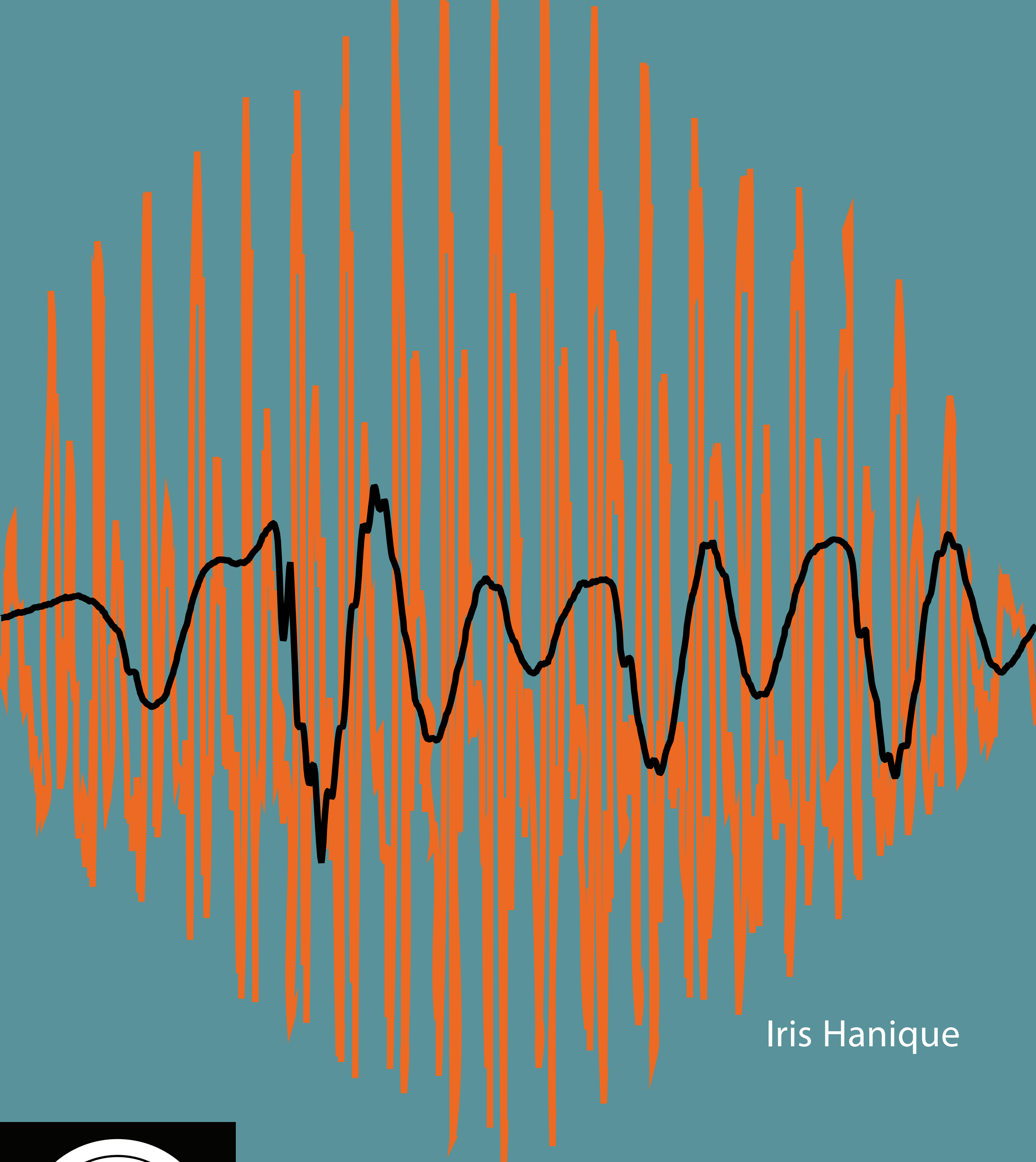

Iris Hanique

Mental representation and processing of reduced words in casual speech

Max Planck Institute for Psycholinguistics

Series

Mental representation and processing of reduced words in casual speech

ISBN: 978-94-6191-903-8
Printed and bound by Ipskamp Drukkers, Nijmegen

Mental representation and processing of reduced words in casual speech

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 13 november 2013
om 16.30 uur precies

door

Iris Arnolda Maria Hanique

geboren op 18 augustus 1985
te Eindhoven

Promotoren:
Prof. dr. L. Boves
Prof. dr. A. Cutler
Prof. dr. M. Ernestus

Manuscriptcommissie:

Prof. dr. A. van den Bosch	(voorzitter)
Dr. R.J.J.H. van Son	(Nederlands Kanker Instituut / Universiteit van Amsterdam)
Prof. dr. N. Warner	(Universiteit van Arizona, VS)

Contents

Introduction

Chapter 1

Everyday conversational speech is characterized by extensive variation in the pronunciations of words. One type of variation is due to the acoustic reduction of words: words are often not produced as their full forms with all segments carefully pronounced, but as reduced forms with altered, shortened, or even absent segments. For instance, the English word *apparently* has the full form /əpʰɛ˞rəntli/, but may be pronounced as [pʰɛ˞ɾi] or [pʰɛ˞] (Johnson, 2004). Similarly, in Dutch, the word *natuurlijk* 'of course' may be pronounced as for instance [ntyrlək], [tylək], or [tyk] instead of /natyrlək/ (Ernestus, 2000). These reduction phenomena occur frequently: for conversational American English, Johnson (2004) showed that in six percent of all content words at least one syllable is absent and that over sixty percent of all words deviate from their full forms. In the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000), high reduction rates have also been found, that is, over forty percent of words are not produced in their full form and almost twenty percent of words are missing a syllable (Schuppler, Ernestus, Scharenborg, & Boves, 2011). Already in the eighties, Mehta & Cutler (1988) showed that the comprehension of conversational speech makes very different demands than the comprehension of carefully articulated, read aloud speech. This dissertation investigates how speakers and listeners process acoustic reductions in casual Dutch. The studies in this dissertation are based on and will help improve models of speech production and speech comprehension. In particular, they address the question how words are represented in the mental lexicon. Furthermore, these studies examine the processing of reduction phenomena by means of different quantitative and experimental methods.

Psycholinguistic models

Psycholinguistic models range on a continuum from purely abstractionist models to entirely exemplar-based models, and differ mostly in their assumptions about the number of lexical representations per word and the amount of detailed information stored in the mental lexicon.

Abstractionist models (e.g., Chomsky & Halle, 1968; Levelt, Roelofs, & Meyer, 1999; Pinker, 1991) assume that the pronunciations of words are represented in the mental lexicon as strings of abstract phonological units, for instance, as strings of phonemes. The far end of the continuum is represented by abstractionist models assuming that for each word the mental lexicon stores exactly one pronunciation. Hence, the word *apparently* has one phonological representation, although it may be pronounced in various ways. A pronunciation that differs from the stored representation results from the application of general phonetic and phonological rules to this stored representation. During speech comprehension, these rules are reversed to map a reduced pronunciation onto the stored abstract representation. Rules may state that a given segment may be adapted or deleted in a certain phonological context. For instance, there may be a rule stating that schwa might be deleted in the Dutch word /xəwon/ *gewoon* 'just'.

At a less extreme point on the continuum, abstractionist models assume that the mental lexicon stores multiple pronunciation variants for each word. According to these models, each pronunciation variant is recognized via its own abstract representation. The production of a word with absent segments may result from two types of processes. On the one hand, speakers may select a word's full form and reduce it during articulation, which results in shortened and, in extreme cases, absent segments (Browman & Goldstein, 1990). Importantly, the shortening and absence of a segment then result from the same gradient process. On the other hand, speakers may retrieve a pronunciation variant with missing segments from the mental lexicon. The missing segments are then always completely (categorically) absent.

Exemplar-based models (e.g., Bybee, 2001, p. 35; Goldinger, 1998) assume the storage of all occurrences of a word that a language user has ever pronounced or heard. Most or all of these occurrences are stored in the form of exemplars, which are acoustically detailed representations. Exemplars representing the same word together form one cloud. The pronunciation of a word results from the retrieval of one or the average of multiple exemplars that represent a certain pronunciation variant. The recognition of a word occurs by mapping a pronunciation onto a cloud of exemplars.

Abstractionist and exemplar-based models are combined in hybrid models, located in the middle of the continuum (e.g., Goldinger, 2007; Hawkins, 2003; McLennan, Luce, & Charles-Luce, 2003; Pierrehumbert, 2002). They combine the advantages of abstractionist and exemplar-based models by assuming that words are stored with both abstract representations and exemplars. The various versions of hybrid models differ in the exact roles of the two types of representations during production and recognition, and how these representations interact. As hybrid models are currently accepted by many researchers and as the literature provides evidence for both abstract representations and exemplars, this dissertation is based on this type of model.

Evidence for multiple representations

Several studies have found support for the assumption that at least some words are represented with more than one pronunciation variant in the mental lexicon. Two studies focusing on speech comprehension report that a pronunciation variant is recognized more accurately and quickly if this variant occurs more often. Pitt, Dilley, & Tat (2011) showed this for the comprehension of English words with a word-medial /t/ that was produced as [t], [ɾ], or [ʔ], or was absent. Ranbom & Connine (2007) showed this for the comprehension of American English words with /nt/ that are pronounced with a nasal flap (e.g., [dʒɛɾ̃l] /dʒɛntl/ *gentle*). Since the variant frequencies play a role in speech comprehension, they must be represented in the mental lexicon. The authors of the two papers assume that the frequencies are stored with the full and reduced pronunciation variants themselves. An alternative assumption is that the variant frequencies specify how often each word's full form is subject to (a series of) reduction processes. The mental lexicon then specifies how often the full form occurs and how often all its variants occur. This latter assumption is experimentally indistinguishable from the first assumption.

Research on speech production has also provided data supporting the assumption that pronunciation variants are stored in the mental lexicon. Bürki, Ernestus, & Frauenfelder (2010) showed that the frequency of a pronunciation variant not only affects comprehension but also production speed. This study was based on French words with word-medial schwa that participants produced in isolation or that were preceded by possessive determiners, with or without schwa (e.g., [ʃəmiz] versus [ʃmiz] *chemise* 'shirt'). A number of studies have adopted a different method. The reason is that a role for variant frequency in speech production can only unambiguously be demonstrated if words are produced in isolation or very short clauses (as in Bürki et al., 2010), but most reduced pronunciation variants only occur in the middle of longer sentences. These studies argue that if a pronunciation variant is stored in the mental lexicon, the absence of a segment is not necessarily the result of extreme shortening. As a consequence, a variant with an absent segment may occur under conditions that do not favor this segment's shortening. Differences in conditions have been found on the basis of words extracted from corpora of spontaneous speech, for example, the absence of French word-internal schwa, as in [ʃmiz] /ʃəmiz/ *chemise* 'shirt' (Bürki, Ernestus, Gendrot, Fougeron, & Frauenfelder, 2011), and the absence of /e/ in the word combination [stɛ] /setɛ/ *c'était* 'it was' (Torreira & Ernestus, 2011).

There is also clear counter-evidence for the assumption that reduced pronunciation variants are stored in the lexicon: extremely reduced words are often difficult to recognize without any context (Ernestus, Baayen, & Schreuder, 2002). However, this observation can also be explained by the assumption that lexical representations include information about the phonological or semantic context in which the variant occurs (Hawkins, 2003). Representations can only be accessed if the variant occurs in the specified context. If it does not, for instance because it is artificially presented out of context, the variant cannot be processed via this representation.

With the exception of Pitt et al. (2011), the studies described above investigated reduction of highly frequent word combinations (Torreira & Ernestus, 2011) and phenomena that are frequent in casual, connected speech as well as in formal speech or in isolated words (Bürki et al., 2010; Bürki, Ernestus, et al., 2011; Ranbom & Connine, 2007). There are, however, many reduction processes that affect many different word types and that are restricted to casual, connected speech. The question arises whether the pronunciation variants resulting from these reduction processes are stored in the mental lexicon as well. If so, the lexicon contains many more pronunciation representations than assumed by most abstractionist models. Pitt et al. (2011) investigated this question by examining how listeners recognize words with medial /t/ in American English. He presented listeners with these words out of their natural context. In Chapters 2 and 3 of this dissertation, we further address this question by studying words within their context. These chapters focus on the realization of schwa in prefixes and word-final /t/ in Dutch past participles within context (e.g., in /xəlœystərt/ *geluisterd* 'listened'). The reduction of these segments occurs in many different word types but is restricted to casual, connected speech. Like Bürki, Ernestus, et al. (2011) and Torreira & Ernestus (2011), in both chapters, we investigate the production of reduced segments by comparing the conditions of segment shortening and segment absence.

Evidence for exemplars

Several researchers have argued that the mental lexicon stores many exemplars for each word. Johnson (1997), for instance, argued that with exemplars we can account for how listeners adapt to different speakers without assuming a process of speaker normalization. A few years later, Johnson (2004) argued that models with abstract representations or with a single representation for each word cannot account for pronunciations with gradient reductions, and that models therefore have to represent words as clouds of exemplars. Finally, Port (2007) argued that acoustically detailed lexical representations are necessary to account for, among other things, how the pronunciations of some words gradually change over time.

Several series of experiments have investigated how exemplars may affect speech processing. Goldinger (1998), for instance, performed shadowing experiments, showing that listeners often mimic the acoustic detail of the stimulus, which implies that participants must have stored this acoustic detail during the comprehension process.

Furthermore, several researchers studied the role of exemplars in speech comprehension by means of long-term identity priming experiments. In these experiments, the first occurrence of a word (the prime) was acoustically identical to or different from the second occurrence (the target): primes and targets were produced by the same speaker or different speakers, represented the same or a different realization of a certain segment (e.g., intervocalic /t,d/ produced as [t,d] or as a flap), or were produced at the same or a different speech rate (e.g., Bradlow, Nygaard, & Pisoni, 1999; Craik & Kirsner, 1974; Goh, 2005; Goldinger, 1996; Mattys & Liss, 2008; McLennan et al., 2003; McLennan & Luce, 2005; Palmeri, Goldinger, & Pisoni, 1993). If a prime creates an exemplar, priming effects should be larger if a prime and target are highly similar. In contrast, if both occurrences of a word are processed via the same single abstract representation, the size of the priming effect should be independent of the detailed acoustic similarity between the prime and target. Most experiments demonstrated that priming effects are larger if primes and targets are identical, suggesting that acoustic details are stored in the form of exemplars.

Interestingly, some of these priming experiments suggested that exemplar-based priming effects only arise under certain conditions, but not consistently. For example, McLennan et al. (2003) found exemplar effects but only if processing was relatively fast, whereas McLennan & Luce (2005) and Mattys & Liss (2008) found effects only if processing was relatively slow. Similarly, Goldinger (1996) reported that if the participants had to indicate for each target whether it had also occurred before, exemplar effects were present after a day if they had heard stimuli produced by two or ten voices but not by six voices. This raises the question about how robust exemplar effects are.

In Chapter 4, we investigate this question further. It focuses on schwa in Dutch word-initial syllables, which is often reduced in casual speech (see Chapter 2). Chapter 4 presents four long-term priming experiments, in which we study the robustness of exemplars in three ways. First, we examine whether exemplar effects arise for different speakers, by using two very different speakers. Second, we investigate whether listeners also use exemplars if, in contrast to the priming studies mentioned above, it is less clear for them that words are repeated. Finally, we examine whether exemplar effects can also be found if listeners hear not one but two types of variation (i.e., pronunciation variation and speaker voice).

Morphology

A large proportion of reduced words are morphologically complex, as this type of word often contains unstressed syllables that can be easily reduced. A long-standing debate on complex words concerns the question of whether they are stored in the mental lexicon, which affects the production and comprehension process. According to some theories (e.g., Chomsky & Halle, 1968; Pinker, 1991; Taft & Ardasinski, 2006), the mental lexicon only stores irregular complex words and the processing of regular ones involves combining their morphemes. Other theories (e.g., Bybee, 2001; Skousen, 1989) assume that complex words are stored. These words are then processed like simple words and the two word types should show similar behavior. Finally, several theories combine the two ways of storing and processing words (e.g., Levelt et al., 1999; Schreuder & Baayen, 1995).

Many studies have investigated whether morphological structure plays a role in language processing. If so, this would support the assumption that morphologically complex words are processed via their morphemes. Most studies focused on reading, but several investigated speech processing (e.g., Bybee, 1995; Cohen-Goldberg, 2013; Marslen-Wilson & Zhou, 1999; Seidenberg & Gonnerman, 2000; Taft, 2004). So far, findings are mixed. The study by Cohen-Goldberg (2013), for example, reported an effect in speech production. This researcher showed that there is competition between phones in English words when they are read aloud in isolation, but that this competition appears smaller if the phones are separated by a morpheme boundary. The author therefore argued that morphologically complex words are processed via their morphemes, at least during reading.

If morphology is important in speech processing, it should also influence acoustic reduction. In the literature, a number of studies are found that support this idea, as they, for example, indicate that segments forming an affix on their own tend to be less reduced than when the same segments are part of a longer morpheme. This has been reported for several word-final segments in English, namely for /t/ and /d/, as in *rapped* versus *rapt* (Bybee, 2002; Guy, 1994; Labov, 2004; Losiewicz, 1992; Sugahara & Turk, 2009), for /s/, as in *tacks* versus *tax* (Schwarzlose & Bradlow, 2001; Sugahara & Turk, 2009), and for /ɪŋ,ɪn/, as in *raising* versus *raisin* (Sugahara & Turk, 2009). Additionally, this has been found for Dutch word-final /t/, as in /bit/ *biedt* 'offers' versus /bit/ *bied* 'offer' (Schuppler, Van Dommelen, Koreman, & Ernestus, 2012).

Chapter 5 of this dissertation presents a review of articles that have reported effects of a word's morphological structure on reduction. We discuss and re-analyze data (amongst others, from Losiewicz, 1992, and Schuppler et al., 2012), and address the question whether the effects reported can also be explained by the assumption that segments that are more important for the identification of a word (i.e., that have a higher information load) are less reduced, independently of their morphological status.

Individual differences

So far, psycholinguistic theories have been designed to account for speech processing in the ideal or average language user and do not account for differences among users. Nevertheless, there is evidence that speakers differ in how often and to what extent they reduce words in casual speech.

First, several studies on socio-linguistics have shown that social groups may differ in their reductions. In general, younger speakers reduce more than older speakers. This has been reported, for instance, for the absence of word-final /t/ and /d/ in American English (Guy, 1980), the absence of segments in Dutch (Strik, Van Doremalen, & Cucciarini, 2008), and the production of /t/ as [ʃ] in Panama Spanish (Cedergren, 1987). Furthermore, men tend to reduce more than women, which has been shown for the absence of word-final /t/ and /d/ (Guy, 1980) and glides (Phillips, 1994) in American English, and for the reduction of suffix *-lijk* in Dutch (as in /moxələk/ *mogelijk* 'possibly'; Keune, Ernestus, Van Hout, & Baayen, 2005). Finally, the Dutch suffix *-lijk* is less often reduced by speakers from Flanders than speakers from the Netherlands, and by highly educated Flemish speakers compared to less highly educated Flemish speakers (Keune et al., 2005).

Second, speakers forming a homogeneous social group may also differ in how they reduce words. However, the evidence for individual differences is scarce, as it only comes from Ernestus (2000, p. 143). In her corpus of spontaneous Dutch, she observed that multiple speakers used the pronunciation [tyk] for /natyrlək/ *natuurlijk* 'of course' in the middle of Intonational Phrases, but only one of the twenty speakers also used it in other positions and even in isolation.

In Chapter 6, we investigate whether individual speakers with a similar social background differ in which words they produce and how they reduce them. We study whether a classifier can distinguish between speakers on the basis of words and pronunciation patterns in spontaneous speech. If so, the speakers have to differ in the words and pronunciations they produce. In addition, we inspect which words and pronunciations are relevant for the classifier and thus display differences between speakers.

Methodology

The studies described in this dissertation use a range of research methods. In Chapters 2 and 6, we investigate the speech production process by analyzing several hours of casual conversations recorded by Ernestus (2000). She placed the speakers in front of a microphone, which often leads to unnatural speech, but since the speakers knew each other very well and talked about everyday topics, they quickly habituated to the situation, forgot the microphone, and spoke very casually. As a result, the speech recorded represents casual, spontaneous conversations, which allows us to study speech as it occurs in everyday life. We analyzed the conversations using two different methods: Chapter 2 presents analyses on the basis of regression models, while Chapter 6 concerns a classification study.

The segments, words, and phrases produced in casual conversations often differ from each other in the speech rate at which they were uttered, in their position in the sentence, in the preceding and following segments, etcetera. Segments, words, or phrases therefore never differ from each other in only one aspect, which makes it difficult to investigate why exactly these units differ from each other in the way they do. To investigate the effect of a factor of interest (e.g., word frequency), corpus studies should preferably be complemented with psycholinguistic experiments that manipulate this one factor and control for all other factors.

It is still not clear how casual speech can easily be elicited in psycholinguistic experiments. Chapter 3 presents a production experiment in which participants repeated sentences that they heard. This shadowing task enabled us to elicit the same words from almost all the participants, and to control for the target words' positions in the sentences and their preceding and following phones. We address the question whether this task provides natural casual speech, by comparing patterns in the speech elicited with patterns in natural casual conversations reported in Chapter 2. Highly similar results would not only show that the speech obtained with the shadowing task is natural; it would also strengthen our conclusions based on the corpus data in Chapter 2.

This dissertation also reports a series of psycholinguistic experiments studying the comprehension of reduced pronunciation variants. Previous studies have investigated the comprehension of reduced pronunciation variants by means of several paradigms, including cross-model priming, lexical decision, implicit semantic priming, and identification tasks, in which words were presented in isolation (Tucker & Warner, 2007; Tucker, 2011; Van de Ven, Tucker, & Ernestus, 2011; Warner, Fountain, & Tucker, 2009). Since reduced pronunciation variants typically occur within sentence context, these experiments do not present these variants in their natural environment, which may affect speech processing. Bard, Shillcock, & Altmann (1988) and Van de Ven, Ernestus, & Schreuder (2012) presented reduced pronunciation variants in their context, but in experiments providing offline measures only (i.e., gating and cloze tests). Finally, Brouwer and colleagues (e.g., Brouwer, Mitterer, & Heuttig, 2012) also presented the words in their natural context, in a visual world paradigm with printed words. These printed words may have activated the full pronunciations of the words, possibly affecting the speech comprehension process. This overview shows that we are still in need of an experimental paradigm that presents reduced pronunciation variants in their natural context, without presenting their orthographic forms and providing online measures.

The comprehension experiments of Chapter 4 focus on a question that has been addressed before for several types of pronunciation variation (e.g., speaker voice or /nt/ produced as a flap; Goldinger, 1996; McLennan et al., 2003). We chose to use the same paradigm as these previous studies, that is, the long-term identity priming paradigm, since this appears to be the best paradigm to investigate this question. Unfortunately, this implied that we presented reduced pronunciation variants out of context.

In this dissertation, we report on the presence versus absence and the durations of segments in a large number of word tokens. We chose not to segment the word tokens by hand, since we would then not be able to study a very large number of word tokens. Moreover, the resulting segmentation may be subjective and inconsistent. Instead, we used an automatic speech recognition (ASR) system to transcribe and segment the word tokens. This ASR system is based on the Hidden Markov Model Toolkit (HTK; Young et al., 2002) and uses four types of input: the speech signal itself, orthographic transcriptions of the speech, a pronunciation lexicon that, for each word in the transcriptions (e.g., *natuurlijk* ‘of course’), provides the full phonetic form (/natyrlək/) and possible pronunciation variants (amongst others, /ntylək/ and /tyk/), and 37 three-state monophone acoustic models, with 32 Gaussians per state (Hämäläinen, Gubian, ten Bosch, & Boves, 2009). These phone models were trained on the read speech component from the Spoken Dutch Corpus (Oostdijk, 2002) and were used to determine, for each word in the orthographic transcriptions, which pronunciation variant from the lexicon best matched the speech signal. As we report in Chapters 2, 3, and 6, the resulting transcriptions and word segmentations are as reliable as transcriptions created by human listeners.

Outline

This dissertation presents five studies investigating how speakers and listeners process acoustic reductions in casual Dutch. Chapters 2 and 3 focus on speech production. In these chapters, we investigate whether schwa and word-final /t/ are more likely to be absent under conditions that also favor shortening of these segments. If so, the absence of these segments may result from extreme gradient reduction. If not, the segments may also be absent in the pronunciations stored in the mental lexicon, which would support models assuming that words are stored with multiple pronunciation representations. The research described in Chapter 2 concerns speech from a corpus of casual Dutch and the work reported in Chapter 3 concerns speech obtained from a shadowing experiment. Comparison of the results of these chapters sheds light on the question whether casual speech can be elicited in an experimental setting (see Chapters 3 and 7).

In Chapters 4, 5, and 6, we further investigate the processes underlying the production and comprehension of reduced pronunciation variants. Chapter 4 reports a study on the role of exemplars in speech comprehension. It focuses on schwa reduction in Dutch and presents four experiments with a long-term identity priming paradigm. The work in Chapter 5 concerns the role of morphological structure in speech reduction. It provides a review and re-analyses of data from papers, which argue that morphological structure is relevant. In this chapter, we examine whether the reported effects can also be explained by the hypothesis that segments that are more important for word identification tend to be less reduced. In Chapter 6, we investigate to what extent speakers differ in the words that they use and in how they pronounce these words. This chapter focuses on recordings of casual speech produced by twenty speakers with a similar social background.

In Chapter 7, we summarize and discuss the results of these five studies. We focus on what these studies have taught us about how to investigate the production and comprehension of reduced pronunciation variants. Moreover, we look at the implications of the results for psycholinguistic models of speech production and comprehension.

Processes underlying acoustic reduction: a corpus study

Chapter 2

This chapter is a reformatted version of:
Iris Hanique, Mirjam Ernestus, and Barbara Schuppler (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *Journal of the Acoustical Society of America*, *133(3)*, 1644–1655.

Abstract

This chapter investigates the nature of reduction phenomena in informal speech. It addresses the question whether reduction processes that affect many word types, but only if they occur in connected informal speech, may be categorical in nature. The focus is on reduction of schwa in the prefixes and on word-final /t/ in Dutch past participles. More than 2000 tokens of past participles from the Ernestus Corpus of Spontaneous Dutch and the Spoken Dutch Corpus (both from the interview and read speech component) were transcribed automatically. The results demonstrate that the presence and duration of /t/ are affected by approximately the same phonetic variables, indicating that the absence of /t/ is the extreme result of shortening, and thus results from a gradient reduction process. Also for schwa, the data show that mainly phonetic variables influence its reduction, but its presence is affected by different and more variables than its duration, which suggests that the absence of schwa may result from gradient as well as categorical processes. These conclusions are supported by the distributions of the segments' durations. These findings provide evidence that reduction phenomena which affect many words in informal conversations may also result from categorical reduction processes.

Introduction

In conversational speech, words are often not produced in their full forms, that is in the form that is used in formal situations (e.g., Ernestus, 2000; Johnson, 2004). Segments may be very short, altered (e.g., a voiced plosive may be pronounced as a glide), or even be completely absent. The English word *particular*, for example, has the full form /pətɪkjələ˞/, but in conversational speech may also be produced as /ptɪkə˞/. Similarly, the Dutch word *gewenst* 'wished' may be produced as /xwɛns/ instead of /xəwɛnst/. Recently, researchers have started to investigate the processes underlying this type of pronunciation variation. This chapter contributes to this line of research by studying the processes underlying the two most frequent reduction phenomena in Dutch, which affect many words but only if they occur within a sentence in informal speech.

Browman & Goldstein (1990) hypothesized that all pronunciation variants that typically occur in casual speech result from a gradient increase in overlap and a gradient decrease in the magnitude of articulatory gestures. Gestural overlap may result in blending if two neighboring segments share an articulator. One or both segments are then pronounced incompletely. An example of this blending process is the pronunciation of English *this shop* as /ðɪʃːɒp/, with a long /ʃ/, but without /s/. Gestural overlap may result in the complete hiding of a segment if two neighboring segments are pronounced with different articulators. For instance, the closure of a bilabial stop may hide a preceding /t/ by making the release of the alveolar constriction inaudible.

Browman and Goldstein found support for their hypothesis in several articulatory studies (see Browman & Goldstein, 1992, for an overview). In addition, their hypothesis is supported by several acoustic studies. Davidson (2006) for instance, studied the absence of pretonic schwa in consonant-schwa-consonant sequences in American English (e.g., in the words *tomato* and *support*). She found that schwa is less often absent if the preceding segment is voiced or is a voiceless stop which can only overlap partially with schwa. In addition, if schwa is absent in /#səp/ sequences, /p/ is often aspirated, which suggests that the articulatory gestures for schwa are still present. Furthermore, Davidson (2006) also showed that preceding voiceless fricatives and following /l/s tend to be longer if schwa is absent. All these results indicate that the absence of schwa in the acoustic signal is due to co-articulation with surrounding consonants.

Another acoustic study supporting Browman and Goldstein's hypothesis was conducted by Torreira & Ernestus (2010b). They studied vowel devoicing in French, a reduction phenomenon that is mostly restricted to fast and casual speech. They showed that partial devoicing is affected by the same predictors as complete devoicing (speech rate, manner of articulation of the preceding consonant, distance to the following Accentual Phrase boundary). These results suggest that both partial and complete devoicing result from the same reduction process, which is consequently gradient in nature.

The pronunciation variants resulting from co-articulation may be stored in the mental lexicon. These representations then have to specify the acoustic details of the variants, for instance the exact durations of their segments. They therefore differ from the abstract representations in the form of phonemes, which are assumed by most models of speech production (e.g., Levelt et al., 1999). Since gradient reduction by definition results in a continuum of pronunciation variants, the storage of these variants also implies a complex mental lexicon.

Importantly, several authors assume that the absence of segments may not only result from gradient processes, but also from categorical higher-level processes. One such categorical process is a phonological rule that deletes a segment completely (e.g., Chomsky & Halle, 1968; McCarthy & Prince, 1993). In addition, the mental lexicon may contain representations for multiple pronunciations of a word (e.g., Goldinger, 1998), and the absence of a segment may result from the selection of a variant with missing segments.

Support for the categorical nature of some reduction phenomena has been found in corpus-based and psycholinguistic studies. For instance, Torreira & Ernestus (2011) studied the processes underlying /e/-reduction in the highly frequent word combination *c'était* /setɛ/ 'it was' using the Nijmegen Corpus of Casual French. The duration of /s(e)/ showed a bimodal distribution instead of a unimodal distribution, with one mode containing most tokens with /e/ and the other containing most tokens without /e/. Furthermore, the presence and duration of /e/ were conditioned by different variables: whereas the presence of /e/ was conditioned by speech rate and the distance to the end of the Accentual Phrase, the duration of /e/ was conditioned by position in the Intonational Phrase and the duration of preceding /s/. Given these results, the authors concluded that the absence and shortening of /e/ can result from different processes, and that the absence of /e/ in French *c'était* sometimes results from a categorical process rather than from extreme vowel shortening.

Bürki and colleagues investigated the nature of word-internal schwa deletion in French. This deletion phenomenon is very frequent in many varieties of French and can also occur in formal speech and when words are uttered without context. Bürki, Fougeron, Gendrot, & Frauenfelder (2011) showed that this schwa deletion phenomenon can result from a gradient shortening process, as the duration of schwa varies from very long to very short, like the duration of other vowels. In addition, it can result from categorical processes; Bürki et al. (2010) demonstrated in several production studies that speakers of French produce a pronunciation variant (e.g., *fenêtre*) faster the more frequent it is in comparison with the other variant (e.g., *fnêtre*), which suggests that both variants are stored. In a corpus study, Bürki, Ernestus, et al. (2011) found differences between the sets of variables that condition word-internal schwa duration and its absence. Schwa duration was influenced by voicing of the preceding and following consonants and by the word's predictability given the preceding word, whereas its absence was conditioned by word position in the utterance, the number of consonants in the sequence that results from the absence of schwa, and whether this sequence adheres to the sonority principle. The authors of these last two studies concluded that the absence of schwa is not always the extreme result of schwa shortening, but can also result from a categorical process.

So far, evidence for the categorical absence of segments is restricted to the types of phenomena discussed above. They represent reduction in highly frequent (function) words or word combinations and reduction that also occurs in formal speech and in words uttered in isolation. This raises the question whether reduction processes that are restricted to informal speech and affect many different word types may also be categorical. If so, this has consequences for our view of the speech production process: it would indicate that the mental lexicon contains an immense number of pronunciation variants or that there are many more categorical (i.e., phonological) reduction processes than previously assumed, which are specified for the conditions in which they can be applied.

In this chapter, we investigate this question by studying two reduction processes in Dutch that affect many words, mainly in informal connected speech. Although both processes have been well documented, little is known about their underlying nature.

The first process is the reduction of schwa, which has been studied before in several languages, like English (e.g., Dalby, 1984; Davidson, 2006; Patterson, LoCasto, & Connine, 2003), French (e.g., Bürki, Ernestus, et al., 2011), and Dutch (e.g., Pluymaekers, Ernestus, & Baayen, 2005; Van Bergem, 1994). Schuppler et al. (2011) reported that in Dutch 44.7% of all schwas are absent. Generally, the studies on Dutch have shown that schwa tends to be more reduced in faster speech, in words that occur more frequently, after fricatives and before plosives, and in words in the middle of utterances.

The second process that we will examine is the reduction of /t/, which has been studied, among others, in Dutch (e.g., Booij, 1995; Cho & McQueen, 2005; Mitterer & Ernestus, 2006; Schuppler et al., 2012) and English (e.g., Gahl & Garnsey, 2004; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999). In careful Dutch, /t/ is usually produced with a closure and a burst. In informal Dutch, the closure and burst of /t/ may be weakened, shortened, and even be completely absent (Mitterer & Ernestus, 2006). For this speech register, /t/ has been reported to be absent in 11.9% of all word tokens and in one-third of the word tokens in which it is in final position (Schuppler et al., 2011). Overall, previous research has shown that in Dutch /t/ is more likely to be reduced in more informal speech, in unstressed syllables, in non-final sentence position, in more predictable word pairs, and if preceded by /s/ or followed by a bilabial consonant than by vowels or other consonants. In some highly frequent words, such as *niet*, /t/ can also be absent in more formal speech registers.

By studying these two segments in the same words and with the same methods, we can reveal differences that can be directly ascribed to their phonetic characteristics, to their position in the word, or, more importantly, to the nature (categorical versus gradient) of the underlying processes. Previous studies have investigated the reduction of different segments within one word (e.g., Pluymaekers et al., 2005), but only a few, if any, have addressed the question of whether these segments were reduced as a consequence of the same processes.

Our study focuses on the reduction of schwa and /t/ in Dutch past participles, since both segments occur in regular past participles. Most Dutch regular past participles consist of a verb base and the circumfix /xə/ + /t/ (e.g., /mak/ *maak* 'make': /xə+mak+t/ *gemaakt* 'made'). Past participles of verbs that start with one of the prefixes *be-*, *er-*, *ge-*, *her-*, *ont-*, or *ver-* generally do not have an additional prefix *ge-* (e.g., /bətal/ *betaal* 'pay': /bətal+t/ *betaald* 'paid'). Irregular past participles typically deviate from regular past participles in that they end in /ə(n)/ instead of /t/ (e.g., /lop/ *loop* 'walk': /xə+lop+ə(n)/ *gelopen* 'walked'), or their base shows a vowel change (e.g., /krœyp/ *kruip* 'crawl': /xə+krop+ə(n)/ *gekropen* 'crawled').

This chapter presents two corpus studies. Study 1 investigates the reduction of schwa, focusing on the initial syllables *ge-*, *be-*, and *ver-*, with the full forms /xə/, /bə/, and /vər/, respectively. These initial syllables all start with a consonant and given that the word is a past participle, the following schwa is almost completely predictable, and therefore probably prone to reduction. Importantly, the three syllables occur in many word types of various frequencies, which allow us to examine word frequency effects. In Study 2, we focus on reduction of word-final /t/.

For both schwa and /t/, we examined which variables condition their acoustic presence versus absence and which variables condition their durations. If a segment's absence and duration are conditioned by the same variables, we hypothesize that they result from the same shortening process, which in extreme cases results in the absence of a segment. If its absence is conditioned by different variables to its duration, we then hypothesize that its absence results from a categorical process. We studied variables that are known to play important roles in pronunciation variation, such as speech register (e.g., Van Bael, Van den Heuvel, & Strik, 2004; Van Son & Pols, 1999), the position of a word in the utterance (e.g., Bell, Jurafsky, Fosler-Lussier, Girand, & Gildea, 2003; Cambier-Langeveld, 2000), speech rate (e.g., Dalby, 1984; Davidson, 2006), the characteristics of the surrounding consonants (e.g., Mitterer & Ernestus, 2006; Van Bergem, 1994), and word predictability (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Pluymaekers et al., 2005). In order to be able to study register, we based our studies on conversational speech from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000) and interview and read speech from the Spoken Dutch Corpus (Oostdijk, 2002).

Following Bürki, Fougeron, et al. (2011) and Torreira & Ernestus (2011), we also examined the distribution of the duration of the segment. If the absence of a segment only results from extreme gradient shortening, the duration is expected to show a unimodal distribution. In contrast, if the absence of that segment can result from both a gradient and a categorical process, this distribution is expected to be bimodal with one peak at zero ms (since the absence of the segment can result from categorical processes and extreme gradient shortening) and one peak around the segment's average duration.

In summary, the present corpus study extends the research on pronunciation variation by investigating the nature of the processes underlying two highly frequent phenomena (i.e., reduction of schwa and /t/ in Dutch) that affect many word types, but that hardly occur in formal speech or words uttered in isolation. By comparing the variables that affect a segment's presence and duration, and by inspecting the distribution of these segments' durations, we investigate whether the absence of a segment results only from extreme shortening or can also result from a categorical deletion process.

Study 1: Schwa reduction

Method

Materials

We extracted past participles from three subcorpora with different speech registers. The first corpus is the Ernestus Corpus of Spontaneous Dutch (ECSD; Ernestus, 2000), which consists of 15 hours of casual dialogues between ten pairs of speakers. All 20 speakers were native speakers of standard Dutch, born and raised in the west of the Netherlands, aged between 21 and 55, and they all held academic degrees. The second and third subcorpora are components of the Spoken Dutch Corpus (CGN; Oostdijk, 2002): we used those parts of the 43 hours of interviews with 681 speakers and the 64 hours of lively read passages of the library of the blind produced by 324 speakers, that were phonetically transcribed (see below) at the moment of this study.

We extracted all past participles produced without background noise or overlapping speech from other speakers. Because *ge-* and *be-* consist of a consonant plus schwa, we only extracted tokens of *ver-* where /r/ was absent (which is the most common pronunciation of *ver-* and these tokens therefore form the majority of tokens). Furthermore, tokens directly preceded or followed by hesitations or with an extremely short (i.e., shorter than 15 ms) or long (i.e., longer than 478 ms) consonant preceding schwa were discarded. Since it is difficult to determine the presence and duration of schwa if it is followed by a vowel, we only used words for our study whose second syllable start with a consonant. All bi-syllabic and most longer past participles were stressed on the second syllable (e.g., /xəˈdɑnst/ *gedanst* 'danced', /bəˈvɛstəxt/ *bevestigd* 'confirmed', /xəˈwɑndəlt/ *gewandeld* 'walked'). Table 2.1 presents an overview of the final data set of schwa in word-initial syllables. The number of tokens per word type ranged from 1 to 151. Compared to the frequent words and word combinations discussed in the literature on the processes underlying reduction, the words in our data set are all of a low frequency. The most frequent word in our study was *geweest* 'been', which occurs 54 times per 100000 word tokens (based on counts from the data used for this study), whereas the highly frequent bigram *c'était* 'it was' studied by Torreira & Ernestus (2011) occurs 280 times per

Table 2.1: Number of tokens and types (between parentheses) in the data set for Study 1, divided by the three types of initial syllable and the three speech registers, namely conversational speech (CS), interviews (IN), and read speech (RS).

	Total	CS	IN	RS
Total	2150 (562)	738 (195)	539 (205)	873 (393)
ge-	1695 (387)	624 (145)	426 (145)	645 (268)
be-	242 (86)	57 (22)	67 (33)	118 (63)
ver-	213 (88)	57 (28)	46 (27)	110 (62)

100000 bigrams (based on the Nijmegen Corpus of Casual French). Interestingly, the number of word types in read speech is twice as high as the number in conversational and interview speech, indicating a difference in the type-token ratio.

Measurements

We determined the presence versus absence of schwa and its duration on the basis of automatically generated broad phonetic transcriptions. Automatic transcriptions are more consistent than human transcriptions and can be easily generated for large quantities of speech. For all materials used, we followed the forced alignment procedure described by Schuppler et al. (2011). For each word token in the orthographic transcriptions an automatic speech recognition (ASR) system retrieved the full phonetic form and possible pronunciation variants from a pronunciation lexicon. For instance, for the word *gemaakt* 'made' it retrieved the full form /xəmakt/ and the variants /xmakt/, /xəmak/, and /xmak/. The possible pronunciation variants resulted from the application of 32 reduction rules to the full pronunciations of the words. These rules were based on earlier observations of reduced variants in casual Dutch (e.g., Ernestus, 2000), and deleted, among others, schwa in the syllables *ge-*, *be-*, or *ver-*, and word-final /t/. From all retrieved pronunciation variants, the ASR system HTK (Hidden Marcov Model Toolkit; Young et al., 2002) selected the variant that best matched the speech signal, using 37 monophone models. These 32-Gaussian tri-state models (Hämäläinen et al., 2009) had been trained on the read speech component of the CGN. Because the acoustic models consisted of at least three emitting states and the frame shift was 5 ms, segments were assigned a minimum duration of 15 ms (even if the actual duration was shorter). On average, schwa was assigned a duration of 48 ms (range 15 - 225 ms) and /t/ a duration of 85 ms (range 15 - 355 ms).

In order to validate the automatic transcriptions, two of the authors manually transcribed 148 schwas in the ECSD data set. We chose ECSD for the validation, because its speech style is the hardest to transcribe. The manual transcriptions were based on the audio and the waveform of the target word and a few surrounding words. The two transcribers first decided on transcription criteria and thereafter transcribed schwa and the preceding and following segments independently from each other using the speech analysis software package *Praat* (Boersma, 2001). They did so without being aware of the results of the ASR system.

As the schwas that were transcribed belonged to the syllables *ge-*, *be-*, or *ver-*, they were preceded by /x/, /b/, /p/, /v/, or /f/. The transcribers placed the boundary between a fricative and schwa at the offset of frication noise. The boundary in a /bə/ sequence was located directly after the burst of the plosive. Schwa could be followed by different kinds of consonants. If it was a fricative, the boundary was formed by the onset of frication noise. If it was a plosive, the boundary was located at the onset of the closure. Finally, if schwa was followed by a sonorant, the transcribers defined the transition as sudden changes in the audio or qualitative changes in the shape or complexity of the periodicity of the waveform.

Table 2.2 compares the presence of schwa in the word-initial syllables as annotated by the ASR system and the two transcribers. We found an agreement of 82.4% and a Cohen's kappa of 0.64 between the two human transcribers, and agreements of 75.7% and 77.0% and kappas of 0.51 and 0.53 between the ASR system and each of the human transcribers. A regression model with the dependent variable *agreement on the presence of schwa*, the independent variable *transcriber pair* and the random variable *word type* showed no significant effects (A-H1 versus A-H2: β = 0.01, *t* = 0.30, *n.s.*; A-H1 versus H1-H2: β = 0.07, *t* = 1.48, *n.s.*; A-H2 versus H1-H2: β = 0.05, *t* = 1.19, *n.s.*). This indicates that the agreements between the ASR and each human transcriber do not differ substantially from the agreement between the two human transcribers. These agreements are also similar to other transcriber agreements obtained for informal speech. For example, Kipp, Wesenick, & Schiel (1997) reported agreements between three human-made transcriptions of spontaneous German of at least 78.8%, and Pitt, Johnson, Hume, Kiesling, & Raymond (2005) presented an

Table 2.2: Comparison of the automatically generated (A) and the human-made (H1 and H2) transcriptions.

	Presence of schwa			Duration of schwa		
Comparison	N	Agreement	κ	N	Mean difference	95% confidence interval
A - H1	148	75.7%	0.51	62	18.0 ms	10.6 - 25.4 ms
A - H2	148	77.0%	0.53	63	8.5 ms	1.5 - 15.5 ms
H1 - H2	148	82.4%	0.64	69	8.2 ms	5.3 - 11.1 ms

overall agreement of 80.3% between human-made transcriptions of conversational American English.

We also compared the durations of those schwas that were transcribed as present by the human transcribers or the ASR system. For this analysis, three outliers with durations longer than 210 ms were excluded (as they were in the duration analysis described below). Table 2.2 presents the results (including mean duration differences) without these outliers. A regression model with *transcriber* as the independent factor and *word type* as the random factor indicated that the durations assigned by the different transcribers differed significantly (1-A: β = 0.02, t = 6.23, $p < 0.0001$; 1-2: β = 0.01, t = 2.04, $p < 0.05$; 2-A: β = 0.02, t = 4.22, $p < 0.0001$). Those of the ASR system are the longest, which is unsurprising as these are minimally 15 ms. Those of Transcriber 2 are longer than those of Transcriber 1. Further comparison of the durations assigned by the ASR system and the human transcribers showed that the differences were mainly caused by three outliers (5% of the data points). Removal of these data points resulted in mean differences of 12.8 ms and 3.7 ms between the ASR system and Transcriber 1 and 2, respectively. These average differences are similar to those obtained in other studies. For example, Wesenick & Kipp (1996) reported that 96% of the segment boundaries determined by three human transcribers for 64 read sentences of German differ by less than 20 ms. Similarly, Raymond et al. (2002) presented an average alignment difference of 16.4 ms between human transcribers of the Buckeye corpus.

This evaluation shows that our automatically obtained transcriptions are comparable to human-made transcriptions. As it is currently not possible to easily obtain significantly better transcriptions for a large number of tokens, we accept our automatic transcriptions to be valid. Given the average duration of schwa (48 ms) and the mean differences in duration assigned by the different transcribers and the ASR system (3.7 and 12.8 ms), our data will, however, have to be interpreted with care.

The validation described above did not include tokens based on /h/-initial verbal bases (e.g., *gehaast* ‘rushed’), because manual inspection showed that the automatic transcriptions of these tokens were not reliable. We therefore decided not to rely on the automatic transcriptions, but have the first syllable and /h/ of 210 past participles transcribed by the same two transcribers. They used the same acoustic criteria as described above and identified boundaries between schwa and /h/ as transitions in the waveform and audio. The two transcribers agreed on the presence of schwa for approximately 73.3% of the tokens (κ: 0.46), and the differences in schwa duration were only 0.7 ms on average (with a 95% confidence interval from -3.3 to 1.9 ms) and non-significant. We included only those tokens in the data set for which the transcribers agreed on the presence of schwa, and we used the average durations from their transcriptions.

Predictors

We tested the role of several predictors that have previously been shown to condition segment reduction. A first predictor was *speech register*, since reduction has been shown to be more prominent in more informal speech (e.g., Van Bael et al., 2004; Van Son & Pols, 1999). This factor differentiated between the conversational speech (CS) of the ECSD, the interviews (IN) of the CGN, and the read speech (RS) of the CGN. A second variable was *speech rate*, since words tend to be more reduced in faster speech (e.g., Dalby, 1984; Davidson, 2006). We defined *speech rate* as the number of syllables per second in the continuous uninterrupted stretch of speech that contained the target word assuming that all words had been produced in their full forms. This variable ranged from 2.36 to 10.67 syllables per second.

In addition, the segmental context of schwa may affect its pronunciation (Davidson, 2006). We therefore tested the role of the identity of the first *syllable*. This predictor contrasted the syllable *ge-* with the syllables *be-* and *ver-.* The syllables *be-* and *ver-* were pooled, since there appeared to be fewer tokens of *be-* (*n* = 242) and *ver-* (*n* = 213) than of *ge-* (*n* = 1695), and because preliminary results indicated that *be-* and *ver-* show similar behavior. Note that this predictor also distinguishes between the different preceding consonants (i.e., /x/ versus /b/ and /v/). Since schwa may be co-articulated with the preceding consonant, which then may result in a longer preceding consonant and a shortened or absent schwa, another potential conditioning variable is the duration of the preceding segment (*duration preceding segment*). We used the log-transformed duration (see below), which ranged from 4.00 to 8.83. Another predictor indicated whether the places of articulation of the consonant preceding and the consonant following schwa are identical (*identical place*; identical: *n* = 530, different: *n* = 1620). If the surrounding consonants share their place of articulation, schwa tends to be longer due to the physiological difficulty of gesture repetition (e.g., Walter, 2007).

Several prosodic variables have also been shown to affect speech reduction. We tested the role of *word length*, since segments tend to be shorter if they are followed by more syllables (Lindblom, 1963; Nooteboom, 1972). This binary factor distinguished bi-syllabic words (*n* = 1142) from words with more than two syllables (*n* = 1008). Another prosodic variable is *chunk finality*, which indicates whether the word was the last word of a chunk of speech (final: *n* = 872; non-final: *n* = 1278). A chunk is surrounded by pauses, and may contain a part of a sentence, or one or multiple complete sentences. Especially in read speech, a chunk corresponds to a prosodic unit, and words at the end of a chunk may therefore undergo prosodic lengthening (e.g., Cambier-Langeveld, 2000).

Finally, we added three predictability variables because words that occur more often or are more likely to occur given their context have been found to be more reduced (e.g., Bell et al., 2009; Pluymaekers et al., 2005). We tested (1) the log-transformed *word frequency*, (2) the conditional probability of the target word (w_{target}) given the preceding word ($w_{preceding}$), which was estimated by

$$log_2(\frac{frequency(w_{preceding}, w_{target})}{frequency(w_{preceding})}), \quad (2.1)$$

and (3) the conditional probability of the target word given the following word ($w_{following}$), calculated with

$$log_2(\frac{frequency(w_{target}, w_{following})}{frequency(w_{following})}). \quad (2.2)$$

All frequency measures used for the calculation of these predictors are based on counts from all CGN components.

Analyses

We investigated the influences of the predictors on the presence of schwa and its duration by means of mixed effects regression models (Baayen, Davidson, & Bates, 2008). In all models *speaker* and *word type* were considered as crossed random factors. Further, we used contrast coding, where for each qualitative variable, one level is treated as the standard level to which all other levels are compared. In order to create a normal distribution for the duration of schwa, we applied a log-transformation to this duration (ms) and also transformed the duration of the preceding segment. In the duration analyses, segments differing more than 2.5 times the residual standard errors from the values predicted by the statistical models were considered outliers, thus removed, and the models were refitted.

Most studies assume that a predictor influences reduction if it is statistically significant in a statistical model. As a consequence, these studies also assume predictors to be relevant if they hardly improve or even worsen the model's relative goodness of fit. We initially applied this procedure, but then decided to abandon it as our corpus study is based on many data points (i.e., 2150) and we found many significant effects, including high-order interactions. Most of these effects had very small effect sizes and several were not interpretable. We therefore adopted a more conservative approach based on the Akaike Information Criterion (AIC; Akaike, 1973). The AIC measure is defined as $-2 \log \mathcal{L} + 2p$, where $\mathcal{L}$ is the maximum of the likelihood for the model, and p is the number of parameters in the model. We added predictors one by one to a regression model. A predictor or interaction was only retained if it showed a significant effect and if the model with this predictor or interaction had a lower absolute AIC value than the same model without this predictor or interaction. Only predictors that passed these criteria are presented in the results section.

Before we added a predictor to a model, we checked for possible correlations with the predictors already present in that model. When predictor A correlated with predictor B, we first orthogonalized them. That is, we replaced predictor A by the residuals of a linear regression model that predicted this predictor A as a function of predictor B, or vice versa.

Following Bürki, Fougeron, et al. (2011) and Torreira & Ernestus (2011), we also examined the distribution of the duration of schwa. We restricted the analysis to those tokens of schwa that are not highly likely to be absent due to co-articulation with the surrounding segments. If we find a peak at zero ms, this peak can therefore be interpreted as resulting from categorical reduction processes.

Results

The statistical results are presented in Table 2.3. The analysis of the presence of schwa was based on 2150 tokens of past participles, 68.4% of which were produced with schwa. The statistical model showed significant effects of *speech register* and *syllable*, indicating that schwa was more often absent in conversational speech (52.0%) than in interviews (35.3%), and least often in read speech (12.1%), and that it was more often absent in the syllable *ge-* (35.5%) compared to *be-* and *ver-* (17.4%). In addition, there were main effects of *word length* and *speech rate*, which were mod-

Table 2.3: Results for the presence and duration of schwa. Only those comparisons between levels of a factor that differ significantly are presented.

Presence of schwa	β	z	$p <$
Speech rate	-0.43	-6.00	0.0001
Syllable	-0.71	-3.61	0.001
Speech register			
RS - CS	-2.21	-12.06	0.0001
RS - IN	-1.29	-7.39	0.0001
IN - CS	-0.92	-5.11	0.0001
Word length	-1.66	-2.86	0.01
Word length $\times$ speech rate	0.27	2.74	0.01
Duration preceding segment	-0.16	-2.09	0.05
Duration of schwa	β	t	$p <$
Speech rate	-0.10	-6.68	0.0001
Syllable	-0.26	-4.94	0.0001
Speech register			
RS - CS	-0.36	-3.89	0.001
RS - IN	-0.23	-2.58	0.01
Syllable $\times$ speech register			
RS - CS	0.31	2.89	0.01
RS - IN	0.32	2.99	0.01
Identical place	0.29	5.72	0.0001

ulated by an interaction of these two variables. Schwa was more likely to be absent in bi-syllabic past participles, especially at higher speech rates (see Figure 2.1). Finally, *duration preceding segment* was significant, showing that schwa was more likely to be absent if preceded by a longer segment. The model estimated the standard deviation of the random variable *word type* at 0.82 and of *speaker* at 0.42.

The analysis of the duration of schwa was based on the 1470 schwas that were present. We found significant effects of *speech rate* and *identical place*: schwa was shorter at higher speech rates and if the preceding and following consonant had different places of articulation (mean duration for different place: 44 ms; for identical place: 54 ms). Also, the effects of *syllable* and *speech register* were significant, and they were modulated by their interaction: in read speech, schwa tended to be shorter in *ge-* (mean: 43 ms) than in *be-* and *ver-* (mean: 52 ms). Finally, the model contained random effects involving *word type*: the standard deviation (in log ms) for the intercept was 0.22, the estimated standard deviations for interviews was 0.36, and for conversational speech 0.31, with estimated correlations of -0.20 (IN) and -0.11 (CS). The residual standard deviation was 0.52.

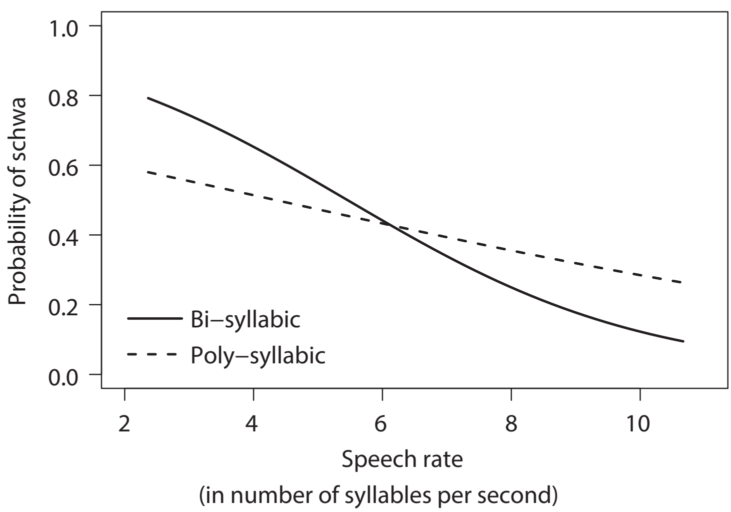

Figure 2.1: The effects of *speech rate* and *word length* on the presence of schwa. The curves are adjusted for the means of all other predictors, and depict past participles from the ECSD carrying the prefix *ge-*.

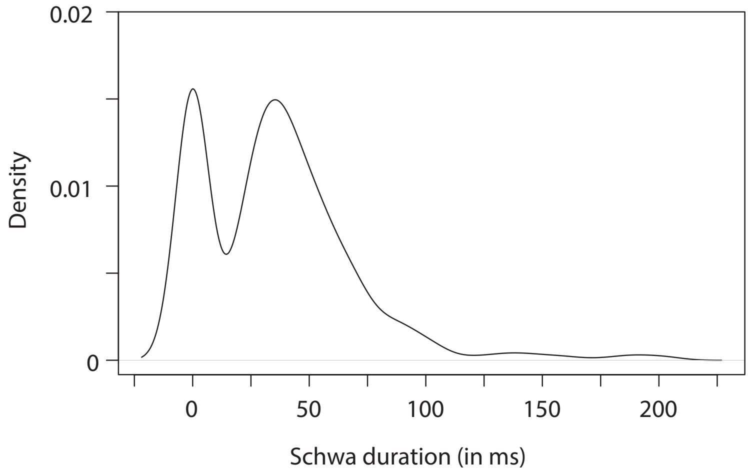

Figure 2.2: The distribution of the duration of schwa.

Figure 2.2 shows the kernel density plot for the duration of schwa. This plot is an estimation of the duration density and can be viewed as a smoothed version of a histogram. The plot is based on all data points that are not outliers (i.e., schwas shorter than 210 ms) and for which we expect no strong co-articulation (i.e., schwas not followed by a voiceless obstruent or a voiced fricative; Ernestus, 2000). Importantly, this plot clearly shows two peaks. This is as expected assuming that the absence of schwa can result from a categorical process and from a gradient shortening process.

Study 2: /t/ reduction

Method

Materials

From the data set used in Study 1, we selected all past participles that ended in /t/. We excluded those that were directly followed by a plosive, as /t/ is then often unreleased (55.3% according to Schuppler et al., 2011), which makes it difficult to determine the boundary between /t/ and the following plosive. The data set for Study 2 consisted of 1166 word tokens, representing 355 word types (404 word tokens representing 133 types from CS, 292 tokens representing 132 types from IN, and 470 tokens representing 227 types from RS).

Measurements, predictors, and analyses

To determine the presence versus absence of /t/ and its duration, we used the method described in Study 1. In connected speech, /t/ may have no oral closure and thus exist of weak frication only. Since the /t/ phone model of the ASR system was trained on both full and reduced /t/s; it is also able to recognize reduced /t/s (see, for instance, Figure 2.3).

The reduction of /t/ was analyzed using the same method as in Study 1, and the predictors used in these analyses were also the same with a few exceptions: we did not include the variables *syllable* and *identical place*, as these are specific for the initial syllable. Instead, we added a factor that identified the *type of segment preceding* /t/. Since /t/ is often reduced after obstruents, especially /s/ (e.g., Ernestus, 2000), this factor distinguished between /s/ (*n* = 188), /x,f/ (*n* = 191), /p,k/ (*n* = 191), and other segments (*n* = 787). Preliminary analyses showed that these other segments (vowels and sonorants) had the same effect.

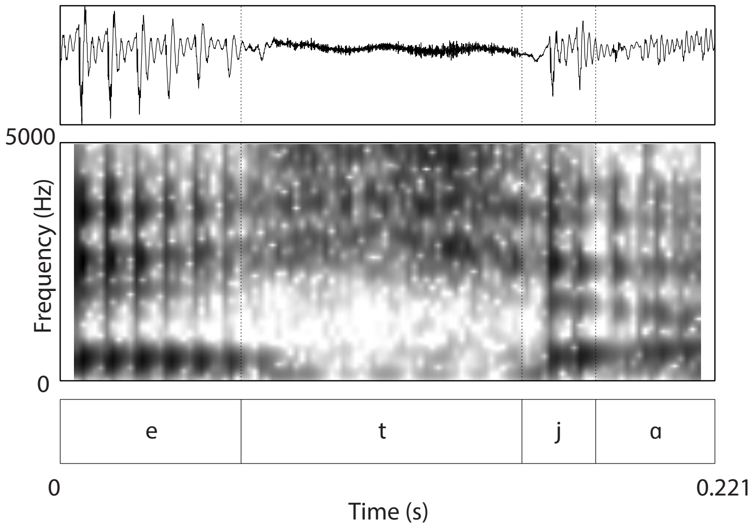

Figure 2.3: An example of a reduced /t/ transcribed by the automatic system.

Results

Table 2.4 presents the results of Study 2. The analysis of the presence of word-final /t/ was based on 1166 past participles, of which 77.9% were produced with /t/. The statistical model yielded significant main effects of *speech rate*, *speech register*, and *type of preceding segment*: word-final /t/ was more likely to be absent at higher speech rates and in more informal speech (CS: 29.0%; IN: 29.1%; RS: 11.9%). In addition, it was most often absent if preceded by a fricative (/s/: 44.1%; /x,f/: 40.3%), than if preceded by a plosive (25.1%), and least often in other cases (8.4%). The standard deviation of the random effect *word type* was estimated at 0.49.

Table 2.4: Results for the presence and duration of /t/. Only those comparisons between levels of a factor that differ significantly are presented.

Presence of /t/	β	z	$p <$
Speech rate	-0.20	-2.66	0.05
Speech register			
RS - CS	-1.28	-6.31	0.0001
RS - IN	-1.04	-4.73	0.0001
Type of preceding segment			
other - /x,f/	-2.03	-7.64	0.0001
other - /s/	-2.14	-6.62	0.0001
other - /p,k/	-1.18	-4.18	0.0001
/x,f/ - /p,k/	0.85	2.78	0.01
/s/ - /p,k/	0.96	2.69	0.01
Duration of /t/	β	t	$p <$
Speech rate	-0.08	-4.32	0.0001
Speech register			
RS - IN	-0.18	-2.33	0.05
IN - CS	0.22	2.75	0.01
Type of preceding segment			
other - /x,f/	-0.22	-3.60	0.01
other - /s/	-0.46	-7.54	0.0001
/x,f/ - /s/	-0.25	-3.14	0.01
/s/ - /p,k/	0.35	4.79	0.0001
Chunk finality	0.75	12.53	0.0001
Chunk finality × speech register			
RS - CS	-0.53	-5.52	0.0001
RS - IN	-0.34	-3.34	0.01

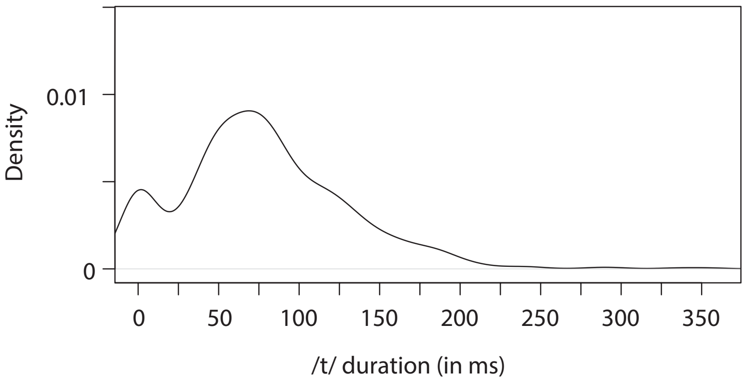

Figure 2.4: The distribution of the duration of /t/.

The statistical model for the durations of the 908 /t/s that were present showed significant effects of *speech rate* and *type of preceding segment*: word-final /t/ tended to be shorter at higher speech rates, and if the preceding segment was a fricative, especially /s/ (mean durations: /s/ 69 ms; /x,f/ 78 ms; /p,k/ 85 ms; other 90 ms). Furthermore, *chunk finality* and *speech register* showed significant main effects and an interaction: past participles tended to end in longer /t/s in chunk final than in chunk initial or medial position, especially in less informal speech (mean durations: CS final: 88 ms, CS non-final: 74 ms, IN final: 80 ms, IN non-final: 58 ms, RS final: 114 ms, RS non-final: 70 ms). The estimated standard deviation of *speaker* was 0.17 log ms, and the residual standard deviation was 0.55 log ms.

Figure 2.4 shows the kernel density plot of the distribution of /t/ duration. It is based on those tokens in which strong co-articulation is not very likely (i.e., if /t/ is not preceded or followed by a bilabial or alveolar segment nor by a fricative). This plot shows two peaks, but the peak at zero is small and not very distinctive, suggesting that if a categorical process plays a role in /t/ duration, it is a very minor one.

General discussion

This chapter investigated whether highly frequent reduction processes that affect many word types, but only in informal speech, can be categorical in nature. In particular, we investigated the presence versus absence and duration of schwa and /t/ in Dutch past participles. If a segment's absence and duration are conditioned by the same variables, they probably result from the same gradient shortening process. In contrast, if the absence of a segment is conditioned by different variables than its shortening, we hypothesize that the absence may result from categorical processes, for instance, a phonological rule or the selection of a reduced pronunciation variant from the mental lexicon.

We extracted past participles from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000) and the Spoken Dutch Corpus (Oostdijk, 2002). Study 1 analyzed 2150 schwas in the word-initial syllables *ge*-, *be*-, and *ver*-, and Study 2 focused on the subset of 1166 past participles ending in /t/. We determined a segment's presence and duration on the basis of automatically generated broad phonetic transcriptions, and found that schwa and /t/ were present in 68.4% and 77.9% of the tokens, respectively. An evaluation of the automatic transcriptions revealed that they were comparable to human-made transcriptions.

The reduction of schwa and /t/ have been hypothesized to occur more often in more informal speech (e.g., Van Bael et al., 2004; Van Son & Pols, 1999). Our data confirm this hypothesis. In our conversational speech, we found that schwa and /t/ were absent in 44.9% and 29.0% of the tokens, respectively, whereas in the read speech, they were absent in only 12.1% and 11.9% of the tokens, although the read speech that we investigated consists of lively read passages from the library of the blind, and is not as formal as, for instance, speeches or lectures.

Our data present three types of evidence that the absence of /t/ is mostly due to extreme gradient shortening. The first evidence comes from Figure 2.4, which shows the distribution of the duration of /t/ for those tokens that are surrounded by segments with which it is not easily co-articulated. This distribution shows only one clear peak, which suggests the absence of /t/ results from gradient shortening. The small peak at zero ms, however, leaves open the possibility that /t/ may occasionally also be absent due to categorical processes. Possibly, /t/ may be categorically absent in a small number of (high frequency) words.

The second type of evidence comes from the types of variables that affect /t/ reduction (see Table 2.5), since the effects of all these variables may result from co-articulation. Like all co-articulation processes, /t/ was more often reduced in more informal speech and at higher speech rates. Moreover, it was also more reduced if preceded by a voiceless obstruent, particularly /s/. If articulatory gestures are decreased in magnitude (Browman & Goldstein, 1990), a /t/ may be realized with

Table 2.5: Significant predictors for the presence and duration of /t/.

Presence of /t/	Duration of /t/
Speech rate	Speech rate
Speech register	Speech register
Type preceding segment	Type preceding segment
	Chunk finality
	Chunk finality × speech register

an incomplete closure and consequently be difficult to distinguish from a preceding obstruent. This is especially the case if this preceding obstruent is /s/, which shares its place of articulation with /t/. Finally, we found that /t/s were longer in chunk final position, especially in less informal speech. The chunks in read speech more often coincided with the ends of prosodic units (e.g., intonational phrases) than the chunks in conversational speech. As a consequence, the end of a chunk in read speech is more likely to show final lengthening (see Cambier-Langeveld, 2000).

Thirdly, the duration and presence of /t/ are influenced by exactly the same variables. The only exception is the position of the word in the chunk, which affected only /t/, duration. We probably did not find an effect of this variable on the presence of /t/ because analyses of binomial factors usually have less statistical power than analyses of continuous measures. Also this similarity in variables suggests that the absence of /t/ mainly results from gradient shortening.

The results for schwa reduction are similar to those of /t/ reduction in that most effects on the presence and duration of schwa (see Table 2.6) can be interpreted as resulting from co-articulation. First, like other co-articulation processes, schwa reduction was more frequent in more informal speech and at higher speech rates. Second, schwa was more often absent after longer consonants, which strongly suggests that its absence often resulted from the articulatory overlap of the preceding consonant and schwa (Browman & Goldstein, 1990). Third, schwa tended to be longer if the places of articulation of the preceding and following consonants were identical. This finding is in line with findings by Walter (2007), who argues that the production of two subsequent consonants with the same place of articulation requires more effort than the production of a non-repetition sequence, which leads to longer vowels between two consonants with the same place of articulation.

Fourth, we found a clear effect of the different word-initial syllables. In general, schwa tended to be more often present in *be-* and *ver-* than in *ge-*. We found the same effect for the duration of schwa, but only in read speech. The difference be-

Table 2.6: Significant predictors for the presence and duration of schwa.

Presence of schwa	Duration of schwa
Speech register	Speech register
Speech rate	Speech rate
Syllable	Syllable
	Syllable × speech register
	Identical place
Word length	
Word length × speech rate	
Duration preceding C	

tween the syllables probably results from their morphological functions. The syllables *be-* and *ver-* are derivational prefixes and therefore contain crucial semantic information about the identity of the word, as illustrated by the minimal word pair /bətalt/ *betaald* 'paid' versus /vərtalt/ *vertaald* 'translated'. In contrast, the syllable *ge-* is an inflectional prefix in 375 of the 387 examined word types, indicating that the word is a past participle (as in /xəfitst/ *gefietst* 'cycled', whose stem does not occur with other prefixes), which is usually expected given the context. As a consequence, the syllables *be-* and *ver-* tend to be less redundant than *ge-*, and this may result in differences in reduction (see Chapter 5 of this dissertation for an overview of the role of predictability in reduction). Probably, the effect on duration is only demonstrated in read speech since, in this register, speakers are more aware of the listener and their speech tends to be more listener-driven than in less formal registers (Lindblom, 1990).

Finally, we found an interaction of word length with speech rate, with schwa being more often absent in bi-syllabic than in longer words, especially at higher speech rates. The direction of this effect is unexpected given previous studies which have shown that vowels tend to be shorter if they are followed by more syllables in the word (Lindblom, 1963; Nooteboom, 1972). We therefore conducted further analyses and found that bi-syllabic and longer words did not only differ in terms of the number of syllables, but that they also differed in several measures of contextual predictability: for instance, the bi-syllabic words are significantly more frequent (mean log frequency: 9.06) than the longer words (mean log frequency: 7.68; $t = 10.94$, $p < 0.0001$). Since more predictable words are more likely to be reduced, especially at higher speech rates (e.g., Pluymaekers et al., 2005), the effect of word length may actually be an effect of contextual predictability.

The variables affecting schwa reduction thus indicate that the absence of schwa also results from gradient shortening. In addition, the absence of schwa may result from categorical processes, as is indicated by two other types of results. First, the density plot of the duration of schwa (Figure 2.2) clearly shows two modes: one mode representing the absent schwas and the other mode containing the present schwas. Second, if the absence of a segment only results from gradient shortening, we expect that its presence and duration would be affected by the same variables, as we found for /t/ above, and as has been found in several other studies (e.g., Torreira & Ernestus, 2010b; Chapter 3 of this dissertation). In contrast to this expectation, we found that the presence of schwa, but not its duration, is affected by word length, an interaction of word length with speech rate, and the duration of the preceding consonant, whereas its duration, but not its presence, is affected by an interaction of the identity of the first syllable with speech register and by whether the surrounding consonants share their place of articulation (see Table 2.6). Moreover, we found that the presence of schwa was influenced by more variables than its duration. Given that analyses of continuous variables generally have greater statistical power than analyses of binomial factors, this difference is unexpected if the presence and duration of schwa result from the same articulatory process.

There is, however, one possible explanation for the greater number of predictors for the absence of schwa versus its duration. The duration data set contained 31.6% fewer data points than the data set on the presence of schwa, and the duration analysis may therefore have had less statistical power. However, for /t/ the duration data set also contained 22.1% fewer data points than the data set on its presence, and for /t/ we did not find that its duration was conditioned by fewer variables than its presence. It seems therefore unlikely that the difference in number of data points can explain the difference in the number of significant effects in the analyses of the presence and duration of schwa.

Furthermore, we examined whether a difference in the sizes of the data sets explains why the presence of schwa was affected by more variables than its duration, whereas this is not the case for /t/. We restricted the schwa data set to those 1162 tokens that also occur in the /t/ data set. Due to the smaller sample size, this data set showed fewer effects on the presence of schwa than the original data set (see the Appendix). Importantly, the variables that condition the presence and duration of schwa are still different. For /t/, we expanded the data set in order to obtain a data set that was as large as the original schwa data set. Since the corpus did not contain past participles that met all restrictions and that were not yet part of the data set, we expanded the data set with other word types (e.g., adjectives and other verb forms). This new data set consisted of 1993 tokens in total. The presence and duration of these new tokens of /t/ were determined with the same procedures described in Study 2. This analysis yielded three additional effects influencing *both* the presence and duration of /t/ (see the Appendix). Overall, these re-analyses show the same patterns as the original analyses. Apparently, statistical power cannot explain the differences in conditioning variables.

On the basis of these findings, we conclude that the absence of schwa in Dutch can result both from gradient and categorical reduction processes. In future research, the evidence for categorical reduction may be extended with more detailed acoustic analyses of our data, like the analyses reported on French *c'était* 'it was' by Torreira & Ernestus (2011).

Thus far, categorical reduction processes have only been identified for highly frequent words or word combinations (e.g., /e/-deletion in French *c'était*, Torreira & Ernestus, 2011) and for phenomena that do not only occur in informal conversations, but also in formal speech or in words uttered in isolation (e.g., word-internal schwa deletion in French, Bürki, Ernestus, et al., 2011). Gradient reduction has been observed for reduction that affects many words, but only if uttered in connected informal speech (e.g., vowel devoicing in French, Torreira & Ernestus, 2010b). In line with this pattern, we found that reduction of word-final /t/ in Dutch, which affects many words in connected informal speech, is mainly gradient in nature. In contrast, the results for schwa point to a more complex picture. The comparison of the variables conditioning the absence and duration of schwa as well as the distribution of schwa duration suggest that extreme vowel shortening is not the only process that can result in its absence. Hence, categorical reduction is not restricted to highly frequent word combinations or phenomena that also occur in formal speech registers.

The categorical process that results in the absence of schwa may consist of either the application of phonological rules or the selection of a reduced pronunciation variant from the mental lexicon during speech production. If the absence of schwa results from phonological rules, we expect these rules to apply to all words equally often (word-specific rules would be similar to the storage of the reduced pronunciation variants together with their frequencies). Our analyses showed that the random factor *word type* has a standard deviation of 0.81, which indicates that words differ substantially in their probability that schwa is absent (approximately 20%). This suggests that words are not reduced equally often. The categorical absence of schwa is therefore not likely to result from phonological rules, but from variant selection. This implies that for many words, the mental lexicon contains at least two pronunciation variants.

Our findings raise the question why the absence of /t/ results from a gradient process, whereas the absence of schwa can result from both a gradient and a categorical process. One possible explanation is that the schwa and /t/ in past participles differ in their perceptual prominence. Whereas schwa is part of the initial syllable, the /t/ is in word-final position. In addition, schwa is part of a prefix that contributes to the meaning of the past participle or signals that the word is a past participle, and therefore carries information. The /t/, in contrast, only indicates the past participle function, which in most cases is also signaled by the prefix (if it is *ge-*) or by the syntactic structure of the sentence. Finally, schwa is probably more prominent because of its acoustic characteristics: a reduced schwa is most likely to be clearly audible, whereas a /t/ with a weakened burst may be difficult to perceive. As a consequence of this difference in perceptual prominence, language users may more easily detect the absence of schwa in the prefix of past participles, and build lexical representations of pronunciation variants without schwa. They may less often notice the absence of word-final /t/ and may interpret its absence as resulting from its low prominence. Accordingly, they may be less inclined to create lexical representations for past participles without /t/.

Unexpectedly, our data showed no effects of the word's (contextual) predictability on schwa and /t/ reduction. We believe that this is due to our conservative analysis method: we only retained those variables in our statistical models that were significant *and* improved the model's AIC value. If we do not apply this AIC criterium, effects of a word's frequency of occurrence and its predictability given the following word emerge. Unfortunately, it is currently not possible to draw general conclusions about the role of predictability in reduction. On the one hand, predictability effects reported in earlier studies may disappear if our conservative analysis method is applied to these studies. On the other hand, since past participles tend to be more predictable than other types of content words, predictability may play no substantial role in the production of these particular verb forms.

An advantage of using corpus data over experimental data is that the examined speech is more natural. Our results suggest that corpus data can indeed be used to infer insights on the processes that drive pronunciation variation. More reliable phonetic transcriptions will even further increase the usefulness of corpus data. The findings based on corpus data should nevertheless be confirmed by data obtained from controlled experiments. It is currently not clear which experimental task would allow us to investigate the production of casual speech. Chapter 3 of this dissertation demonstrated, for instance, that one experimental paradigm that may seem promising, the shadowing task, elicits fast speech that cannot be compared to casual speech.

In summary, our results show that the absence of word-final /t/ results from extreme shortening. The absence of schwa, in contrast, can result from gradient vowel shortening as well as from a categorical process. We therefore conclude that reduction phenomena that mostly occur in informal conversations and that affect many different words can also result from categorical processes. This categorical process is probably the selection of a variant without schwa from the mental lexicon, which implies that for many words, the mental lexicon contains at least two pronunciation variants.

Appendix

To examine whether the difference in power explains why the presence of schwa was affected by more variables than its duration, we re-analyzed schwa and /t/ with better matched data sets. The statistical models for schwa showed the same significant effects as Study 1, but the main effects of syllable and word length and the interaction between word length and speech rate on the presence of schwa were absent.

For /t/, the statistical models showed the same effects as in Study 2 plus some additional effects (see Table 2.7 for these additional effects): word-final /t/ tended to be more often absent and shorter in word types other than past participles, and in chunk non-final position, particularly in words other than past participles. Furthermore, /t/ was more likely to be reduced in chunk non-final position if not preceded by fricatives.

Table 2.7: Additional effects found in the re-analyses for the presence and duration of /t/. Only those comparisons between levels of a factor that differ significantly are presented.

Presence of /t/	β	*z*	$p <$
Past participle	0.53	2.97	0.05
Chunk finality	1.07	3.35	0.01
Chunk finality × past participle	-0.65	-1.98	0.05
Chunk finality × type preceding segment			
other - /x,f/	-1.26	-3.49	0.001
Duration of /t/	β	*t*	$p <$
Past participle	0.20	4.68	0.0001
Chunk finality	0.94	14.52	0.0001
Chunk finality × past participle	-0.79	-2.52	0.05
Chunk finality × type preceding segment			
other - /s/	-0.24	-2.16	0.05
other - /p,k/	-0.22	-2.28	0.05

Processes underlying acoustic reduction: a production experiment

Chapter 3

This chapter is a reformatted version of:
Iris Hanique and Mirjam Ernestus (2012). The processes underlying two frequent casual speech phenomena in Dutch: A production experiment. *Proceedings of Interspeech 2012*, Portland, Oregon

Abstract

This chapter investigated whether a shadowing task can provide insights in the nature of reduction processes that are typical of casual speech. We focused on the shortening and presence versus absence of schwa and /t/ in Dutch past participles. Results showed that the absence of these segments was affected by the same variables as their shortening, suggesting that absence mostly resulted from extreme gradient shortening. This contrasts with results based on recordings of spontaneous conversations. We hypothesize that this difference is due to non-casual fast speech elicited by a shadowing task.

Introduction

One main characteristic of casual speech is that many words are not produced in their full forms, but in reduced pronunciation variants. For example, the English word *hilarious* may be produced as /hlɛrɛs/ instead of /hılɛriʌs/ (Johnson, 2004). This production study investigates the processes underlying two of these reduction phenomena in Dutch and compares the results to those of a corpus study. This chapter contributes to our knowledge of speech reduction and shows the advantages and disadvantages of two very different research methods for the study of casual speech phenomena.

Processes underlying speech reduction may be gradient. The altered pronunciation or even absence of a segment then originates from gradient overlap and decrease in magnitude of articulatory gestures (Browman & Goldstein, 1992). For instance, /t/ in *must be* may be partly or even completely hidden by the closure of the following bilabial stop, which results in /mʌsbi/.

In addition, the absence of segments may result from categorical deletion processes, which may be phonological rules or the selection of a reduced pronunciation variant from the variants stored in the mental lexicon. So far, two types of reduction phenomena have been found to be categorical: processes that affect only highly frequent words or word combinations (e.g., /e/-deletion in French *c'était* /setɛ/ 'it was'; Torreira & Ernestus, 2011), and phenomena that not only occur in casual, connected speech, but also in words produced in isolation or formal speech (e.g., word-internal schwa deletion in French; Bürki et al., 2010).

So far, only one study has investigated the nature of reduction phenomena that affect many different words and are restricted to connected informal speech (Chapter 2 of this dissertation). It examined schwa and /t/ reduction in past participles in two speech corpora of Dutch. The results showed that the presence and duration of /t/ are affected by roughly the same phonetic variables, suggesting that absence of /t/ results from the same gradient process as its shortening. Also the presence and duration of schwa were mainly influenced by phonetic variables, but the presence of schwa was affected by more and different variables than its duration. The authors therefore argued that schwa reduction can result from gradient as well as categorical processes.

A disadvantage of corpus studies is that they are often restricted to highly frequent word types, and that these are represented by a widely varying number of tokens. Also, the segmental context of units under study cannot be well-controlled for. These disadvantages do not apply to controlled production experiments, but it is currently unclear which experimental task can elicit casual speech.

We conducted a shadowing experiment, in which, like in Chapter 2, we focused on schwa and /t/ reduction in Dutch past participles. These words usually consist of /xə/, the verbal base, and /t/ (e.g., *gedanst* /xə+dɑns+t/ 'danced'). We examined whether this controlled experiment produces results similar to those of the corpus study mentioned above.

Method

Participants

We tested 35 Dutch native speakers aged between 18 and 27 (mean 20 years).

Materials

Our experiment consisted of 180 target past participles starting with *ge-* (140 end in /t/) and 100 filler past participles starting with *ver-* or *be-* (88 end in /t/). These past participles consisted of two or three syllables and the schwa in the initial syllable was followed by a consonant. They spanned the entire range of frequency of occurrence, which was based on the Spoken Dutch Corpus (Oostdijk, 2002).

Each past participle was embedded in the middle of a sentence (which on average consisted of 10 words). Sentence accent was never on the past participle. Also, past participles were never preceded by a fricative, and those ending in /t/ were never followed by /t/ or /d/. Whereas we created one sentence for each filler, we created two sentences for each target: in one sentence the past participle was followed by a vowel, and in the other by a consonant. For example, the sentences for the target *getankt* /xətɛnkt/ 'refueled' were (1) *Ze had per ongeluk diesel getankt in plaats van benzine* 'She accidentally refueled diesel instead of gas', and (2) *Hij heeft voornamelijk getankt waar de brandstof goedkoop is* 'He mainly refueled where fuel was cheap'. All sentences were recorded by a native Dutch female speaker in a casual (average sentence duration: 2019 ms) and careful way (average sentence duration: 2208 ms). On the basis of automatically generated transcriptions (using the same procedure as described below), we observed that schwa was absent in 125 of the 180 casually produced sentences, while it was never absent in the carefully produced sentences. The average durations of the present schwas were 28.1 ms (sd: 9.7 ms) in the casual and 49.2 ms (sd: 12.9 ms) in the careful condition. Participants heard every past participle only once, that is, only in the casual or in the careful condition, and followed either by a consonant or a vowel.

The experiment started with a practice block of 10 filler trials followed by four experimental blocks, each consisting of 45 target and 25 filler sentences. The first block after the practice trials started with seven fillers, while the other blocks started with three fillers. Within one block, we presented either casual or careful sentences. If block one and two contained casual sentences, blocks three and four consisted of careful sentences, and vice versa.

We created three pseudo-randomizations of all stimuli, in which no more than three target sentences occurred in succession. On the basis of each randomization, we created four lists. In each list, half of the target past participles were followed by a vowel, and the other half by a consonant. Moreover, half of the stimuli in a list had been produced in a casual way, and the other half in a careful way. Together all four lists contained all sentences in both speech styles.

Procedure

Each participant was tested individually in a sound-attenuated booth. We presented sentences via headphones, and asked participants to repeat these sentences as quickly and accurately as possible, and to start repeating as soon as possible. We recorded responses on an R-09 Edirol recorder. Each trial started with a fixation point shown for 500 ms on a computer screen, and after an interval of 100 ms the stimulus was presented. The next trial started 1500 ms after the end of this stimulus. Each session lasted approximately 30 minutes.

Data processing

For all target sentences produced, we created automatic broad phonetic transcriptions by means of forced alignment as described in Chapter 2 of this dissertation. The automatic system selected the variant for each word that best matched the speech signal from a lexicon that contained, among others, pronunciation variants in which schwa and /t/ were present or absent. Since the acoustic models consisted of three emitting states and the frame shift was 5 ms, the system assigned to each segment a duration of at least 15 ms. If a segment was in reality shorter than 15 ms, its boundaries were placed within the neighboring segments (which were consequently assigned a shorter duration than they really had). We validated the automatic transcriptions by manually transcribing 100 target schwas, and found that the agreement on the presence of schwa between two transcribers (90%) was very similar to those between each transcriber and the forced alignment (85% and 87%). The average differences in duration were smaller than 13 ms, which is usually considered as acceptable for this type of speech (Wesenick & Kipp, 1996). As expected, given the automatic method used, the durations assigned by the automatic system were generally greater than the durations assigned by the human transcribers.

We excluded 35% of the 6300 trials in two steps. First, 1866 trials were excluded on the basis of the transcriptions. We excluded transcriptions that indicated a silence directly before or after the target word, since in these trials the target words are not embedded in stretches of connected speech. In addition, we excluded transcriptions that were likely to be incorrect. We identified these incorrect transcriptions by means of a set of criteria, which we determined by checking 200 automatically generated transcriptions. Transcriptions were excluded that contained three or more segments of 15 ms in the preceding, target, or following word, which typically indicates that the sentence was produced incompletely or with a wrong word order, or that the forced alignment system had selected a non-suitable pronunciation variant. Further, we also excluded trials if the longest phoneme was the schwa or shorter than 50 ms, which are highly likely to be transcription errors. Finally, we excluded those targets in which the longest phoneme was suspiciously long, as this often indicates that multiple phonemes are transcribed as one long phoneme. We set the boundary for suspiciously long plosives at 175 ms, for vowels and fricatives at 165 ms, and for other consonants at 155 ms. In the second step, we listened to all remaining trials, and excluded from further analyses those 339 trials in which the speaker had not produced the target word or the directly preceding or following word fluently. The fact that we discarded 2205 out of 6300 trials shows that the task was difficult. Interestingly, the number of disregarded trials varied among participants from 23 to 104 trials.

Predictors

We tested the influence of several predictors on the presence and duration of schwa and /t/. Three of these predictors were defined on the basis of the experimental design. The first predictor is the register (careful or casual) in which the stimulus was presented. The second predictor tested whether there were differences between the blocks of the experiment. The third predictor is the duration of the sentence presented to the participant.

Further, we added other predictors to our statistical models that the literature has shown to be relevant. One of these is speech rate, since segments tend to be more reduced in faster speech (e.g., Dalby, 1984). We defined speech rate as the number of syllables per second in the whole sentence produced by the participant. In addition, we tested a factor word length, which indicates whether the past participle consisted of two or three syllables, as segments are often shorter if they are followed by more syllables (Nooteboom, 1972).

We also examined three measures of word predictability, since words tend to be more reduced if they are more predictable (e.g., Bell et al., 2009). One measure was the log-transformed word's frequency of occurrence. The second and third measures were the conditional probabilities (CP) of the target word (w_{target}) given the preceding word ($w_{preceding}$) or the following word ($w_{following}$), which were calculated with formulae 3.1 and 3.2, respectively. The frequencies used for these predictability measures were based on all components of the Spoken Dutch Corpus (Oostdijk, 2002).

$$log_2(\frac{Frequency(w_{preceding}, w_{target}) + 1}{Frequency(w_{preceding}) + 1}) \quad (3.1)$$

$$log_2(\frac{Frequency(w_{target}, w_{following}) + 1}{(w_{following}) + 1}) \quad (3.2)$$

Finally, the surrounding segments may affect reduction (e.g., Dalby, 1984). For schwa, we therefore tested the log-transformed durations of the preceding and following consonants, the place and manner of articulation of the following consonant, as well as its voicing, and whether it was velar. For /t/, we examined the place and manner of articulation of the preceding segment, the log-transformed durations of the preceding and following segments, and whether the following segment was a vowel or a consonant (henceforth *type*).

Analyses

We used mixed effects regression models with contrast coding (i.e., for factors, one level is placed on the intercept, and all other levels are compared to this default level). In order to account for differences between individual stimuli and participants, the models contained *target word* and *participant* as crossed random effects. Each predictor was added individually to a model, and only remained in that model if it was significant *and* improved the AIC value. Duration analyses were based on present segments only.

Results and discussion for schwa

Table 3.1 presents the two final statistical models for the presence and duration of schwa. As expected, schwa was more often absent and shorter if the stimulus was produced in a casual (29.8%; 42 ms) compared to a careful way (22.9%; 45 ms). However, this difference in duration was much smaller than the difference between the stimuli in the two conditions.

Replicating earlier findings (Bell et al., 2009), schwa was more often absent and shorter in past participles with higher frequencies of occurrence. Highly frequent words are produced more often, and their production has therefore become more automatized and efficient. This typically results in more overlapping gestures.

Also, the consonants surrounding schwa affected its reduction. The main effect of the manner of articulation of the following consonant showed that schwa was significantly longer if this consonant was a plosive (47 ms), and shorter if it was a fricative (38 ms; other segments: 44 ms). We hypothesize that schwa can more easily be co-articulated with a following fricative than a following plosive, since fricatives are continuants.

In addition, we found an effects of the duration of the following consonant. Since the duration of the following consonant correlated with its manner of articulation, we orthogonalized these variables: we replaced this duration by the residuals of a model that predicted duration as a function of manner of articulation. Schwa tended to be shorter if the preceding and following consonants were longer. Schwa was also more likely to be absent if the preceding consonant was longer, especially in block 1 and 3. Moreover, this vowel was more often absent if the following consonant was longer, especially if this consonant was voiced. The effect of the duration of the surrounding consonants can also be explained by co-articulation: schwa appears shortened or completely hidden by the preceding or following consonants, which are

Table 3.1: Results for schwa: those for its presence are above the thick line, and those for its duration are below the thick line. For the factors Stimulus register, Voicing following consonant, and Block, the levels on the intercept are casual, voiced, and block 4, respectively. For the random effects, the table reports the estimated standard deviations (in ms for duration). The preceding and following consonant are denoted by prec C and foll C, respectively.

Presence of schwa (N = 4095)			
Fixed effects	F	df	$p <$
Stimulus register	26.79	1,3864	0.0001
Word frequency	6.51	1,3864	0.05
Duration foll C	36.16	1,3864	0.0001
Duration prec C	12.80	1,3864	0.001
Voicing foll C	40.33	1,3864	0.0001
Block	2.76	3,3864	0.05
Duration foll C × Voicing foll C	18.45	1,3864	0.0001
Duration prec C × Block	3.22	3,3864	0.05
Random effects	Word	Participant	
Intercept	0.46	0.94	
Duration of schwa (N = 3016)			
Fixed effects	F	df	$p <$
Stimulus register	43.16	1,2678	0.0001
Word frequency	12.27	1,2678	0.0005
Duration foll C	117.72	1,2678	0.0001
Duration prec C	100.37	1,2678	0.0001
Manner foll C	32.05	2,2678	0.0001
Speech rate	4.67	1,2678	0.05
Speech rate × Word length	7.49	1,2678	0.01
Random effects	Word	Participant	
Intercept	14.73	4.74	
Duration prec C	2.43		

then assigned longer durations. Schwa may often be absent especially before long consonants that are voiced, since it is more difficult to observe (for both humans and automatic speech recognizers) a short, co-articulated, voiced vowel next to a voiced rather than a voiceless consonant.

Finally, speech rate correlated with stimulus register, which we had therefore orthogonalized. Schwa was more likely to be shorter at higher speech rates, but only in bi-syllabic past participles. A possible explanation is that vowels tend to be longer in the initial syllables of bi-syllabic than tri-syllabic words (Nooteboom, 1972). Consequently, schwa in bi-syllabic words can show more variation in its duration (as shortening is less likely to result in deletion), and may therefore be more easily affected by gradient reduction processes like speech rate.

Results and discussion for /t/

The final statistical models for the presence and duration of /t/ are presented in Table 3.2. Since the duration and type (vowel versus consonant) of the following segment were correlated, we had orthogonalized them. First, as expected, the models show that /t/ was shorter at higher speech rates. In addition, /t/ was more likely to be absent and shortened if it was followed by a consonant (28.3%; 58 ms) than a vowel (10.5%; 66 ms). The articulatory gestures of /t/ are more similar to those of other consonants than to those of vowels, since vowels require a relatively open vocal tract whereas consonants typically involve a (almost) closed one. Therefore, /t/s may more easily overlap with and be hidden by other consonants. If so, they may be difficult to distinguish from these overlapping consonants, and appear acoustically shortened or even absent.

Word-final /t/ was also more likely to be absent and shorter if the following segment was longer, but only if this was a consonant. An explanation may be that if /t/ overlaps with a following consonant, (part of) its duration may be attributed to this following consonant. For the presence of /t/, the effects of the following consonant were also greater if the word was more predictable given the following word. Word combinations that are often used together are more automatized, can thus be produced more easily, and are consequently more likely to show effects of co-articulation.

Further, we investigated the roles of the duration of the preceding segment and its manner of articulation. Since these two predictors were correlated, they were orthogonalized (following the method for orthogonalization described above). Word-final /t/ tended to be shorter if the preceding segment was longer, especially if this segment was a fricative or nasal. If the gestures of /t/ overlap with a preceding fricative or nasal, its closure may be incomplete, and /t/ may therefore be hard to distinguish from this fricative or nasal, which then appears longer.

In addition, the interaction of the durations of the preceding and following segments showed that /t/ was more likely to be absent and shorter if either the preceding or following segment is longer, especially if the other immediately neighboring segment is shorter. This suggests that gestural overlap of /t/ with an adjacent segment is larger if it overlaps less with the other adjacent segment.

Finally, /t/ tended to be shorter and more often absent in words that are more predicable given the following word. These predictability effects are likely the result of more sloppy pronunciations of more often repeated and thus more automatized words or word sequences.

Table 3.2: Results for /t/: those results for its presence are above the thick line, and those for its duration are below the thick line. For the factors Type following segment and Manner preceding segment, the levels on the intercept are consonant and fricative, respectively. For the random effects, the table reports the estimated standard deviations (in ms for duration). The preceding and following segment are denoted by prec segm and foll segm, respectively.

Presence of /t/ (N = 3133)			
Fixed effects	F	df	$p <$
Duration foll segm	28.87	1,3123	0.0001
Type foll segm	149.12	1,3123	0.0001
CP foll word	7.85	1,3123	0.05
Duration foll segm × Duration prec segm	13.80	1,3123	0.001
Duration foll segm × Type foll segm	22.66	1,3123	0.0001
Duration foll segm × CP foll word	30.94	1,3123	0.0001
Duration foll segm × Type foll segm × CP foll word	29.33	1,3123	0.0001
Random effects	Word	Participant	
Intercept	1.92	0.84	
Duration foll segm	0.79	0.39	
Duration prec segm	0.04		
Duration of /t/ (N = 2472)			
Fixed effects	F	df	$p <$
Duration foll segm	42.83	1,2389	0.0001
Type foll segm	109.36	1,2389	0.0001
Duration prec segm	21.53	1,2389	0.0001
Manner prec segm	13.01	4,2389	0.0001
CP foll word	24.04	1,2389	0.0001
Speech rate	28.40	1,2389	0.0001
Duration foll segm × Type foll segm	70.10	1,2389	0.0001
Duration foll segm × Duration prec segm	7.81	1,2389	0.01
Duration prec segm × Manner prec segm	5.37	1,2389	0.001
Random effects	Word	Participant	
Intercept	5.83	6.87	
Duration foll segm	5.58	2.07	

General discussion

We demonstrated that the shortening and absence of schwa and /t/ show patterns that can easily be interpreted as resulting from co-articulation. Furthermore, we found that their presence and duration were conditioned by similar variables, suggesting that the absence of these segments is the extreme result of their shortening, and thus of a gradient underlying process. Note that we did find slightly more effects for the duration measures, as expected, since analyses of a continuous variable have generally more statistical power than analyses of a factor.

We expected participants to repeat the pronunciation variants that were presented, and thus that many more schwas would be absent in the casual than in the careful condition. However, these percentages were relatively low in both conditions (29.8% and 22.9%, respectively). Our results thus suggest that participants did not aim at repeating the variant they heard, but at producing the word's full form. This would explain why we did not find evidence for categorical absence of schwa, as reported in the corpus study of Chapter 2 of this dissertation. Our shadowing task did elicit reduction phenomena resulting from co-articulation. Apparently, this task evokes non-casual fast speech. This hypothesis is supported by the fact that several words which are often drastically reduced in casual speech (e.g., *eigenlijk* /ɛɪxələk/ is often reduced to /ɛɪk/, *allemaal* /ɑləmal/ to /ɑm/, and *helemaal* /heləmal/ to /hemə/) are never produced in these extremely reduced forms by the participants. We therefore recommend, when studying natural speech, to always use this task in combination with another research method.

In conclusion, our production experiment shows that the shadowing task elicits non-casual fast speech, in which reduction of schwa and /t/ in Dutch past participles is only affected by gradient co-articulation.

Exemplar effects in word comprehension

Chapter 4

This chapter is a reformatted version of:
Iris Hanique, Ellen Aalders, and Mirjam Ernestus (submitted). How robust are exemplar effects in word comprehension?

Abstract

This chapter studies the robustness of exemplar effects in word comprehension by means of four long-term priming experiments with lexical decision tasks in Dutch. A prime and target represented the same word type and were presented with the same or different degree of reduction. In Experiment 1, participants heard only a small number of trials, a large proportion of repeated words, and stimuli produced by only one speaker. They recognized targets more quickly if these represented the same degree of reduction as their primes, which forms additional evidence for the exemplar effects reported in the literature. Similar effects were found for two speakers who differ in their pronunciations. In Experiment 2, with a smaller proportion of repeated words and more trials between prime and target, participants recognized targets preceded by primes with the same or a different degree of reduction equally quickly. Also, in Experiments 3 and 4, in which listeners were not exposed to one but two types of pronunciation variation (reduction degree and speaker voice), no exemplar effects arose. We conclude that the role of exemplars in speech comprehension during natural conversations, which typically involve several speakers and few repeated content words, may be smaller than previously assumed.

Introduction

Several models of speech comprehension assume that the mental lexicon stores the pronunciation of a word with two types of representations, namely abstract representations and exemplars (e.g., Goldinger, 2007; McLennan et al., 2003). Abstract representations are strings of sound symbols like phonemes or phonological features, which only contain information about acoustic properties that distinguish between these symbols. In contrast, clouds of exemplars represent many occurrences of words that the language user has uttered or heard. Each exemplar is a detailed representation corresponding to the speech signal of one occurrence and thus contains subtle acoustic information, for example about the word's exact pronunciation or the speaker's voice. Many articles in the literature point to a role of exemplars in word comprehension. This study investigates the robustness of these exemplar effects.

Exemplar effects have been established in several priming experiments (e.g., Bradlow et al., 1999; Craik & Kirsner, 1974; Goh, 2005; Goldinger, 1996; Janse, 2008; Mattys & Liss, 2008; McLennan et al., 2003; McLennan & Luce, 2005; Palmeri et al., 1993). These experiments contained repeated words and the comprehension of the second occurrence of a word (the target) is expected to be facilitated by the first occurrence (the prime). Primes and targets were completely identical, that is the same token, or they differed in speech rate, time-compression, the realization of a certain segment (e.g., intervocalic /t,d/ produced as [t,d] or as a flap in American English), or the speaker's voice. Most experiments showed that participants reacted more quickly or produced fewer errors on the target if it was identical to the prime. Presumably, participants stored primes with all their acoustic detail and, if targets were acoustically identical to these primes, they could quickly recognize them via these exemplars formed by the primes.

Not all experiments showed these exemplar effects. McLennan et al. (2003) studied allophonic variability and found exemplar effects only when participants processed stimuli relatively fast. Conversely, for indexical variability (e.g., variability in speaker voice and speech rate) McLennan & Luce (2005) only observed exemplar effects if processing was slow. To account for this, McLennan and Luce suggest that more abstract features are generally dominant early in processing and show effects when participants are fast, while surface features (e.g., indexical details) dominate later stages and show effects when participants are slow. However, surface representations can still show effects at an early stage if they represent variants that are relatively frequent (e.g., representations containing flaps instead of underlying /t/ and /d/).

Palmeri et al. (1993) also observed that exemplar effects do not always occur. In an old-new judgment task they investigated whether exemplar effects remain if primes and targets are separated by a large number of words, and therefore they varied the lag between primes and targets (1, 2, 4, 8, 16, 32, or 64 words). In addition, the authors examined whether exemplar effects are influenced by the number of speakers heard (1, 2, 6, 12, or 20 speakers). Primes and targets were produced by either the same or a different speaker. Their results suggest that exemplar effects are only present at lags smaller than 64 words. Exemplar effects did not differ for the different numbers of speakers.

Goldinger (1996) investigated similar issues. He studied the extent to which exemplar effects decrease if the time interval between primes and targets increased (from five minutes, to one day, to a week). In the same experiment, Goldinger investigated whether exemplar effects arise if stimuli were produced by two, six, or ten speakers. Speaker voice for a given prime and corresponding target was either the same or different and participants performed one of two tasks (identifying the words in white noise or judging whether the word has been presented before). The identification task showed exemplar effects for all time intervals, yet for the old-new judgments a week's interval was enough to block these effects. This provides additional evidence that exemplar effects become more difficult to access over time. The effect of the number of speakers in the experiment is less clear.

Our study also investigates the issue of when exemplar effects arise in speech comprehension. More specifically, this study investigates whether exemplar effects are robust under more natural conditions than those typically tested in the literature, providing us with information about the role of exemplars in the comprehension process. Following McLennan & Luce (2005), we conducted four long-term priming experiments using lexical decision only. This way, participants had to process words completely - unlike in, for example, phoneme monitoring or shadowing - and did not have to rely on explicit memory (as in old-new judgment).

Our targets were Dutch verbs that start with the unstressed prefixes *be-* or *ver-*. Words of this type are common in Dutch and their prefixes often contain reduced schwas in casual speech (Chapter 2 of this dissertation; Pluymaekers et al., 2005). In all four experiments, primes and targets could differ in their degree of reduction: segments in the reduced tokens were shorter than in the unreduced tokens, and some segments were completely missing. Our unreduced tokens therefore represent tokens that are typically found in slow speech, while our reduced tokens represent tokens that can be found at a high speech rate in casual speech. We hypothesized that if participants react more quickly to targets showing the same degree of reduction as their primes, participants must have accessed the exemplars of these primes.

First, we examined whether exemplar effects arise for different speakers, by using two very different speakers. Speakers tend to differ in whether and how they reduce words at high speech rates in casual conversations (e.g., Chapter 6 of this dissertation). In Experiments 1 and 2, we investigated whether exemplar effects were larger if the difference in degree of reduction between the reduced and unreduced tokens was larger. Both experiments consisted of two subexperiments that were identical except for speaker voice.

Second, we investigated whether exemplar effects also occur if the repetition of words is less clear for participants in our experiment than in experiments reported in the literature. In the experiments in the literature, the number of trials varied from 48

(McLennan & Luce, 2005) to 436 (Craik & Kirsner, 1974) and between 50% and 100% of the words were repeated. Furthermore, the majority of these experiments used an explicit memory task (old-new judgments). Since it was clear to the participants that many words were repeated, they may have used a strategy in which they directly accessed exemplars. In Experiments 1 and 3 of our study, participants listened to 288 trials in which 33% of the words were repeated. In Experiments 2 and 4, the number of trials was increased to 800 and 864 respectively, and the percentage of repeated words decreased to approximately 12%.

Third, in previous experiments, listeners only heard one type of variation in the speech signal. For instance, in Bradlow et al. (1999), speech rate, amplitude, and speaker were varied, but each participant only heard one of these variations. In Experiments 3 and 4, we investigated whether exemplar effects also arise if the stimuli in the experiment differ in two indexical properties: degree of reduction and speaker voice.

Finally, our experiments differ from previous experiments in that the prime and the target were never completely identical. We chose to always have different productions of the same word in order to obtain results that are ecologically more valid. In real life, listeners are very unlikely to hear a given word produced twice in exactly the same way.

Experiment 1

Method

Participants

We tested 48 native speakers of Dutch aged 18 to 28 (mean 21 years). Nine were left-handed and ten were male. In this experiment, as in all other experiments presented in this paper, none of the participants reported any hearing impairment, all were paid for their participation, and they had not participated in any of the other experiments in this study.

Materials

The materials consisted of an equal number of existing Dutch words and pseudo-words; all were tri-syllabic infinitives. Half of them started with the prefix *be-* and the other half with *ver-*, (e.g., *beschrijven* ‘to describe’ and *vertolken* ‘to interpret’). The pseudo-infinitives did not contain phonotactically illegal phoneme sequences. All primes were existing infinitives and primes and targets represented the same word types.

Table 4.1 presents an overview of the number of the different types of stimuli. The experiment contained 48 pairs of primes and targets. We wished to keep the number of trials intervening between primes and targets small so that, at least in this respect,

Table 4.1: The number of stimuli presented in Experiment 1. The stimuli are broken down for prefix (*be-* or *ver-*), whether they function as primes, targets, or foils (which are subdivided in repeated and non-repeated foils), whether they are existing words or pseudo-words, whether they occur in Part 1 or Part 2 of the experiment, and whether within this part they occurred in Block 1 (B1) or Block 2 (B2).

Part	Prefix	Primes	Targets	Repeated foils		Foils		Total	
		Existing		Pseudo		Existing	Pseudo	Existing	Pseudo
		B1	B2	B1	B2	B1/B2	B1/B2	B1/B2	B1/B2
1	*be-*	12	12	12	12	12	12	36	36
	ver-	12	12	12	12	12	12	36	36
2	*be-*	12	12	12	12	12	12	36	36
	ver-	12	12	12	12	12	12	36	36
Total		48	48	48	48	48	48	144	144

our experiment resembled the experiments in the literature that showed exemplar effects. These 48 prime-target pairs were therefore divided over two parts. Each part consisted of two blocks: the first block had 24 primes and 48 foils and the second block contained the corresponding 24 targets and 48 foils. Each word type occurred in only one part of the experiment. In order to better hide the aim of the experiment, in the second block of each part, we repeated existing words (the targets) as well as 24 pseudo-infinitives (foils).

We used two pronunciation variants for the primes, targets, and foils: an unreduced one, carefully articulated at a slow speech rate, and a reduced one, with shorter and possibly absent segments. A prime and target represented either the same or a different pronunciation variant. All stimuli were recorded by two Dutch native speakers: one male (henceforth *Speaker A*) and one female (*Speaker B*). Stimuli were recorded over the course of multiple recording sessions. Since speakers typically do not produce casual speech in front of a microphone, we had to tell our speakers that the reduced stimuli had to sound as if uttered in casual speech. The instructions given to the speakers determined whether tokens were categorized as reduced or unreduced. For each word type that occurred as prime and target or as repeated foil, each speaker recorded several unreduced and reduced tokens (see Figure 4.1 for an example). From these tokens we selected the two best tokens for each pronunciation variant for a given speaker, so that primes and targets (and repeated foils) were always different tokens. For the remaining foils, we recorded either reduced or unreduced variants and selected the best token for a given speaker.

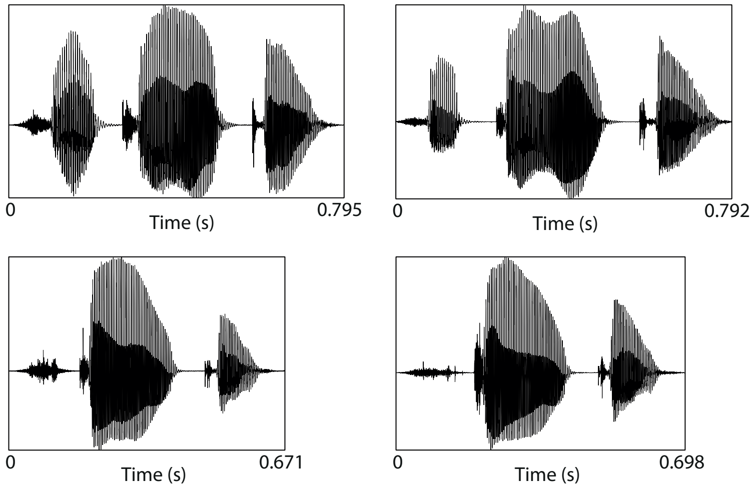

Figure 4.1: Examples of recorded stimuli: two unreduced and two reduced variants of *vertolken* /vərtɔlkə/ 'to interpret', produced by Speaker B.

We analyzed the recordings to examine whether the reduced and unreduced stimuli differed in degree of reduction, and whether the speakers varied in the degree of the pronunciation difference between reduced and unreduced stimuli. For the 384 recorded primes and target tokens, we created broad phonetic transcriptions using the forced alignment procedure described by Schuppler et al. (2011). From these transcriptions we extracted the duration of the whole word and determined whether schwa was present. The averages are presented in Table 4.2. Subsequently, we analyzed these two measures as dependent variable. For the presence of schwa, we fitted logistic mixed effects regression models and for word duration we fitted mixed effect regression models, with *prefix* (*be-* vs. *ver-*), *speaker* (Speaker A vs. B), and *variant* (reduced vs. unreduced) as fixed effects and *word type* (e.g., *vertolken* or *beschrijven*) as random effect. Table 4.3 shows the resulting models. As shown in Figure 4.2, reduced stimuli were significantly shorter than unreduced stimuli. This difference was larger for stimuli produced by Speaker B. Similarly, schwa was more often absent in reduced stimuli and in stimuli produced by Speaker B. The automatically generated transcriptions suggest that schwa was even frequently absent in Speaker B's unreduced realizations (10.4%). In general, our analyses clearly demonstrate that the reduced tokens are more reduced than the unreduced tokens. In addition, Speaker B shows a larger difference between the reduced and unreduced tokens than Speaker A.

Table 4.2: Average word duration and average percentages of word tokens produced with schwa split for speaker and pronunciation variant.

Measure	Speaker A		Speaker B	
	Reduced	Unreduced	Reduced	Unreduced
Word duration	588 ms	664 ms	485 ms	616 ms
Schwa presence	52.1%	100%	13.5%	89.6%

Table 4.3: Statistical models of the phonetic analysis of the recorded primes and targets.

	Word duration			Presence of schwa		
	β	t	$p <$	β	z	$p <$
Prefix (*ver-*)	56.78	4.01	0.0001	0.75	1.97	0.05
Speaker (Speaker A)	101.51	22.14	0.0001	2.33	6.21	0.0001
Variant (unreduced)	131.21	28.96	0.0001	4.69	10.11	0.0001
Speaker $\times$ variant	-53.20	-8.23	0.0001	-	-	*n.s.*

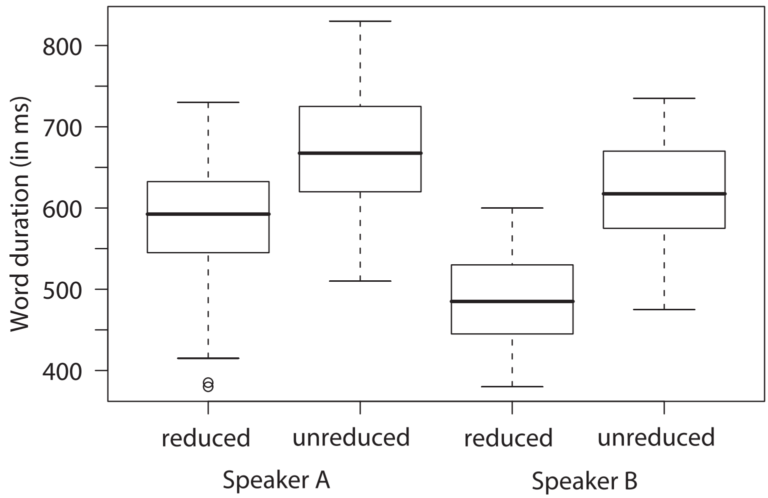

Figure 4.2: Boxplot of word duration split to speaker and pronunciation variant.

To test whether the differences between the unreduced and reduced tokens and between two speakers could also be perceived by naïve listeners, we conducted a rating experiment. We asked 50 participants aged between 18 and 29 (mean 21) to rate 60 foils and all primes and targets on a 6-point scale ranging from very unintelligible (rating score 1) to very intelligible (rating score 6). We created eight different pseudo-randomized orders of the stimuli, so that together the eight lists contained every token of each word (i.e., two primes and two targets produced by each speaker). Each participant heard one list in which each word type occurred once and which contained both reduced and unreduced stimuli produced by Speaker A as well as Speaker B. Since the rating scores were not normally distributed, we converted the scores to a factor in which scores 1, 2, and 3 were treated as unintelligible and scores 4, 5, and 6 as intelligible. We then fitted a logistic linear mixed effects model (Faraway, 2006) based on all primes and targets with *word type* and *participant* as crossed random effects. This model yielded significant effects of *speaker* (β = 2.89, z = 10.51, $p < 0.0001$) and *pronunciation variant* (β = 3.86, z = 9.46, $p < 0.0001$) and their interaction (β = -2.39, z = -3.51, $p < 0.001$), indicating that reduced items were less intelligible than unreduced items, especially if the items were produced by Speaker B (Speaker A: reduced 96.7% intelligible and unreduced 99.2%; Speaker B: reduced 72.5% and unreduced 98.7%). The two speakers clearly differed in their pronunciation and intelligibility of the reduced variants and it is therefore interesting to compare exemplar effects for these two speakers.

In the main experiment, the order in which the stimuli were presented to the participants was identical for those listening to Speakers A and those listening to Speaker B. We created four master lists for each speaker which tested priming for a given word only once. In each of the blocks of these lists, half of the primes or targets and approximately half of the foils were unreduced, and the other half were reduced. The four lists represented four different pseudo-randomizations of the trials. These randomizations had to obey four restrictions: (1) each block started with at least one foil; (2) each prime and target was followed by at least one foil; (3) at most eight words or eight pseudo-words occurred in succession; (4) prime and target were separated by a maximum of 100 trials (average: 67; range: 19 to 100). Trials with primes and targets were randomly assigned to one of the four possible combinations of the prime and target's pronunciation types: unreduced prime and unreduced target, unreduced prime and reduced target, reduced prime and unreduced target, and reduced prime and reduced target. For each master list, we created three other lists with the same words in the same order: together the four lists formed a set that represents all four possible combinations of the prime and target's pronunciation variants for each word. The combination of these four sets of four lists for each speaker resulted in 32 stimulus lists. Each list was randomly assigned to one or two participants with half of the participants receiving lists with Speaker A and the other half receiving lists with Speaker B.

Procedure

The experiment was conducted in a sound-attenuated booth, and participants were tested individually. Participants listened to the stimuli over headphones and performed a lexical decision task. They responded by pressing buttons on a button box; yes-responses were always given with the dominant hand and no-responses with the other hand. In each trial, one stimulus was presented and the next trial was initiated one second after a response was given or 3.5 seconds after the end of the stimulus. There was a pause between the two parts of the experiment, and one session lasted approximately 15 minutes.

Analyses

We analyzed the accuracy of the answers to the targets by means of logistic mixed effects models and the log-transformed response times to the targets by means of mixed effects regression models, with *word type* and *participant* as crossed random effects. Random slopes were tested for all fixed effects. The analysis of the response times was based only on those trials that received a correct response and for which the corresponding prime had also elicited a correct response. Response times for which the residual standard errors deviated more than 2.5 times from the values predicted by the statistical model were regarded as outliers and discarded. Subsequently, the model was refitted.

We tested the influences of three predictors of interest, namely *variant match*, which indicated whether the prime and target represent the same (i.e., match) or a different pronunciation variant (i.e., mismatch), *speaker* (Speaker A vs. B), and the distance in trials between the prime and target. In addition, we added several control predictors to the statistical models which, in earlier studies, have been shown to affect speech processing (e.g., Van de Ven et al., 2011): *trial number*, *experiment part* (part 1 vs. 2), the *pronunciation variant* of the target (reduced vs. unreduced), *prefix* (*be-* vs. *ver-*), the log-transformed *target duration*, the log-transformed response times to the prime (*RT prime*) and to the preceding trial (*RT preceding*), and the log-transformed *word frequency* (based on counts from the Spoken Dutch Corpus; Oostdijk, 2002). Interactions were tested for the predictors of interest only. All non-significant effects and random slopes were excluded from the model.

All correlating variables were orthogonalized before they were added to our statistical model: If a continuous predictor A was correlated with predictor B, we replaced predictor A by the residuals of a linear regression model predicting predictor A as a function of predictor B. If the correlation involved two continuous predictors, the influence of the least interesting one (in the example above, predictor B) was partialled out. Thus, in Experiment 1 we had four residualized predictors in our model: *frequency* (correlated with *prefix*), *target duration* (correlated with *speaker* and *prefix*), *RT preceding* (correlated with *speaker*), and *RT prime* (correlated with *RT preceding*, *speaker*, and *prefix*).

Results and discussion

Participants made errors in 5% of the target trials. Analysis of these trials did not show an effect of any of the variables of interest. The same holds for the errors in Experiment 2, 3, and 4.

As none of the participants made errors in more than 20% of the trials, none were excluded from our analyses of the response times. We restricted our analyses to those target words for which more than 80% of the responses were correct, which led to the exclusion of the word *bekransen* 'to garland'. Table 4.4 shows the statistical model based on the remaining 1980 trials (85.9% of all trials). Response times measured from word onset were 943 ms on average and ranged from 522 to 2375 ms. The effects of our control predictors showed that responses were faster to words carrying the prefix *be-* (mean: 913 ms) than *ver-* (971 ms); to words produced by Speaker B (mean: 879 ms) compared to Speaker A (1003 ms); and to words with a higher frequency of occurrence. In addition, responses were faster if the word itself or its prime was shorter. Finally, responses were faster the faster the response to the prime or the preceding trial.

Importantly, we found a significant main effect of *variant match*, which indicated that responses were faster if the prime and the target represented the same pronunciation variant (mean: 933 ms) compared to different variants (952 ms). *Variant match* did

Table 4.4: Statistical models for the response times of Experiments 1 and 2. Estimated standard deviation is indicated by *sd*.

	Experiment 1			Experiment 2		
Fixed effects	β	*t*	$p <$	β	*t*	$p <$
Prefix (*ver-*)	0.06	5.9	0.0001	0.08	7.23	0.0001
Speaker (Speaker A)	0.13	4.9	0.0001	0.08	4.91	0.0001
Word frequency	-0.01	-2.3	0.001	-0.02	-3.63	0.0001
Target duration	0.39	14.2	0.0001	0.40	19.26	0.0001
RT prime	0.16	8.5	0.0001	0.07	6.61	0.0001
RT preceding trial	0.21	10.2	0.0001	0.12	12.37	0.0001
Variant match (mismatch)	0.02	3.5	0.0001	-	-	*n.s.*
Random effects			*sd*			*sd*
Word type	intercept		0.03	intercept		0.03
Word type	RT preceding trial		0.08			-
Participant	intercept		0.09	intercept		0.10
Participant			-	target duration		0.09
Residual			0.15			0.14

not significantly interact with the random effects *word type* or *participant*, suggesting that the effect does not depend on a subset of word types or participants. Furthermore, the interaction between *variant match* and *speaker* was not significant, which suggests that the effect of *variant match* does not differ for the two speakers. We examined whether the effect was also significant for the two speakers separately and it was (Speaker A: β = 0.018, t = 2.29, $p < 0.05$; Speaker B: β = 0.021, t = 2.61, $p < 0.05$).

The effect of *variant match* appears to be reliable and indicates that exemplar effects arise even in experiments in which only 33% of the words are repeated. Although our phonetic analyses and the rating study clearly showed differences between the stimuli produced by the two speakers, the effect of *variant match* is similar for both speakers. A possible explanation is that each participant heard only one speaker. As listeners typically adapt very rapidly to a new speaker (e.g., Dahan, Drucker, & Scarborough, 2008), participants had probably already adapted to the speaker during the first block. Consequently, the differences between the speakers did not play a substantial role.

To further investigate the robustness of exemplars, in Experiment 2 we increased the number of non-repeated foils. This experiment consisted of 800 trials. As only 12% of the words were repeated, this setup closely approximates natural conversations, in which speakers avoid repetition by often replacing content words by pronouns. As we were not able to create large numbers of stimuli using the prefixes *be-* and *ver-* only, Experiments 2 also contained foils with the prefixes *in-*, *aan-*, and *ont-*. In addition, we increased the average number of trials between primes and targets. Since we thought these manipulations would make it harder to find exemplar effects, we tested more participants.

Experiment 2

Method

Participants

All 130 participants were native speakers of Dutch (21 male), aged between 18 and 31 (mean 21); 14 were left-handed.

Materials

We used the same stimuli as in Experiment 1 plus additional foils. The Dutch lexicon contains approximately 500 tri-syllabic infinitives with the prefix *be-* or *ver-* and a unique stem (Celex; Baayen, Piepenbrock, & Gulikers, 1995), including very low frequency infinitives (e.g., *verzoeten* ‘to sweeten’, *bewolken* ‘to cloud over’, and *verzagen* ‘to saw up’). The additional foils therefore also represented three other prefixes: *in-*, *aan-*, and *ont-*. In order to ensure that each of the five prefixes was presented

Table 4.5: The number of stimuli presented in Experiment 2. The stimuli are broken down for prefix (*be*-, *ver*-, *in*-, *aan*-, or *ont*-), whether they function as primes, targets, or foils (which are subdivided in repeated and non-repeated foils), whether they are existing words or pseudo-words, and whether they occurred in Block 1 (B1) or Block 2 (B2).

Prefix	Primes	Targets	Repeated foils				Foils		Total	
	Existing		Existing		Pseudo		Existing	Pseudo	Existing	Pseudo
	B1	B2	B1	B2	B1	B2	B1/B2	B1/B2	B1/B2	B1/B2
be-	24	24			24	24	32	32	80	80
ver-	24	24			24	24	32	32	80	80
in-			5	5	5	5	76	64	86	74
aan-			5	5	5	5	76	64	86	74
ont-			5	5	5	5	58	83	68	92
Total	48	48	15	15	63	63	274	274	400	400

160 times in the entire experiment, we added 32 infinitives starting with *be*- and *ver*- and 480 infinitives carrying either the prefix *in*-, *aan*-, or *ont*- (see Table 4.5). The number of existing and pseudo-infinitives starting with *in*-, *aan*-, or *ont*- was unequal, as only a limited number of existing *ont*-infinitives are available. To avoid repetition of the prefixes *be*- and *ver*- only, 30 foils with the prefix *in*-, *aan*-, and *ont*- were also presented twice.

Furthermore, we increased the number of trials between primes and targets and presented all stimuli in one part with two blocks. The first block consisted of 48 primes and 352 foils, and the second of 48 targets and 352 foils. The number of trials between the primes and targets was entirely random (average: 405; range: 79 to 765 trials).

Both speakers recorded all new foils only once, in either a reduced or unreduced pronunciation variant. As each participant only heard stimuli from one speaker, both occurrences of the 30 repeated fillers with the prefixes *in*-, *aan*-, and *ont*- were the same recording (token).

Procedure

We used the same procedure as in Experiment 1, except that all stimuli were presented in one part with a pause between the two blocks. A session lasted approximately 37 minutes.

Results and discussion

We analyzed the response times of Experiment 2 with the same method and predictors[1] as used in Experiment 1 except for the predictor *experiment part*. Three participants and the word *bekransen* 'to garland' were excluded from analyses as their error rates were above 20%. Table 4.4 shows the statistical model based on the remaining 5111 trials (81.9% of all trials). The average response time was 981 ms (range: 525 to 2108 ms). All control predictors that were significant in Experiment 1 were also significant in this experiment.

Although the statistical power of Experiment 2 was greater than that of Experiment 1 (due to the larger number of participants), we found no main effect of *variant match* or an interaction between *variant match* and *speaker*. Hence, in an experimental setting with a smaller proportion of repeated words and more trials between prime and target, targets preceded by primes representing the same or a different pronunciation variant are recognized equally quickly. This experiment therefore suggests that the exemplar effects found in Experiment 1 only arise in short experiments with little variation.

In Experiment 3, we further investigated the robustness of exemplar effects. We returned to the stimuli and set up of Experiment 1 and investigated whether exemplar effects are found if the prime and target may differ in two, instead of one, indexical property. The experiment tested four experimental conditions: (1) speaker match and variant match between primes and targets, (2) speaker match and variant mismatch, (3) speaker mismatch and variant match, (4) speaker mismatch and variant mismatch.

Experiment 3

Method

Participants

All 49 participants were native speakers of Dutch (six male), aged between 18 and 26 (mean 20); four were left-handed.

Materials and procedure

We used the same stimuli and recordings as in Experiment 1. In contrast to the previous experiments, half of the trials in a stimulus list were produced by Speaker A and the other half by Speaker B. Furthermore, whereas primes were either reduced or unreduced, targets were always reduced. We created three different pseudo-randomizations of the trials with the same restrictions as Experiment 1. We created

[1] We residualized the following predictors: *frequency* (correlated with *prefix*), *target duration* (correlated with *speaker* and *prefix*), and *RT preceding* (correlated with *speaker*).

eight lists for each randomization by varying the variant and speaker of the prime and the speaker of the target, which resulted in 24 different stimulus lists. A prime and target were again separated by 67 trials on average. The procedure and duration of a session were identical to those of Experiment 1.

Analyses

Except for the *pronunciation variant* of the target, which was always reduced, we used the same predictors[2] as in Experiment 1; note that *speaker* refers to the speaker of the target. In addition, we used the new predictor of interest *speaker match* (match vs. mismatch between the speaker of the prime and target).

Results and discussion

As the error rates of all participants were lower than 20%, no participants were excluded from analyses. The word *bekransen* 'to garland' had an error rate above 20% and was again omitted from further analyses. The statistical model of Experiment 3 was based on the remaining 2004 trials (85.2% of all trials) and is shown in Table 4.6. Response times were, on average, 943 ms (range: 540 to 1943 ms). The control predictors that were significant in the preceding experiments were again significant and showed similar effects. In addition, we found a difference between the two experiment parts, indicating that responses obtained in the second part (mean: 921 ms) were faster than in the first part (966 ms).

Importantly, neither *variant match* nor *speaker match* showed a significant effect, nor did they interact with each other or with *speaker*. An effect of *variant match* is absent although the statistical power of Experiment 1 was similar to the power of Experiment 3 (Experiment 1: 24 match responses from 24 participants for both speakers; Experiment 3: 12 match responses from 49 participants also for both speakers).

The difference found between Experiments 1 and 3 was supported by an analysis of the combined dataset of the responses to Speaker A in Experiment 1 and the responses in Experiment 3, which shows an interaction between *experiment* and *variant match* (β = -0.031, t = -2.7, $p < 0.05$): whereas *variant match* explains variance in Speaker A's part of Experiment 1, it does not in Experiment 3. An analysis of the combined dataset of the responses to Speaker B in Experiment 1 and the responses in Experiment 3 gave the same result (β = -0.023, t = -2.1, $p < 0.05$). These results suggest that exemplar effects were greater in Experiment 1, if they were present in Experiment 3 at all.

Neither Experiment 2 nor Experiment 3 showed exemplar effects. Nevertheless, we decided to conduct Experiment 4, which is a combination of Experiments 2 and 3:

[2] We residualized the following predictors: *frequency* (correlated with *prefix* and *experiment part*), *target duration* (correlated with *speaker*, *prefix*, and *experiment part*), RT preceding (correlated with prefix and experiment part), and RT prime (correlated with RT preceding and experiment part).

Table 4.6: Statistical models for the response times of Experiments 3 and 4. Estimated standard deviation is indicated by *sd*.

	Experiment 3			Experiment 4		
Fixed effects	β	*t*	$p <$	β	*t*	$p <$
Prefix (*ver-*)	0.05	3.9	0.0001	0.05	4.07	0.0001
Speaker (Speaker A)	0.08	7.9	0.0001	0.08	11.09	0.0001
Word frequency	-0.02	-2.3	0.05	-	-	*n.s.*
Target duration	0.40	6.3	0.0001	0.29	5.27	0.0001
RT prime	0.12	6.2	0.0001	0.11	3.76	0.001
RT preceding trial	0.14	8.4	0.0001	0.22	13.27	0.0001
Experiment part (part 2)	-0.05	-4.6	0.0001	-	-	*n.s.*
Random effects			*sd*			*sd*
Word type	intercept		0.05	intercept		0.04
Word type	speaker		0.05	RT prime		0.11
Participant	intercept		0.09	intercept		0.08
Participant	experiment part		0.06	RT prime		0.14
Residual			0.14			0.17

participants heard the targets in the same four conditions as in Experiment 3, while the experiment was identical to Experiment 2 in the number and diversity of the foils and in the distances between primes and targets. If the null results in Experiments 2 and 3 were due to Type II errors, we would expect to find exemplar effects in Experiment 4. Moreover, we can combine the results from Experiment 4 with those from Experiments 2 and 3 to see whether these increased datasets present evidence for exemplar effects.

Experiment 4

Method

Participants

The participants were 68 native speakers of Dutch (17 male), aged between 18 and 27 years (mean 21); seven were left-handed.

Materials and procedure

The stimuli were the same set as those presented in Experiment 2. In line with Experiment 3, half of the stimuli presented to each participant were produced by Speaker

A and the other half by Speaker B. Furthermore, as in Experiment 3, all targets were reduced. To make sure that not all reoccurring stimuli were reduced, we added 32 foils (16 existing and 16 pseudo-words) with the prefixes *be-* and *ver-* that also reoccurred and were unreduced in Block 2 (these were reduced or unreduced in Block 1). Each participant was presented with a stimulus list of 864 trials. The procedure was identical to that of Experiment 2. One session lasted approximately 40 minutes.

Results and discussion

To analyze the response times, we used the predictors[3] from Experiment 3, except for *experiment part*. All participants and target words were included in the analyses, except for the target words *bekransen* 'to garland' and *beschaven* 'to civilize', as they had error rates higher than 20%. The statistical model of Experiment 4, based on the remaining 2459 trials (75.3% of all trials), is presented in Table 4.6. The average response time was 956 ms (range: 549 to 2444 ms). The same significant control effects were found as in Experiment 3, with the exception of *word frequency*.

Similar to Experiments 2 and 3, we found no effects of the predictors of interest. Hence, in the experimental setting with the most variation and in which only a small proportion of stimuli were primed, no evidence for the use of exemplars was found.

In an analysis of the combined data of Experiments 2, 3, and 4, we found no main effect of *variant match* nor an interaction of *variant match* with *experiment*. This indicates that exemplars did not play a substantial role in any of these experiments.

Additional analysis of all experimental data

So far, we investigated the presence of exemplar effects by analyzing the datasets with two categorical predictors (*variant match* and *speaker match*). The variation between a reduced and an unreduced realization differs between speakers (see Figure 4.2), word types, and word tokens. We therefore also analyzed all datasets with a continuous predictor indicating the similarity in reduction between the prime and target, namely the absolute difference between the log-transformed duration of the prime and the log-transformed duration of the target. Only the analysis of Experiment 1 showed a significant main effect of this continuous predictor ($\beta = 0.17$, $t = 2.2$, $p < 0.05$), indicating that responses were faster if the duration difference between prime and target was smaller. In addition, none of the experiments showed an interaction between this predictor and speaker. These results indicate that even a more sensitive predictor shows no exemplar effects in Experiments 2, 3, and 4, and thus confirm the results obtained with the categorical predictors *variant match* and *speaker match*.

[3] We residualized the following predictors: *target duration* (correlated with *speaker* and *prefix*) and *RT prime* (correlated with *RT preceding* and *prefix*).

General discussion

In this chapter, we investigated exemplar effects in a series of priming experiments with lexical decision tasks, in which primes and targets represented the same or a different pronunciation variant. We examined the robustness of exemplar effects under more natural conditions than in experiments reported in the literature so far (e.g., Craik & Kirsner, 1974; McLennan & Luce, 2005; Palmeri et al., 1993) and did so in four ways. First, we studied the generalizability of exemplar effects over two very different speakers. Second, we investigated whether exemplar effects arise if the repetition of words is less clear for participants than in experiments showing exemplar effects reported in the literature. Third, we investigated if exemplar effects arise when listeners are exposed to not one but two types of pronunciation variation in the experiment (i.e., pronunciation variant and speaker voice). Finally, in contrast to earlier studies, primes and targets were never completely identical.

In Experiment 1, 33% of the 288 words were repeated and each participant only listened to one of the two speakers. This experiment showed a clear exemplar effect: responses were faster to targets that represented the same pronunciation variant as their primes. In contrast to earlier studies (e.g., Mattys & Liss, 2008; McLennan et al., 2003; McLennan & Luce, 2005; Palmeri et al., 1993), in our experiments, primes and targets were always different recordings, even when they represented the same pronunciation variant produced by the same speaker. The results of Experiment 1 thus show that even if the target is not completely identical to the prime, its processing can be facilitated by the exemplar formed by its prime.

The exemplar effects arose regardless of the number of trials intervening between the prime and target. This shows that the priming effects remained constant during the first five minutes following the presentation of a prime. In this respect, our results differ from those obtained by Palmeri et al. (1993), who found that exemplar effects were only present if the interval between prime and target was smaller than 64 trials. A likely explanation for this difference in results is that Palmeri et al. (1993) used an old-new judgment task while we used lexical decision.

The exemplar effect was significant for both speakers, who clearly differ in their pronunciations and intelligibility. Hence, exemplar effects can be found for very different speakers. Our results appear to contrast with those obtained by Mattys & Liss (2008), who found that the size of exemplar effects depends on the level of intelligibility of the speakers: participants who listened to dysarthric speakers showed longer response times and larger exemplar effects. Following McLennan & Luce (2005), the authors argue that exemplar effects are larger if performance latencies are longer. In our study, the less intelligible speaker did not elicit longer response times. Therefore, these authors would correctly predict similar exemplar effects for both speakers.

Experiment 1 provides data that are informative about speech processing in natural conditions. The percentage of words repeated within an interval of 100 words in lectures and classes from the Spoken Dutch Corpus (i.e., component n of the corpus; 53045 words) is as high as 46.6% (18.8% if only content words are taken into account). Our results thus hold for a substantial number of word tokens that listeners hear during classes and when listening to, for instance, news bulletins.

While the repetition of words in Experiment 1 was already less obvious than in previous studies, participants may still have noticed it, which may have increased their use of exemplars. This was further investigated in Experiment 2, where the repetition of words was made less obvious by simultaneously increasing the number of trials between a prime and target from 67 to 405 on average and reducing the proportion of reoccurring words to 12%. This more closely approximates natural conversations, in which the frequent replacement of content words by pronouns decreases the repetition of content words. Although the statistical power of Experiment 2 was greater than Experiment 1 (due to the larger number of participants), Experiment 2 showed no exemplar effects. This indicates that exemplars effects are negligible when the repetition of words is less clear for participants.

The difference in delay between primes and targets in Experiments 1 and 2 may explain why we found priming effects in Experiment 1 but not in Experiment 2. The decay of the primes' exemplars (or of their activation) may have been too large at the moment the target was presented in Experiment 2. Only a small percentage of prime-target pairs (1.1%) were separated by maximally 100 trials and only 9.8% was separated by maximally 180 trials. Moreover, the block of primes was separated from the block of targets by a pause. Earlier findings that exemplar effects can be present even after one week (Goldinger, 1996) seem to contradict this explanation. However, exemplars may contain information about the context in which the occurrence was heard (e.g., the laboratory). If so, words presented in the laboratory after one week are more similar to exemplars with the same context information than exemplars encountered in a different context in that intervening week. When participants re-entered the laboratory after a week in Goldinger's experiment, they may have reactivated the exemplars specific to that laboratory. Consequently, at the moment a target word was presented, the number of different activated exemplars was probably larger in Experiment 2 than in Goldinger's experiment after one week, resulting in smaller priming effects. Further research is necessary to test this explanation.

Like Experiment 1, Experiment 2 did not show an effect of the distance between the prime and the target. This was probably due to the high number of prime-target pairs that were separated by a large number of trials. These pairs may not have shown priming effects at all, precluding a main effect of or interaction with the distance between prime and targets.

Experiment 3 studied the role of exemplars if the speech signal contained more than one type of variation (i.e., degree of reduction and speaker voice). Although the proportion of reoccurring words was the same as in Experiment 1, we found no effect of the similarity in pronunciation variant nor of the similarity in speaker voice. The statistical powers of Experiment 1 and 3 were the same, as were the average response latencies. A possible explanation comes from the earlier finding that if memory load is higher, listeners tend to use less acoustic detail in speech comprehension (e.g., Mattys & Wiget, 2011). The combination of two types of variation in Experiment 3 made Experiment 3 more demanding than Experiment 1, since the greater variation made linking the acoustic signal to semantic representations more effortful for the participants. As a consequence, participants may have paid less attention to acoustic similarity.

Another possible explanation for the absence of exemplar effects has to do with the difference in reduction patterns between our speakers. As illustrated in Figure 4.2, a reduced pronunciation produced by Speaker A may be very similar in word duration to an unreduced pronunciation produced by Speaker B. Primes and targets which constitute a variant mismatch may therefore be very similar, whereas those that constitute a match may be very dissimilar. This may explain why we did not find an effect of variant match. In order to test this hypothesis, we conducted additional analyses (presented above, after Experiment 4) investigating whether the difference in word duration between the prime and target predicts reaction times. This appeared not to be case for Experiments 2, 3, and 4, which may be taken as evidence that the absence of an effect of variant match in Experiment 3 is not due to differences in reduction patterns between the two speakers. However, the speakers may not only differ in their speech rate in the two pronunciation variants, but also in their exact realization of the different segments of these variants. For instance, Speaker B may always weaken liquids after vowels, whereas Speaker A may produce them very clearly, at least in the unreduced tokens. Therefore, the absence of an effect of pronunciation variant in Experiment 3 may be due to substantial differences between the tokens representing one single variant produced by the two speakers. This explanation implies that listeners do not classify a given word token as unreduced or reduced depending on the speaker, which is in line with models assuming acoustically detailed representations for pronunciation variants.

This second possible explanation can also account for why Experiment 3 did not show a main effect of speaker match. If substantial exemplar effects only arise if a speaker match is combined with a variant match, they can only be expected in one of the four conditions in the experiment. Possibly, our experiment had too little power to show the difference between this condition and the three other ones. Future research has to show which of these explanations is most likely.

Regardless of the underlying cause, the absence of exemplar effects in Experiment 3 raises the question of what role exemplars play in speech comprehension in daily life. Most speech that people perceive is produced in spontaneous conversations involving several speakers in which degree of reduction varies greatly. The absence of exemplar effects in Experiment 3 suggests that, under these conditions, abstract lexical representations play a more important role than exemplars.

Finally, to complete the series of experiments, in Experiment 4 we examined the two types of variation simultaneously, in an experiment in which only a small proportion of stimuli reoccurred. In line with the results of Experiments 2 and 3, this experiment also showed no exemplar effects. These results confirm our findings that exemplar effects are absent in experimental setups like those of Experiments 2 and 3.

The results of this study have implications for modelling spoken word comprehension. The absence of exemplar effects in Experiments 2 to 4 disqualifies pure exemplar models but leaves hybrid models a viable option. Hybrid models do require further specification to explain under which conditions exemplars can affect comprehension. Our findings can also be accounted for in a model assuming only abstract lexical representations, provided that it assumes domain-general episodic memory. The exemplar effect found in Experiment 1 should then be reinterpreted as an episodic effect that arose because it was so obvious to participants that many words were repeated: participants were encouraged to base decisions on episodic rather than abstract representations.

In conclusion, we conducted four priming experiments, and found exemplar effects in only the simplest experiment with no speaker variation and the largest proportion of repeated words. In spontaneous conversations, listeners may hear more than one speaker and content words are often replaced by pronouns. Hence, this paper suggests that, in a situation where more variation is available to the listener, like natural conversation, exemplars play a smaller role than has been previously assumed.

The role of morphology in acoustic reduction

Chapter 5

This chapter is a reformatted version of:
Iris Hanique and Mirjam Ernestus (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio*, *2012(2)*, 147–164.

Abstract

This chapter examines the role of morphological structure in the reduced pronunciation of morphologically complex words by discussing and re-analyzing data from the literature. Acoustic reduction refers to the phenomenon that, in spontaneous speech, phonemes may be shorter or absent. We review studies investigating effects of the repetition of a morpheme, of whether a segment plays a crucial role in the identification of its morpheme, and of a word's morphological decomposability. We conclude that these studies report either no effects of morphological structure or effects that are open to alternative interpretations. Our analysis also reveals the need for a uniform definition of morphological decomposability. Furthermore, we examine whether the reduction of segments in morphologically complex words correlates with these segments' contribution to the identification of the *whole* word, and discuss previous studies and new analyses supporting this hypothesis. We conclude that the data show no convincing evidence that morphological structure conditions reduction, which contrasts with the expectations of several models of speech production and of morphological processing (e.g., WEAVER++ and dual-route models). The data collected so far support psycholinguistic models which assume that all morphologically complex words are processed as complete units.

Introduction

One common type of variation in spontaneous speech is the reduced pronunciation of words. For instance, *ordinary* may be pronounced as [ɑnɹi] and *apparently* as [pʰɛɾĩ]. As reported by Johnson (2004), over sixty percent of the words in spontaneous American English deviate from their full form in at least one segment, and six percent of the words are pronounced with at least one syllable missing. This highly frequent phenomenon also occurs in other languages (see Ernestus & Warner, 2011, for an introduction to the phenomenon).

Whether segments are reduced depends on many variables, like the rate of speech (e.g., Dalby, 1984), the speech style (e.g., Van Son & Pols, 1999), the surrounding segments (e.g., Van Bergem, 1994), and the predictability of the segment or word (e.g., Bell et al., 2009). Knowledge about these variables informs us about the speech production process and the nature of the mental lexicon.

In this chapter, we investigate whether the production of reduced pronunciation variants in conversational speech is also affected by morphological structure. We formulate several hypotheses about how morphological structure may affect the degree of reduction and discuss new analyses and data from the literature bearing on these hypotheses.

Our first hypothesis concerns the effect of the repetition of morphemes on their degree of reduction. Previous studies have shown that words tend to be more reduced the more often they are repeated in a conversation (e.g., Fowler & Housum, 1987). If morphological structure plays a role in acoustic reduction, this may also hold for the repetition of morphemes. That is, morphemes may be more reduced if they are repeated, even in different words.

Our second hypothesis is based on previous studies which show that segments tend to be less reduced if they are more relevant for the identification of a linguistic unit (e.g., Van Son & Pols, 2003). If morphemes play a role in acoustic reduction, we expect that a segment's degree of reduction also depends on its contribution to the identification of its morpheme in morphologically complex words. A segment is more important for the identification of a morpheme if it is its only segment (i.e., for single segment affixes, like the plural *s* in *words*) or its initial segment than if it is its final segment. Segments forming single segment affixes as well as morpheme-initial segments are therefore expected to be less reduced than segments at the end of longer morphemes (Losiewicz, 1992).

Importantly, words differ from each other in their morphological decomposability. The effect of a word's morphological structure on acoustic reduction is expected to depend on this morphological decomposability. Semantically opaque words (e.g., *department*) are more difficult to parse in their morphemes than semantically transparent words (e.g., *attachment*). Hay (2003), however, argues that also semantically transparent words can differ in their decomposability. The effect of a word's morphological structure on acoustic reduction is expected to be greater for words that are more easily decomposable.

Hay (2003) quantified decomposability as the relation of the frequency of the word to the sum of frequencies of the word's base and what she called its inflectional variants (e.g., for *softly*, the sum of frequencies of *soft*, *softer*, and *softest*). If the word frequency is higher than the inflectional variants frequency, Hay assumes the word to be less decomposable and that its morphological structure only plays a minor role in processing. In contrast, if the inflectional variants frequency is higher than the word frequency, Hay assumes that the word is more easily decomposable. Then, morphological structure could play a more important role in speech production, and segments that are crucial for the identification of morphemes would be less likely to be reduced. This hypothesis has also been addressed by other researchers (Bürki, Ernestus, et al., 2011; Schuppler et al., 2012), who all use different quantifications of decomposability.

If, contrary to the hypotheses formulated above, morphological structure has no role in reduction, the reduction of morphologically complex words is expected to be conditioned by the same factors as monomorphemic words. One such factor is a segment's importance in distinguishing the *complete* word from its competitors (henceforth *word information load*; e.g., Van Son & Pols, 2003). Segments that are more relevant for the identification of the *complete* word are hypothesized to be less reduced.

After having discussed data bearing on these hypotheses, we discuss the implications for models of speech production. Models differ in the way they assume words are represented in the mental lexicon and processed during speech production: Some models assume that complex words are stored and processed as complete units (e.g., Bybee, 2001; Skousen, 1989), whereas other models assume that regular complex words are processed on the basis of their morphemes (e.g., Chomsky & Halle, 1968; Pinker, 1991; Taft & Ardasinski, 2006), or they combine the two ways of processing (e.g., Levelt et al., 1999; Schreuder & Baayen, 1995). We examined which existing psycholinguistic model best explains the available data on the role of morphological structure on acoustic reduction.

In short, this paper provides an overview and discussion of studies that investigated the role of morphological structure in acoustic reduction. We first examine the question whether morphemes tend to be more reduced if they are repeated, even in different word types. Then, we investigate the hypothesis that a segment is less reduced if it forms an affix by itself, and subsequently focus on the hypothesis that reduction is influenced by the word's morphological decomposability. We thereafter discuss data supporting the contrasting hypothesis that acoustic reduction is not conditioned by morphological structure, but rather by a segment's word information load. In addition to reviewing existing literature, we present new analyses of our own data. Finally, we discuss the implications of our findings for several psycholinguistic models of speech production.

The repetition of morphemes

Several studies have reported that words tend to be shorter if they are repeated in a spontaneous conversation (e.g., Fowler & Housum, 1987). If morphological structure plays an important role in acoustic reduction, we expect that this also holds for morphemes within complex words.

Viebahn, Ernestus, & McQueen (2012) examined whether Dutch past participles that occur closely together (but do not necessarily directly succeed each other) in casual speech tend to be reduced in the same degree. Importantly, they investigated if the similarity in reduction between these co-occuring words is affected by the similarity of their morphological structure. The authors differentiated three degrees of morphological similarity: identical words, different words starting with the same prefix, and completely different words also carrying different prefixes. This study focused on schwa in the prefix of 1848 past participles extracted from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000) and the interview and read speech components of the Spoken Dutch Corpus (Oostdijk, 2002). Results showed that schwa tended to be short if the schwa in the preceding past participle was also short, but only if the two co-occurring past participles were tokens of the same word. That is, the schwa in, for instance, the past participle *gelopen* 'walked' tended to be as short as in a preceding token of *gelopen*, but not as short as in another preceding past participle, like *gefietst* 'cycled'. Hence, repetition of just the prefix does not influence its degree of reduction.

The unit that plays an important role in speech production appears to be the complete complex word. This finding challenges the hypothesis that the morphological structure of complex words plays a large role in their pronunciation.

Single segment affixes

As explained in the Introduction, if morphological structure plays a role in acoustic reduction, we would expect that a segment tends to be less reduced if it forms an affix by itself than if it is positioned at the end of a longer morpheme. Losiewicz (1992) tested this hypothesis in a production experiment studying the correlation between the reduction of English word-final /t/ and /d/ and these segments' morphological status. The stimuli consisted of six minimal word pairs (see Table 5.1): one word of each pair was a regular past tense verb ending in the suffix *-ed* (either pronounced as [t] as in *tacked* or as [d] as in *swayed*), in which /t/ or /d/ functioned as a single segment affix, while the other word was a monomorphemic homophone ending in stem-final /t/ or /d/. These twelve words were placed in lists and read aloud once by sixteen participants. Losiewicz found that the morphemic /t/ or /d/ of the regular past tense verbs was on average five milliseconds longer than the non-morphemic /t/ or /d/, which suggests that a segment's duration is affected by its morphological status (i.e., whether it forms a single segment affix or is the final segment of the stem).

Table 5.1: Word pairs studied by Losiewicz (1992) with the accompanying word frequencies from CELEX and the difference in duration between the suffix and the stem-final segment. For the word pairs printed in bold, the word with the shortest final segment is also the word with the highest frequency of occurrence.

Word pair		Word frequency		Duration difference
swayed	**suede**	**25**	**58**	**+11**
tacked	**tact**	**5**	**76**	**+11**
spayed	**spade**	**2**	**51**	**+4**
rapped	**rapt**	**9**	**36**	**+3**
massed	mast	11	0	+3
bussed	bust	0	2	-2

Losiewicz selected her stimuli so that they had frequencies below 10 tokens in Carroll, Davies, & Richman (1971) and Francis & Kucera (1982). These frequency lists are based on relatively small corpora: one and five million word tokens, respectively. Currently, frequency lists are available based on larger corpora and consequently are more reliable. For instance, the CELEX database (Baayen et al., 1995) provides counts based on 17.9 million word tokens. In Table 5.1, we list the word frequencies from CELEX. These counts show that for four of the six word pairs (in bold), the word with the shortest final segment has the highest frequency of occurrence[4]. Since many studies have shown that more frequent words tend to be more reduced (e.g., Bell et al., 2009; Pluymaekers et al., 2005), the morphological effect found by Losiewicz may also be a word frequency effect.

A second study that examined the role of a segment's morphological status in acoustic reduction is Schuppler et al. (2012). Initially, the authors focused on word-final /t/ in all Dutch content words and extracted 5130 word tokens ending in /t/ from the Ernestus Corpus of Spontaneous Dutch. In contrast to Losiewicz, they did not find that the morphological status of /t/ significantly influenced its presence or absence.

Schuppler et al. (2012) subsequently focused only on Dutch singular present tense verb forms ending in /t/. In regular Dutch verbs, the first person singular is equal to the stem of the verb (e.g., /fits/ *fiets* 'cycle'), and the third person singular is created by adding the suffix /t/ (e.g., /fits+t/ *fietst* 'cycles'). Importantly, if a stem ends in /t/ or /d/, this stem-final segment and the third person singular suffix /t/ are degeminated, which results in the pronunciation of a single /t/ (e.g., /nit+t/, [nit] *niet* 'staples'; /bid+t/ [bit] *biedt* 'offers'). If the stem ends in /d/, the third person

[4] Frequency counts based on the 85 million words of the spoken component of the Corpus of Contemporary American English (Davies, 2008) also show that for four word pairs the most frequent word is the one with the shortest final segment.

singular is spelled with *dt*. This form therefore differs from the first person singular in orthography and the two forms can be easily identified in a corpus of spontaneous speech with orthographic transcriptions. Note that the crucial difference between the first and third person singular verb forms is that all first person forms end in a [t] that is part of the stem only (e.g., /bid/ [bit] *bied* 'offer'), whereas all corresponding third person forms end in a [t] that is part of the stem and is a suffix (e.g., /bid+t/ [bit] *biedt* 'offers'). Schuppler et al. (2012) analyzed 366 tokens. In agreement with Losiewicz (1992), they showed that suffix /t/ was significantly less likely to be absent than non-suffix /t/.

However, like Losiewicz's results, this finding by Schuppler et al. (2012) should be interpreted carefully. It is again based on a dataset consisting of a small number of tokens, which represent nine word types only. Moreover, this dataset is not balanced, as for the third person singular present tense verb forms *antwoordt* 'answers', *biedt* 'offers', and *snijdt* 'cuts', it does not contain the corresponding first person forms (i.e., *antwoord*, *bied*, and *snijd*). In addition, the word *wordt* is often followed by a /d/-initial function word (e.g., *wordt dat* 'becomes that'). Due to degemination, this type of sequence is nearly always produced with only one alveolar stop, independently of whether the stop of *wordt* is present. Furthermore, the /t/ of the first person verb form *houd* 'love' is never produced in casual speech, which is reflected in informal writing (as in *ik hou van je* 'I love you'). Finally, the word type *vind* 'find' often occurs in highly frequent word combinations (e.g., *vind ik* 'believe I'), in which the /t/ is often absent, which is sometimes also reflected in writing (e.g., *vinnik*). These word combinations are likely to be lexicalized items, in which word-final /t/ is no longer present. Due to these problems with the dataset, Schuppler and colleagues' (2012) findings do not convincingly support the hypothesis that the morphological status of a segment influences its degree of reduction.

We also investigated the role of a segment's morphological status using the dataset from Chapter 2 of this dissertation, which consists of Dutch past participles. Most Dutch regular past participles consist of the prefix *ge-* /xə/, a verbal stem, and the suffix /t/ (e.g., /xə+wɛns+t/ *gewenst* 'wished'). In past participles of verbal stems ending in /t/, /t/ is analyzed as resulting from degemination of the stem-final /t/ and suffix /t/ (e.g., /xə+prat+t/ [xəprat] *gepraat* 'talked'). We focused on word-final /t/ and analyzed 1166 tokens of past participles ending in /t/ (in 165 tokens, /t/ was part of the stem and the suffix, as in /xə+prat+t/ [xəprat] *gepraat* 'talked', and in 1001 tokens, /t/ only represented the suffix, as in /xə+wɛns+t/ *gewenst* 'wished'). All tokens were extracted from the Ernestus Corpus of Spontaneous Dutch and the interview and read speech components of the Spoken Dutch Corpus. Analyses showed that the morphological status of /t/ did not affect its duration, but it did affect its presence: word-final /t/ was more likely to be absent if it was not part of the stem but only represented the suffix. This is unexpected given the hypothesis that segments forming single segment affixes are generally less likely to be reduced. We will return to this finding in the section on word information load.

This overview shows that, so far, no study has convincingly demonstrated an effect of a segment's morphological status on acoustic reduction. Some studies showed no effects of morphological status at all (the overall dataset of Schuppler et al. (2012), and our new analyses reported in this section), while the effects observed in other studies are open to alternative interpretations (Losiewicz, 1992; the verb form dataset in Schuppler et al., 2012).

Morphological decomposability

As explained in the Introduction, morphological structure may only play a role in the pronunciation of words that are easily decomposable. Morphemes are more important in easily decomposable words, and only in these words, may segments at morpheme boundaries be expected to be less often reduced. In order to investigate this hypothesis, researchers have quantified decomposability in several ways.

According to Hay (2003), a word's morphological decomposability is reflected by the relation between its own frequency and the cumulative frequency of its inflectional variants. Hay hypothesized that if a word occurs more frequently than its variants, it is less decomposable and its segments at morpheme boundaries are therefore more likely to be reduced. Hay tested this hypothesis on the basis of /t/ in pairs of English adverbs (see Table 5.2), of which one word is more frequent than its inflectional variants (e.g., the frequency of *swiftly* is higher than the cumulative frequency of *swift*, *swifter*, and *swiftest*), and the other word is less frequent than its variants (e.g., the frequency of *softly* is lower than the cumulative frequency of *soft*, *softer*, and *softest*). These adverbs were placed at the end of sentences, which were read aloud four times by six participants. For each participant, the duration of /t/ in the first production of a word was ranked with respect to the duration of /t/ in the first production of the other word in the pair. The second, third, and fourth productions were ranked similarly,

Table 5.2: Word pairs used by Hay (2003). The accompanying word frequencies and cumulative frequencies of the inflectional variants are based on Celex.

Word	Word Frequency	Variants Frequency	Word	Word Frequency	Variants Frequency
diligently	35	31	arrogantly	17	116
frequently	1036	396	recently	1676	1814
swiftly	268	221	softly	440	1464
exactly	2535	532	directly	1278	1472
listless	42	19	tasteless	30	1072

and then average rankings were calculated for each word. Analysis of these average rankings showed that /t/ was more reduced in words that are less decomposable.

There are, however, several reasons for interpreting these results with caution. First, as shown in Table 5.2, this study is only based on five word pairs. Second, the analyses are based on rankings of the durations rather than on the durations themselves. Third, it is unclear why Hay's decomposability measure does not take into account the frequency with which the stem of the adverb occurs in derived words (e.g., for *soft*, the frequency of *softness* was not taken into account).

Following Hay (2003), Schuppler et al. (2012) also investigated the correlation between decomposability and reduction. Their study analyzed the presence versus absence of affixal /t/ in Dutch third person singular present tense verb forms (e.g., /ʊon+t/ *woont* 'lives'). Their dataset consisted of 2110 verb forms extracted from the Ernestus Corpus of Spontaneous Dutch. They quantified a word's decomposability as the ratio of its word frequency (e.g., *woont* 'lives') and the frequency with which its stem occurs without affixal /t/, that is, the frequency of the first person singular present tense (e.g., *woon* 'live'). In contrast to Hay (2003), they found that /t/ was *more* likely to be absent in words that are more decomposable.

In order to investigate whether this dataset would also show an effect if the quantification of decomposability was more in line with Hay's definition, we re-analyzed the data. We quantified the decomposability of the third person singular present tense as the frequency of this word divided by the cumulative frequency of all words in the verbal paradigm (e.g., we investigated the effect of the frequency of *woont* relative to the sum of frequencies of *gewoond, wonen, wonend, wonende, woon, woonde, woonden,* and *woont*). Since it is unclear whether decomposability is only sensitive to word forms containing exactly the same form of the stem, we calculated two measures of decomposability for irregular verbs: one based on only those verb forms that contain exactly the same form of the stem (e.g., *zoekt* 'searches' versus *zoek, zoeken, zoekend, zoekende,* and *zoekt*), and one based on all verb forms of the verbal paradigm (e.g., *zoekt* 'searches' versus *gezocht, gezochte, zocht, zoek, zoeken, zoekend, zoekende,* and *zoekt*). None of these relative frequencies significantly correlated with the presence versus absence of /t/. This raises the question whether decomposition does play a role, as claimed by Schuppler et al. (2012).

We observed that the decomposability measure defined by Schuppler et al. (2012, e.g., the frequency of *zoekt* relative to the frequency of *zoek*) correlates (r = -0.19) with the stem frequency of the verb form, that is with the sum of frequencies of the verb forms containing exactly the same stem (e.g., the sum of frequencies of *zoek, zoeken, zoekend, zoekende,* and *zoekt*). We examined which frequency measure (their decomposability measure or stem frequency) better predicts the absence of /t/ by conducting a regression analysis with both measures as predictors. We orthogonalized these measures by replacing stem frequency with the residuals of a regression model predicting stem frequency as a function of the decomposability measure. The model predicting absence of /t/ showed significant effects of both measures.

We then orthogonalized the measures by replacing the decomposability measure with the residuals of a regression model that predicted this decomposability measure as a function of stem frequency. Absence of /t/ correlated with stem frequency but not with the residuals of the decomposability measure, which suggests that stem frequency better predicts reduction of /t/ than the decomposability measure. The stem frequency effect shows that /t/ is more likely to be absent at the end of words with higher stem frequencies. An explanation may be that all words belonging to highly frequent paradigms can be retrieved more easily and quickly, resulting in more reduced pronunciations.

Another reason for why the effect reported by Schuppler et al. (2012) may in fact not be driven by decomposability is that, of the 155 word types in their dataset, many of the third person singular present tense verb forms are homophones of the corresponding first person singular forms (e.g., /ʋet/ [ʋet] *weet* 'know' versus /ʋet+t/ [ʋet] *weet* 'knows') or differed substantially from the first person singular form (e.g., /hɛb/ [hɛp] *heb* 'have' versus /heft/ [heft] *heeft* 'has'). Schuppler et al. (2012) did not differentiate between the different syntactic functions of a form, and considered the frequency of each syntactic function to be the total frequency of the corresponding form. For homophones, the decomposability measure therefore does not divide just the frequency of the third person singular by the frequency of the first person singular, but the total frequency of the form, that is, of the first, second, and third person singular by exactly this same sum frequency. Hence, it cannot be compared to the measure for non-homophone verb forms. For the irregular third person forms, such as *heeft*, there is another problem with the decomposability measure: word-final /t/ is completely predictable given the preceding part of the word and therefore does not itself carry morphological information. If both the homophone and irregular items are excluded from the dataset, the decomposability measure used by Schuppler et al. (2012) is no longer significant. We conclude that whether the data analyzed by Schuppler et al. (2012) show morphological effects remains open to discussion.

We also examined possible effects of decomposability on the basis of word-final /t/ in Dutch past participles (Chapter 2 of this dissertation). We used the two decomposability measures with which we re-analyzed the data provided by Schuppler et al. (2012, see above). The presence versus absence and duration of /t/ were not significantly affected by either of these relative frequencies.

Finally, Bürki, Ernestus, et al. (2011) studied the influences of a word's morphological decomposability on the reduction of French word-internal schwa (as in /ʁə+diʁ/ *redire* 'to say again'). The authors extracted 4294 words with an internal schwa from the ESTER corpus (Galliano et al., 2005), which contains radio-broadcast news. Their measure for decomposability was different from the measures described above: five native speakers of French were asked to assess each schwa on a five-point scale from being clearly morpheme-internal (as in /ʃəmɛ̃/ *chemin* 'way') to being clearly at a morpheme boundary (as in /ʁə+diʁ/ *redire* 'to say again'). For each word, the average value over assessors was used as a measure of the decomposability of the word. Bürki, Ernestus, et al. (2011) observed that the presence versus absence and duration of schwa did not correlate with this decomposability measure.

In conclusion, so far, only two studies have reported an effect of decomposability and they did so by using different quantifications of decomposability. Since there are no clear theoretical reasons to prefer one quantification over another and since the effects observed may in fact not be morphological in nature, we conclude that so far the literature does not provide any convincing evidence for the role of a word's morphological decomposability in reduction.

Word information load

Since we have, so far, found no convincing evidence that morphological structure plays a role in the reduced pronunciation of morphologically complex words, this section examines whether an important factor conditioning reduction in monomorphemic words may also be relevant for complex words. Previous studies (e.g., Van Son & Pols, 2003) have shown that a segment tends to be less reduced the more it contributes to the identification of a word, and therefore to distinguishing the word from other words. If morphological structure does not play a role, this word information load hypothesis should also hold for segments in morphologically complex words: the less a segment contributes to distinguishing the complete word from other words, the more it may be reduced.

In the section on single segment affixes, we presented data that support this hypothesis. We reported that word-final /t/ in Dutch past participles is more reduced if it only represents a suffix (i.e., it is not both part of the stem and a suffix). This would be unexpected if morphological structure played a role in reduction, since the /t/ is the affix's only segment, and reduction of /t/ would therefore lead to weaker acoustic cues to morphological structure. In contrast, this finding is expected if word information load is relevant: as most regular Dutch past participles end in suffixal /t/, this segment is highly predictable and thus does not contribute much to distinguishing the word from its competitors.

Pluymaekers, Ernestus, Baayen, & Booij (2010) investigated whether morphological structure or word information load is a better predictor for reduction of the cluster /xh/ in the Dutch word-final string /əxhɛit/ *-igheid*. Dutch words ending in *-igheid* either have the morphological structure stem+*igheid* (e.g., *vast+igheid* 'security') or stem+*heid*, in which case the stem ends in *-ig* (e.g., *zuinig+heid* 'thriftiness'). If morphological structure plays a role in reduction, /xh/ is expected to be less reduced in *+heid* words than in *+igheid* words, since in *+heid* words, the cluster /xh/ contains a morphological boundary between the stem and *-heid*, whereas in *+igheid* words, /xh/ is a cluster within the morpheme (Booij, 1995). If word information load plays a more important role, the opposite reduction pattern is expected. In *+igheid* words (like *vastigheid*), the cluster /xh/ eliminates inflectional variants (like *vaste* 'solid') and compounds (like *vasteland* 'main land' and *vastenavond* 'Mardi Gras'). In contrast, since stems ending in *-ig* (like *zuinig*) do not occur in compounds, *+heid* words based on these stems solely compete with a few inflectional variants (like *zuinige* 'thrifty' and *zuinigste* 'most thrifty'). As a consequence, the cluster /xh/ has a much lower word information load in stem+*heid* than in stem+*igheid* and is therefore hypothesized to be more often reduced in *+heid* words. Pluymaekers et al. (2010) extracted 432 tokens ending in *-igheid* from the read speech component of the Spoken Dutch Corpus and demonstrated that the duration of /xh/ was significantly shorter in *+heid* than in *+igheid* words. This finding supports the word information load hypothesis.

A final study that we reviewed investigated the role of morphological structure and word information load in the weakening of word-final /s/ in Spanish (Torreira & Ernestus, 2012). The authors distinguished three types of /s/. First, /s/ can be part of a stem as in *martes* 'Tuesday', in which case /s/ distinguishes the word from other words (e.g., *marte* 'Mars'). Second, /s/ can be a suffix that is redundant given the context, like the plural suffix in *años* 'years' when preceded by *cuatro* 'four'. In this case, the suffix's word information load is low. Third, /s/ can be a non-redundant suffix carrying new information, as the plural suffix in *quiero cosas* 'I want things'. Torreira & Ernestus (2012) extracted 930 tokens of word-final /s/ from the Nijmegen Corpus of Casual Spanish (Torreira & Ernestus, 2010a). Type of /s/ neither affected the maximal difference in high-frequency intensity between the midpoint of /s/ and the beginning of the following vowel nor the duration of the low-frequency intensity dip in vowel-/s/-vowel sequences, but it significantly influenced voicing. Word-final /s/ tended to be voiced less often if it was part of the stem (48% voicing) or a non-redundant suffix (50%) compared to a redundant suffix (56%). This difference between redundant and non-redundant suffixes is best explained by word information load. It is also in line with previous studies showing that context plays a role in reduction. For example, Bell et al. (2009) have shown that words tend to be more reduced if they are more likely to occur given their context.

To summarize, one study directly addressed the question whether word information load (i.e., a segment's contribution to the identification of a word) or morphological structure better predicts acoustic reduction, and provided evidence in favor of an important role for word information load (Pluymaekers et al., 2010). The results of two other studies also appear to be better explained by the word information load hypothesis (Torreira & Ernestus, 2012; our new analyses discussed in the section on single segment affixes). We therefore conclude that word information load is more important in predicting degree of reduction than morphological structure.

General discussion

This paper provides an overview of studies from the literature and adds new analyses on the role of morphological structure in acoustic reduction. In this section, we discuss our findings and relate them to models of speech processing.

We first examined the role of morphological structure by investigating whether a morpheme tends to be more reduced if it is repeated. Previous studies have shown that complete words tend to be more reduced the more often they occur in the conversation (Fowler & Housum, 1987). The only study investigating whether this also holds for morphemes in complex words is by Viebahn et al. (2012). This study did not find a similar repetition effect for inflectional prefixes, unless the complete word was repeated. This suggests that the pronunciation process does not analyze morphologically complex words into morphemes and consequently that morphological structure does not play a major role in speech production.

This conclusion can also be drawn on the basis of the data discussed in the section on single segment affixes. We examined whether segments forming single segment affixes are less likely to be acoustically reduced than segments at the end of longer morphemes, as segments forming single segment affixes are more important for the identification of the word's morphological structure. We carefully reviewed each study that has investigated this hypothesis in either production experiments or on the basis of speech corpora, including Losiewicz (1992), Schuppler et al. (2012), and our new analyses of data reported in Chapter 2 of this dissertation. We concluded that none of these provided clear evidence that a segment's morphological status affects its degree of reduction. In some studies, no effects of morphological status were found at all, and in others the effects found may not be morphological in nature.

We then examined whether acoustic reduction was influenced by a word's morphological decomposability. Since words differ in how easily they may be decomposed into their morphemes, morphological structure may not affect the reduction of all words to the same extent. For instance, single segment affixes may only be less reduced than stem final segments if these affixes are part of highly decomposable words. Several studies have investigated this hypothesis. They quantified a word's decomposability by comparing the frequency of the word with the frequencies of all or some words that are morphologically related, or by asking native speakers to rate the word's decomposability. Two studies (Bürki, Ernestus, et al., 2011, and our new analyses of the data reported in Chapter 2 of this dissertation) did not observe decomposability effects. The other studies (Hay, 2003; Schuppler et al., 2012) reported an effect of decomposability, but in both studies the effects observed are open to alternative interpretations.

None of the four studies on morphological decomposability discussed in this paper provided a clear argumentation for why they operationalized decomposability in the way they did, and it is therefore difficult to choose between the measures. This is important, as the two datasets that show a correlation between decomposability and reduction (Hay, 2003; Schuppler et al., 2012) only do so if they quantify decomposability in different ways (i.e., decomposability was quantified as the word's frequency relative either to the frequency of all inflectional variants or to the frequency of the stem). To determine the role of decomposability in acoustic reduction, we therefore need a uniform definition of decomposability that is well-grounded in morphological theory and that can be easily quantified. Since this definition is currently not available and the experimental and corpus-based studies conducted do not provide convincing evidence for a role of decomposability in reduction, we conclude that, so far, the data on acoustic reduction do not support the hypothesis that morphological decomposability affects speech production.

Finally, we examined whether a segment's importance in identifying a word (i.e., its word information load) plays a more important role in acoustic reduction than a segment's importance for identifying a word's morphological structure. The studies that we found related to this issue (i.e., Pluymaekers et al., 2010; Torreira & Ernestus, 2012) support the word information load hypothesis. This again suggests that the important units in speech production are words rather than morphemes.

Our review of the literature and our new analyses strongly suggest that morphemes do not play a major role in speech production. This challenges traditional models of morphological processing that assume that regular complex words are not stored in the mental lexicon, but are computed from their morphemes on the basis of rules (e.g., Chomsky & Halle, 1968; Pinker, 1991; Taft & Ardasinski, 2006). In addition, our results challenge speech production models such as WEAVER++ (Levelt et al., 1999). In this model, the word to be produced (*lemma*) is first selected from the mental lexicon. If this word is morphologically complex, it is either treated as if it is monomorphemic (e.g., semantically opaque words), or it is processed via its morphemes (many completely regular complex words). In this model, morphological structure is therefore important in speech production.

This lack of clear evidence for the role of a word's decomposability in speech production has implications for theories assuming that a speaker tailors a word's degree of reduction to the listener's needs. Hay (2003) based her hypothesis that a word's decomposability affects its reduction on the dual-route models for word comprehension (e.g, Schreuder & Baayen, 1995). These models assume that complex words can be recognized via two routes: a decomposition route in which the word is decomposed into its morphemes, and a direct route in which the entire word is directly retrieved from the mental lexicon. Hay assumed that words are more likely to be processed via the decomposition route the easier they are to decompose. Decomposition would be facilitated if a word's pronunciation contained clearer cues to its morphological structure. Hay hypothesized that speakers therefore mark morphological structure more clearly in words that can best be recognized via the decomposition route, that is, in words whose stem often occurs in inflectional variants. The absence of convincing evidence for a role of a word's decomposability in reduction means that there is no clear evidence that speakers adapt their pronunciations of morphologically complex words in order to facilitate the listener's task.

The absence of data convincingly supporting a role of a word's decomposability in acoustic reduction also has consequences for the Hyper- and Hypospeech theory (Lindblom, 1990). This theory assumes that speakers prefer to hypo-articulate, unless this is harmful for the listener. The lack of clear evidence that speakers fine-tune the degree of reduction to their listener's needs strongly suggests that this theory does not hold for how speakers produce every individual segment in a word. This conclusion is in line with the finding by Bard et al. (2000), that speakers reduce repeated words independently of whether the current listener has heard the previous token.

We conclude that the words of the language are stored as complete units in the mental lexicon and are accessed directly. How easily the production of a segment is planned depends on several factors, including its predictability given the preceding part of the word and the preceding wider context (i.e., its information load). If this process of planning only takes a little time, the speaker can produce the segment more quickly, which may result in more reduced pronunciations (Bell et al., 2009). This view supports analogical models of speech production (e.g., Bybee, 2001; Skousen, 1989), in which morphological structures are not highly relevant for speech production. The connections between related words reveal the morphological structure, which the speaker can use in the formation of new morphologically complex words.

Importantly, our overview not only demonstrates that there is currently no evidence that morphological structure plays a substantial role in acoustic reduction, it also makes clear that several studies which present support for a role of morphological structure seem to do so because the effect that they report is not only correlated with morphological structure but also with, for instance, word information load. This again shows that researchers should carefully consider conflicting explanations for the patterns in their data before concluding that one of them plays a role. Moreover, this shows that progress in research can only be made if researchers make their datasets available to other researchers. In this paper, we only re-analyzed datasets on which we had worked ourselves, as these were the only ones available to us.

In sum, the studies conducted so far show no convincing evidence for a role of morphological structure in acoustic reduction. Future studies first have to provide a better definition of morphological decomposability before we further investigate the role of morphological structure in speech processing. Moreover, they have to avoid confounds of morphological structure and word information load. Should these future studies not provide clear evidence for a role of morphological structure, this will have serious implications for the existing psycholinguistic models of speech production.

Individual differences in the choice and pronunciation of words

Chapter 6

This chapter is a reformatted version of:
Iris Hanique, Mirjam Ernestus, and Lou Boves (submitted). Choice and pronunciation of words: Individual differences within a homogeneous group of speakers.

Abstract

This paper investigates whether individual speakers forming a homogeneous group differ in their choice and pronunciation of words when engaged in casual conversation, and if so, how they differ. More specifically, it examines whether the Balanced Winnow classifier is able to distinguish between the twenty speakers of the Ernestus Corpus of Spontaneous Dutch, who all have the same social background. To examine differences in choice and pronunciation of words, instead of characteristics of the speech signal itself, classification was based on lexical and pronunciation features extracted from hand-made orthographic and automatically generated broad phonetic transcriptions. The lexical features consisted of words and two-word combinations. The pronunciation features represented pronunciation variations at the word and phone level that are typical for casual speech. The best classifier achieved a performance of 79.9% and was based on the lexical features and on the pronunciation features representing single phones and triphones. The speakers must thus differ from each other in these features. Inspection of the relevant features indicated that, among other things, the words relevant for classification generally do not contain much semantic content, and that speakers differ not only from each other in the use of these words but also in their pronunciation.

Introduction

Language users have a multitude of different words at their disposal, and individuals may differ in their choice of words. For instance, some people may prefer the word *start* to the word *begin* or may use *big* instead of *large*. In speech, an additional type of variation is the exact pronunciation of words. Many words produced in casual speech show a range of possible pronunciations from the full pronunciation variant to highly reduced ones, in which phones are replaced by others or are completely missing. For instance, *probably* may be pronounced as [prɑbəbli], [prɑbli], [prɑli], and [prɑ]. Deviations in one phone from the full form occur in over 60% of the word tokens in casual American English, and two or more phones deviate in 28% of the tokens (Johnson, 2004). Similar numbers have been found for other languages (Ernestus & Warner, 2011). In this paper, we focus on a socially homogeneous group of speakers and investigate whether these speakers differ in their choice and pronunciation of words in casual conversations, and if so, how they differ. Research on individual differences in conversational speech will improve our understanding of the speech production process, and help us improve psycholinguistic models of speech processing.

Previous research on individual differences in word choice has focused on written text and function words (e.g., Ebrahimpour et al., 2013; Koppel, Schler, & Argamon, 2009; Stamatatos, 2009). Content words, such as *table* and *sleeping*, and word combinations, such as *old tree*, are very context dependent. In contrast, function words, such as *that* and *but*, are not likely to vary greatly with the topic of the text and can consequently more easily reveal topic independent individual differences in word choice. In the present study, we investigate the roles of both function words and content words, henceforth *unigrams*, and also of combinations of two words, henceforth *bigrams*.

Differences in acoustic reduction have been shown between groups of speakers. Several studies have reported effects of gender; for example, in Dutch, men tend to reduce words ending in the suffix /lək/ *-lijk* more often than women (e.g., in /moxələk/ *mogelijk* ‘possible’; Keune et al., 2005), and, in American English, they more often delete word-final /d/ and /t/ (Guy, 1980) and glides (Phillips, 1994). Further, younger speakers tend to reduce more than older speakers. This has been demonstrated, for instance, for the absence of word-final /d/ and /t/ in American English (Guy, 1980) and for the absence of segments in spontaneous Dutch (Strik et al., 2008). Finally, speakers of Dutch in Flanders tend to reduce less than speakers of Dutch in the Netherlands (Keune et al., 2005).

There is also some evidence that individual speakers may differ from each other in their reduction of words, even if they are members of the same social group. Ernestus (2000, p. 143) studied a group of twenty speakers who were all highly educated men aged between 21 and 55, and who were all born and raised in the western part of the Netherlands. She observed differences in the pronunciation of the Dutch word /natyrlək/ *natuurlijk* 'of course'. Whereas most speakers only produced the extremely reduced variant [tyk] in the middle of Intonational Phrases, one speaker also pronounced [tyk] in the initial and final positions of the Intonational Phrase and even in isolation. This raises the question whether differences between individual speakers can also be observed for other reduction phenomena that are typical for casual conversations.

To study differences in the choice and reduction of words between individual speakers, we applied a classification algorithm, in which speech fragments are attributed to their speakers on the basis of lexical and pronunciation patterns. If classification results in high performance scores, this would indicate that speakers differ in their speech habits. To examine how speakers differ, we inspected which words and pronunciation variants were important for distinguishing a speaker from others.

Our study is based on human-made orthographic and automatically generated broad phonetic transcriptions. These show the words that were used and how these words were pronounced at the phone level. By using broad phonetic transcriptions, we ignore all detailed information in the spectro-temporal representation of the speech. We do so because this spectro-temporal representation not only contains linguistically relevant information about how words were exactly articulated, but also paralinguistic information including voice quality, and these two types of information cannot easily be separated.

We are not the first to apply speaker classification to phonetic transcriptions. Van Bael & Van Halteren (2007) studied the effects of the speaker's age, gender, regional background, and level of education on word choice and pronunciation variation by classifying speakers belonging to groups differing in these characteristics. Using automatically generated broad phonetic transcriptions of the telephone dialogues of the Spoken Dutch Corpus (Oostdijk, 2002), the authors generated two sets of classification features: one set of approximately 150000 lexical features, including average utterance length, part-of-speech tags, and uni-, bi-, and trigram counts, and another set of 94 pronunciation features representing phone differences between full pronunciations of words and their actual phonetic transcriptions. The classification algorithm that they used was able to classify speakers according to their age, gender, and regional background on the basis of lexical features. Interestingly, classification was hardly effective on the basis of the pronunciation features. The authors suggested that this may be due to the broadness of the phonetic transcriptions, the limited set of pronunciation features, or the heterogeneity within their speaker groups.

In this paper, we study Dutch speakers who have the same regional background, gender, and educational level; that is, a homogeneous set of speakers. We investigate whether these speakers show individual differences in their choice and pronunciation of words in casual conversations, and if so, how these speakers differ.

Method

Speech data

For our study, we used the Ernestus Corpus of Spontaneous Dutch (ECSD; Ernestus, 2000), which consists of 15 hours of casual dialogues produced by ten pairs of speakers. These twenty speakers together uttered 155294 word tokens representing 9044 word types. On average, each speaker produced 7765 word tokens (ranging from 5419 to 10936 tokens). The speakers form a very homogeneous group: they are all males who hold academic degrees. Further, they are all native speakers of standard Dutch born and raised in the western part of the Netherlands. The main characteristic in which these speakers vary is their age, which ranges from 21 to 55 years.

Schuppler et al. (2011) generated broad phonetic transcriptions for the ECSD using an automatic speech recognition (ASR) system based on the Hidden Markov Model Toolkit (Young et al., 2002). An ASR system uses speech fragments and orthographic transcriptions of these fragments as input. In addition, it requires a pronunciation lexicon containing the full form and possible pronunciation variants for each word in the corpus (e.g., for the Dutch word *gewoon* 'just' the lexicon contained the full form /xəwon/ and the variants /xwon/ and /xon/). These pronunciation variants were created with 32 rules that had been formulated on the basis of earlier observations of pronunciation variation and that insert, alter, or delete phones. Finally, the ASR system uses 37 monophone acoustic models consisting of three states with 32 Gaussians per state (Hämäläinen et al., 2009). On the basis of these phone models, which had been trained on the read speech component of the Spoken Dutch Corpus, the ASR system determined for each word in the orthographic transcriptions which variant from the pronunciation lexicon best matched the speech signal.

Schuppler et al. (2011) validated this transcription procedure by comparing its output for the IFA corpus (Van Son, Binnenpoorte, Van den Heuvel, & Pols, 2001) with manual transcriptions of this corpus. They calculated how often phones in the automatic transcriptions deviated from those in the manual transcriptions in terms of insertions, replacements, and deletions. They observed an overall agreement of 86.0%, which is similar to agreements among human transcribers reported in the literature (e.g., Kipp et al., 1997, reported agreements between human-made transcriptions of spontaneous German of 78.8%, 79.9%, and 82.6%; for more information on agreements typically obtained for phonetic transcriptions see Ernestus & Baayen, 2011). Chapter 2 validated the automatically generated transcriptions of the ECSD with human-made transcriptions on the basis of 148 schwas in the initial syllables of past participles (as in /xəmɪst/ *gemist* 'missed'). Two human transcribers agreed on the presence versus absence of schwa in 82.4% of the tokens, while they agreed with the ASR system in 75.7% and 77.0% of the tokens. These agreements did not differ significantly from each other. Given these evaluations, and since obtaining better transcriptions for such a large corpus is difficult, we accepted these automatic transcriptions as being valid.

As automatic phonetic transcriptions can only be created for uninterrupted speech (i.e., without, for instance, overlapping speech or laughter), the number of transcribed words is lower than the number of words in the entire corpus. Our transcriptions contain 95173 word tokens and 6965 word types, ranging from 1 to 3459 word tokens per word type. The most frequent word types were *ik* 'I' and *dat* 'that' with 3459 and 3402 tokens respectively. On average, for each speaker 4759 word tokens were transcribed, representing 944 word types.

For our classification tests (see below), we used the automatically generated phonetic transcriptions and the corresponding orthographic transcriptions. Moreover, we divided the transcriptions of each speaker into ten equally sized fragments. The size of these fragments was different for each speaker: it varied between 375 and 742 word tokens, and had an average of 479 tokens.

Classification Features

Classification algorithms distinguish between classes (in our case speakers) on the basis of features which represent properties of these classes (e.g., single words such as *window* or the absence of a phone such as /t/). We represented each of the 200 fragments in our dataset (10 fragments per speaker) as a list of features that are based on the fragment's orthographic and phonetic transcription.

To investigate word choice, we extracted all unigrams and bigrams from the transcriptions. If a word was not preceded or followed by another word, a bigram was created including a silence before or after the word. We then selected those unigrams and bigrams that occurred more than twenty times in the entire corpus and that were produced by at least two speakers. This resulted in 403 unigrams, each of which on average occurred 195 times in the entire corpus (range: 21 to 3459) and in 61 fragments (range: 12 to 200). The total number of bigrams was 642, each of which on average occurred 75 times in the corpus (range: 21 to 1931) and in 44 fragments (range: 12 to 199). Following Van der Sijs (2002), we considered prepositions, conjunctions, determiners, pronouns, and numerals as function words, and nouns, verbs, adjectives, adverbs, and interjections as content words. The selected unigrams consisted of 158 function word types and 245 content word types.

To study individual differences in the pronunciation of words, we included pronunciation features. Since we focused on how speakers reduce words in casual speech, we ignored three well-known types of variation in Dutch pronunciation, because these occur as often in read speech as in casual speech. First, we ignored the variation in the pronunciation of Dutch word-final *-en*, which is mostly pronounced as [ə] and sometimes as [ən] except in the east of the Netherlands. Second, we also ignored the insertion of phones, for example, the pronunciation of /mɛlk/ *melk* 'milk' as [mɛlək], as this is not reduction.

Third, we ignored the variation in the pronunciation of obstruents as voiced or voiceless (e.g., /v/ versus /f/). The voicing of obstruents is highly variable in Dutch, especially as spoken in the western part of the Netherlands. Like most speakers from this area, our speakers often replaced voiced fricatives by their voiceless counterparts (Schuppler et al., 2011) and frequently applied regressive and progressive voice assimilation (Ernestus, Lahey, Verhees, & Baayen, 2006) Moreover, the automatic transcriptions that were used in our study have been generated by means of acoustic models that may not reliably encode voicing for obstruents, since they had been trained with speech for which the voicing of obstruents may not have been reliably transcribed. We therefore collapsed the members of the obstruent pairs /z,s/, /v,f/, /d,t/, /g,k/, /b,p/, /ʒ,ʃ/, and /ɣ,x/.

We examined pronunciation variation at the phone level, henceforth *phone features*, and at the word level, henceforth *word pronunciation features*. Our phone features provided information whether a phone was produced as it would be if the word had been produced in careful speech. For each phone in a word token, we determined whether it was unreduced, i.e., produced as in the full pronunciation of the word, was replaced by another phone, or was completely absent. For example, the Dutch word /heləmal/ *helemaal* 'completely' may be pronounced as [həlmal], in which we considered all consonants and the final vowel as being unreduced, the first vowel as being replaced by schwa, and the second vowel as being absent.

For each phone, we defined four types of possible features. One type was the phone itself without any neighboring phone, henceforth *uniphones*; for example, /e/ replaced by schwa. Two possible feature types included either the preceding or following phone from the full form, henceforth *biphones*; for example, /e/ replaced by schwa and followed by /l/. The final possible feature type represented the phone and both neighboring phones from the full form, henceforth *triphones*; for example, /e/ replaced by schwa in the sequence /hel/. If the phone was positioned at a word boundary, we took the neighboring phone from the neighboring word, for example, in [həlmalni] *helemaal niet*, the phone following the final [l] of *helemaal* was /n/.

Each possible feature was only used if it met the following two criteria. First, it had to occur more than twenty times in the entire corpus. Second, there had to be at least one other pronunciation of that phone (sequence) that occurs at least twice in the entire corpus; for example, the feature *replacement of* /e/ *by schwa followed by* /l/ was only used if another variant, that is a *present* or *absent* /e/ *followed by* /l/, occurred at least twice. In total, we used 955411 phone features that represent 2394 phone feature types. Table 6.1 presents the numbers of the phone features split to uniphones, biphones, and triphones, and whether the phone was unreduced, replaced, or absent. On average, a phone feature occurred 399 times (ranging from 21 to 41799 times) and in 87 fragments (ranging from 4 to 200 fragments).

Word pronunciation features represent pronunciation variation at the word level. For instance, the word *mensen* 'people' can be produced in the unreduced form [mɛnsə] or in a reduced variant, such as [mɛns] or [mɛs]. We selected as word

Table 6.1: Number of tokens and types (in brackets) of the uniphone, biphone, and triphone features split into phones that are unreduced, replaced, or absent.

	unreduced	replaced	absent
uniphones	267704 (31)	8770 (23)	32957 (27)
biphones preceding	213445 (327)	6676 (82)	31180 (150)
biphones following	211486 (319)	7822 (58)	31763 (143)
triphones	112478 (829)	5242 (54)	25888 (351)

pronunciation features, word pronunciations that occurred more than twenty times in the entire corpus and produced by at least two speakers. In addition, the words they represented had to have at least one other pronunciation variant that occurred at least twice in the entire corpus. This is to ensure that the features could indeed capture pronunciation variation and not just lexical information, which would certainly be the case if there is only one pronunciation variant. Note that if the other pronunciation variant does not meet the selection criteria, this other variant is not used as a feature. In total, 290 word pronunciation variants met these restrictions, and these variants represent 157 word types (52 word types had one pronunciation variant; 86 word types had two variants; eleven word types had three variants; seven word types had four variants; and one word type had five variants). These variants represented 128 unreduced forms, for example [mɛnsə] *mensen* 'people', and 162 reduced variants, for example [mɛs], a variant of *mensen* 'people'.

Classification algorithm

We used the Balanced Winnow classifier (Dagan, Karov, & Roth, 1997; Littlestone, 1988) implemented in the Linguistic Classification System (LCS; Koster, Seutter, & Beney, 2003; Koster & Beney, 2009). This algorithm assigns two weights (w^+ and w^-) to each feature for every speaker, and the overall (Winnow) weight is their difference. Features of a certain speaker with a positive overall weight are used to classify a fragment as belonging to that speaker, whereas those with a negative overall weight are used for classification as not belonging to that speaker. The value of an overall weight indicates how useful the feature is to distinguish the speaker from all other speakers in the corpus. The output for each speaker from the LCS is a model, henceforth *speaker profile*, which consists of two lists, one with positive overall weights and one with negative overall weights. We used them to identify the speaker of a new fragment and thus to test how well the classifier performed. In addition, we used the profiles to characterize the differences among speakers: in a certain speaker profile, features with a positive overall weight are assumed to be characteristic for that speaker, whereas those with a negative overall weight are assumed to be uncharacteristic for that speaker.

The classifier created speaker profiles in the training phase in which it receives all (training) fragments as input, labeled as a positive or negative example for a given speaker. A fragment produced by a certain speaker is a positive example for only that speaker and is a negative example for all other speakers. The classifier first constructs initial speaker profiles on the basis of all fragments. Each initial profile consists of a list of all features with their initial weights, calculated through the LTC algoritm (Salton & Buckley, 1988). Subsequently, these speaker profiles are adapted during multiple training iterations, in each of which all (training) fragments are again presented to the classifier. For each fragment, the classifier calculates the correspondences between that fragment and every speaker based on the fragment's features and the weights for these features in each of the speaker profiles. Balanced Winnow is a mistake-driven classifier, which means that weights in a speaker profile are only updated if a fragment is classified incorrectly during training. If a fragment belonging to a speaker scores for that speaker above threshold θ^{+}, the fragment is correctly classified as a positive example and the weights remain unchanged. In contrast, if it scores below this threshold, the classification is treated as a mistake and the weights in the speaker profile are updated by multiplying the positive weights of the active features, i.e. those that occur in both the fragment and the speaker profile, with parameter α and the negative weights of the active features with parameter β. In addition, if a fragment that does not belong to a given speaker has a score for that speaker below another threshold, θ^{-}, it is correctly classified as a negative example, whereas if this fragment scores above the threshold, the weights of the active features in that speaker's profile are updated by multiplying positive weights with β and negative weights with α. After testing values for the parameters around the default settings of LCS, we used those settings that resulted in the highest performance, namely α = 1.05, β = 0.98, θ^{+} = 0.8, θ^{-} = -0.8, and a maximum of 20 training iterations.

Classification tests

To test how well the classifier performed, we used ten-fold-cross-validation: the classifier was trained on 180 fragments (nine from each speaker) and tested with the remaining 20 fragments (one from each speaker). This procedure of training and testing was repeated ten times, so that each fragment was used as a test fragment exactly once. Each fragment used in this study belonged to only one speaker, and therefore our tests are mono-classifications.

The order in which the training module of the classifier processes the fragments cannot be controlled and is entirely random. As a consequence, running the classifier twice does not usually lead to exactly the same results. We therefore ran each ten-fold-cross-validation 100 times. From the 100 performances of each classification test, we determined the lowest, highest, and average performance. The difference between the lowest and highest performance was on average 7.24% (range: 5.5% to 9.5%). Average performances are reported in the Results. To compare the performance of different classification tests, we performed several unpaired t-tests on obtained performance scores (see below). As we performed multiple t-tests, we applied Bonferroni correction and used an alpha level of 0.0045.

To obtain information about which features contribute to the identification of speakers and thus in what aspects of choice and pronunciation of words speakers differ from each other, we manually inspected speaker profiles. For this manual inspection, we trained the classifier with the best performing combination of features (see below). Furthermore, we used all 200 fragments for training and thus obtained speaker profiles that are based on as much data as possible. As these speaker profiles are created by running the classifier only once, running it again will probably result in slightly different speaker profiles. To use only those features that are robust, i.e., likely to be part of the speaker profiles if we run the classifier again, we examined only those features that have an overall Winnow weight that is larger than the median weight, which was calculated separately for positive and negative weights. The positive overall weights ranged from 0.00025 to 2.54 with a median of 0.08, and the negative overall weights varied between -0.005 and -6.43 with a median of -0.15. We focused on the features that are characteristic for only a few speakers and thus provide information about differences between speakers. We therefore determined which features have a positive overall weight larger than the positive median for only one to four speakers and a negative weight larger than the negative median for 15 or more speakers.

Results and discussion

As it is only meaningful to investigate differences between speakers with a well-performing classifier, we first investigated how well our best classifier performs in general. We therefore examined how often it correctly classified a fragment. Moreover, we investigated whether speakers in the same conversation were often confused with each other, which may be due to the topic of the conversation or to interactive speech alignment (Pickering & Garrod, 2004). Thereafter, we investigated the relevance of the different types of features by comparing classifiers trained and tested with various combinations of feature sets. Furthermore, for each feature type, we examined which features are especially relevant for characterizing individual speakers. As a performance measure for these classifiers, we used the harmonic means of their precision and recall (F_1), presented in Table 6.2.

General performance

The best performing classifier made use of both lexical features and of the uniphone and triphone features. The average performance of this classifier is as high as

Table 6.2: Averaged harmonic means (F_1) obtained in our classification tests with different sets of features. A plus sign indicates that the feature type was included in training and testing the classifier. The columns *Bi.prec*, *Bi.foll.*, and *Word* denote the target phone features with the preceding phone, the target phone features with the following phone, and the word pronunciation features, respectively.

	Unigrams	Bigrams	Uniphones	Bi.prec.	Bi.foll.	Triphones	Word	F_1
1	+							51.7%
2		+						59.6%
3	+	+						73.9%
4	+	+	+					75.5%
5	+	+		+				77.9%
6	+	+			+			76.2%
7	+	+				+		79.2%
8	+	+	+	+				79.0%
9	+	+	+		+			77.3%
10	+	+	+			+		79.9%
11	+	+	+	+	+			77.8%
12	+	+		+	+			77.4%
13	+	+		+	+	+		79.1%
14	+	+	+	+	+	+		78.9%
15	+	+					+	75.9%
16	+	+	+				+	76.3%
17	+	+		+			+	75.7%
18	+	+			+		+	77.7%
19	+	+				+	+	78.0%
20	+	+	+	+			+	76.2%
21	+	+	+		+		+	78.0%
22	+	+	+			+	+	78.4%
23	+	+	+	+	+		+	75.9%
24	+	+		+	+		+	75.8%
25	+	+		+	+	+	+	78.9%
26	+	+	+	+	+	+	+	78.9%

79.86%, which may be surprising given the homogeneity of the group of speakers. As shown in the confusion matrix (Table 6.3), the percentage of correct classifications for individual speakers ranged from 43.7% to 100%. Some speakers (e.g., Speakers J, N, and O) were seldom confused with other speakers. In contrast, a substantial number of fragments were incorrectly attributed to Speakers G and T, as were fragments of Speakers E and I to other speakers. Apparently, some speakers were more difficult to classify than others. Note that the classifier was still able to correctly classify the more difficult speakers well above chance level (i.e., 5%), suggesting that they were not indistinctive in their word choice or pronunciations.

We examined whether speakers who participated in the same conversation were more often confused with each other than with other speakers. Speakers in the same conversation discuss the same topics, which inevitably results in the use of the same content words. Moreover, several studies have shown that speakers tend also to align their speech on other levels, including syntactic and phonological levels (interactive speech alignment, e.g., Pickering & Garrod, 2004). As a consequence, speakers in the same conversation may be more confused with each other than with other speakers. This was only the case for six out of the twenty speakers. As shown by the underlined numbers in Table 6.3, Speakers A, E, I, K, Q, and U were more often classified as the other speaker in the conversation than as speakers from other conversations. Interestingly, the partners for Speakers A, I, K, and U (i.e., Speakers B, S, L, and V respectively) were not more often classified as the other speaker in the conversation, which suggests that these speakers also display more idiosyncratic properties.

Only the speakers of the pair E-Q were often confused with each other. Twenty-one percent of these two speakers' top 50 positive features concern a content word, which is either *eens* 'once', *is* 'is', *ja* 'yes', *kunnen* 'can', *maken* 'make', *natuurlijk* 'of course', *nu* 'now', *vind* 'find', *weet* 'know', or *wil* 'want'. These content words are not very informative about what the speakers talked about. It is therefore unlikely that confusion between Speakers E and Q is the result of the topic of the conversation. It probably results from alignment at various linguistic levels or coincidental similarities between these speakers.

Lexical features

In order to investigate differences in speakers' choice of words, we first trained and tested the classifier on the basis of lexical features only, namely uni- and bigrams. These tests are presented in the first three rows of Table 6.2. T-tests showed that including both unigrams and bigrams resulted in a significantly better performance than using only unigrams ($t(198.0) = -104.4$, $p < 0.0001$) or bigrams ($t(197.5) = -69.6$, $p < 0.0001$). The performance with both unigrams and bigrams was approximately 74% (row 3), indicating that speakers greatly differ in their choice of words.

Table 6.4 shows features that are characteristic for only a few speakers. The majority of the characteristic bigrams contain a silence (e.g., ∅ *dus* '∅ so'), indicating that speakers differ especially in which words they produce directly before or after a pause. Furthermore, the majority of the features are function words (e.g., *want* 'because' and *we* 'we') and those that are content words are highly frequent and semantically relatively weak (e.g., *goed* 'good' and *nee* 'no').

To further investigate the contribution of content and function words to the classifi-

Table 6.3: Confusion matrix with the number of classifications based on lexical features, uniphone features, and triphone features. Underlined numbers are combinations of speakers that participated in the same conversation.

Actual	Classified speaker																			
speaker	A	B	E	Q	F	G	H	R	I	S	J	T	K	L	M	N	O	P	U	V
A	647	<u>100</u>	0	0	0	46	75	55	0	17	7	0	49	0	0	0	3	0	0	1
B	<u>0</u>	830	0	9	40	1	29	0	0	0	0	0	24	10	0	56	0	0	0	1
E	0	2	437	<u>281</u>	25	0	77	13	118	0	9	21	13	0	1	0	0	0	0	3
Q	0	0	<u>123</u>	709	81	0	0	0	0	0	0	0	0	2	0	7	0	0	0	78
F	5	14	1	0	834	<u>0</u>	0	0	0	0	31	0	1	0	0	15	0	0	0	99
G	4	0	0	0	<u>1</u>	980	0	0	0	2	12	0	0	0	0	0	0	1	0	0
H	99	1	0	0	0	0	877	<u>21</u>	0	0	0	0	0	0	0	0	0	0	2	0
R	0	0	51	0	0	1	<u>33</u>	888	0	0	24	0	0	0	0	0	0	0	3	0
I	0	13	7	10	1	0	9	0	564	<u>134</u>	0	34	46	47	0	0	30	1	1	103
S	38	39	27	0	1	7	2	0	<u>0</u>	761	0	86	0	0	0	0	10	8	21	0
J	0	0	0	0	0	0	25	0	0	0	969	<u>0</u>	0	0	6	0	0	0	0	0
T	0	0	0	4	0	0	1	0	0	0	<u>0</u>	984	0	0	0	0	0	11	0	0
K	0	0	0	0	9	48	2	10	7	0	44	36	669	<u>167</u>	6	0	2	0	0	0
L	0	25	0	0	0	0	0	5	16	1	14	0	<u>12</u>	924	1	2	0	0	0	0
M	0	71	0	0	0	0	0	0	0	0	0	0	100	0	775	<u>2</u>	0	0	52	0
N	8	14	1	0	0	0	0	0	0	0	15	0	1	0	<u>0</u>	959	0	2	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1000	<u>0</u>	0	0
P	0	0	0	0	0	100	0	0	0	0	0	77	0	0	0	0	<u>1</u>	822	0	0
U	0	3	0	1	39	0	0	67	0	0	0	3	40	2	13	0	0	0	735	<u>97</u>
V	0	0	0	8	0	36	6	0	84	0	0	2	73	0	0	6	0	98	<u>79</u>	608
Total	801	1112	647	1022	1031	1219	1136	1059	789	915	1125	1243	1028	1152	802	1047	1046	943	893	990

Table 6.4: The lexical features that are characteristic for only a few speakers. Within a bigram, ∅ indicates a silence.

unigrams:
ben 'am', *bij* 'by', *goed* 'good', *hè* 'isn't it?', *heel* 'very', *jij* 'you', *mij* 'me', *nee* 'no', *om* 'to', *toch* 'still', *want* 'because', *we* 'we', *ze* 'they', *zo* 'so'
bigrams:
∅ *de* '∅ the', ∅ *dus* '∅ so', ∅ *een* '∅ an', ∅ *nee* '∅ no', *dat* ∅ 'that ∅', *een* ∅ 'an ∅', *en dan* 'and then', *het* ∅ 'it ∅', *in de* 'in the', *maar* ∅ 'but ∅', *niet* ∅ 'niet ∅', *nou* ∅ 'well ∅'

cation performance, we also trained and tested the classifier on unigrams only including either of these word types. Based on function words only, we obtained a performance of 26.0%, whereas classification based on content words only resulted in a performance of 39.5%. Both results were significantly lower than the classification performance of 51.7% based on all unigrams (function words: $t(198.0) = 121.2$, $p < 0.0001$; content words: $t(196.8) = 59.6$, $p < 0.0001$), suggesting that neither of the word types can solely account for the performance based on all unigrams. Interestingly, classification based on content words performed significantly better than the classification based on function words ($t(197.2) = 66.0$, $p < 0.0001$). Importantly, the number of features cannot explain this difference in classification performance, since the number of content word features (27381) was lower than the number of function word features (51377). These results suggest that speakers differ from each other especially in their use of semantically weak content words, such as discourse markers.

Phone features

We refrained from testing models with pronunciation features only, as these would mainly signal lexical variation instead of pronunciation variation. For instance, when using this type of classifier, if we found that the word pronunciation variant /tyk/ for *natuurlijk* 'of course' is characteristic for a few speakers, this would probably be because these speakers more often produced this word and not because they pronounced it differently from the other speakers. All our classifiers therefore included lexical features.

We first combined the lexical features with phone features. The best performance was obtained with the classification including the uniphone and triphone features (compare row 10 of Table 6.2 to rows 4 to 9 and 11 to 14). This set of features improved performance by approximately 6% compared to the classification that was based on lexical features only (row 3 vs. 10: $t(197.9) = -28.4$, $p < 0.0001$).

As classification including biphones resulted in lower performances than classification including triphones, information about either neighboring segment is apparently less helpful than information about both neighboring phones. The probable explanation is that the exact pronunciation of a phone depends on both the preceding and following phone (e.g., /ə/ may especially be absent after /x/ and before a sonorant); both neighboring phones are therefore necessary for variation to be meaningful.

Table 6.5 presents the phone features that are characteristic for only a few speakers. There are two reasons for not being able to directly conclude that these differences among speakers reflect individual differences in reduction. First, if only one variant is incorporated as a phone feature, speaker differences in the use of these phone features cannot represent individual differences in pronunciation; they show that speakers differ in how often they produce words with these phones. Second, if speakers differ in the words in which these phones occur, differences among these speakers may reflect differences in word choice rather than in phone realization. Research has shown that a segment is more likely to be reduced in one word than in another depending on the word's frequency of occurrence, the preceding or following segment, and whether the segment carries stress (e.g., Bell et al., 2009; Cho & McQueen, 2005; Mitterer & Ernestus, 2006).

Two-thirds of the features in Table 6.5 represent unreduced phones with their neighboring phones. For 51.1% of these unreduced characteristic features (which is 35.4% of all features in Table 6.5), counterparts with an absent or replaced phone do not occur frequently enough in the corpus to be included in the phone features. For instance, the unreduced feature /ty̲r/ was characteristic for Speaker I, who produced it in the words /natyrlək/ *natuurlijk* 'of course', /prosədyrə(s)/ *procedure(s)* 'procedure(s)', /litərɑtyr/ *literatuur* 'literature', /dyr/ *duur* 'expensive', /dyrə/ *dure* 'expensive', /dyrt/ *duurt* 'lasts', and /dyrdə/ *duurde* 'lasted' (remember that voiced and

Table 6.5: The uniphone and triphone features that are characteristic for only a few speakers. The target phone is underlined and a silence is indicated by ∅. A replacement of phone A by phone B is denoted by A → B.

unreduced phones:
ø, xə̲t, kə̲∅, pə̲t, tə̲f, tə̲p, ma̲l, ∅ɑ̲l, kɑ̲n, tɑ̲l, wɑ̲n, nɑ̲ʊ̲∅, xe̲n, me̲r, we̲t, pɛ̲n, tɛ̲n, ti̲t, nɪ̲k, tɪ̲k, so̲n, kɔ̲m, xu̲t, ty̲r, ɛɪx̲ə, ɑl̲ə, ɑl̲s, el̲ə, an̲t, ɑn̲s, en̲p, en̲s, ɛn̲s, ɪn̲ə, ɛr̲k, ɔr̲t, ns̲ə, ət̲∅, ət̲ə, at̲ə, it̲s, xt̲ə, lt̲ə, nt̲∅ əw̲o,

absent phones:
x̲, i̲, s̲, xə̲l, xə̲w, jə̲t, tə̲∅, ∅ɪ̲k, əl̲ə, ɑl̲s, en̲s

replaced phones:
a̲ → ə̲, ɑ̲ → ə̲, o̲ → ə̲, a̲ → ɑ̲, nə̲t → nɛ̲t, ∅e̲n → ∅ə̲n, re̲n → rə̲n, se̲n → sə̲n, te̲n → tə̲n

voiceless obstruents have been collapsed in the pronunciation features). Its counterpart in which /y/ is absent occurs only twice in the entire corpus (both occurrences concern an inflection of /kʏltyrel/ *cultureel* 'cultural' produced by Speaker K) and the replacement of this segment occurs only once (in the pronunciation [natʏr] /natyr/ *natuur* 'nature' produced by Speaker H). Since this phone sequence is only represented with its full pronunciations in the feature sets, its relevance cannot be attributed to pronunciation variation but probably results from variation in word choice.

In contrast, the remaining 48.9% of the unreduced characteristic features appear to represent pronunciation variation. For example, an unreduced schwa in the sequence /xə̱t/ is characteristic for Speakers P and T, who produced it mainly in /xədan/ *gedaan* 'done', /zɛxə/ *zeggen* 'say' followed by /t/ or /d/, /texə/ *tegen* 'against' followed by /t/ or /d/, and /xədeltə/ *gedeelte* 'part'. In contrast, the absence of this schwa is characteristic for Speakers K, L, P, R, and V. These speakers produced the sequence /xə̱t/ mainly in the words /xədan/ *gedaan* 'done', /zɛxə/ *zeggen* 'say' followed by /t/ or /d/, and /xətɑl/ *getal* 'number'. As the sequence /xə̱t/ occurs in approximately the same words for the two speaker groups, the difference in pronunciation between these two groups is not likely the result of differences in word choice but in genuine pronunciation variation.

A second example represents the unreduced pronunciation of the sequence /ɑḻə/, which is characteristic for Speaker V only. He produced this sequence in the words /ɑlə/ *alle* 'all', /ɑləs/ *alles* 'everything', /ɑlərlɛɪ/ *allerlei* 'all kinds of', /ɑləmal/ *allemaal* 'all', /antɑlə/ *aantallen* 'numbers', and /xəvɑl/ *geval* 'case' followed by a schwa. The counterpart in which /l/ was absent was characteristic for six other speakers, namely Speakers E, H, M, P, R, and T. These speakers produced this counterpart in the word /ɑləmal/ *allemaal* 'all' only. Speaker V therefore differs from these other speakers in how he pronounces the sequence /ɑḻə/ in the semantically weak word *allemaal*.

The other third of all phone features in Table 6.5 represent absent and replaced phones in approximately the same numbers. All but one (i.e., se̱n → sə̱n) of these features have unreduced counterparts in the feature set and are therefore likely to represent genuine pronunciation variation. The absence of schwa in /jə̱t/ is an example of a feature that clearly represents pronunciation variation. It is characteristic for Speaker H and uncharacteristic for sixteen other speakers. Sixteen of this feature's occurrences (84.2%) represent the word /jə/ *je* 'you' followed by a /t/ or /d/, while the remaining occurrences represent the word types /spœytjə/ *spuitje* 'little syringe', /festjə/ *feestje* 'little party', and /betjə/ *beetje* 'a little' followed by /t/ or /d/. The unreduced pronunciation of /jə̱t/ was characteristic for five different speakers (i.e., Speakers E, F, K, L, and Q) and uncharacteristic for ten speakers (i.e., B, H, I, M, N, O, R, S, T, and U). These five speakers produced this sequence also mainly in the word /jə/ *je* 'you' followed by a /t/ or /d/ (83.9%). This indicates that the absence of schwa in /jə̱t/ represents pronunciation variation, primarily in the semantically weak word *je*, and is not likely to result from lexical choice. Our data thus show that our socially homogeneous group of speakers are not at all homogeneous in their pronunciation of the phone sequence /jə̠t/.

Since some phone features appear to represent occurrence of phone sequences rather than pronunciation variation, the question arises whether genuine pronunciation variation really contributes to speaker identification. In order to answer this question, we ran an additional classifier based only on lexical features and the full pronunciations of the phone features. For instance, the uniphone feature *absence of* /x/ was converted to *unreduced* /x/ and *replacement of* /a/ *by* /ə/ was replaced by *unreduced* /a/. All unreduced features remained unchanged. This classifier thus contained no features representing pronunciation variation. Consequently, if the best performing classifier so far uses information about pronunciation variation, the classifier without pronunciation variation should perform worse. The performance of the classifier without pronunciation variation was 70.0% (not incorporated in Table 6.2), which is significantly less than the classification based on lexical features and uniphone and triphone features (row 10 of Table 6.2; $t(198.0) = -48.1$, $p < 0.0001$). Interestingly, this performance is also worse than the performance of the classifier based on lexical features only (row 3; $t(197.9) = -19.0$, $p < 0.0001$). We conclude that pronunciation variation contributes to classification, and speakers thus differ in how they pronounce phones and phone sequences.

Interestingly, the classifier achieved better performance if it was not only based on lexical and triphone features but also on uniphone features (Table 6.2 row 7 vs. 10: $t(197.9) = -3.4$, $p < 0.001$). This suggests that variation in the single phones, regardless of the context, also reflects speaker-specific behavior. Most of the relevant uniphones represent reductions (see Table 6.5). For instance, the absence of /x/ is characteristic for Speaker V. This speaker did not produce /x/ 33 times out of 422 times, and this was mainly in the semantically weak words /nɔx/ *nog* ‘yet’ and /tɔx/ *toch* ‘still’ (also in /xəxan/ *gegaan* ‘went’, /ɛɪxələk/ *eigenlijk* ‘actually’, /ɔnxəver/ *ongeveer* ‘approximately’). As only one of the characteristic uniphones is an unreduced pronunciation, we conclude that speakers hardly differ in how often they produce most single phones. In contrast, they differ in how they reduce single phones.

Comparison of the relevant features that regard vowels and consonants (Table 6.5) showed that for both unreduced and absent uniphones and triphones, the numbers are approximately the same (i.e., unreduced: 24 vowels and 21 consonants; absent: 6 vowels and 5 consonants). This is unsurprising for the uniphones, given the similarity in the numbers of vowels and consonants among the unreduced and absent uniphones provided to the classifier (i.e., unreduced: 16 vowels and 15 consonants; absent: 14 vowels and 13 consonants). In contrast, the numbers of vowels and consonants differ among triphones in the input (i.e., unreduced: 450 vowels and 379 consonants; absent: 202 vowels and 149 consonants). The fact that approximately the same number of triphones with vowels and consonants characterize speakers therefore suggests that, for both the unreduced and absent segments, a larger proportion of the consonants than vowels is speakers specific.

Furthermore, the replaced phones that distinguish speakers from each other are all vowels. This is in line with the number of replaced vowel and consonant features in the input of the classifier: 22 uniphone and 47 triphone replacements concerned a vowel, but only 1 uniphone and 7 triphone replacements concerned a consonant which all concerned the pronunciation of [m] instead of /b/ or /p/. A likely explanation is that vowels are easily reduced to schwa, while consonants are not often replaced by another consonant (aside from the consonant that has the same characteristics except for voicing; e.g., /v/ is often replaced by /f/). As explained in the Method, variation in voicing is not specific for casual speech and was therefore ignored.

In conclusion, the classification including phone features showed that speakers differ in their pronunciations of single phones and of sequences of three phones. Interestingly, we found that several triphone features mainly originate from semantically weak words, which shows that speakers differ in the pronunciations of these words.

Word pronunciation features

Finally, we ran a classifier that used all types of features, that is, lexical features, phone features, and word pronunciation features. To avoid words being represented by both lexical and word pronunciation features, words that could be represented by both were only included as word pronunciation features. For instance, in the classification tests without word pronunciation features, the word *mensen* ‘people’ was represented as unigram, in five bigrams (i.e., *mensen die* ‘people who’, *die mensen* ‘those people’, *de mensen* ‘the people’, *mensen* ∅ ‘people ∅’, and ∅ *mensen* ‘∅ people’), and in the phone features. As [mɛnsə], [mɛsə], [mɛns], and [mɛs] were part of the word pronunciation features, when including these, we replaced each occurrence of *mensen* that was produced as one of these variants by its actual pronunciation. If an occurrence of *mensen* was pronounced differently from the pronunciations that are part of the word pronunciation features (e.g., as [mɛnsən] or [mɛsn]), the occurrence was not replaced. For bigram features, this means that none of the words, both of the words, or either of the words could be replaced. This procedure gave a preference to the word pronunciation features over the lexical features and thus gave the word pronunciation features every chance to distinguish between speakers.

The best performance including word pronunciation features was achieved with the combination of all lexical features, phone features, and word pronunciation features (78.9%; final row of Table 6.2). Importantly, this performance is significantly worse than the best performance with lexical and phone features only (row 10; $t(196.9) = 4.9$, $p < 0.001$). Hence, the addition of word pronunciation features does not result in an improvement. One explanation may be that speakers do not differ in their pronunciations for complete words. This is however unlikely as classification based on lexical and word pronunciation features only (i.e., without any phone features) resulted in an improved performance compared to classification with only lexical features (row 15 vs. row 3 of Table 6.2; $t(186.5) = -10.3$, $p < 0.0001$). An alternative and more probable explanation is that the variation in the pronunciation of entire words is already captured by the phone features. This is supported by the triphone features discussed in the previous section which upon closer inspection appeared to represent a small number of words (e.g., reduced /jə̲t/ mainly represented the word *je* 'you' followed by /t/ or /d/ and unreduced /ɑl̲ə/ represented mainly the word *allemaal* 'all'). Moreover, all words are also represented by uniphone and triphone sequences. For example, the pronunciation [mɛs] for /mɛnsə/ *mensen* 'people' is presented as a word pronunciation variant, but is also represented in the uniphones /m/, /ɛ/, absence of /n/, /s/, and absence of /ə/, and in the triphones /?m̲ɛ/ (in which ? indicates any possible preceding segment), /mɛ̲n/, absence of /n/ in /ɛn̲s/, /ns̲ə/, and absence of /ə/ in /sə̲?/ if these features met the restrictions described in the Method. The phone features outperformed the word pronunciation features probably because they additionally capture variation that spans word boundaries (e.g., resulting from cross-word assimilation), and they therefore contain more information about the speaker's pronunciation habits than word pronunciation features.

General discussion

This paper investigated whether individual speakers sampled from a socially homogeneous group differ in their choice and pronunciation of words when engaged in casual conversations, and if so how. We studied the homogeneous group of twenty male speakers of the ECSD and tested whether a classification algorithm was able to distinguish between these speakers. In order to focus on the speakers' choice and pronunciation of words rather than characteristics of the speech signal (including voice characteristics), we trained and tested the classifier on the basis of features extracted from hand-made orthographic and from automatically generated broad phonetic transcriptions. We hypothesized that if the classifier was able to distinguish between these speakers, they have to differ in their choice and pronunciation of words. To study how speakers differ from each other, we inspected which features in the speaker profiles created by the classifier were relevant for classification.

Our classification tests based on only lexical features resulted in a high performance (73.9%), indicating that the speakers differed in the words they used. Two types of words appeared to be relevant in distinguishing between speakers. The first are function words (e.g., *want* 'because' and *dat* 'that'), which is as expected as they are often regarded as useful features in research on authorship attribution (e.g., Koppel et al., 2009; Stamatatos, 2009). The second type are highly frequent content words that are semantically relatively weak (e.g., *goed* 'good' and *no* 'nee'). In authorship attribution, content words are often argued to be topic dependent and thus less suitable for distinguishing between writers. Our results suggest that including semantically weak content words may provide more information than including function words only and may thus be beneficial.

Classification tests including features that represent pronunciation variation typical for casual speech performed better than tests based on lexical features alone. However, inspection of the relevant pronunciation features showed that some of them probably represent lexical information rather than pronunciation variation. Others represent pronunciation variation and, importantly, they are the ones responsible for the increase in classifier performance.

These pronunciation features show that speakers differ in how often they reduce certain phones and phone sequences. Closer inspection of the speaker specific triphone features showed that some of them mainly originate from a few semantically weak words, including *allemaal* 'all' and the pronoun *je* 'you'. This suggests that speakers differ in how they realize phones not only given the immediate phone context but also given the carrier word. Moreover, it shows that semantically weak words contribute to speaker classification at two levels: speakers differ in their use and pronunciation of these words.

Interestingly, the performance of classification with uniphones and triphones was higher than that of classifications including biphones. Apparently, the pronunciation of single phones by themselves or with two neighboring segments is more informative than the pronunciation of these phones given only one neighboring segment. A likely explanation is that generally a phone's pronunciation does not only depend on one of the neighboring phones but on both neighbors. Moreover, as mentioned above, speakers differ from each other in how much they reduce certain semantically weak words and a given word is better identified by a triphone than by a biphone.

Another type of pronunciation feature that did not increase classification performance is formed by the word pronunciations. In the classification including pronunciation features at the word level, we replaced lexical features that were also represented as word pronunciation features by their actual pronunciation. By doing so, we favored the word pronunciation features. Nevertheless, all classifications including word pronunciation features performed worse than the classification with lexical features and uniphone and triphone features. Probably, the triphone features contained all information present in the word pronunciation features, in addition to pronunciation variation spanning word boundaries.

Previous research has shown that speakers participating in the same conversation tend to align their speech at for instance lexical, syntactic, and phonological levels (e.g., Pickering & Garrod, 2004). We expected that this alignment would demonstrate itself in that speakers within a conversation would be more often confused with each other than with other speakers in general. This expectation was borne out for only 30% of the speakers. This low number indicates that the properties of the speech produced by a given speaker at a given moment is colored more by idiosyncratic speech habits than by processes of speech alignment with the conversation partner.

Van Bael & Van Halteren (2007) investigated classification of groups of speakers on the basis of phonetic transcriptions and reported that classification based on pronunciation features performed poorly. The authors discussed several possible explanations for this finding. Our results indicate which of their explanations is most likely. First, the authors noted that their pronunciation feature set may have been too small as it contained only 94 pronunciation features. This is a likely explanation since we used many more pronunciation features and our classification improved when we added any type of pronunciation features to the lexical features (see Table 6.2). A second explanation provided by Van Bael & Van Halteren (2007) concerns the heterogeneity within their classes. They classified speakers in terms of social groups defined by, for instance, regional background and age. As a consequence, one class contained multiple speakers who may have had different pronunciation habits. Our findings suggest that this is also a highly probable explanation, as our classifier was able to distinguish between speakers within the same social group.

Van Bael & Van Halteren (2007) also suggested that their phonetic transcriptions may not have been sufficiently detailed to capture pronunciation differences among speakers. Since the transcriptions that we used were also broad phonetic transcriptions without any fine-phonetic detail, our results suggest that this is an unlikely explanation. However, we agree with these authors that individual differences may be larger if also fine-phonetic detail is taken into account. Whereas our study only investigated whether speakers differ in which segments they realize or substitute by other segments (i.e., categorical reduction), previous research has shown that reduction may also be gradient in nature (e.g., Browman & Goldstein, 1990; Davidson, 2006; Torreira & Ernestus, 2010b, Chapter 2): only a part of a segment may be reduced or segments may not be reduced sufficiently to be identified as different phones. Future studies focusing on both categorical and gradient reduction may report larger individual differences.

Currently, psycholinguistic models explain how the average speaker produces his speech. Our finding that individual speakers differ in their choice and pronunciation of words needs to be incorporated in these models. For instance, in a production model like WEAVER++ (Levelt et al., 1999), speakers should differ in the resting activation levels of words and perhaps even of pronunciation variants. Exemplar-based production models, such as the one described by Goldinger (1998), should assume that speakers differ in their number of exemplars for some words and pronunciation variants. Furthermore, speech comprehension models may assume that a listener adapts to the specific pronunciations of the speaker he is listening to.

In conclusion, the speakers that we investigated belonged to a homogeneous group, and may therefore be expected to show similar speech habits. Nevertheless, our classification tests showed that these speakers differ in the words they use, as well as in how they pronounce the words in casual conversations. Individual differences between speakers' pronunciations can be observed, even if these speakers have the same social background.

General discussion and conclusions

Chapter 7

The aim of this dissertation was to improve our understanding of the way language users produce and recognize acoustically reduced words in casual speech. It presented a series of studies on reduction processes typical for natural Dutch conversations. In this chapter, we discuss what these empirical chapters have taught us about how to investigate the production and comprehension of reduced speech. Moreover, we examine the results of these studies and discuss the implications for psycholinguistic models of speech production and comprehension.

Methodology

While conducting our studies, we encountered three main types of methodological challenges. The first of these concerns the analysis of corpus data and experimental data. Our analyses were difficult, as variation in the data could be explained by many different predictors, which were not necessarily those we were interested in. For instance, in Chapter 4, we had no special interest in the effects of word duration, although it is known to be an important predictor for response times. If we had excluded word duration from our analyses, any variation it could have captured may then have been captured by other variables, including the predictors of interest. As a consequence, any effects of the predictors of interest might then have (partly) been caused by word duration. We therefore had to include a large number of predictors. However, the inclusion of many predictors can lead to overfitting, which has to be prevented. The second challenge is the method that we used in Chapter 6, namely classifying speakers on the basis of their choice and pronunciation of words. As this method has not often been used to study individual differences in pronunciation variation, we had to develop a new approach to adapt this method for our research question. The third challenge concerns the fact that this dissertation focuses on reduction processes in casual speech, while speakers usually do not produce this type of speech when placed in front of a microphone in a laboratory setting.

Analyses

In Chapters 2, 3, and 4, we statistically analyzed our corpus data or experimental data using mixed effects regression models. We first fitted a model with all fixed effects, namely predictors of interest and control predictors, and eliminated non-significant effects (backward elimination). Subsequently, we tested for interactions between the remaining fixed effects and predictors of interest (forward selection). Furthermore, we tested the included fixed effects as random slopes, as recommended by Barr, Levy, Scheepers, & Tily (2013), and kept those random slopes that were significant. In our studies, this procedure led to the inclusion of many fixed effects and interactions; some of these had very small effect sizes, some were uninterpretable higher-order interactions, and others were not replicable and probably resulted from overfitting. To improve the reliability of our results, we adopted more conservative statistical approaches with respect to the inclusion of fixed effects. Which conservative approach was used depended on the aim of the analyses.

In Chapters 2 and 3, the aim was to find out which variables predict the degree of reduction of a segment and to compare the predictors that influence the shortening of a segment with those that influence its absence. In order to base this comparison on reliable predictors, we included only those predictors that were statistically significant and improved the quality of the model. The model's relative quality was determined by means of the Akaike Information Criterion (Akaike, 1973), which takes the goodness of fit of the model and the complexity of the predictors into account. This procedure led to a manageable number of interpretable predictors.

In Chapter 4, we set out to examine how robust priming effects are that provide evidence in favor of the storage of exemplars. In these analyses, only a few predictors were of specific interest. One predictor allowed us to investigate generalizability over the two speakers who recorded the stimuli. Two other predictors indicated whether primes and targets were similar in their degree of reduction, and whether they were produced by the same speaker. We tested for main effects of these predictors of interest as well as several control predictors, but restricted our analyses to interactions between the predictors of interest. In this way, we obtained clear and interpretable results.

In Chapter 5, we reviewed several studies that have argued that segments important for a word's morphological structure tend to be less reduced. This review showed that not only the selection of the best statistical model has to be conducted with care, but also the interpretation of this model. We questioned the interpretations of data patterns revealed in several of the studies. For instance, some studies based their conclusions on the comparison of two small sets of words that not only differed in morphological structure but also in other properties, such as frequency of occurrence. These differences in other properties instead of difference in morphological structure may be responsible for the difference in pronunciation observed between these two sets of words. Furthermore, several researchers investigated the ease with which a morphologically complex word can be decomposed, and used different quantifications of this decomposability, which resulted in opposite findings. Since the researchers did not provide theoretical justifications for their quantification methods, these findings are difficult to interpret.

This dissertation thus demonstrates that many statistical approaches can be used to analyze the same speech data and that these do not necessarily lead to the same results. There is currently no standard statistical approach and this dissertation can be viewed as a contribution to the development of such a standard. Furthermore, it demonstrates that some results can be explained by multiple hypotheses, and that these hypotheses should all be carefully considered before strong conclusions are drawn.

Classification

Chapter 6 presented the results of a study on individual differences in the choice and pronunciation of words. We classified speakers using lexical and pronunciation features. Examples of lexical features are the single word *allemaal* ‘all’ and the word combination *ze zeggen* ‘they say’, whereas examples of pronunciation features are the pronunciation of /ɛɪk/ for the word /ɛɪxələk/ *eigenlijk* ‘actually’ and the absence of the segment /x/. The classifier provided insights into which features are relevant for the identification of the speaker of a speech fragment. We inspected the relevant features, however, as running the classifier twice resulted in slightly different results and thus slightly different relevant features, not all features identified as relevant by a given classifier were reliable. We therefore used the weights assigned to these features which indicate how relevant a feature is relative to the other features, and inspected only those features with a weight higher than the median weight.

The original aim of Chapter 6 was to focus on differences in pronunciation only. However, we had to consider differences in word choice as well because some pronunciation features appeared to reflect lexical differences rather than pronunciation differences. For instance, the sequence /tyr/ did not occur in the feature set as a reduced sequence, in which /y/ is absent or replaced, but only occurred as an unreduced sequence. The speaker for whom this unreduced feature was characteristic therefore differed from other speakers in the frequency with which he produced the sequence, regardless of the pronunciation. To determine whether speakers really differed in degree of reduction, our classifiers were also based on lexical features, representing differences in word choice. As a consequence, the pronunciation features are more likely to really capture pronunciation variation. Moreover, we examined every relevant pronunciation feature as to whether it really reflected differences in pronunciation, by determining by hand whether the speakers not only differed in the pronunciation of the relevant phonemes, but also in the words containing these phonemes.

Casual versus laboratory speech

In a substantial part of this dissertation, we focused on the question of whether reduced pronunciation variants have their own lexical representations. This has previously been examined by means of psycholinguistic experiments investigating whether the frequencies of these variants influence speech processing (e.g., Bürki et al., 2010; Ranbom & Connine, 2007; Pitt et al., 2011). These researchers argue that if these frequencies influence speech processing, they have to be stored, together with the variants they belong to. Unfortunately, the role of variant frequency in speech processing can only be studied well in words presented in isolation, but many reductions only occur in words in the middle of sentences.

In Chapter 2, we therefore used a different method to investigate whether reduced pronunciation variants have their own lexical representations. We based our investigations on recordings of spontaneous conversations between speakers who knew each other well. As these conversations took place in an informal setting in which the speakers were able to talk freely, the speech of these recordings mirrors everyday casual speech. We focused on two reduction processes in Dutch that affect many word types and only occur in casual, connected speech, namely reduction of schwa and /t/. Following Bürki, Ernestus, et al. (2011) and Torreira & Ernestus (2011), for each of these segments we compared the sets of predictors that influence the shortening and absence of the segment. If a segment's absence is influenced by the same set of predictors as its shortening, its absence is likely to be the extreme result of its shortening. In contrast, if the sets of predictors differ, the segment's absence is not likely to be the result of extreme shortening; rather, it results from a process that makes the segment categorically absent, such as the retrieval of a pronunciation variant in which this segment is completely missing. As the data were extracted from spontaneous conversations, the segments investigated differed from each other in many aspects, which could all influence their shortening and absence but which could not all easily be established and tested (e.g., prosodic position, speech rate, frequency of the carrier word, and surrounding segments). This raises the question of whether it is possible at all to find the same sets of predictors in this type of data and thus whether our research method is valid. Torreira & Ernestus (2010b) also studied highly casual speech and found that complete and partial devoicing of vowels in spontaneous French are influenced by the same sets of predictors. Moreover, in our own data, we found that the absence and shortening of /t/ are influenced by the same sets of predictors. Finally, Chapter 3 showed that finding the same sets of predictors is possible on the basis of less variable data. This demonstrates that if differences in the sets of predictors are found (as we did for schwa in Chapter 2), this cannot be ascribed to chance, but that it is meaningful.

The segments studied in Chapter 2 differed from each other in many aspects, including those that could not easily be determined, such as prosodic position. We therefore tried to replicate our results on the basis of speech in which the segments studied only differed in easily determinable aspects, whereas other aspects were kept the same. We constructed sentences in which these segments, for instance, differed in the preceding segment but occurred in the same prosodic position. One speaker recorded these sentences twice: once in a formal speech style and once in a casual speech style, including many reductions. In a shadowing experiment, we then asked participants to repeat these sentences as quickly as possible. We hoped that this procedure would result in casual realizations of the sentences. Unfortunately, it did not. The speech elicited did not contain any extremely reduced forms that are typical of casual speech, such as /ɛɪk/ for /ɛɪxələk/ *eigenlijk* 'actually' or /ɑm/ for /ɑləmal/ *allemaal* 'all'. Moreover, the number of absent schwas in the participants' responses was relatively small regardless of the degree of reduction of the sentence they heard, indicating that participants did not repeat the pronunciation variants they heard, but aimed at producing full forms. These results suggest that the shadowing task is inappropriate for eliciting casual speech.

During the preparation of the experiments described in Chapter 4, we also experienced how difficult it is to elicit a certain type of speech. We asked two speakers to record words in a reduced and an unreduced way. As one of these speakers normally produces speech with a larger degree of reduction than the other speaker, we informed the speakers of the type of stimuli we wanted to obtain, hoping that this would result in similar reduced stimuli as well as similar unreduced stimuli. The speaker who normally produces highly reduced speech made an effort to clearly and slowly pronounce the unreduced stimuli, and the speaker whose speech is typically not very reduced made an effort to produce the reduced stimuli by pronouncing them quickly and sloppily. Unfortunately, the degree of reduction of the reduced and unreduced stimuli was still larger for the speaker who tends to reduce more.

In sum, corpus studies allow us to investigate speech phenomena as they occur in natural conversations, but the data differ from each other in many aspects, making it more difficult to investigate why one speech unit differs from another. Moreover, corpus data cannot easily provide information about the time course of the production process and the predictors that influence this time course. This is, however, relevant for our understanding of the speech production process. The data from experimental studies tend to vary in fewer aspects and may provide information about the time-course of speech processing. However, as the shadowing task used in Chapter 3 was unable to elicit casual speech, still more research is required in order to find an experimental paradigm that allows us to study the production of this type of speech. Currently, research needs to use both corpus and experimental studies complementarily, which ideally will provide converging evidence.

Experimental results

Mental representations

Chapters 2 and 3 reported on the processes underlying the absence of a segment. A segment's absence can result from extreme gradient shortening of the segment during articulation. Alternatively, it may result from a categorical process, for instance, the retrieval of a stored pronunciation variant in which segments are completely (categorically) absent. In both chapters, we investigated whether the absence of schwa and /t/ in Dutch are gradient or categorical by comparing the predictors that influence a segment's shortening and absence.

In Chapter 2, we examined natural casual conversations, and in Chapter 3, we studied speech elicited by a shadowing task. As the speech studied in Chapter 3 appeared to be less casual than the speech studied in Chapter 2; we in fact, studied both segments in two types of speech. The results for /t/ indicate that the shortening and absence of this segment are influenced by the same set of predictors in both types of speech. This was also found for schwa but only in the less casual speech elicited by the shadowing task. The relevant predictors are phonetic in nature; for example, schwa tended to be shorter and more often absent if the preceding segment was longer, and /t/ tended to be shorter and more often absent if the preceding segment was a fricative rather than another preceding segment. Both examples suggest that schwa and /t/ may be co-articulated with their preceding segment. The absence of these segments therefore appears to be the extreme result of segment shortening, and thus of a gradient process. This result is in line with earlier papers reporting gradient reduction processes underlying absence of English /t/ and schwa (Browman & Goldstein, 1990, 1992), absence of schwa in English consonant-schwa-consonant sequences (Davidson, 2006), absence of French schwa (Bürki et al., 2010), and French vowel devoicing (Torreira & Ernestus, 2010b).

The sets of predictors for the absence and for the shortening of schwa in casual conversations showed similarities as well as clear differences (Chapter 2). Both schwa's absence and shortening were influenced by, for instance, speech rate and the syllable that the schwa belonged to. These predictors suggest that, at high speech rates, schwa can be co-articulated with its surrounding consonants, and its absence can thus result from extreme gradient shortening. Predictors that influenced either schwa's absence or shortening include word length and duration of the preceding consonant, which influenced schwa's absence but not its shortening, and the place of articulation of the surrounding consonants, which only influenced schwa's shortening. These results strongly suggest that the absence of schwa may also result from a categorical process. Earlier support for categorical reduction is based on processes that only affect highly frequent word combinations (e.g., /e/ in French /setɛ/ *c'était* 'it was'; Torreira & Ernestus, 2011) or affect words both in casual and formal speech or both in context and in isolation (e.g., French word-internal schwa; Bürki, Ernestus, et al., 2011). Our results contribute to these findings by showing that reduction processes may be categorical, if they affect many word types and only occur in casual, connected speech.

Further, our finding that the absence of schwa may result from a categorical process in casual speech, whereas it typically results from extreme shortening in less casual speech, raises questions about the production processes underlying different speech styles. If other reduction processes that are categorical in casual, connected speech are gradient in less casual speech, the production processes underlying different speech styles may be qualitatively different: the role that lexical representations for reduced pronunciation variants play during speech production then depends on the speech style.

In Chapter 4, we focused on the role of exemplars in the comprehension of reduced speech. A role for exemplars is supported by earlier studies reporting priming effects if primes and targets were acoustically identical, whereas these effects were smaller or even absent if primes and targets differed in, for instance, the speaker or the realization of a certain segment (e.g., Bradlow et al., 1999; McLennan et al., 2003). We investigated how robust these exemplar effects are by focusing on reduction and examining whether they also arise in conditions more similar to natural conversations. Most reduced stimuli used in this study contained acoustic traces of all their segments and thus represented gradient reduction. As these pronunciation variants cannot be stored in abstract representations, any priming effects would therefore provide evidence for exemplars.

Only one of the four priming experiments showed exemplar effects. This was the simplest experiment in which participants were presented with one type of variation (i.e., degree of reduction), participants only heard one speaker, the lag between primes and targets was small, and a large proportion of the words were primed, which probably made the repetition of words clear to the participants. This experiment was most similar to earlier priming experiments showing exemplar effects. The other three experiments tested exemplar effects under conditions more similar to those of natural conversations and did not show any exemplar effects. In Experiments 2 and 4, a smaller proportion of words were primed and the lag between primes and targets was longer; in Experiments 3 and 4, participants were presented with differences in both degree of reduction and speaker voice (participants heard two speakers). Overall, in Chapter 4, we suggest that in casual conversations, exemplars do not play a large role in speech comprehension. Perhaps exemplars are only relevant in situations where a stimulus that created an exemplar is likely to be repeated, and where accessing exemplars may be helpful for word recognition.

Furthermore, exemplars are possibly not stored in the mental lexicon but only in episodic memory. Whereas the mental lexicon represents procedural knowledge stored in the neocortex, episodic memory is stored in the hippocampus and contains detailed, contextual knowledge about past events; for instance, it not only contains the acoustic details of a given utterance but also knowledge about when and where the event happened, and who else was present. These episodic memories are not very stable and disappear after a short period of time (often estimated at several hours up to four or five days; e.g., Vertes, 2004). The fact that episodic memory is not very stable can explain our findings that priming effects are not very robust. More research is needed to investigate this further.

Morphology

In Chapter 5, we reviewed several papers that suggested that the degree of reduction of a word is influenced by its morphological structure, and we re-analyzed some of the data. Moreover, we contrasted this morphological structure hypothesis with the word information load hypothesis that states that reduction is influenced by a segment's relevance for the identification of the entire word, regardless of its morphological structure. For instance, we examined studies on whether a segment tends to be less reduced if it forms a morpheme on its own than if it is part of a longer morpheme. The data on these single segment affixes as well as the other data reviewed provided no clear evidence for a substantial role for morphological structure, but demonstrated that word information load is more important in the production of reduced words. We concluded that, so far, there is no convincing evidence that morphological structure influences acoustic reduction.

This conclusion is in line with data from Chapter 2 on the reduction of schwa in the prefixes *ge-*, *be-*, and *ver-* of Dutch past participles. Although schwa is equally relevant in these prefixes for the identification of the prefix and therefore for the identification of the word's morphological structure, it tended to be more reduced in the prefix *ge-* than in the prefixes *be-* and *ver-*. This difference can be explained by the prefixes' relevance for the identification of the complete word: whereas the prefix *ge-* only indicates the word's syntactic function, namely that it is a past participle, the prefixes *be-* and *ver-* contain semantic information that is crucial for identifying the meaning of the entire word. For example, the stem *taal* can be combined with both prefixes *be-* and *ver-*, but the meanings of the two resulting verbs are completely unrelated: the past participle *betaald* means 'paid' and the past participle *vertaald* means 'translated'.

Future research should further determine whether there is any role for morphological structure in the processing of casual speech. One line of research may be based on the study by Cohen-Goldberg (2013), who reported that during the production process the phones of a word compete with each other, and that this competition is smaller across morpheme boundaries. His study was based on words read aloud in isolation and the question arises whether such effects can also be found for words produced in more natural conditions, such as casual, connected speech.

Individual differences

In Chapter 6, we reported on whether individual speakers differ in their choice and degree of reduction of words. Our study focused on a homogeneous group of speakers with the same social background. Earlier studies (e.g., Guy, 1980; Keune et al., 2005) have shown that social groups of speakers, for instance based on gender or educational level, differ in the way they pronounce words. Speakers within a social group should therefore display similar speech patterns. Nevertheless, we found clear differences between individual speakers that belong to the same social group.

These speakers differed in the words they used (i.e., lexical choice) and in how they pronounced these words. Differences in lexical choice concerned function words, such as *om* 'to' and *ze* 'they', and semantically weak content words, such as *nee* 'no' and *goed* 'good'. In the literature on authorship attribution (see Koppel et al., 2009; Stamatatos, 2009), content words are often regarded as being topic dependent and therefore as being less useful for the identification of the author of a text. However, we suggest that the usage of certain content words, namely semantically weak ones, may also be beneficial for research on authorship attribution.

Differences in the speakers' pronunciations were found for single phones (e.g., reduction of /x/) and for phones given both the preceding and following phone (e.g., the pronunciation of /l/ in the sequence /ɑlə/). The pronunciation differences that we found were often only represented by a few words, mostly semantically weak ones. For instance, differences in the reduction of /x/ were mainly based on /nɔx/ *nog* 'yet' and /tɔx/ *toch* 'still' and variation in the sequence /ɑlə/ mainly resulted from variation in the word /ɑləmal/ *allemaal* 'all'. Speakers did not differ in their pronunciation of phones given either the preceding or following phone. Because combinations of two phones occur in many more different word types than combinations of three phones, taken together our results suggest that these speakers did not differ so much in their pronunciations of certain phone sequences, but mainly in their pronunciations of certain semantically weak words.

The study described in Chapter 6 demonstrates that clear differences can be observed in the words that speakers use and the way they pronounce these words, even if these speakers are members of the same social group. To deal with this quite substantial variation, listeners probably adapt to this variation similarly to the way they adapt to speakers with a (foreign) accent (e.g., Dahan et al., 2008; Witteman, 2013). Furthermore, our study demonstrates that differences in degree of reduction can be used for the identification of the speaker of a certain speech fragment, which may be useful in the field of forensic science. More research is needed to establish the usefulness of individual differences in reduction for forensic science.

Psycholinguistic models

The results of our studies provided us with information about speech processes and lexical representations and can be used to improve psycholinguistic models of speech production and comprehension. These models vary mostly in their assumptions about the number of pronunciation variants that are stored and whether the lexical representations contain acoustic detail. They range on a continuum from abstractionist models storing only a word's full pronunciation without acoustic details (e.g., Chomsky & Halle, 1968; Levelt et al., 1999; Pinker, 1991) to exemplar models storing many acoustically detailed (combinations of) tokens for each word (e.g., Bybee, 2001; Goldinger, 1998). In the middle, the continuum contains abstractionist models storing multiple pronunciation representations per word (e.g., Ranbom & Connine, 2007) and hybrid models storing both abstract representations and exemplars (e.g., Goldinger, 2007; Hawkins, 2003; McLennan et al., 2003; Pierrehumbert, 2002).

In Chapter 2, we provided evidence that schwa may be absent categorically in many different words produced in casual, connected Dutch. This result supports models that, for a large number of words, do not store just one but multiple pronunciation variants per word. These lexical representations may be strings of phonemes or be acoustically detailed.

In Chapter 4, we have shown that the role of exemplars may be smaller in the comprehension of casual, connected speech than previously assumed. This study thus provides only little support for models that assign an important role to exemplars in speech processing. Above, we argued that acoustic traces of words are perhaps not part of the mental lexicon, but are stored in episodic memory.

As the results of Chapters 2 and 4 indicate that the mental lexicon stores multiple pronunciation variants for each word, but not in the form of exemplars, the question arises regarding the exact nature of these representations. Purely abstract representations cannot account for gradual changes over time of the pronunciation of some words (Port, 2007), nor can they easily account for the role of the fine acoustic details in speech comprehension (Johnson, 2004). Perhaps the representations do not consist of purely abstract strings of speech units, such as phonemes, but contain some acoustic details that are not token-specific, as in exemplar-based models, but that are generalizations over many tokens. For our understanding of speech processing, it is important that future research investigates the exact nature of mental representations and memory systems that are involved in the storage of multiple pronunciation variants.

Results presented in Chapter 5 showed that there is currently no convincing evidence for a substantial role of morphological structure in speech production. This suggests that morphologically complex words are not computed from their morphemes during speech production, but are retrieved as single units. Complex words are thus stored in the mental lexicon. As the lexicon also contains several pronunciation variants for each word (e.g., Chapter 2; Bürki et al., 2010; Bürki, Ernestus, et al., 2011; Ranbom & Connine, 2007; Pitt et al., 2011; Torreira & Ernestus, 2011), the storage of complex words implies that the lexicon contains many more words and pronunciation variants than previously assumed (e.g., Chomsky & Halle, 1968; Levelt et al., 1999; Pinker, 1991; Schreuder & Baayen, 1995; Taft & Ardasinski, 2006). This presents a challenge for production and comprehension models, as they have to explain how speakers and listeners select the correct word and pronunciation variant.

Our finding that morphological structure does not play a substantial role in speech production raises the question of how speakers create new morphologically complex words and how these words are understood by listeners. One possibility is that the words that share form and meaning (i.e., morphologically related words) form networks in the mental lexicon (Bybee, 2001). New words may be created and understood on the basis of these networks.

Finally, in Chapter 6, we demonstrated that individual speakers differ in the words they produce and how they pronounce these words. To incorporate these results, psycholinguistic models need to adopt the assumption that the mental lexicon may not be equal for all speakers. Models including abstract representations may assume that speakers differ in the resting activation levels of each of the stored pronunciation variants. If future research shows a clear role for exemplars, which would be unexpected given the findings described in Chapter 4, models storing exemplars may assume that speakers differ in the number of exemplars they have stored for some words and pronunciation variants.

Conclusions

In this dissertation, we focused on reduction processes that are typical for casual, conversational Dutch. We addressed the difficulties of studying how speakers produce and how listeners understand this type of speech and showed the importance of conducting corpus studies and experimental research complementarily. The results of our studies provide important information about how speakers and listeners process speech in general, more specifically about the nature and the role of lexical representations. Research on casual speech is therefore not only beneficial for gaining a better understanding of how speakers and listeners process casual speech, but it is also beneficial for our understanding of speech processing in general.

English summary

One characteristic of casual, connected speech is the acoustic reduction of words. Words are often not produced as their full forms with all segments carefully articulated, but are pronounced with altered, shortened, or even absent segments. For instance, the English word *apparently* has the full form /əpʰɛ˞rəntli/, but may be pronounced as [pʰɛ˞ɾi] or [pʰɛ˞] (Johnson, 2004). Similarly, the Dutch word *natuurlijk* ‘of course’ with the full form /natyrlək/ may be reduced to for instance [ntyrlək], [tylək], or [tyk] (Ernestus, 2000). Acoustic reduction is a highly frequent phenomenon (e.g., Johnson, 2004; Schuppler et al., 2011) that occurs in many different languages (Ernestus & Warner, 2011). This dissertation investigates how speakers and listeners process acoustic reductions, on the basis of casual Dutch.

Processes underlying acoustic reduction

In Chapters 2 and 3, we aimed at improving our knowledge of the processes underlying the production of reduced pronunciation variants. Two different types of processes may result in the absence of segments. On the one hand, segments may be absent as the extreme result of gradient processes: Gradient overlap and decrease in magnitude of articulatory gestures may result in the shortening of segments and in extreme cases in their absence (e.g., Browman & Goldstein, 1990, 1992). On the other hand, segments may be absent as a result of categorical higher-level processes. A phonological rule may completely delete segments from a word’s full representation (e.g., Chomsky & Halle, 1968; McCarthy & Prince, 1993), or the mental lexicon may contain multiple pronunciation variants for each word and a variant in which segments are completely (categorically) absent may be retrieved directly (e.g., Bürki et al., 2010).

So far, categorical reduction has been identified for processes that only affect highly frequent words or word combinations (e.g., /e/-deletion in French *c’était* /setɛ/ ‘it was’; Torreira & Ernestus, 2011) or affect words in both casual and formal speech or both in context and in isolation (e.g., word-internal schwa deletion in French; Bürki et al., 2010). We investigated whether highly frequent reduction processes that affect many word types, but only in casual, connected speech, can also be categorical in nature. We focused on reduction of schwa in prefixes and on word-final /t/ in Dutch past participles (e.g., *gewenst* /xəwɛnst/ ‘wished’). If a segment’s absence and duration are conditioned by the same variables, we hypothesize that they result from the same shortening process, which in extreme cases results in the absence of a segment. If its absence is conditioned by different variables than its duration, we hypothesize that its absence results from categorical processes.

Chapter 2 presents a corpus study based on tokens of past participles extracted from recordings of casual conversations. Its results demonstrate that the absence and duration of /t/ are affected by approximately the same variables. For schwa, the data show that its absence is affected by different and more variables than its duration. These results strongly suggest that the absence of schwa, a reduction phenomenon that occurs in many word types and only in casual, connected speech, may also result from a categorical process.

This categorical process could be a phonological rule that deletes the schwa. Because our analyses show that words differ substantially in how often they occur without schwa, this rule then has to apply to different words with different frequencies. Word-specific rules or word-specific application frequencies are computationally difficult to distinguish from the assumption that the variants without schwa are stored in the mental lexicon, together with their frequencies of occurrence. Our results therefore support psycholinguistic models assuming that, for a large number of words, the mental lexicon does not store just one, but multiple pronunciation variations per word.

In Chapter 3, we aimed at replicating the results of our corpus study in a controlled production experiment. We conducted a shadowing experiment, in which participants listened to sentences produced in a careful or casual way and repeated these sentences. In the participants' productions of the sentences, the absence of /t/ was conditioned by the same variables as its duration, suggesting that its absence mostly resulted from extreme shortening. These results support those obtained in the corpus study. The same results were obtained for the absence of schwa, which is not in line with the data from the corpus study. This raised the question what could account for this discrepancy in results for schwa. Further inspection of the recordings indicated that participants did not aim at repeating the pronunciation variants of the words they heard, but at producing the words' full forms. In addition, participants never produced drastically reduced forms that are typical of casual, connected speech (e.g., *eigenlijk* /ɛɪxələk/ was never reduced to [ɛɪk]). This indicates that the shadowing experiment elicited non-casual fast speech, which may explain why we did not find evidence for categorical absence of schwa, as reported in the corpus study.

Our finding that the absence of schwa may result from a categorical process in casual speech, whereas it typically results from extreme shortening in less casual speech, raises questions about the production processes underlying different speech styles. If other reduction processes that are categorical in casual, connected speech are gradient in less casual speech, the production processes underlying different speech styles may be qualitatively different: the role that lexical representations for reduced pronunciation variants play during speech production then depends on the speech style.

Further, this finding is in line with exemplar-based models (e.g., Bybee, 2001, p. 35; Goldinger, 1998) and hybrid models (e.g., Goldinger, 2007; McLennan et al., 2003), which both assume that the mental lexicon contains representations of all occurrences of a word that a language user has ever encountered (i.e., exemplars).

In Chapter 4, we further examined evidence for these models. A role for exemplars in speech processing is supported by earlier studies of speech comprehension using long-term identity priming. These studies report larger priming effects for words that are primed by acoustically identical primes than by primes that differ in, for instance, the speaker or the realization of a certain segment (e.g., Bradlow et al., 1999; McLennan et al., 2003). We investigated how robust these exemplar effects are for reduced pronunciation variants by studying whether exemplar effects also arise in conditions more similar to natural conversations than in earlier experiments.

Our study consists of four long-term priming experiments with lexical decision tasks. A prime and target represented the same word type and were presented with the same or a different degree of reduction, but were never acoustically identical. Only Experiment 1 showed exemplar effects. This was the most basic experiment in which participants were presented with one type of variation (i.e., degree of reduction), participants heard only one speaker, the lag between primes and targets was small (maximally 100 trials), and a large proportion of the words were primed (33%), which probably made it clear to the participants that many words in the experiment were repeated. This experiment was most similar to earlier priming experiments showing exemplar effects.

The other three experiments tested exemplar effects under conditions more similar to those of natural conversations. In Experiments 2 and 4, a smaller proportion of words were primed (12%) and the lag between primes and targets was longer (maximally 805 trials); in Experiments 3 and 4, participants were presented with differences in both degree of reduction and speaker voice (participants heard two speakers). None of these experiments showed any exemplar effects. This suggest that, in a situation where more variation is available to the listener, as in natural conversations, exemplars play a smaller role than has been previously assumed. Exemplars appear only relevant in situations where a stimulus is likely to be repeated, and where accessing exemplars may thus be helpful for word recognition. A possible explanation of the findings in our study and in the other studies reporting exemplar effects is that exemplars are not part of the mental lexicon but are stored in episodic memory. The fact that episodic memories are not very stable is in line with our findings that exemplar effects are not very robust.

Morphology

In Chapter 5, we examined whether a word's morphological structure affects the reduction degree of its segments. For instance, English /t/ and /d/ have been reported to be less reduced when they form an affix on their own, as in *rapped*, than when they are part of longer morphemes, as in *rapt* (e.g., Losiewicz, 1992). We examined this question by discussing and re-analyzing data from the literature. Moreover, we contrasted this morphological structure hypothesis with the word information load hypothesis that states that a segment's reduction degree is influenced by its relevance for the identification of the entire word, regardless of its morphological structure.

Our review discussed studies investigating effects of the repetition of a morpheme (Viebahn et al., 2012), of whether a segment plays a crucial role in the identification of its morpheme (Losiewicz, 1992; Schuppler et al., 2012; new analyses based on the data of Chapter 2), and of a word's morphological decomposability (Hay, 2003; Schuppler et al., 2012; Bürki, Ernestus, et al., 2011; Chapter 2). Our analyses revealed that the studies reviewed either do not report convincing evidence that morphological structure influenced acoustic reduction, or they report data that are open to alternative interpretations. Word information load appears to play a more important role in the production of reduced words. Furthermore, the studies that investigated the role of a word's morphological decomposability used different measures for decomposability, without providing clear argumentation for their measures. To understand the role of decomposability in acoustic reduction, we need a uniform definition of morphological decomposability.

These findings contrast with the expectations of several models of morphological processing, such as traditional models assuming that regular complex words are always computed from their morphemes (e.g., Chomsky & Halle, 1968; Pinker, 1991; Taft & Ardasinski, 2006) or models that assume that words can be processed via a direct and a decomposition route (e.g., Schreuder & Baayen, 1995). The data collected so far support psycholinguistic models which assume that all morphologically complex words are processed as complete units (e.g., Bybee, 2001; Skousen, 1989).

Individual differences

Language users have a multitude of different words at their disposal, and individuals may differ in their choice of words (e.g., the words *begin* and *start* can often be interchanged). In addition to the choice of words, speakers may also differ in the degree to which they reduce these words. In Chapter 6, we studied a homogeneous group of speakers with the same social background and investigated whether and how these twenty speakers differ in their choice and degree of reduction of words when engaged in casual conversation. Earlier studies (e.g., Guy, 1980; Keune et al., 2005) have shown that social groups of speakers, for instance based on gender or educational level, differ in the way they pronounce words. Speakers within a social group should therefore display similar speech patterns. Nevertheless, we found that a classifier was able to correctly classify 79.9% of unseen speech fragments as belonging to one of the speakers, which indicates that clear differences may exist between individual speakers belonging to the same social group.

These speakers differed in the words they used (i.e., lexical choice) and in how they pronounced these words. Differences in lexical choice concerned mainly function words, such as *om* 'to' and *ze* 'they', and content words that are semantically weak as they provide no indication of the topic of the conversation, such as *nee* 'no' and *goed* 'good'. In the literature on authorship attribution (see Koppel et al., 2009; Stamatatos, 2009), content words are often regarded as being topic dependent and therefore as being less useful for the identification of the author of a text. Our results suggest that the usage of certain content words, namely semantically weak ones, may also be beneficial for authorship attribution.

Differences in the speakers' pronunciations were found for single phones (e.g., pronunciation of /x/) and for phones given both the preceding and following phone (e.g., the pronunciation of /l/ in the sequence /ɑlə/). The pronunciation differences that we found were often only represented by a few words, mostly semantically weak ones. For instance, differences in the pronunciation of /x/ were mainly based on /nɔx/ *nog* 'yet' and /tɔx/ *toch* 'still' and variation in the sequence /ɑlə/ mainly resulted from variation in the word /ɑləmal/ *allemaal* 'all'. Hence, speakers differ not only from each other in the use of semantically weak words but also in the degree to which they reduce these words, even if these speakers belong to the same social group.

Current psycholinguistic models explain how the average speaker produces speech. To incorporate our results, these models need to adopt the assumption that speech production processes and the mental lexicon may not be equal for all speakers.

Methodology

While conducting the studies described above, we encountered three main types of methodological challenges. The first of these concerns the statistical analysis of corpus data and experimental data. In Chapters 2, 3, and 4, we used mixed effects regression models. Analyses first led to the inclusion of a high number of fixed effects and interactions, many of which had very small effect sizes, were uninterpretable higher-order interactions, or were not replicable and probably resulted from overfitting. To improve the reliability of our results, we adopted more conservative statistical approaches: In Chapters 2 and 3, we used the Akaike Information Criterion for the selection of fixed effects, and in Chapter 4, we restricted our analyses to interactions between the predictors of interest. In Chapter 5, we also encountered statistical challenges: we showed that not only the selection of the best statistical model has to be conducted with care, but also the interpretation of this model. We questioned the interpretations of data patterns revealed reported in the literature. For instance, some studies based their conclusions on the comparison of two small sets of words that not only differed in morphological structure but also in other properties, such as frequency of occurrence. These differences in other properties may be responsible for the difference in pronunciation observed between these two sets of words. Alternative hypotheses should therefore be carefully considered before strong conclusions are drawn.

We encountered our second methodological challenge in Chapter 6, in which we classified speakers on the basis of their choice and pronunciation of words. The classification method is informative as it provides insights in the differences between the speaker's choice and pronunciation of words, but this method has not often been used to study individual differences in pronunciation variation. We therefore had to develop a new approach to adapt this method for our research question.

The third challenge that we encountered concerns the type of speech that we investigated, namely casual, connected speech. Corpus studies allow us to investigate speech phenomena as they occur in natural conversations. However, casual conversations typically contain many high frequency words and very few low frequency words and the word types that can be investigated on the basis of a corpus is therefore limited. Moreover, the tokens of words or segment sequences that a researcher may want to compare may occur in very different contexts or differ in many other respects. For instance, /t/ occurs in different positions in the words *tree*, *notable*, and *accept*, and these words also differ in, for instance, their prosodic structure, length, word type, and word frequency. These differences may be difficult to determine, making it difficult to investigate why one speech unit differs from another. Researchers therefore prefer to work with data elicited in experiments which are designed to contain the exact speech units under investigation and to elicit them under very similar conditions. However, as the shadowing task used in one of our experiments (Chapter 3) was unable to elicit casual speech, still more research is required in order to find an experimental paradigm that allows us to study casual, connected speech. Currently, research needs to use both corpus and experimental studies complementarily, which ideally will provide converging evidence.

Concluding remarks

This dissertation focused on reduction processes that are typical for casual conversational Dutch. We addressed the difficulties of studying how speakers produce and how listeners understand this type of speech and showed the importance of conducting corpus studies and experimental research complementarily.

The results of our studies provide important information about how speakers and listeners process reduced speech. They show that reduction phenomena that occur in many word types and only in casual, connected speech may result from gradient as well as categorical reduction processes. Furthermore, they show that the role of exemplars in the comprehension of reduced speech may be smaller than previously assumed. Our results also indicate that there is currently no convincing evidence that morphological structure plays a substantial role in the production of reduced speech. Finally, they demonstrate that individual speakers differ in the words they produce and how they pronounce these words.

These findings also provide information about speech processing in general, more specifically about the nature and the role of lexical representations. They suggest that the mental lexicon stores morphologically complex words as entire units and that it stores multiple pronunciation variants per word, but not in the form of exemplars. Finally, the mental lexicon and speech processes may not be equal for all speakers. Research on casual, connected speech is thus not only beneficial for gaining a better understanding of how speakers and listeners process casual, connected speech, but also for our understanding of speech processing in general.

Nederlandse samenvatting

In alledaagse spontane spraak worden veel klanken niet volledig uitgesproken. Sprekers kunnen klanken veranderen, verkorten of helemaal weglaten. Ze kunnen bijvoorbeeld *natuurlijk* uitspreken als *ntuurlijk*, *tuulijk* of *tuuk* en *eigenlijk* als *eiglijk*, *eigek* of *eik* (Ernestus, 2000). Dit is een veelvoorkomend fenomeen (bijv. Johnson, 2004; Schuppler et al., 2011) dat akoestische reductie genoemd wordt en in vele verschillende talen voorkomt (Ernestus & Warner, 2011). Mijn proefschrift bevat verschillende studies naar akoestisch gereduceerde spraak.

Alle studies in dit proefschrift onderzoeken spontane lopende spraak in het Nederlands. Het doel van deze studies was om meer te weten te komen over hoe sprekers en luisteraars met gereduceerde spraak omgaan. Hoe produceren sprekers gereduceerde woorden en hoe worden deze door luisteraars verwerkt? Worden gereduceerde woorden ook opgeslagen in ons mentale woordenboek? Als dat zo is, hoe zien de representaties van deze gereduceerde woorden er uit? Antwoorden op deze vragen helpen ons bij het verbeteren van theorieën over spreken en luisteren in het algemeen.

Reductieprocessen

Het doel van Hoofdstuk 2 en 3 was om meer inzicht te krijgen in de manier waarop sprekers woorden en klanken reduceren. Er zijn grofweg twee verschillende processen die ervoor kunnen zorgen dat klanken niet uitgesproken worden. Het eerste proces is een gradueel proces en houdt in dat een spreker een klank verkort door hem gedeeltelijk tegelijkertijd met omringende klanken uit te spreken of door hemzelf te verkorten (bijv. Browman & Goldstein, 1990, 1992). In sommige gevallen leidt dit ertoe dat een klank zo verkort wordt dat hij helemaal niet meer te horen is. Een spreker kan bijvoorbeeld het woord *gewenst* beginnen uit te spreken en de *uh*-klank in *ge* zo verkorten dat het woord uiteindelijk klinkt als *gwenst*. De afwezigheid van de *uh*-klank is dan het extreme gevolg van een gradueel verkortingsproces.

Het tweede proces houdt in dat een klank niet tijdens de uitspraak verloren gaat, maar al op een eerder moment. Een klank wordt dan niet extreem verkort, maar is volledig afwezig doordat hij door de spreker niet gepland is. Dit wordt categoriale reductie genoemd. Categoriale reductie kan ontstaan doordat een spreker een woord uit zijn mentale woordenboek ophaalt en regels toepast die klanken weghalen (bijv.

Chomsky & Halle, 1968). Als de spreker het woord *gewenst* ophaalt, kan een regel bijvoorbeeld stellen dat de *uh*-klank verwijderd kan worden. Een andere mogelijkheid is dat sprekers kunnen kiezen uit verschillende uitspraken opgeslagen in hun mentale woordenboek. Ze hebben bijvoorbeeld *gewenst* opgeslagen, maar ook *gwenst* en *gwens*. Sprekers kunnen dan een uitspraak ophalen, zoals *gwenst*, waarin één of meerdere klanken volledig afwezig zijn (bijv. Bürki et al., 2010).

We hebben onderzocht of categoriale reductie voorkomt in Nederlandse voltooid deelwoorden, zoals *gewenst*, als deze uitgesproken worden in spontane lopende spraak. In deze woorden worden de *uh*-klank in *ge* en de laatste *t*-klank vaak gereduceerd. We hebben onderzocht in welke omstandigheden deze klanken worden verkort en in welke omstandigheden ze afwezig zijn. In een verzameling opnames van spontane gesprekken en in opnames van minder spontane spraak uitgelokt in een experiment, was de t-klank afwezig in dezelfde omstandigheden als de omstandigheden waarin hij verkort werd. Dit geldt ook voor de *uh*-klank in de minder spontane spraak van ons experiment. Omdat de klanken afwezig zijn in dezelfde omstandigheden als waarin ze verkort worden, concluderen we dat dezelfde processen de reductie veroorzaken. In deze gevallen is een klank dus voornamelijk afwezig als gevolg van graduele verkortingsprocessen.

Daarentegen was de *uh*-klank in spontane spraak vaak afwezig in omstandigheden waarin hij niet erg verkort wordt. De *uh*-klank is bijvoorbeeld vaak afwezig als de voorafgaande *g*-klank erg lang was, terwijl het voor de verkorting van de *uh*-klank niet uitmaakt wat de lengte van de voorafgaande *g*-klank is. Dit geeft aan dat de *uh*-klank in spontane spraak onderhevig kan zijn aan zowel graduele als categoriale processen. Verdere analyses toonden aan dat deze categoriale processen waarschijnlijk geen regels zijn die toegepast worden, maar dat een gereduceerde uitspraak uit ons mentale woordenboek gehaald kan worden. Het woord *gewenst* is dus niet alleen opgeslagen als *gewenst* in ons mentale woordenboek, maar ook als *gwenst* en we gebruiken *gwenst* vooral in spontane conversaties.

Deze bevinding komt overeen met theorieën die aannemen dat een aparte representatie van een woord opgeslagen wordt elke keer dat een spreker het woord zegt of een luisteraar het woord hoort. Deze opgeslagen representaties bevatten informatie over de precieze uitspraak van het woord en worden *exemplars* genoemd. Verschillende studies vermelden dat exemplars een rol spelen in het herkennen van spraak (bijv. Bradlow et al., 1999; Craik & Kirsner, 1974; Goh, 2005; McLennan et al., 2003). Sommige van deze studies suggereren dat de rol van exemplars niet in alle omstandigheden duidelijk naar voren komt (bijv. Bradlow et al., 1999; McLennan et al., 2003).

In Hoofdstuk 4 hebben we onderzocht hoe robuust de rol van exemplars is in de herkenning van gereduceerde woorden. Dit hebben we gedaan door de rol van exemplars te onderzoeken in vier experimenten, die verschillen in de mate waarop ze op natuurlijke conversaties lijken. In de experimenten die meer op natuurlijke conversaties lijken, hoorden de proefpersonen bijvoorbeeld niet één maar twee sprekers en meer verschillende woorden. In slechts één van deze experimenten vinden we aanwijzingen voor exemplars. Dit experiment komt het meest overeen met eerdere studies en het minst met natuurlijke conversaties. Omdat exemplars alleen een duidelijke rol spelen in het meest uitgeklede experiment, is de rol van exemplars in de herkenning van gereduceerde woorden minder robuust dan eerder werd gedacht.

Morfologie

In Hoofdstuk 5 hebben we onderzocht of de reductie van klanken in een woord beïnvloed wordt door de morfologische structuur van het woord. Morfologie is de studie naar de kleinste betekenisvolle eenheden van een taal, die *morfemen* worden genoemd. Elk woord bestaat uit één of meerdere morfemen. Het woord *tafel* bijvoorbeeld bestaat uit één morfeem en het woord *eettafel* uit twee morfemen (*eet* en *tafel*). Omdat niet elk morfeem zelfstandig gebruikt kan worden, zijn morfemen niet per definitie gelijk aan woorden. Bijvoorbeeld, in het woord *tafels* geeft het morfeem *s* de meervoudsvorm aan, maar dit morfeem kan niet zelfstandig gebruikt worden.

Verschillende studies suggereren dat de morfologische structuur van een woord invloed heeft op de reductie van de klanken in dit woord (bijv. Losiewicz, 1992; Hay, 2003; Schuppler et al., 2012; Bürki, Ernestus, et al., 2011). De hypothese van deze studies is dat een klank die belangrijk is voor de herkenning van een morfeem, minder snel gereduceerd zal worden. In regelmatige Nederlandse voltooid deelwoorden, zoals *gewenst*, vormt de laatste *t*-klank een eigen morfeem en zou daarom niet snel gereduceerd moeten worden. In Hoofdstuk 5 hebben we een aantal studies besproken en laten we zien dat er op dit moment geen overtuigend bewijs is voor deze hypothese. Onze resultaten toonden aan dat een klank minder snel gereduceerd wordt als deze belangrijk is voor de herkenning van het gehele woord, ongeacht de morfologische structuur van het woord. De *t*-klank in regelmatige voltooid deelwoorden wordt bijvoorbeeld vaker gereduceerd in *gewenst* dan in *gepraat*. In *gewenst* geeft de *t*-klank alleen maar aan dat het een voltooid deelwoord betreft en is niet erg van belang voor de herkenning van het hele woord. In *gepraat* is deze klank niet alleen van belang voor de herkenning van het type woord (voltooid deelwoord), maar ook voor de herkenning van de stam *praat*. Het is daarom relevant voor de herkenning van het woord in zijn geheel.

Morfologie speelt dus geen grote rol in akoestische reductie. Dit komt niet overeen met theorieën die veel belang hechten aan morfologie en die stellen dat sprekers bijvoorbeeld *gewenst* kunnen zeggen door de morfemen *ge*, *wens* en *t* op te halen uit hun mentale woordenboek (bijv. Chomsky & Halle, 1968; Pinker, 1991; Taft & Ardasinski, 2006; Schreuder & Baayen, 1995). Onze bevindingen komen wel overeen met theorieën die stellen dat alle woorden in hun geheel opgehaald worden uit het mentale woordenboek (bijv. Bybee, 2001; Skousen, 1989). Niet alleen simpele woorden zoals *eet*, *tafel* en *wens* zijn opgeslagen en worden in hun geheel verwerkt, maar ook complexe woorden als *eettafel* en *gewenst*.

Individuele verschillen

Alle sprekers reduceren hun spraak in spontane gesprekken. Eerder onderzoek heeft laten zien dat sprekers die tot verschillende sociale groepen behoren, verschillen in de mate waarin ze reduceren. Woorden als *moeilijk* en *natuurlijk* worden bijvoorbeeld vaker tot *moeik* en *tuuk* gereduceerd door mannen dan door vrouwen (Keune et al., 2005). Om dit verschil tussen mannen en vrouwen te kunnen vinden, moeten er overeenkomsten zijn in de spraak van de sprekers binnen zo'n groep. In Hoofdstuk 6 hebben we onderzocht of er binnen zo'n groep ook verschillen te vinden zijn: we hebben onderzocht of een groep van twintig Nederlandse mannen met dezelfde sociale achtergrond verschillen in de woorden die ze gebruiken en de manier waarop ze deze woorden reduceren.

Op basis van een grote verzameling spraakopnames hebben we automatisch bepaald welke woorden deze sprekers uitgesproken hebben. Daarnaast hebben we voor deze woorden bepaald welke klanken uitgesproken, veranderd of afwezig zijn. Op basis van de woorden en uitspraken heeft een computerprogramma geleerd om voor een onbekend spraakfragment te bepalen welke spreker het uitgesproken heeft. Hoewel onze sprekers dezelfde achtergrond hebben en hun spraak op elkaar zou moeten lijken, was dit programma in staat om dit voor 79.9% van de spraakfragmenten goed te doen. Het computerprogramma kan dus onderscheid maken tussen de sprekers, wat aan geeft dat de sprekers onder andere van elkaar moeten verschillen in de mate waarin ze woorden reduceren. Een voorbeeld waarin ze verschillen is de uitspraak van het woord *allemaal*. Sommige sprekers verkorten dit woord meestal tot *amal*, terwijl andere sprekers voorkeur hebben voor een minder gereduceerde variant als *almal* of *allemaal*.

Deze studie toont aan dat sprekers duidelijk van elkaar verschillen in de woorden die ze gebruiken en de mate waarin ze woorden reduceren, ook al hebben ze dezelfde sociale achtergrond. Theorieën beschrijven gewoonlijk hoe spraak geproduceerd wordt in algemene zin en houden daarbij geen rekening met individuele verschillen tussen sprekers. Om onze bevindingen op te nemen, zouden deze theorieën moeten aannemen dat het mentale woordenboek en de spraakproductie-processen niet per definitie voor alle sprekers gelijk zijn.

Algemene conclusies

Dit proefschrift beschrijft onderzoek naar hoe sprekers en luisteraars verkorte uitspraakvarianten van woorden verwerken. We toonden aan dat de afwezigheid van een klank het resultaat kan zijn van zowel graduele processen als categoriale processen: een spreker kan een klank extreem verkorten, waardoor deze niet meer hoorbaar is, en kan ook een gereduceerd woord uit zijn mentale woordenboek ophalen. Verder lieten we zien dat luisteraars bij de herkenning van gereduceerde woorden minder gebruik maken van exemplars (oftewel opgeslagen woorden met details over de precieze uitspraak) dan eerder gedacht werd. Daarnaast worden sprekers niet sterk beïnvloed door morfologie wanneer ze woorden in een gereduceerde vorm uitspreken. Ten slotte laten onze resultaten zien dat individuele sprekers van elkaar verschillen in de mate waarin ze woorden reduceren.

Onze bevindingen verschaffen ook informatie over het verwerken van spraak in algemene zin en over de representaties van woorden in ons mentale woordenboek. Ze suggereren dat ons mentale lexicon bestaat uit in hun geheel opgeslagen woorden en dat het meerdere uitspraken per woord bevat, maar niet in de vorm van exemplars. Een woord als *gewenst* is dus niet opgeslagen als *ge*, *wens* en *t*, maar als *gewenst*. Verder zijn er uitspraken opgeslagen zoals *gewenst* en *gwenst*, maar is er niet een aparte representatie voor elke keer dat een spreker *gwenst* heeft uitgesproken of gehoord. Ten slotte zijn de mentale woordenboeken van sprekers niet per definitie identiek en kunnen sprekers verschillen in de manier waarop ze spreken. Onderzoek naar spontane spraak is dus niet alleen van belang om een beter begrip te krijgen van hoe sprekers en luisteraars reducties in spontane spraak verwerken, maar ook voor ons begrip van het verwerken van spraak in het algemeen.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of isit-1973* (pp. 267–281). Budapest.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (cd-rom)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22.

Bard, E., Shillcock, R., & Altmann, G. (1988). The recognition of words after theii acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, *44*(5), 395–408.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

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

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of Acoustical Society of America*, *113*(2), 1001–1024.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–355.

Booij, G. E. (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

Bradlow, A., Nygaard, L., & Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics*, *61*, 206–219.

Brouwer, S., Mitterer, H., & Heuttig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, *27*(4), 539–571.

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Between the grammar and physics of speech [papers on laboratory of phonology 1]* (pp. 341–376). Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Bürki, A., Ernestus, M., & Frauenfelder, U. (2010). Is there only one 'fenêtre' in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, *62*(4), 421–437.

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *Journal of Acoustical Society of America*, *130*(6), 3980–3991.

Bürki, A., Fougeron, C., Gendrot, C., & Frauenfelder, U. (2011). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics*, *39*(3), 279–288.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. Unpublished doctoral dissertation, LOT, Utrecht. (pp. 171)

Carroll, J., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Boston: Houghton Mifflin.

Cedergren, H. (1987). The spread of language change: Verifying inferences of linguistic diffusion. In P. Lowenberg (Ed.), *Language spread and language policy: Issues, implications and case studies* (pp. 45–60). Washington, DC: Georgetown University Press.

Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*(2), 121–157.

Chomsky, N., & Halle, M. (1968). *The sound pattern of english*. New York: Harper and Row.

Cohen-Goldberg, A. (2013). Towards a theory of multimorphemic word production: The heterogeneity of processing hypothesis. *Lanuage and Cognitive Processes*, 1–29.

Craik, F., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*(2), 274–284.

Dagan, I., Karov, Y., & Roth, D. (1997). Mistake-driven learning in text categorization. In *Proceedings of 2nd conference on empirical methods in natural language processing* (pp. 55–63). Providence, RI.

Dahan, D., Drucker, S., & Scarborough, R. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*, 710–718.

Dalby, J. (1984). *Phonetic structure of fast speech in American English*. Unpublished doctoral dissertation, Indiana University.

Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica*, *63*, 79–112.

Davies, M. (2008). *The corpus of contemporary american english (coca): 425 million words, 1990 – present.* (Available online at http://corpus.byu.edu/coca/)

Ebrahimpour, M., Putnin, T., Berryman, M., Allison, A., Ng, B.-H., & Abbott, D. (2013). Automated authorship attribution using advanced signal classification techniques. *PLOS ONE*, *8*, 1–12.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface*. Unpublished doctoral dissertation, LOT, Utrecht.

Ernestus, M., & Baayen, R. (2011). Corpora and exemplars in phonology. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory (2nd ed.)* (pp. 374–400). Chichester, West Sussex: Wiley-Blackwell.

Ernestus, M., Baayen, R., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, *120*(2), 1040–1051.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*, 253–260.

Faraway, J. (2006). *Extending linear models with R: generalized linear mixed effects and nonparametric regression models*. Boca Raton, FL: Chapman and Hall/CRC.

Fowler, C., & Housum, J. (1987). Talker's signalling of 'new' and 'old' words in speech and listener's perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489 – 504.

Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Gahl, S., & Garnsey, S. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, *80*, 748–775.

Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J., & Gravier, G. (2005). ESTER phase ii evaluation campaign for the rich transcription of French broadcast news. In *Proceedings of Interspeech-2005* (pp. 1149 – 1152). Lisboa, Portugal.

Goh, W. (2005). Talker variability and recognition memory: instance-specific and voice specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 40–53.

Goldinger, S. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychogical Review*, *105*(2), 251–279.

Goldinger, S. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 49–54). Saarbrücken.

Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistics Society*, *35*, 151–166.

Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York: Academic Press.

Guy, G. (1994). Contextual conditioning in variable lexical phonology. *Language Variation and Change*, *3*(2), 223–239.

Hämäläinen, A., Gubian, M., ten Bosch, L., & Boves, L. (2009). Analysis of acoustic reduction using spectral similarity measures. *Journal of Acoustical Society of America*, *126*, 3227–3235.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405.

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

Janse, E. (2008). Spoken-word processing in aphasia: Effects of item overlap and item repetition. *Brain and Language*, *105*, 185–198.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press, Inc.

Johnson, K. (2004). Massive reduction in conversational American English. In *Proceedings of the 10th international symposium on spontaneous speech: Data and analysis* (pp. 29–54). Tokyo, Japan.

Keune, K., Ernestus, M., Van Hout, R., & Baayen, R. (2005). Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok'. *Corpus Linguistics and Linguistic Theory*, *1*, 183–223.

Kipp, A., Wesenick, M., & Schiel, F. (1997). Pronunciation modeling applied to automatic segmentation of spontaneous speech. In *Proceedings of eurospeech-1997* (pp. 1023–1026). Rhodes, Greece.

Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology*, *60*, 9–26.

Koster, C., & Beney, J. (2009). Phrase-based document categorization revisited. In *Proceedings of the pair workshop at cikm-2009* (pp. 49–55). Hong Kong, China.

Koster, C., Seutter, M., & Beney, J. (2003). Multi-classification of patent applications with winnow. In *Lecture notes in computer science* (pp. 545–555). Berlin / Heidelberg: Springer.

Labov, W. (2004). Quantitative analysis of linguistic variation. In U. Ammon, N. Dittmer, K. Mattheier, & P. Trudgill (Eds.), *Sociolinguistics : An international handbook of the science of language and society* (pp. 6–21). De Gruyter.

Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of Acoustical Society of America*, *35*(11), 1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H & H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Dordrecht: Kluwer Academic Publishers.

Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, *2*, 285–318.

Losiewicz, B. (1992). *The effect of frequency on linguistic morphology*. Unpublished doctoral dissertation, University of Texas.

Marslen-Wilson, W., & Zhou, X. (1999). Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes*, *14*(4), 321–352.

Mattys, S., & Liss, J. (2008). On building models of spoken-word recognition: When there is as much to learn from natural 'oddities' as artificial normality. *Perception and Psychophysics*, *70*(7), 1235–1242.

Mattys, S., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*, 145–160.

McCarthy, J. J., & Prince, A. (1993). Generalized alignment. In G. E. Booij & J. Van Marle (Eds.), *Yearbook of morphology* (pp. 79–154). Dordrecht: Kluwer Academic Publishers.

McLennan, C., & Luce, P. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321.

McLennan, C., Luce, P., & Charles-Luce, J. (2003). Representation of lexical form. *Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 539–553.

Mehta, G., & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, *31*(2), 135–156.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73–103.

Nooteboom, S. (1972). *Production and perception of vowel duration: A study of durational properties of vowels in Dutch*. Unpublished doctoral dissertation, University of Utrecht.

Oostdijk, N. (2002). The design of the Spoken Dutch Corpus. In P. Peters, P. Collins, & A. Smith (Eds.), *New frontiers of corpus research* (pp. 105–112). Amsterdam: Rodopi.

Palmeri, T., Goldinger, S., & Pisoni, D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309-328.

Patterson, D., LoCasto, P., & Connine, C. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, *60*, 45–69.

Phillips, B. (1994). Southern English glide deletion revisited. *American Speech*, *69*(2), 15-127.

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–226.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology vii* (pp. 101–140). Berlin: Mouton de Gruyter.

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530 – 535.

Pitt, M., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, *39*, 304–311.

Pitt, M., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*, 89–95.

Pluymaekers, M., Ernestus, M., & Baayen, R. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of Acoustical Society of America*, *118*(4), 2561–2569.

Pluymaekers, M., Ernestus, M., Baayen, R., & Booij, G. (2010). Effects of morphological structure on fine phonetic detail: The case of Dutch -igheid. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 511–531). Berlin: Mouton de Gruyter.

Port, R. (2007). How are words stored in memory? beyond phones and phonemes. *New Ideas in Psychology*, *25*, 143–170.

Ranbom, L., & Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298.

Raymond, W., Pitt, M., Johnson, K., Hume, E., Makashay, M., Dautricourt, R., & Hilts, C. (2002). An analysis of transcription consistency in spontaneous speech from the Buckeye corpus. In *Proceedings of icslp-2002* (pp. 1125–1128). Denver, Colorado.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing Management*, *24*(5), 513–523.

Schreuder, R., & Baayen, R. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 131–154). Hillsdale, New Jersey: Lawrence Erlbaum.

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*(1), 96–109.

Schuppler, B., Van Dommelen, W., Koreman, J., & Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, *40*, 595–607.

Schwarzlose, R., & Bradlow, A. (2001). What happens to segment durations at the end of a word? *Journal of the Acoustical Society of America*, *109*(5), 2292-2292.

Seidenberg, M., & Gonnerman, L. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361.

Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, *60*(3), 538–556.

Strik, H., Van Doremalen, J., & Cucciarini, C. (2008). Pronunciation reduction: How it relates to speech style, gender, and age. In *Proceedings of Interspeech-2008* (pp. 1477–1480). Brisbane.

Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, *26*(3), 477–524.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745–765.

Taft, M., & Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *The Mental Lexicon*, *1*(2), 183 – 199.

Torreira, F., & Ernestus, M. (2010a). The Nijmegen Corpus of Casual Spanish. In *Proceedings of lrec 2010* (pp. 2981 – 2985).

Torreira, F., & Ernestus, M. (2010b). Phrase-medial vowel devoicing in spontaneous French. In *Proceedings of Interspeech-2010* (p. 2006-2009). Makuhari, Japan.

Torreira, F., & Ernestus, M. (2011). Vowel elision in casual French: The case of vowel /e/ in the word c'était. *Journal of Phonetics*, *39*, 50–58.

Torreira, F., & Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica*, *69*, 124–148.

Tucker, B. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, *39*(3), 312–318.

Tucker, B., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 1949–1952). Saarbrücken.

Van Bael, C., & Van Halteren, H. (2007). Speaker classification by means of orthographic and broad phonetic transcriptions of speech. In C. Müller (Ed.), *Speaker classification ii* (pp. 293–307). Berlin / Heidelberg: Springer.

Van Bael, C., Van den Heuvel, H., & Strik, H. (2004). Investigating speech style specific pronunciationvvariation in large spoken language corpora. In *Proceedings of Interspeech-2004* (pp. 2793–2796). Jeju, Korea.

Van Bergem, D. (1994). A model of coarticulatory effects on the schwa. *Speech Communication*, *14*, 143–162.

Van der Sijs, N. (2002). *Chronologisch woordenboek. de ouderdom en herkomst van onze woorden en betekenissen*. Veen.

Van de Ven, M., Ernestus, M., & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology*, *3*, 455–481.

Van de Ven, M., Tucker, B., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition*, *39*(7), 1301–1316.

Van Son, R., Binnenpoorte, D., Van den Heuvel, H., & Pols, L. (2001). The IFA corpus: A phonemically segmented Dutch 'open source' speech database. In *Proceedings of eurospeech-2001* (pp. 2051–2054). Aalborg, Denmark.

Van Son, R., & Pols, L. (1999). An acoustic description of consonant reduction. *Speech Communication*, *28*, 125–140.

Van Son, R., & Pols, L. (2003). Information structure and efficiency in speech production. In *Proceedings of eurospeech-2003* (pp. 769–772). Geneva, Switzerland.

Vertes, R. (2004). Memory consolidation in sleep: Dream or reality. *Neuron*, *44*(1), 135–148.

Viebahn, M., Ernestus, M., & McQueen, J. (2012). Co-occurrence of reduced word forms in natural speech. In *Proceedings of Interspeech-2012.* Portland, Oregon.

Walter, M. (2007). *Repetition avoidance in human language*. Unpublished doctoral dissertation, MIT. (Online available at: http://faculty.wcas.northwestern.edu/~maw962/docs/walter-dissertation.pdf, date last viewed 7 November 2012)

Warner, N., Fountain, A., & Tucker, B. (2009). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America*, *125*(5), 3317–3327.

Wesenick, M., & Kipp, A. (1996). Estimating the quality of phonetic transcriptions and segmentations of speech signals. In *Proceedings of icslp-1996* (pp. 129–132). Philadelphia, PA.

Witteman, M. (2013). *Lexical processing of foreign-accented speech: Rapid and flexible adaptation*. Unpublished doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., … Woodland, P. (2002). *The HTK Book 3.2*. Cambridge: Entropic.

Acknowledgments

Now that my dissertation is finished, it's finally time to thank everyone who has been there for me and helped me during this journey. I won't be able to name everyone individually, but I am grateful to each and every one of you!

The person I am most indebted to is Mirjam Ernestus. Thank you Mirjam, for always making time for questions and discussions, for providing critical but constructive feedback, for reading papers over and over again as if you'd never seen them before, for always being optimistic and enthusiastic, and for providing a great and inspiring working environment. I would not have been able to write this dissertation without you.

I am also very thankful for all the help I received from my second promoter Lou Boves. Whenever I got stuck and needed someone who was not too involved in the analyses or other stages of the research process, you were available with fresh ideas and new insights. You were also typically the first person besides other authors to provide feedback on my papers. In addition, your knowledge on classification and classifiers was highly valuable when I was conducting the individual differences study.

Also my third promoter, Anne Cutler, deserves many thanks: my PhD journey would not have started if you hadn't provided funding for the first year of the project. Furthermore, I am grateful for your comments on the chapters of my dissertation and for the opportunity to attend many of your bi-weekly Language Comprehension Group meetings at the Max Planck Institute, which have taught me a great deal.

Besides my promoters, I also want to express my gratitude to numerous other people. Thanks to my colleagues at the CLSM research group, the Linguistics department, and the Max Planck Institute for creating a pleasant atmosphere to work in. I want to extend my appreciation to Roger for improving my English and to the Technical Group of the Max Planck Institute for keeping all systems up and running.

Furthermore, I thank Barbara and Ellen. I was very happy with our collaborations and believe that both were very fruitful and resulted in nice papers. Barbara, especially thanks for providing automatically generated transcriptions of Mirjam's corpus and part of the CGN, and for helping me create my own transcriptions. This made the extraction of data a lot easier!

Without student assistants, I would not have been able to conduct all the experiments within a relatively short time frame. Thank you, Corine, Ferdy, Henriette, Ilse, Ingrid, Jens, Lieke, Martijn, Nadia, and Tijl, for running these experiments with in total

380(!) participants. Many thanks to Corine and Jens for their patience with repetitively recording large numbers of stimuli for these experiments.

Additionally, I'd like to thank Annika, Anton, Arina, Dan, Esther, Esther, Helen, Huib, Jessamyn, Juliane, Karin, Kim, Laura, Louis, Malte, Marco, Mirjam, Mybeth, Nienke, Rian, Sho, Sophie, Thordis, Wieke, and Xaver for our daily lunches. They were welcome breaks in my daily life as a PhD student. Special thanks to the people I shared an office with: Ellen, Francisco, Kim, Malte, Marco, and Mybeth. You were great office mates, who were always in for short chats or extensive discussions. Linsey and Merel, thanks for all our *gezellige* lunches and dinners. I enjoyed comparing the progress of our projects, discussing our frustrations and successes, and chatting about everything else unrelated to research. Thank you, Linsey and Mybeth, for agreeing to be my paranymphs and supporting me during the final steps of my PhD journey.

I would also like to thank Marcel and Mirjam for their inspiration and feedback on the design of the cover. For those of you who don't recognize the wave forms: The cover depicts two pronunciations of schwa, which has been the focus of multiple studies described in this dissertation. The orange waveform illustrates a clearly produced and unreduced schwa, and contrasts with the black waveform that depicts a present, but heavily reduced schwa.

Ten slotte nog een paar bedankjes in het Nederlands: Pap, mam en overige familie, bedankt voor al jullie vertrouwen en interesse in mijn promotieproject en voor jullie geduld en begrip, wanneer ik weer eens niet op bezoek kon komen omdat ik het te druk had. Last but definitely not least: Marcel, heel erg bedankt voor je steun, liefde en vertrouwen en vooral dat je ervoor gezorgd hebt dat ik me ook met andere dingen bezig hield dan mijn proefschrift!

Iris Hanique
September, 2013

Curriculum Vitae

Iris Hanique was born in Eindhoven, the Netherlands, on 18 August 1985. After obtaining her vwo diploma from St. Willibrord Gymnasium in Deurne in 2004, she studied Artificial Intelligence at Radboud University Nijmegen. As part of her studies, she did an internship at the department of Visual Experiences at Philips Research in Eindhoven and in 2009 she obtained her master's degree (bene meritum). In 2010, Iris started her PhD project at the Centre for Language Studies at Radboud University Nijmegen and the Language Comprehension Group at the Max Planck Institute for Psycholinguistics. Her research was part of the project *Acoustic reduction in European Languages*, which was funded by a European Young Investigator Award from the European Science Foundation awarded to Mirjam Ernestus. Currently, Iris is a consultant at Quintiq in 's Hertogenbosch and is working on solving planning puzzles in the metals industry.

List of Publications

Hanique, I., Ernestus, M., and Boves, L. (submitted). Choice and pronunciation of words: Individual differences within a homogeneous group of speakers.

Hanique, I., Aalders, E., and Ernestus, M. (submitted). How robust are exemplar effects in word comprehension?

Hanique, I., Ernestus, M., and Schuppler, B. (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *Journal of the Acoustical Society of America*, *133(3)*, 1644–1655.

Hanique, I. and Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio*, *2012(2)*, 147–164.

Rajae-Joordens, R. and Hanique, I. (2012). The effect of colored light on arousal and valence in participants primed with colored emotional pictures. In *Proceedings of Experiencing Light 2012*, Eindhoven, the Netherlands.

Hanique, I. and Ernestus, M. (2012). The processes underlying two frequent casual speech phenomena in Dutch: A production experiment. In *Proceedings of Interspeech-2012*, Portland, Oregon.

Hanique, I. and Ernestus, M. (2011). Final /t/ reduction in Dutch past-participles: the role of word predictability and morphological decomposability. In *Proceedings of Interspeech-2011*, (pp. 2849–2852), Florence, Italy.

Hanique, I., Schuppler, B., and Ernestus, M. (2010). Morphological and predictability effects on schwa reduction: The case of Dutch word-initial syllables. In *Proceedings of Interspeech-2010*, (pp. 933–936), Makuhari, Japan.

Lemhöfer, K., Schriefers, H., and Hanique, I. (2010) Native language effects in learning second-language grammatical gender: A training study. *Acta Psychologica*, *135(2)*, 150–158.

MPI series in psycholinguistics

1. The electrophysiology of speaking: investigations on the time course of semantic, syntactic, and phonological processing.
   Miranda I. van Turennout

2. The role of the syllable in speech production: evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography.
   Niels O. Schiller

3. Lexical access in the production of ellipsis and pronouns.
   Bernadette M. Schmitt

4. The open-/closed class distinction in spoken-word recognition.
   Alette Petra Haveman

5. The acquisition of phonetic categories in young infants: a self-organising artificial neural network approach.
   Kay Behnke

6. Gesture and speech production.
   Jan-Peter de Ruiter

7. Comparative intonational phonology: English and German.
   Esther Grabe

8. Finiteness in adult and child German.
   Ingeborg Lasser

9. Language input for word discovery.
   Joost van de Weijer

10. Inherent complement verbs revisited: towards an understanding of argument structure in Ewe.
    James Essegbey

11. Producing past and plural inflections.
    Dirk J. Janssen

12. Valence and transitivity in Saliba: an Oceanic language of Papua New Guinea.
 Anna Margetts

13. From speech to words.
 Arie H. van der Lugt

14. Simple and complex verbs in Jaminjung: a study of event categorisation in an Australian language.
 Eva Schultze-Berndt

15. Interpreting indefinites: an experimental study of children's language comprehension.
 Irene Krämer

16. Language-specific listening: the case of phonetic sequences.
 Andrea Christine Weber

17. Moving eyes and naming objects.
 Femke Frederike van der Meulen

18. Analogy in morphology: the selection of linking elements in dutch compounds.
 Andrea Krott

19. Morphology in speech comprehension.
 Kerstin Mauth

20. Morphological families in the mental lexicon.
 Nivja Helena de Jong

21. Fixed expressions and the production of idioms.
 Simone Annegret Sprenger

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria).
 Birgit Hellwig

23. Paradigmatic structures in morphological processing: computational and cross-linguistic experimental studies.
 Fermín Moscoso del Prado Martín

24. Contextual influences on spoken-word processing: an electrophysiological approach.
 Danielle van den Brink

25. Perceptual relevance of prevoicing in Dutch.
 Petra Martine van Alphen

26. Syllables in speech production : effects of syllable preparation and syllable frequency.
    Joana Cholin

27. Producing complex spoken numerals for time and space.
    Marjolein Henriëtte Wilhelmina Meeuwissen

28. Morphology in auditory lexical processing : sensitivity to fine phonetic detail and insensitivity to suffix reduction.
    Rachèl Jenny Judith Karin Kemps

29. At te same time....: the expression of simultaneity in learner varieties.
    Barbara Schmiedtová

30. A grammar of Jalonke argument structure.
    Friederike Lüpke

31. Agrammatic comprehension : an electrophysiological approach.
    Marijtje Elizabeth Debora Wassenaar

32. The structure and use of shape-based noun classes in Miraña (North West Amazon).
    Frank Seifart

33. Prosodically-conditioned detail in the recognition of spoken words.
    Anne Pier Salverda

34. Phonetic and lexical processing in a second language.
    Mirjam Elisabeth Broersma

35. Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension.
    Oliver Müller

36. Lexically-guided perceptual learning in speech processing.
    Frank Eisner

37. Sensitivity to detailed acoustic information in word recognition.
    Keren Batya Shatzman

38. The relationship between spoken word production and comprehension.
    Rebecca Özdemir

39. Disfluency: interrupting speech and gesture.
    Mandana Seyfeddinipur

40. The acquisition of phonological structure: distinguishing contrastive from non-constrative variation.
    Christiane Dietrich

41. Cognitive cladistics and the relativity of spatial cognition.
    Daniel Haun

42. The acquisition of auditory categories.
    Martijn Bastiaan Goudbeek

43. Affix reduction in spoken Dutch: probabilistic effects in production and perception.
    Mark Pluymaekers

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence.
    Valesca Madalla Kooijman

45. Space and iconicity in German sign language.
    Pamela M. Perniss

46. On the production of morphologically complex words with special attention to effects of frequency.
    Heidrun Bien

47. Crosslinguistic influence in first and second languages: convergence in speech and gesture.
    Amanda Brown

48. The acquisition of verb compounding in Mandarin Chinese.
    Jidong Chen

49. Phoneme inventories and patterns of speech sound perception.
    Anita Eva Wagner

50. Lexical processing of morphologically complex words: an information-theoretical perspective.
    Victor Kuperman

51. A grammar of Savosavo: a Papuan language of the Solomon Islands.
    Claudia Ursula Wegener

52. Prosodic structure in speech production and perception.
    Claudia Kuzla

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: investigating knowledge of forms and functions in production and comprehension.
Sarah Schimke

54. Studies on intonation and information structure in child and adult German.
Laura de Ruiter

55. Processing the fine temporal structure of spoken words.
Eva Reinisch

56. Semantics and (ir)regular inflection in morphological processing.
Wieke Tabak

57. Processing strongly reduced forms in casual speech.
Susanne Brouwer

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch.
Miriam Ellert

59. Lexical interactions in non-native speech comprehension: evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging.
Ian FitzPatrick

60. Processing casual speech in native and non-native language.
Annelie Tuinman

61. Split intransitivity in Rotokas, a Papuan language of Bougainville.
Stuart Payton Robinson

62. Evidentiality and intersubjectivity in Yurakaré: an interactional account.
Sonja Gipper

63. The influence of information structure on language comprehension: a neurocognitive perspective.
Lin Wang

64. The meaning and use of ideophones in Siwu.
Mark Dingemanse

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants.
Marco van de Ven

66. Speech reduction in spontaneous French and Spanish.
Francisco Torreira

67. The relevance of early word recognition: insights from the infant brain.
Caroline Mary Magteld Junge

68. Adjusting to different speakers: extrinsic normalization in vowel perception.
Matthias Johannes Sjerps

69. Structuring language: contributions to the neurocognition of syntax.
Katrien Rachel Segaert

70. Infants' appreciation of others' mental states in prelinguistic communication: a second person approach to mindreading.
Birgit Knudsen

71. Gaze behavior in face-to-face interaction.
Federico Rossano

72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing place.
Connie de Vos

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning.
Attila Andics

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation.
Marijt Witteman

75. The use of deictic versus representational gestures in infancy.
Daniel Puccini

76. Territories of knowledge in Japanse conversation.
Kaoru Hayano

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective.
Kimberley Mulder

78. Contributions of executive control to individual differences in word production.
Zeshu Shao

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing.
Patrick van der Zande

80. High pitches and thick voices. The role of language in space-pitch associations.
Sarah Dolscheid

81. Seeing what's next: Processing and anticipating language referring to objects.
    Joost Rommers

82. Mental representation and processing of reduced words in casual speech.
    Iris Hanique