nennen
doktor (U/R)	doktor	empfangen (CT)	passen
sendung (CT)	doktor	passen (U/R)	passen
ehrlich (U/R)	ehrlich	schauen (U/R)	schauen
günstig (CT)	ehrlich	unternehmen (CT)	schauen
eigentlich (U/R)	eigentlich	schlagen (U/R)	schlagen
maximal (CT)	eigentlich	unterbringen (CT)	schlagen
fahren (U/R)	fahren	raum (CT)	sinn
notieren (CT)	fahren	sinn (U/R)	sinn
finden (U/R)	finden	schreiben (CT)	tragen
reden (CT)	finden	tragen (U/R)	tragen
gehen (U/R)	gehen	glück (CT)	verfügung
verstehen (CT)	gehen	verfügung (U/R)	verfügung
bericht (CT)	glas	hervorragend (CT)	vielleicht
glas (U/R)	glas	vielleicht (U/R)	vielleicht
angriff (CT)	hotel	fach (CT)	vorschlag
hotel (U/R)	hotel	vorschlag (U/R)	vorschlag
ideal (U/R)	ideal	tisch (CT)	woche
übrig (CT)	ideal	woche (U/R)	woche
karneval (U/R)	karneval	bekommen (CT)	wollen
sache (CT)	karneval	wollen (U/R)	wollen
halten (CT)	kriegen	über (U/R)	über
kriegen (U/R)	kriegen	zusammen (CT)	über

Appendix H:
Sentence list for the transcription part of experiment 5, with the crucial words in bold face.

- Das Wochenende **vielleicht** 22, 23, 24
- Können sie vielleicht mal nen **Vorschlag** machen?
- Um 9 Uhr im **Hotel** in Stockholm
- Ich muss jetzt noch mal eben genau **schauen**
- Und dann **schlagen** sie doch einen andern Termin vor
- Ich werde Ihnen einfach mal die Termine **nennen**, an denen ich also Zeit habe
- Moment jetzt Dienstag oder Mittwoch wir **brauchen** nur noch einen Tag
- Und da habe ich wieder **eigentlich**, um ehrlich zu sein, den ganzen Tag zur Disposition hier
- Also in der **Woche** vom 17 bis zum 22. Oktober
- Einfach mal auf ein **Glas** Wein oder so
- Nur zwischen dem ersten und 16. 12. 93 habe ich überhaupt Tage zur **Verfügung**
- Das wird mir leider gar nicht gut **passen**, weil ich da unterwegs bin
- Zum **Beispiel** am Mittwoch Nachmittag, dem 9. Februar
- Und dann kriegen wir bestimmt auch noch ein **bisschen** was geschafft
- Das ist **ideal**, 10 bis 12
- Ein Kaffe trinken und dann abends nen lauen Lenz. Wie **finden** Sie das?
- Ich würde dazu vorschlagen, dass wir vielleicht auch doch mal in den deutschen Osten **fahren**
- Ja **natürlich**, das mache ich
- Sonst **nehmen** wir vielleicht zwei andere Tage besser
- Wenn wir den busstag freihalten **wollen**
- **Vielleicht** am 6. und 7. Dezember?
- Wenn sie vielleicht mal n **Vorschlag** machen könnten?
- München macht keinen **Sinn**, ich geh lieber nach Freiburg
- Welche Zeit **schlagen** Sie vor?
- Ja Rosenmontag schaue ich mir ja immer gern die Sendung im Fernsehen an, **Karneval** aus Mainz
- Und da habe ich wieder eigentlich, um **ehrlich** zu sein, den ganzen Tag zur Disposition hier
- **Nennen** sie mir irgend einen Termin
- Lassen Sie uns da mal versuchen, auf 15 Uhr zu **gehen**
- Zwei Stunden werden wir **brauchen**
- Hier ist der Herr **Doktor** Müller Lüdenscheidt
- So lange wird das ja vielleicht gar nicht **dauern**, vielleicht ein Tag
- Nein, im Februar habe ich leider **über, über** ... Faschingszeit überhaupt keine Zeit
- Dann **kriegen** wir das glaube ich ganz gut hin

¬ Das heißt, **Moment**, allerdings erst nachmittags
¬ Bis 19 uhr würde ich dann zur **Verfügung** stehen
¬ Wir müssten einen Termin **finden** zwischen dem 12. Und dem 16. März
¬ Dann machen wir am 3. Juli das Vorbereitungstreffen und **fahren** dann danach gleich los
¬ Dann **tragen** wir schon mal den Montag den 6. ein
¬ Wir müssen **natürlich** jetzt aufpassen
¬ Das heißt die könnte man also diese Termine **nehmen**
¬ Sie können mir ja dann noch genauer sagen, welches **Hotel** sie aufsuchen werden
¬ Müssten Sie halt mal **schauen**, bei mir ist schon einiges belegt
¬ Ich denke, dass es im **Sinn** unserer Arbeit sehr nützlich wäre
¬ Oder nach **Karneval**, ich weiß nicht, wie heftig sie feiern
¬ Ganz **ehrlich**, also wenn wir schon eine Woche miteinander verbringen
¬ Mich würde das zwar nicht stören, aber dann **gehen** wir auf den zweiten Mai
¬ Herr **Doktor** Bergemer, wie sieht das aus?
¬ Kommt natürlich darauf an, wann wir jetzt **eigentlich** losfahren
¬ Ich denke aber, in der darauf folgenden **Woche**, das würde mir ganz gut passen
¬ Bei einem Dia Abend, bei einem **Glas** Wein oder Bier
¬ Also ich denke sie wird 5 Tage **dauern**
¬ Wir müssen ja noch einen Bericht abfassen, **über** diese Reise
¬ Ob sie ihre Freunde in Bonn und Mannheim am Samstag und Sonntag aus den Betten **kriegen**
¬ **Moment**, jetzt Dienstag oder Mittwoch, wir brauchen nur noch einen Tag
¬ Und zwar würde es mir gut **passen**, am Wochenende, 6. oder 7. Mai
¬ Vom Dienstag den 18. bis Freitag den 21. könnte ich zum **Beispiel** erstmal anbieten
¬ Das wäre mir also n **bisschen** zu spät muss ich sagen
¬ Mittwoch Vormittag scheint mir **ideal**
¬ Soweit die Füße **tragen**
¬ **Wollen** wir's sonst da machen?

«…life is infinitely stranger than anything which the mind of man could invent.»
***Sherlock Holmes** (Arthur Conan Doyle, A Case of Identity)*

References

Anttila, A. 2002. Morphologically conditioned phonological alternations. *Natural Language & Linguistic Theory*, 20.1-42.

Anttila, A. and Cho, Y.-M. Y. 1998. Variation and change in Optimality Theory. *Lingua*, 104.31-56.

Anttila, A. and Fong, V. 2004. Variation, ambiguity and noun classes in English. *Lingua*, 114.1253-1290.

Archangeli, D. 1988. Aspects of underspecification theory. *Phonology*, 5.183-207.

Arvaniti, A. 2007. On the relationship between phonology and phonetics (or why phonetics is not phonology). Paper presented at the 16th International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken.

Avery, P. and Rice, K. 1989. Segment structure and coronal underspecification. *Phonology*, 6.179-200.

Baayen, R. H., Piepenbrock, R. and Gulikers, L. 1995. *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baayen, R. H., Davidson, D. J. and Bates, D. M. to appear. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language.*

Bayer, J. and Brandner, E. 2004. Klitisiertes zu im Bairischen und Alemannischen. *Morphologie und Syntax Deutscher Dialekte und Historische Dialektologie des Deutschen*, ed. by F. Patocka and P. Wiesinger, 160-188. Wien: Praesens Edition.

Bell, A. 1984. Language style as audience design. *Language in Society*, 13.145–204.

Benware, W. A. 1986. *Phonetics and Phonology of Modern German*. Washington (D.C.): Georgetown University Press.

Boersma, P. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21, 43-58.

Boersma, P. 1998. *Functional Phonology. Formalizing the Interactions Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Boersma, P. and Weenink, D. 2007. *PRAAT - Doing Phonetics by Computer*. Version 5.0.22.

Bölte, J. 2001. Graded lexical activation by pseudowords in cross-modal semantic priming: Spreading of activation, backward priming, or repair? Paper presented at the 23rd Meeting of the Cognitive Science Society, Mawah (NJ).

Bölte, J. and Coenen, E. 2002. Is phonological information mapped onto semantic information in a one-to-one manner? *Brain and Language*, 81.384-397.

Bradlow, A. R. 1995. A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97.1916-1924.

Brockhaus, W. 1995. *Final Devoicing in the Phonology of German*. Tübingen: Niemeyer.

Brown, C. M. 1990. *Spoken-Word Processing in Context*, University of Nijmegen: Ph D. Thesis.

Bybee, J. 2000a. Lexicalization of sound change and alternating environments. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. by M. B. Broe and J. B. Pierrehumbert, 250-268. Cambridge (UK): Cambridge University Press.

Bybee, J. 2000b. The phonology of the lexicon: Evidence from lexical diffusion. *Usage-Based models of Language*, ed. by M. Barlow and S. Kemmer, 65-87. Stanford (CA): CSLI Publications.

Bybee, J. 2001. *Phonology and Language Use*. Cambridge (UK): Cambridge University Press.

Bybee, J. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisistion*, 24.215-221.

Bybee, J. 2007. *Frequency of Use and the Organization of Language*. Oxford (UK): Oxford University Press.

Bybee, J. and Hopper, P. J. (eds.) 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Byrd, D. 1994. Relations of sex and dialect to reduction. *Speech Communication*, 15.39-54.

Byrd, D. and Tan, C. C. 1996. Saying consonant clusters quickly. *Journal of Phonetics*, 24.263-282.

Caramazza, A., Costa, A., Miozzo, M. and Bi, Y. 2001. The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27.1430-1450.

Carson-Berndson, J. 1998. *Time Map Phonology*. Dordrecht: Kluwer Academic Publishers.

Cedergen, H. J. and Sankoff, D. 1974. Variable rules: Performance as a statistical reflection of competence. *Language*, 50.333-355.

Cho, T., McQueen, J. M. and Cox, E. A. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35.210-243.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English (SPE)*. New York: Harper & Row.

Christophe, A., Gout, A., Peperkamp, S. and Morgan, J. 2003. Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, 31.585-598.

Christophe, A., Peperkamp, S., Pallier, C., Block, E. and Mehler, J. 2004. Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51.523-547.

Clahsen, H. 2006a. Dual-mechanism morphology. *Encyclopedia of Language and Linguistics (Vol. IV)*, ed. by K. Brown, 1-5. Oxford (UK): Elsevier.

Clahsen, H. 2006b. Linguistic perspectives on morphological processing. *Advances in the Theory of the Lexicon*, ed. by D. Wunderlich, 355-388. Berlin: Mouton de Gruyter.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook*, 2.225-253.

Clements, G. N. 2001. Representational economy in constraint-based phonology. *Distinctive Feature Theory*, ed. by T. A. Hall, 71-146. Berlin: Mouton der Gruyter.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology*, 20.287-333.

Clements, G. N. and Hume, E. 1995. The internal organization of speech sounds. *The Handbook of Phonological Theory*, ed. by J. A. Goldsmith, 245-306. Oxford (UK): Blackwell.

Clopper, C. G. and Pierrehumbert, J. B. 2008. Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, 124.1682-1688.

Coenen, E., Zwitserlood, P. and Bölte, J. 2001. Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes*, 16.535-564.

Coetzee, A. W. 2006. Variation as accessing 'non-optimal' candidates. *Phonology*.23.335-385.

Connine, C. M. 2004. It's not what you hear but how often you hear it: in the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, 11.1084-1089.

Connine, C. M., Blasko, D. G. and Titone, D. 1993. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32.193–210.

Crosswhite, K. M. 2004. Vowel reduction. *Phonetically Based Phonology*, ed. by B. Hayes, R. Kirchner and D. Steriade, 191-231. Cambridge (UK): Cambridge University Press.

Cutler, A. 1997. The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes*, 12.839-845.

Cutler, A. 1998. The recognition of spoken words with variable representation. Paper presented at ESCA Workshop on Sound Patterns of Spontaneous Speech, Aix-en-Provence.

Cutler, A., Mehler, J., Norris, D. and Segui, J. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25.385-400.

Cutler, A., Eisner, F., McQueen, J. and Norris, D. 2006. Coping with speaker-related variation via abstract phonemic categories. Paper presented at Labphon, Paris (F).

Dalby, J. M. 1986. *Phonetic Structure of Fast Speech in American English*. Bloomington, IN: Indiana Linguistics Club.

Deelman, T. and Connine, C. M. 2001. Missing information in spoken word recognition: Nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 27.656-663.

Dell, G. S. and Gordon, J. K. 2003. Neighbors in the lexicon: Friends or foes? *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, ed. by N. O. Schiller and A. S. Meyer, 9-37. New York: Mouton de Gruyter.

Dilley, L. C. and Pitt, M. A. 2007. A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, 122.2340-2353.

Dressler, W., Fasching, P., Chromec, E., Wintersberger, W., Leodolter, R., Stark, H., Groll, G., Reinhart, J. and Pohl, H. D. 1972. Phonological fast speech rules in colloquial Viennese German. *Wiener Linguistische Gazette*, 1.1-30.

Dressler, W. U. 1972. Approaches to fast speech rules. *Phonologica*.219-234.

Duez, D. 1995. On spontaneous French speech: Aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics*, 23.407-427.

Ellis, L. and Hardcastle, W. J. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics*, 30.373-396.

Ernestus, M. 2000. *Voice Assimilation and Segment Reduction in Casual Dutch*. Utrecht: LOT.

Ernestus, M. and Baayen, H. 2007. The comprehension of acoustically reduced morphologically complex words: the roles of deletion, duration, and frequency of occurrence. Paper presented at 16th International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken.

Ernestus, M., Baayen, H. and Schreuder, R. 2002. The recognition of reduced word forms. *Brain and Language*, 81.162-173.

Ernestus, M., Lahey, M., Verhees, F. and Baayen, R. H. 2006. Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120.1040-1051.

Fasold, R. 1972. *Tense Marking in Black English*. Arlington (VA): Center for Applied Linguistics.

Flege, J. E. 1988. Effects of speaking rate on tongue position and velocity of movement in vowel production. *Journal of the Acoustical Society of America*, 84.901 – 916.

Forster, K. I. 1990. Lexical processing. *Language: An Invitation to Cognitive Science*, ed. by D. N. Osherson and H. Lasnik, 95-131. Cambridge (MA): The MIT Press.

Fosler-Lussier, E. and Morgan, N. 1999. Effects of speaking rate and word frequency on pronunciation in conversational speech. *Speech Communication*, 29.137-158.

Fougeron, C. and Keating, P. A. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101.3728-3740.

Fougeron, C. and Steriade, D. 1997. Does deletion of French Schwa lead to neutralization of lexical distinctions? Paper presented at Eurospeech 97 - 5th Conference on Speech Communication and Technology, Rhodes (Greece).

Foulkes, P. and Docherty, G. 2006. The social life of phonetics and phonology. *Journal of Phonetics*, 34.409-438.

Fox Tree, J. E. and Clark, H. H. 1997. Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62.151-167.

Frauenfelder, U. H., Mehler, J., Segui, J. and Morton, J. 1982. The word frequency effect and lexical access. *Neuropsychologia*, 20.615-627.

Fukaya, T. and Byrd, D. 2005. An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association*, 35.45-58.

Gahl, S. and Yu, A. C. L. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review*, 23.213-216.

Ghini, M. 2001a. *Asymmetries in the Phonology of Miogliola*. Berlin: Mouton de Gruyter.

Ghini, M. 2001b. Place of articulation first. *Distinctive Feature Theory*, ed. by T. A. Hall, 147-176. Berlin: Mouton de Gruyter.

Ghini, M. 2003. From prosody to place: The development of prosodic contrasts into place of articulation contrast in the history of Miogliola. *Development in Prosodic Systems*, ed. by P. Fikkert and H. Jacobs, 419-455. Berlin/New York: Mouton de Gruyter.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105.251-279.

Goldinger, S. D. 2007. A complementary-systems approach to abstract and episodic speech perception. Paper presented at 16th International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken.

Goldinger, S. D. and Azuma, T. 2003. Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31.305-320.

Goldinger, S. D., Luce, P. A. and Pisoni, D. B. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28.501-518.

Gow, D. W. jr. 2001. Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45.133-159.

Gow, D. W. jr. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28.163-179.

Gow, D. W. jr. 2003. Feature parsing: feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65.575-590.

Gow, D. W. jr. and Hussami, P. 1999. Acoustic modification in English place assimilation. *Journal of the Acoustical Society of America*, 106.2243.

Gow, D. W. jr. and Im, A. M. 2004. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51.279-296.

Greenberg, S. 1999. Speaking in shorthand - A syllable centric perspective for understanding pronunciation variation. *Speech Communication*, 29.159-176.

Greenberg, S. and Fosler-Lussier, E. 2000. The uninvited guest: Information‘s role in guiding the production of spontaneous speech. Paper presented at CREST Workshop on Models of Speech Production: Motor Planing and Articulatory Modeling, Kloster Seeon, Germany.

Greenberg, S., Carvey, H., Hitchcock, L. and Chang, S. 2002. Beyond the Phoneme: A juncture-accent model of spoken language. Paper presented at Proceedings of the Human Language Technology Conference (HLT) 2002, San Diego.

Gumnior, H., Zwitserlood, P. and Bölte, J. 2005. Assimilation in existing and novel German compounds. *Language and Cognitive Processes*, 20.465-488.

Guy, G. R. 1980. Variation in the group and the individual: The case of final stop deletion. *Locating Language in Time an Space*, ed. by W. Labov, 1-36. New York: Academic Press.

Guy, G. R. 1992. Contextual condition in variable lexical phonology. *Language Variation and Change*, 3.223-229.

Hall, T. A. 1992. *Syllable Structure and Syllable Related Processes in German*. Tübingen: Max Niemeyer Verlag.

Hall, T. A. 1999. Phonotactics and the prosodic structure of German function words. *Studies on the Phonological Word*, ed. by T. A. Hall and U. Kleinhenz, 99-131. Amsterdam/ Philadelphia (PA): John Benjamins Publishing Company.

Hall, T. A. 2000. *Phonologie - Eine Einführung*. Berlin, New York: Walter de Gruyter.

Halle, M. 1995. Feature geometry and feature spreading. *Linguistic Inquiry*, 26.1-46.

Halle, M., Vaux, B. and Wolfe, A. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry*, 31.387-444.

Harrington, J. 2006. An acoustic analysis of ‘happy-tensing’ in the Queen’s Christmas broadcasts. *Journal of Phonetics*, 34.439-457.

Hawkins, S. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31.373-405.

Hay, J. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics*, 39.1041-1070.

Hay, J. 2003. *Causes and Consequences of Word Structure*. New York: Routledge.

Hay, J. and Baayen, R. H. 2005. Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9.342-348.

Hayes, B. 2009. *Introductory Phonology*. Chichester (UK): Wiley-Blackwell.

Herold, R. 1990. *Mechanisms of Merger: The Implementation and Distribution of the Low Back Merger in Eastern Pennsylvania*, University of Pennsylvania: Ph D. Thesis.

Hintzman, D. L. 1986. "Schema Abstraction" in a multiple-trace memory model. *Psychological Review*, 93.411-428.

Hockett, C. 1955. *Manual of Phonology*. Bloomington: Indiana University.

IPDS, Institut Für Phonetik Und Digitale Sprachverarbeitung. 1994. *The Kiel Corpus of Spontaneous Speech*. Kiel: IPDS.

Jacoby, L. L. 1983. Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, 22.485-508.

Jakobson, R., Fant, G. M. and Halle, M. 1963. *Preliminaries to Speech Analysis - The distinctive features and their Correlates*. Cambridge (MA): MIT Press.

Janse, E., Nooteboom, S. G. and Quené, H. 2007. Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, 22.161-200.

Jespersen, O. 1922. *Language - Its Nature, Development and Origin*. London (UK): George Allen & Unwin Ltd.

Johnson, K. 1997. Speech perception without speaker normalization. *Talker Variability in Speech Processing*, ed. by K. Johnson and J. W. Mullennix, 145-166. New York: Academic Press.

Johnson, K. 2004a. Massive reduction in conversational American English. Paper presented at Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium, August, 2002, Tokyo.

Johnson, K. 2004b. Aligning phonetic transcriptions with their citation forms. *Acoustics Research Letters On-Line*, 5.19-24.

Johnson, K. 2007. Decisions and mechanisms in exemplar-based phonology. *Experimental Approaches to Phonology*, ed. by M.-J. Solé, P. S. Beddor and M. Ohala, 25-40. Oxford (UK): Oxford University Press.

Johnson, K. and Mullennix, J. W. (eds.) 1997. *Talker Variability in Speech Perception*. San Diego: Academic Press.

Jones, D. 1967. *The Phoneme - its Nature and Use*. Cambridge (UK): Heffer.

Jones, D. 1972. *An Outline of English Phonetics*. Cambridge (UK): Cambridge University Press.

Jun, J. 1995. *Perceptual and Articulatory Factors in Place Assimilation: An Optimality Theoretic Approach*, University of California, Los Angeles: PH. D. Thesis.

Jun, J. 1996. Place assimilation is not the result of gestural overlap: Evidence from Korean and English. *Phonology*, 13.377-407.

Jun, J. 2004. Place assimilation. *Phonetically Based Phonology*, ed. by B. Hayes, R. Kirchner and D. Steriade, 58-86. Cambridge (UK): Cambridge University Press.

Jurafsky, D., Bell, A. and Girand, C. 2002. The role of the lemma in form variation. *Laboratory Phonology 7*, ed. by C. Gussenhoven and N. Warner, 3-34. Berlin: Mouton de Gruyter.

Jurafsky, D., Bell, A., Gregory, M. and Raymond, W. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Frequency and the Emergence of Linguistic Structure*, ed. by J. Bybee and P. Hopper, 229-254. Philadelphia (PA): Benjamins.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C. and Raymond, W. 1998. Reduction of English function words in Switchboard. Paper presented at International Conference on Spoken Language Processing, Sydney.

Jusczyk, P. W. 1997. *The Discovery of Spoken Language*. Cambridge (MA): The MIT Press.

Kabak, B. 2007. Hiatus resolution in Turkish: An underspecification account. *Lingua*, 117.1378-1411.

Kabak, B. and Schiering, R. 2006. The phonology and morphology of function word contractions in German. *Journal of Comparative Germanic Linguistics*, 9.53-99.

Kaisse, E. M. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. Orlando: Academic Press.

Kallmeyer, W. 1981. Aushandlung und Bedeutungskonstitution. Dialogforschung - *Jahrbuch des Instituts für Deutsche Sprache, 1980*, ed. by P. Schröder and H. Steger, 89-127. Mannheim: Pädagogischer Verlag Schwann.

Keating, P. A. 1998. Word-level phonetic variation in large speech corpora. *ZAS Working Papers in Linguistics*, ed. by B. Pompino-Marschal. Berlin: ZAS.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Cambridge (UK): Blackwell Publishers.

Kessinger, R. H. and Blumstein, S. E. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26.117-128.

Kim, H. and Jongman, A. 1996. Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, 24.295-312.

Kingston, J. 2006. Lenition. Proceedings of the Third Conference on Laboratory Approaches to Spanish Phonology, ed. by L. Colantoni and J. Steele: Cascadilla Press.

Kingston, J. 2007. The phonetics-phonology interface. *The Handbook of Phonology*, ed. by P. de Lacy, 401-434. Cambridge (UK): Cambridge University Press.

Kingston, J. and Diehl, R. L. 1994. Phonetic Knowledge. *Language*, 70.419-454.

Kiparsky, P. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10.421-441.

Kiparsky, P. 1982. From Cyclic Phonology to Lexical Phonology. *The Structure of Phonological Representations*, ed. by H. van der Hulst and N. Smith, 131-177. Dordrecht: Foris Publications.

Kirchner, R. 1998. *An Effort-Based Approach to Consonant Lenition*, UCLA: Ph D. Thesis.

Kirchner, R. 2001. *An Effort-Based Approach to Consonant Lenition: Outstanding Dissertations in Linguistics*. New York & London: Routledge.

Kirchner, R. 2004. Consonant Lenition. *Phonetically Based Phonology*, ed. by B. Hayes, R. Kirchner and D. Steriade, 313-345. Cambridge (UK): Cambridge University Press.

Kirchner, R. and Moore, R. K. 2008. Speech production with an exemplar-based lexicon. Paper presented at The 16th Manchester Phonology Meeting (MFM), Manchester (UK).

Klatt, D. H. 1979. Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7.279-312.

Kohler, K. J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. *Speech Production and Speech Modelling*, ed. by W. J. Hardcastle and A. Marchal, 69-92. Dordrecht: Kluwer Academic Publishers.

Kohler, K. J. 1991. The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, 48.180-192.

Kohler, K. J. 1995a. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.

Kohler, K. J. 1995b. Articulatory reduction in different speaking styles. Paper presented at 13th International Congress of Phonetic Sciences (ICPhS XIII), Stockholm.

Kohler, K. J. 1996. Phonetic realization of German /ə/-syllables. *Sound Patterns in Spontaneous Speech*, ed. by K. J. Kohler, C. Rehor and A. P. Simpson, 159-194. Kiel: IPDS.

Kohler, K. J. and Rodgers, J. 2001. Schwa deletion in German read and spontaneous speech. *Sound Patterns in German Read and Spontaneous Speech: Symbolic Structures and Gestural Dynamics*, ed. by K. J. Kohler, 97-123. Kiel: IPDS.

Kohler, K. J., Pätzold, M. and Simpson, A. P. (eds.) 1995. *From Scenario to Segment - the Controlled Elicitation, Transcription, Segmentation and Labelling of Spontaneous Speech*. Kiel: IPDS.

Kraehenmann, A. and Lahiri, A. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *Journal of the Acoustical Society of America*, 123.4446-4455.

Kruschke, J. K. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99.22–44.

Labov, W. 1966. *The Social Stratification of English in New York City*. Washington (D.C.): Center for Applied Linguistics.

Labov, W. 1967. Some sources of reading problems for Negro speakers of non-standard English. *New Directions in Elementary English*, ed. by A. Frazier. Champaign (IL): National Council of Teachers of English.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia (PA): University of Pennsylvania Press.

Labov, W. 1991. The three dialects of English. *Quantitative Analyses of Sound Change*, ed. by P. Eckert, 1-44. New York: Academic Press.

Labov, W. 2001. *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford (UK): Blackwell.

Labov, W. 2006. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, 34.500-515.

Lacerda, F. 1997. Distributed memory representations generate the perceptual-magnet effect. Manuscript. *http://www.ling.su.se/staff/frasse/frasse.html#Publications* (last access: 01. December 2008)

Lahiri, A. 2000a. Phonology: Structure, representation, and process. *Aspects of Language Production*, ed. by L. Wheeldon, 165-226. Philadelphia (PA): Psychology Press.

Lahiri, A. 2000b. Hierarchical restructuring in the creation of verbal morphology in Bengali and Germanic: Evidence from phonology. *Analogy, Levelling, Markedness*, ed. by A. Lahiri, 71-123. Berlin/New York: Mouton de Gruyter.

Lahiri, A. 2007. Non-equivalence between phonology and phonetics. Paper presented at The 16th International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken.

Lahiri, A. and Marslen-Wilson, W. D. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38.245-294.

Lahiri, A. and Evers, V. 1991. Palatalisation and coronality. *The Special Status of Coronals: Internal and External Evidence*, ed. by C. Paradis and J.-F. Prunet, 79-100. San Diego: Academic Press.

Lahiri, A. and Marslen-Wilson, W. D. 1992. Lexical processing and phonological representation. *Papers in Laboratory Phonology II - Gesture, Segment, Prosody*, ed. by G. J. Docherty and R. Ladd, 229-254. Cambridge (UK): Cambridge University Press.

Lahiri, A. and Reetz, H. 2002. Underspecified recognition. *Laboratory Phonology 7*, ed. by C. Gussenhoven and N. Warner, 637-675. Berlin: Mouton de Gruyter.

Lahiri, A. and Kraehenmann, A. 2004. On maintaining and extending contrasts: Notker's Anlautgesetz. *Transactions of the Philological Society*, 102.1-55.

Lahiri, A. and Reetz, H. accepted. Underspecification. *Journal of Phonetics*.

Lahiri, A., Wetterlin, A. and Jönsson-Steiner, E. 2006. Scandinavian lexical tone: prefixes and compounds. Paper presented at Nordic Prosody IX, Lund.

Lamel, L. F., Kassel, R. H. and Seneff, S. 1986. Speech database development: design and analysis of the acoustic-phonetic corpus. Paper presented at DARPA Speech Recognition Workshop.

Langacker, R. W. 1987. *Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites*. Stanford (CA): Stanford University Press.

Langacker, R. W. 2000. A dynamic usage-based model. *Usage-Based Models of Language*, ed. by M. Barlow and S. Kemmer. Stanford (CA): CSLI Publications.

Levelt, W. 1989. *Speaking - from intention to articulation*. Cambridge (MA): MIT Press.

Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6.172-187.

Lieberman, P. 1986. On the genetic basis of linguistic variation. *Invariance and Variability in Speech Processes*, ed. by J. S. Perkell and D. H. Klatt. Hillsdale (NJ): Lawrence Earlbaum Associates.

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

Lively, S. E., Pisoni, D. B. and Goldinger, S. D. 1994. Spoken word recognition. *Handbook of Psycholinguistics*, ed. by M. A. Gernsbacher, 265-301. San Diego: Academic Press.

Local, J. 2003. Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, 31.321-339.

LoCasto, P. C. and Connine, C. M. 2002. Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics*, 64.208-219.

Lodge, K. 1992. Assimilation, deletion paths and underspecification. *Journal of Linguistics*, 28.13-52.

Lodge, K. 1995. Kalenjin phonology and morphology: A further exemplification of underspecification and non-destructive phonology. *Lingua*, 96.29-43.

Luce, P. A. and Pisoni, D. B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, 19.1-36.

Luce, P. A. and McLennan, C. T. 2005. Spoken word recognition: The challenge of variation. *The Handbook of Speech Perception*, ed. by D. B. Pisoni and R. E. Remez, 591-609. Malden (MA): Blackwell.

Luce, P. A., McLennan, C. T. and Charles-Luce, J. 2003. Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. *Rethinking Implicit Memory*, ed. by J. S. Bowers and C. J. Marsolek, 197-214. Oxford (UK): Oxford University Press.

Maekawa, K., Koiso, H., Furui, S. and Isahara, H. 2000. Spontaneous speech corpus of Japanese. Paper presented at Second International Conference on Language Resources and Evaluation (LREC-2000), Athens (Greece).

Manuel, S. Y. 1991. Recovery of “deleted” Schwa. Paper presented at Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS XIV), Stockholm.

Manuel, S. Y. 1995. Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics*, 23.453-476.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R. and Pinker, S. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology*, 29.189-256.

Marslen-Wilson, W. D. 1990. Activation, competition, and frequency in lexical access. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, ed. by G. Altmann, 148-172. Cambridge (MA): MIT Press.

Marslen-Wilson, W. D. and Welsh, A. 1978. Processing interactions and lexical access during word recognition in continous speech. *Cognitive Psychology*, 10.29-63.

Marslen-Wilson, W. D. and Tyler, L. K. 1980. The temporal structure of spoken language understanding. *Cognition*, 8.1-71.

Marslen-Wilson, W. D. and Zwitserlood, P. 1989. Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15.576-585.

McCarthy, J. 2002. *Thematic Guide to Optimality Theory*. Cambridge (UK): Cambridge University Press.

McClelland, J. L. and Elman, J. L. 1986. The TRACE model of speech perception. *Cognitive Psychology*, 18.1-86.

McLennan, C. T. 2006. The time course of variability effects in the perception of spoken language: Changes across lifespan (Variability effects across the lifespan). *Language and Speech*, 49.113-125.

McLennan, C. T. and Luce, P. A. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31.306-321.

McLennan, C. T., Luce, P. A. and Charles-Luce, J. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29.539-553.

McQueen, J. M., Cutler, A. and Norris, D. 2006. Phonological abstraction in the mental lexicon. *Cognitive Science*, 30.1113-1126.

Mehler, J., Dommergues, J. Y. and Frauenfelder, U. 1981. The syllable‘s role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20.298-305.

Meunier, F. and Segui, J. 1999. Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, 41.327-344.

Mitterer, H. and Ernestus, M. 2006. Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34.73-103.

Mitterer, H., Yoneama, K. and Ernestus, M. 2008. How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language*, 59.133-152.

Mohanan, K. P. 1993. Fields of attraction in Phonology. *The Last Phonological Rule*, ed. by J. A. Goldsmith, 61-116. Chicago: The University of Chicago Press.

Nespor, M. and Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.

Neu, H. 1980. Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. *Locating Language in Time and space*, ed. by W. Labov, 37–54. New York: Academic Press.

Newman, R. and Evers, S. 2007. The effect of talker familiarity on stream segregation. *Journal of Phonetics*, 35.85-103.

Nolan, F. 1992. The descriptive role of segments. *Papers in Laboratory Phonology II - Gesture, Segment, Prosody*, ed. by G. J. Docherty and R. Ladd, 261-280. Cambridge (UK): Cambridge University Press.

Nosofsky, R. M. 1988. Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14.700-708.

Nosofsky, R. M. and Palmeri, T. J. 1997. An exemplar-based random walk model of speech classification. *Psychological Review*, 104.266-300.

Obleser, J., Lahiri, A. and Eulitz, C. 2003a. Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage*, 20.1839-1847.

Obleser, J., Lahiri, A. and Eulitz, C. 2004. Magnetic brain response mirrors extraction of phonological features from spoken words. *Journal of Cognitive Neuroscience*, 16.31-39.

Obleser, J., Elbert, T., Lahiri, A. and Eulitz, C. 2003b. Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research*, 15.207-213.

OED. 1989. *Oxford English Dictionary*. Oxford (UK): Oxford University Press.

Ogden, R. 1999. A declarative account of strong and weak auxiliaries in English. *Phonology*, 16.55-92.

Ohala, J. J. 1990. The phonetics and phonology of aspects of assimilation. P*apers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. by J. Kingston and M. E. Beckman, 258-275. Cambridge (UK): Cambridge University Press.

Oostdijk, N. 2000. The spoken Dutch corpus. Overview and first evaluation. Paper presented at Second International Conference on Language Resources and Evaluation (LREC-2000).

Paradis, C. and Prunet, J.-F. (eds.) 1991. *The Special Status of Coronals: Internal and External Evidence. Phonetics and Phonology, Vol. 2*. San Diego: Academic Press.

Passy, P. 1891. *Etude sur les changements phonétiques et leurs caractères généraux*. Paris: Librairie Firmin - Didot.

Patterson, D. and Connine, C. M. 2001. A corpus analysis of variant frequency in American English flap production. *Phonetica*, 58.254–275.

Patterson, D., LoCasto, P. C. and Connine, C. M. 2003. Corpora analyses of frequency of Schwa deletion in conversational American English. *Phonetica*, 60.45-60.

Perkell, J. S. and Klatt, D. H. (eds.) 1986. *lnvariance and Variability in Speech Processes*. Hillsdale (NJ): Lawrence Erlbaum Associates.

Peters, B. 2001. *‚Video Task' oder ‚Daily Soap Szenario' - Ein Neues Verfahren zur Kontrollierten Elizitation von Spontansprache*. Kiel: IPDS.

Phillips, B. S. 2001. Lexical diffusion, lexical frequency, and lexical analysis. *Frequency and the Emergence of Linguistic Structure*, ed. by J. Bybee and P. Hopper, 123-126. Amsterdam: John Benjamins.

Picheny, M. A., Durlach, N. I. and Braida, L. D. 1985. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28.96-103.

Pickett, J. M. and Pollack, I. 1963. Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of context. *Language and Speech*, 6.151-164.

Pierrehumbert, J. B. 2000. What people know about sounds of language. *Studies in Linguistic Sciences*, 29.111-120.

Pierrehumbert, J. B. 2001a. Exemplar Dynamics: Word frequency, lenition and contrast. *Frequency and the Emergence of Linguistic Structure*, ed. by J. Bybee and P. J. Hopper, 137-158. Amsterdam: John Benjamins.

Pierrehumbert, J. B. 2001b. Stochastic Phonology. *GLOT*, 5.1-13.

Pierrehumbert, J. B. 2002. Word-specific phonetics. *Laboratory Phonology 7*, ed. by C. Gussenhoven and N. Warner, 101–139. Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. 2003a. Probabilistic Phonology: Discrimination and robustness. *Probabilistic Linguistics*, ed. by R. Bod, J. Hay and S. Jannedy, 177-228. Cambridge (MA): The MIT Press.

Pierrehumbert, J. B. 2003b. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46.115-154.

Pierrehumbert, J. B. 2006a. The statistical basis of an unnatural alternation. *Laboratory Phonology 8, Varieties of phonological competence*, ed. by L. Goldstein, D. H. Whalen and C. T. Best, 81-107. Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. 2006b. The next toolkit. *Journal of Phonetics*, 34.516-530.

Pinker, S. 1998. Words and rules. *Lingua*, 106.219-242.

Pinker, S. and Prasada, S. 1993. Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, 1.1-56.

Pinker, S. and Prince, A. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 29.73-193.

Pinker, S. and Prince, A. 1994. Regular and irregular morphology and the psychological status of rules of grammar. *The Reality of Linguistic Rules*, ed. by S. D. Lima, R. L. Corrigan and G. K. Iverson, 321-352. Amsterdam: John Benjamins.

Piroth, H. G. and Janker, P. M. 2004. Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics*, 32.81-109.

Pitt, M. A. and Johnson, K. 2003. Using pronunciation data as a starting point in modeling word recognition. Paper presented at paper presented at the15th International Congress of Phonetic Sciences (ICPhS XV).

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S. and Raymond, W. 2003. The ViC Corpus of Conversational Speech. Paper presented at IEEE.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S. and Raymond, W. 2005. The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication*, 45.89-95.

Pitt, M. A., Dilley, L. C., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. 2007. *Buckeye Corpus of Conversational Speech (Final release)*. Columbus (OH): Ohio State University.

Pluymaekers, M., Ernestus, M. and Baayen, R. H. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118.2561-2569.

Pollack, I. and Pickett, J. M. 1963. The intelligibility of excerpts from conversation. *Language and Speech*, 6.165-171.

Port, R. and O'Dell, M. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13.455–471.

Port, R., Mitleb, F. M. and O'Dell, M. 1981. Neutralization of obstruent voicing in German is incomplete. *Journal of the Acoustical Society of America*, 70.S10.

Prince, A. and Smolensky, P. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick: Rutgers Center for Cognitive Science.

Pulleyblank, D. 1988. Underspecification, the feature hierarchy and Tiv vowels. *Phonology*, 5.299-326.

Ranbom, L. J. and Connine, C. M. 2007. Lexical representation of phonological variation in s poken word recognition. *Journal of Memory and Language*, 57.273-298.

Raymond, W. D., Dautricourt, R. and Hume, E. 2006. Word internal /t,d/ deletion in spontaneous speech: Modelling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18.55-97.

Recasens, D. 2004. The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). *Journal of Phonetics*, 32.435-453.

Reetz, H. 1998. *Automatic Speech Recognition with Features*, Universität des Saarlandes: Habilitationsschrift.

Reetz, H. 1999a. Converting speech signals to phonological features. Paper presented at 14th International Congress of Phonetic Sciences (ICPhS XIV), San Francisco.

Reetz, H. 1999b. *Artikulatorische und Akustische Phoneitk.* Trier: Wissenschaftlicher Verlag Trier.

Reetz, H. 2000. Underspecified phonological features for lexical access. *Phonus: Reports in Phonetics*, 5.161-173.

Reetz, H. and Kleinmann, A. 2003. Multi-subject hardware for experiment control and precise reaction time measurement. Paper presented at 15th International Congress of Phonetic Sciences (ICPhS XV), Barcelona.

Reetz, H. and Jongman, A. 2009. *Phonetics - Transcription, Production, Acoustics, and Perception*. Chichester (UK): Wiley-Blackwell.

Rehor, C. 1996. Phonetische Realisierung von Funktionswörtern im Deutschen. *Sound Patterns in Spontaneous Speech*, ed. by K. J. Kohler, C. Rehor and A. P. Simpson, 1-114. Kiel.

Rehor, C. and Pätzold, M. 1996. The phonetic realization of function words in German Spontaneous speech. *Sound Patterns of Connected Speech - Description, Models And Explanation*, ed. by A. S. Simpson and M. Pätzold, 5-11. Kiel: IPDS.

Rodgers, J. 1999. Three influences on glottalization in read and sponataneous German. *Phrase-Level Phonetics and Phonology of German*, ed. by K. J. Kohler, 177-284. Kiel: IPDS.

Samuel, A. G. 1996. Phoneme restoration. *Language and Cognitive Processes*, 11.647-653.

SAS. 2002. *JMP*. Cary (NC): SAS Institute. Version 5.1.0.2.

Scharinger, M. 2006. *The representation of Vocalic Features in Vowel Alternations: Phonological, Morphological and Computational Aspects*, Linguistics Department, University of Konstanz: Ph D. Thesis.

Schiering, R. 2005. Flektierte Präpositionen im Deutschen? Neue Evidenz aus dem Ruhrgebiet. *Zeitschrift für Dialektologie und Linguistik*, 72.52-79.

Schweinberger, S. R. 2001. Human brain potential correlates of voice priming and voice recognition. *Neuropsychologia*, 39.921-936.

Segui, J. and Meunier, F. 1999. Morphological priming effect: The role of surface frequency. *Brain and Language*, 68.54-60.

Selkirk, E. O. 1982. The syllable. *The structure of phonological representations (Part 2)*, ed. by H. van der Hulst and N. Smith. Dordrecht: Foris Publications.

Selkirk, E. O. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge (MA): The MIT Press.

Sheldon, S., Pichora-Fuller, M. K. and Schneider, B. A. 2008. Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America*, 123.489–499.

Shokey, L. 2003. *Sound Patterns of Spoken English*. Cambridge (UK): Blackwell.

Shriberg, E. E. 1999. Phonetic consequences of speech disfluency. Paper presented at The 14th International Congress of Phonetic Sciences (ICPhS XIV), San Francisco.

Simpson, A. P. 1998. *Phonetische Datenbanken des Deutschen in der Empirischen Sprachforschung und der Phonologischen Theoriebildung*. Kiel: IPDS.

Sloman, S. A., Hayman, C. A. G., Ohta, N., Law, J. and Tulving, E. 1988. Forgetting in primed fragment completion. Journal of Experimental Psychology: Learning, *Memory and Cognition*, 14.223-239.

Slowiaczek, L. M. and Dinnsen, D. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, 13.325–341.

Smolka, E., Zwitserlood, P. and Rösler, F. 2007. Stem access in regular and irregular inflection: Evidence from German participles. *Journal of Memory and Language*, 57.325-247.

Snoeren, N. D., Hallé, P. A. and Segui, J. 2006. A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics*, 34.241-268.

Snoeren, N. D., Seguí, J. and Hallé, P. A. 2008. On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. *Cognition*, 108.512-521.

Spinelli, E., McQueen, J. M. and Cutler, A. 2003. Processing resyllabified words in French. *Journal of Memory and Language*, 48.233-254.

Squire, L. R. 1986. Mechanisms of memory. *Science*, 232.1612-1619.

Stampe, D. 1979. *A Dissertation on Natural Phonology*. Bloomington (Indiana): Indiana University Linguistics Club.

Steriade, D. 1995. Underspecification and markedness. *The Handbook of Phonological Theory*, ed. by J. Goldsmith, 114-174. Oxford: Blackwell.

Steriade, D. and Fougeron, C. 1996. Articulatory characteristics of French consonants are maintained after the loss of Schwa. *Journal of the Acoustical Society of America*, 100.2690.

Stevens, K. N. 1998. *Acoustic Phonetics*. Cambridge (MA): The MIT Press.

Sumner, M. and Samuel, A. G. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52.322-338.

Sweet, H. 1877. *A Handbook of Phonetics*. Oxford (UK): Clarendon Press.

Tabain, M. and Perrier, P. 2007. An articulatory and acoustic study of /u/ in preboundary position in French: The interaction of compensatory articulation, neutralization avoidance and featural enhancement. *Journal of Phonetics*, 35.135-161.

Tabossi, P. 1996. Cross-Modal semantic priming. *Language and Cognitive Processes*, 11.569-576.

Tenpenny, P. L. 1995. Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, 2.339-363.

Trudgill, P. 1974. *The Social Differentiation of English in Norwich*. Cambridge (UK): Cambridge University Press.

Tucker, B. V. 2007. *Spoken Word Recognition of the Reduced American English Flap*, Department of Linguistics, University of Arizona: Ph D. Thesis.

Tucker, B. V. and Warner, N. 2007. Inhibition of processing due to reduction of the American English flap. Paper presented at The 16th Internation Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken.

Turk, A. E. and White, L. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27.171-206.

Turk, A. E. and Shattuck-Hufnagel, S. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35.445-472.

Umeda, N. 1991. Multimode database and its preliminary results. *Journal of the Acoustical Society of America*, 89.2010.

Van Bael, C., Van Den Heuvel, H. and Strik, H. 2007. Validation of phonetic transcriptions in the context of automatic speech recognition. *Language Resources & Evaluation*, 41.129-146.

Van Son, R. J. J. H., Binnenpoorte, D., Van Den Heuvel, H. and Pols, L. C. W. 2001. The IFA Corpus: A phonemically segmented Dutch "open source" speech database. Paper presented at Eurospeech 2001, Aalborg, Denmark.

Vater, H. (ed.) 1979. *Phonologische Probleme des Deutschen*. Tübingen: Gunter Narr Verlag.

Ventura, P., Kolinsky, R., Brito-Mendes, C. and Morais, J. 2001. Mental representations of the syllable internal structure are influenced by orthography. *Language and Cognitive Processes*, 16.393-418.

Vitevitch, M. S. and Luce, P. A. 1998. When words compete: Levels of processing in spoken word perception. *Psychological Science*, 9.325-329.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B. and Auer, E. T. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, 68.306-311.

Wade, T. 2007. Implicit rate and speaker normalization in a context-rich normalization in a context-rich phonetic exemplar model. Paper presented at The 16th International Congress of Phonetic Sciences (ICHPS XVI), Saarbrücken.

Wade, T. and Möbius, B. 2008. Detailed Phonetic Memory for Multi-Word and Part-Word Sequences. Paper presented at Labphon 11, Wellington (NZ).

Warner, N. 2002. The phonology of epenthetic stops: implications for the phonetics–phonology interface in optimality theory. *Linguistics*, 40.1-27.

Warner, N. and Weber, A. 2001. Perception of epenthetic stops. *Journal of Phonetics*, 29.53-87.

Warner, N., Jongman, A., Sereno, J. and Kemps, R. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics*, 32.251-276.

Warner, N., Good, E., Jongman, A. and Sereno, J. 2006. Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics*, 34.285-293.

Warren, R. M. and Obusek, C. J. 1971. Speech perception and phonemic restorations. *Perception & Psychophysics*, 9.358-362.

Wassink, A. B., Wright, R. A. and Franklin, A. D. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. J*ournal of Phonetics*, 35.363-379.

Wedel, A. B. 2006. Exemplar models, evolution and language change. *The Linguistic Review*, 23.247-274.

Wesener, T. 1999. The phonetics of function words in German spontaneous speech. *Phrase-Level Phonetics and Phonology of German*, ed. by K. J. Kohler, 327-377. Kiel: IPDS.

Wester, M., Kessens, J. M., Cucchiarini, C. and Strik, H. 2001. Obtaining phonetic transcriptions: A comparison between expert listeners and a continuous speech recognizer. *Language and Speech*, 44.377-403.

Wetterlin, A. 2007. *The Lexical Specification of Norwegian Tonal Word Accents*, Linguistic Department, University of Konstanz: Ph D. Thesis.

Wheeldon, L. and Waksler, R. 2004. Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain And Language*, 90.401-412.

Wiese, R. 1996. *The Phonology of German*. Oxford (UK): Oxford University Press.

Wolfram, W. A. 1967. *A Sociolinguistic Description of Detroit Negro Speech*. Washington, (DC): Center for Applied Linguistics.

Wood, S. A. J. 1996. Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics*, 24.139-164.

Wright, R. 1994. Coda lenition in American English consonants: An EPG study. Paper presented at the 127th Meeting of the Acoustical Society of America, Cambridge, (MAS).

Wurzel, W. U. 1970. *Studien zur Deutschen Lautstruktur*. Berlin: Akademie Verlag.

Ziegler, J., C., Munneaux, M. and Grainger, J. 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, 48.779-793.

Zimmerer, F., Reetz, H. and Lahiri, A. 2008. Harmful reduction? Paper presented at Labroratory Phonology 11, Wellington (NZ).

Zimmerer, F., Reetz, H. and Lahiri, A. 2009. Place assimilation across words in running speech: Corpus analysis and perception. *Journal of the Acoustical Society of America,* 125.2307-2322.

Zue, V. W. and Laferriere, M. 1979. Acoustic study of medial /t,d/ in American English. *Journal of the Acoustical Society of Americ*a, 66.1039-1060.

Zue, V. W., Seneff, S. and Glass, J. 1990. Speech database development at MIT: TIMIT and beyond. *Speech Communication*, 9.351-356.

Zwicky, A. M. 1972. On casual speech. Papers from the 8th Regional Meeting, Chicago Linguistic Society, 607-615.

Curriculum Vitae

Frank Emil Wilhelm Zimmerer

Adresse:	Finkernstrasse 12, 8280 Kreuzlingen, Schweiz
Telefon:	+49(0)176 625 38 621 (mobil)
Email:	zimmerer@em.uni-frankfurt.de
Geburtsdatum:	11.04.1977 in Konstanz
Familienstand:	ledig
Nationalität:	deutsch

Beruflicher Werdegang

10/2006 –	Wissenschaftlicher Angestellter an der Johann Wolfgang Goethe-Universität, Frankfurt am Main, Institut für Phonetik, in einem Projekt des DFG Schwerpunktprogramms (SPP) 1234: «Sprachlautliche Kompetenz: Zwischen Grammatik, Signalverarbeitung und neuronaler Aktivität».
2004 – 2006	Wissenschaftlicher Angestellter an der Universität Konstanz im Sonderforschungsbereich 471: «Variation und Entwicklung im Lexikon», gefördert von der DFG.
1997 – 2003	Magisterstudium: Politikwissenschaft und theoretische Sprachwissenschaft; Thema der Magisterarbeit (Politikwissenschaft): «Angst vor dem Absturz? Staatliche Subventionen für Airlines nach dem 11. September 2001».
2000 – 2001	Integriertes Auslandsstudium (IAS) als Graduate Student (Politikwissenschaft) an der York University, in Toronto, Kanada, gefördert vom DAAD.
1998 – 2003	Wissenschaftliche Hilfskraft, Lehrstuhl Prof. Aditi Lahiri (Universität Konstanz).
1987 – 1996	Gymnasium der Geschwister-Scholl-Schule in Konstanz (Abschluss: Abitur).
1983 – 1987	Grundschule Wollmatingen (Konstanz).

Frankfurt am Main, den 16.12.2008

Hiermit erkläre ich, dass ich die Dissertation mit dem Titel *Reduction in Natural Speech* selbständig verfasst und nur die in der Dissertation angegebenen Hilfsmittel in Anspruch genommen habe, sowie die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, durch Angabe der Quellen kenntlich gemacht habe.

Frankfurt am Main, den 16.12.2008

Frank Zimmerer