The role of acoustic detail and context in the comprehension of reduced pronunciation variants

ISBN/EAN: 978-90-76203-40-9
Cover design: In Zicht Grafisch Ontwerp, Arnhem
Printed and bound by Ipskamp Drukkers, Nijmegen

The role of acoustic detail and context in the comprehension of reduced pronunciation variants

een wetenschappelijke proeve op het gebied van de Letteren

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 2 november 2011
om 13.30 uur precies

door

Marco Adrianus Marinus van de Ven

geboren op 19 januari 1982
te Oss

Promotoren:	Prof. dr. A. Cutler
	Prof. dr. R. Schreuder
Copromotor:	Dr. M. Ernestus

Manuscriptcommissie:

Prof. dr. R. van Hout	(Voorzitter)
Prof. dr. M. Pitt	(Ohio State University)
Dr. R. van Son	(Nederlands Kanker Instituut - Antoni van Leeuwenhoek Ziekenhuis & Universiteit van Amsterdam)

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellshaft zur Förderung der Wissenshaften, München, Germany.

Voor mijn ouders.

Contents

Introduction

Chapter 1

The comprehension of conversational speech has not received much attention in the psycholinguistic literature. Most research on human speech comprehension concerns the comprehension of clearly articulated "laboratory" speech. Nevertheless, more than twenty years ago, researchers already showed that conversational speech makes very different demands on comprehension than formal, read-aloud, speech (Mehta and Cutler, 1988). Conversational speech contains hesitations, false starts, background noises, and many reduced pronunciation variants (i.e. words with weakly articulated or missing segments). For example, the English word *particular* may sound like [p^{h}t^{h}ɪk^{h}ə] (Johnson, 2004). This dissertation tries to extend our knowledge of speech comprehension from isolated words and simple, clearly pronounced sentences to conversational speech, by investigating the role of semantic/syntactic and acoustic information in the processing of these reduced variants.

In everyday speech, many segments and syllables are missing or realised differently. For example, in English spontaneous speech, the fragment *you know what I mean* [juˈnəʊwɒtaɪˈmin] may be realised like [jɔ̃w̃ʌ̃mɪ̃] (Shockey, 2003). Similarly, in spontaneous Dutch, the fragment *blijf ik ongeveer op hetzelfde* 'I stay approximately on the same' [blɛifɪkɔnxəˈverɔpətˈsɛlvdə] may sound like [blɛifkɔxˈfeɔpt ˈsɛldə] (Ernestus, 2000). Previous research has shown that these reduced pronunciations occur highly frequently in spontaneous speech. For example, Johnson (2004) found that, in American English conversational speech, 25% of the segments were changed or missing, while complete syllables were missing in 6% of the word tokens. Similarly, Schuppler, Ernestus, Scharenborg, and Boves (2011) found that, in Dutch conversational speech, segments were changed or missing in 48.2% of the word tokens and complete syllables were missing in approximately 19.4% of the word tokens. Segments are typically missing in unstressed syllables, and most frequently in function words and discourse markers (Ernestus, 2000).

Interestingly, Ernestus, Baayen, and Schreuder (2002) demonstrated that listeners cannot recognise highly reduced pronunciation variants out of linguistic context (ca. 50% correct), and problems remain when listeners hear these variants just within their phonological contexts (including the neighbouring vowels and intervening consonants; ca. 70% correct). Listeners experienced no difficulties recognising these reduced variants embedded in their sentence contexts (more than 90% correct). In comparison, listeners consistently recognised mildly reduced pronunciation variants in all three conditions. These findings show that listeners require information from the sentence context to recognise highly reduced pronunciation variants. It is yet unclear how listeners deal with these reduced variants, and which types of information in the context but also from the word itself contribute to the recognition of these variants.

This dissertation describes four studies investigating the role of semantic/syntactic and acoustic information in the context, and acoustic information in reduced pronunciation variants, in the processing of these variants. Understanding how listeners deal with these reduced variants is crucial for our understanding of how humans process spoken language (e.g. Warner, to appear).

The various psycholinguistic models of speech comprehension propose different accounts for how listeners recognise reduced pronunciation variants. These models can be placed on a continuum from abstractionist to exemplar-based models of speech comprehension.

Abstractionist models assume that only a word's citation form is stored in the mental lexicon (Gaskell, 2003). Listeners match the sounds that they hear with the sounds of the words' citation forms, and the lexical candidates that emerge during this process then compete for recognition. However, the sounds of words in spontaneous speech may not always match the sounds from the words' citation forms. These mismatches are likely to inhibit word recognition, especially if these mismatches occur early on in the word recognition process, before the words' uniqueness points. For example, the Dutch words *verlaten* 'leave' and *flater* 'blunder' with the citation forms [vərˈlatən] and [ˈflatər] may be realised like [ˈflatə] and [ˈflatər], and this may temporarily lead to ambiguity. Abstractionist models of speech comprehension assume context-dependent reconstruction mechanisms that recover the missing sounds, and these mechanisms may be based on fine phonetic cues and/or phonotactics.

Exemplar-based models of speech comprehension (e.g. Goldinger, 1998), on the other hand, assume that listeners store the different pronunciation variants in the mental lexicon. These models predict that listeners have no difficulty recognising these variants. Exemplar-based models can explain the finding that listeners cannot recognise highly reduced pronunciation variants in isolation by assuming that listeners store reduced pronunciation variants together with the contexts that they occurred in (Hawkins and Smith, 2001).

Neither abstractionist nor exemplar-based models of speech comprehension specify which types of information contribute to the recognition of reduced pronunciation variants, and to what extent.

1.1 Aims of the dissertation

This dissertation investigates which types of information contribute to the recognition of reduced pronunciation variants in spontaneous speech, with the general aim to collect information for building ecologically valid speech comprehension models.

First of all, Chapter 2 investigates whether listeners are sensitive to the acoustic information in the context, and whether they can use this information to facilitate their recognition of upcoming reduced pronunciation variants. Various studies on the processing of laboratory speech have shown that listeners can use fine-grained spectral cues to facilitate word recognition (e.g. Marslen-Wilson and Warren, 1994; Gow, 2002). However, in spontaneous speech, segments are typically shorter, realised variably, or not realised at all, and acoustic information may be more difficult for listeners to rely on. Conversely, research suggests that listeners may rely more heavily on any type of information, including acoustic contextual information, in the processing of spontaneous speech (e.g. Heinrich and Hawkins, 2009).

Second, Chapters 2-4 investigate to what extent listeners use semantic/syntactic information in the context to recognise reduced pronunciation variants. Previous studies have shown that higher-level, contextual information (e.g. semantic and syntactic information) plays a significant role in the recognition of clear, laboratory speech (e.g. Blank and Foss, 1978; Marslen-Wilson and Tyler, 1980). These studies mainly focus on words with a high predictability. Many reduced words (e.g. function words and discourse markers), however, are not highly predictable. Chapter 2 addresses the question to what extent semantic/syntactic information can facilitate the recognition of these word types. Chapters 3 and 4 zoom in on the effects of semantic information (separately from syntactic information) in the context. Previous research suggests that pronunciation variation may influence the time course of lexical activation (e.g. Sumner and Samuel, 2005; McLennan, 2006). We investigated the role of semantic information in the processing of reduced pronunciation variants by native and non-native listeners respectively. Chapter 4 focused on semantic effects in non-native speech processing since previous research suggests that non-native listeners have significantly more difficulty using contextual information to facilitate their processing of upcoming words in conversational speech (Bradlow and Alexander, 2007).

Third, Chapter 5 investigates how the different segments of reduced words contribute to the recognition of these words. The focus is on words with reduction in their first syllables, that is, before their uniqueness points, since such words may be difficult for listeners to recognise.

Finally, Chapters 2 and 5 tested the interaction between the roles of semantic/syntactic contextual information and acoustic information during the processing of reduced pronunciation variants.

1.2 Methodology

As the vast majority of speech comprehension research has studied the comprehension of laboratory speech, the experimental paradigms for studying comprehension have been designed for this purpose. The question arises whether the comprehension of conversational speech can be investigated by means of these existing paradigms, since these paradigms assume that all words have been pronounced clearly, and that their recognition can be tested without context or in simple sentences.

First of all, offline experiments such as the cloze task and the gating task can be used for studying the comprehension of conversational speech. In a cloze task, participants are instructed to predict target words on the basis of their context. Gating tasks are similar to cloze tasks, except that participants are also provided with some segmental information from the target words themselves, and this amount of information varies either across or within participants. Although these techniques cannot directly tap in to the time course of lexical activation, they can provide useful insights into the types of information listeners can rely on during the processing of conversational speech.

The auditory lexical decision task is suitable especially for studying the comprehension of isolated words. This task may be adjusted for studying the recognition of sentence-final words, by asking participants to make a lexical decision for the last word in each sentence. However, this task cannot be used with materials extracted from corpora of conversational speech, because lexical decision experiments require pseudowords, and conversational speech does not contain pseudowords. Hence, we need methods to elicit conversational speech before we can use lexical decision experiments with this type of speech.

Cross-modal priming and eye-tracking experiments can more easily be used for studying the comprehension of conversational speech. In cross-modal priming (e.g. Zwitserlood, 1989), participants hear speech fragments, and they have to make lexical decisions for words that are presented visually on the screen. In a recent version of the eye-tracking paradigm (e.g. Brouwer, 2010), participants hear speech fragments while they see four words presented on the screen, and they are asked to click on the target word. Yet, in these two experimental paradigms orthographic transcriptions of the target words are shown on the computer screen, which may activate words' citation forms more strongly than their reduced forms.

We have combined several experimental paradigms to investigate how listeners process conversational speech. The paradigms used in this dissertation include the cloze task, the gating paradigm, and the lexical decision paradigm.

Researchers using materials extracted from corpora of conversational speech also have to face problems with matching stimuli (e.g. for word type, surrounding words, etc.) for the different experimental conditions, since there is no experimental control over the speech produced in these corpora. As a consequence, it becomes more difficult to statistically compare different experimental conditions.

1.3 Outline of the dissertation

This dissertation is structured as follows. Chapter 2 investigates the contributions of semantic/syntactic and acoustic cues in the context to the predictability of reduced pronunciation variants. Chapters 3 and 4 investigate the effects of the semantic context on the processing of unreduced and reduced pronunciation variants in native listeners (Chapter 3) and in non-native listeners (Chapter 4). Chapter 5 describes a study investigating listeners' sensitivity to acoustic information in reduced pronunciation variants. Finally, in Chapter 6 we discuss the conclusions that can be drawn on the basis of this dissertation.

Predicting Acoustically Reduced words in Spontaneous speech

Chapter 2

This chapter has been reformatted from:
Marco van de Ven, Mirjam Ernestus and Robert Schreuder (submitted). Predicting Acoustically Reduced words in Spontaneous speech: The role of Semantic/Syntactic and Acoustic cues in the context.

Abstract

In spontaneous speech, words are often reduced (e.g., English 'yesterday' may be pronounced like [ˈjɛʃeɪ]). Previous research has shown that context is required to understand highly reduced pronunciation variants. We investigated to which extent listeners can predict low predictability reduced words on the basis of the semantic/syntactic and acoustic cues in the context. In four experiments, participants were presented with either the preceding context or the preceding and following context of reduced words, either visually (orthographic transcriptions) or auditorily (extracted from conversational speech). In all sentences, the reduced target words had been completely removed, and listeners were asked to predict these missing reduced words on the basis of the context alone, choosing from four plausible options. Listeners made use of acoustic cues in the context, although casual speech typically has a high speech rate, and acoustic cues are much weaker than in careful speech. Moreover, they relied on semantic/syntactic cues. Whenever there was a conflict between acoustic and semantic/syntactic contextual cues, measured as the word's probability given the surrounding words, listeners relied more heavily on acoustic cues. Further, context appeared generally insufficient to predict the reduced words, underlying the significance of the acoustic characteristics of the reduced words themselves.

2.1 Introduction

In casual speech, words may be pronounced much shorter than their citation forms (e.g. Ernestus, 2000; Johnson, 2004). For example, in casual English the adjectives *ordinary* and *hilarious* with the citation forms [ˈɑrdənɛrɪ] and [hɪˈlɛriəs] can be reduced to [ˈɑnrɪ] and [hˈlɛrɛs], respectively (Johnson, 2004). The absence of single or multiple segments is very common in casual speech. In fact, Johnson (2004) reported that, in English, complete syllables are absent in 6% of the tokens on average. Previous research indicates that listeners cannot well recognise highly reduced pronunciation variants without context (Ernestus, Baayen, and Schreuder, 2002).

We investigated what types of information listeners extract from the context to comprehend reduced variants and to which extent, focusing on the roles of semantic/syntactic cues and acoustic cues in Dutch spontaneous speech. We provided listeners only with contextual information of the reduced words, and these words themselves were never present in the signal (thus only context).

Ernestus et al. (2002) investigated the role of context in the comprehension of pronunciation variants of different degrees of reduction. Dutch listeners were presented with sound fragments extracted from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000), including pronunciation variants of a low, medium, or high degree of reduction. These variants were either presented in isolation, in phonological context (the neighbouring vowels and any intervening consonants), or in sentential context, and the participants were asked to orthographically transcribe the speech fragments presented. The results showed that listeners have difficulty identifying highly reduced variants in isolation (ca. 50% correct). Their performance increased significantly when highly reduced variants were presented within their phonological context, although identification problems remained (ca. 70% correct). Only when presented with the full sentence context, listeners were able to identify highly reduced variants successfully (more than 90% correct). For the variants with a medium degree of reduction, listeners needed only the phonological context for correct identification, while the variants with a low degree of reduction were identified correctly in all three context conditions. This study then suggests that, although phonological contextual information is beneficial, long-distance contextual information is a prerequisite for the identification of highly reduced variants, while this is not the case for less reduced variants.

Kemps, Ernestus, Schreuder, and Baayen (2004) also investigated the role of context in the comprehension of highly reduced variants. They presented Dutch participants with Dutch words ending in the suffix *-lijk*, which contains [l] in its citation form ([lək]), but not in its highly reduced form ([k]). For example, the Dutch word *namelijk* 'namely', which has the citation form [ˈnamələk], can be reduced to

[ˈnamk]. Listeners performed a phoneme monitoring task, in which they had to press a button whenever they heard an [l]. When presented with just the suffix, listeners only pressed the button when there was acoustic evidence for [l] in the signal. When the words were presented within sentential context, however, listeners often (in 52% of the cases) pressed the button also when there was no acoustic evidence for [l] in the signal at all. This result suggests that listeners somehow reconstruct missing phonemes based on the context.

Having discussed relevant literature on the role of context in the comprehension of reduced variants, we now turn to the role of semantic/syntactic cues, and then to the role of acoustic cues in the perception of unreduced laboratory speech.

Several studies have established a role for contextual semantic/syntactic information in the comprehension of carefully pronounced words. In a cross-modal priming study, Zwitserlood (1989) investigated at which stage in the comprehension process the influence of semantic context sets in. Dutch participants were presented with prime words that have a relatively late uniqueness point (e.g. Dutch /kapit/ is consistent with both /kapiˈtɛin/ 'captain' and /kapiˈtal/ 'capital'). These words were embedded in a neutral context (e.g. *They mourned the loss of their* *captain.*) or a biasing context (e.g. *With dampened spirits the men stood around the grave. They mourned the loss of their* *captain.*). Participants were asked to make a lexical decision for visual probes that were either semantically related (for the example above: *ship*) or unrelated (for the example above: *money*) to the prime words in the auditorily presented sentences. In the biasing context condition, there was already significant priming for the semantically congruent word just before the uniqueness point of the auditory prime. These results indicate that semantic information in the preceding context can enhance word recognition already before a word becomes unique.

Similar conclusions were drawn by Van den Brink, Brown, and Hagoort (2001; 2006), who recorded event-related potentials (ERPs) while participants were presented with spoken sentences. The sentences ended either in semantically plausible (e.g. *The painter colored the details with a small* *brush.*) or implausible words (e.g. *The painter colored the details with a small* *pension.*). Contextually incongruent words yielded larger ERPs than contextually congruent words. More importantly, as Van den Brink et al. (2006) showed, the onset of this N400 effect occurs prior to the word's uniqueness point, which indicates that the N400 peak is not simply due to semantic integration difficulty. Rather, listeners unconsciously formulate predictions about the upcoming words, and processing is inhibited if these predictions are incorrect.

In line with these results, several studies have shown that processing is faster when a word is more predictable given its immediately preceding word(s) (e.g. Morton and Long, 1976; McDonald and Shillcock, 2003). For example, in a self-paced reading task by McDonald and Shillcock (2003), participants read a word faster if it had a higher probability given the preceding words, as estimated by means of N-gram probability measures. These measures are corpus-based frequency counts for word sequences (e.g. the word combination "dog" + "walks" can be found *X* times in a given corpus), representing the likelihood of these word sequences. This result indicates that sequences of semantically/syntactically related words are processed faster than sequences of semantically/syntactically unrelated words.

In addition to semantic and syntactic cues in the context, listeners have been shown to use acoustic cues that do not necessarily distinguish phonemes from each other. One such acoustic correlate is segmental duration, which cues, among others, the number of following syllables in the word, since segments tend to be shorter if followed by more syllables (e.g. Nooteboom, 1972). For instance, Salverda, Dahan, and McQueen (2003) showed in an eye-tracking study with unreduced laboratory speech that when listeners hear the syllable *cap*, they are more likely to look at *cap* if the syllable is relatively long, whereas they are more likely to look at *captain* if the syllable is relatively short. Hence, listeners use segment duration to predict whether the word is monosyllabic (i.e. *cap*) or disyllabic (i.e. *captain*). See also Davis, Marslen-Wilson, and Gaskell (2002) and Kemps, Wurm, Ernestus, Schreuder, and Baayen (2005) for similar results using different experimental paradigms.

Segment duration is also a cue to speech rate, which helps listeners interpret the duration of other segments as relatively long or short. Thus, Nooteboom (1980) showed that Dutch listeners identify the first vowel in a word as long (e.g. Dutch *taak* /ˈtaːk/ 'task') or short (e.g. Dutch *tak* /ˈtɑk/ 'branch') by relating its duration to the durations of the preceding segments, which are in the preceding word. These findings demonstrate that listeners use temporal cues to enhance word recognition and that these cues may be located in the word itself, or in its context.

In addition to temporal cues, listeners use fine-grained spectral cues to facilitate word recognition (e.g. Marslen-Wilson and Warren, 1994; Hawkins and Warren, 1994; Gow, 2002; Hawkins, 2003). This appears, for example, from a study by Marslen-Wilson and Warren (1994), who conducted a lexical decision task with cross-spliced existing words, containing acoustic cues in line with the actually presented word (e.g. *job* with a vowel transition announcing a [b]: [ˈdʒɒb]), with a different existing word (e.g. *job* with a vowel transition announcing a [g]: [ˈdʒɒ$_g$b]), or with a nonword (e.g. *job* with a vowel transition announcing a [d]: [ˈdʒɒ$_d$b]). They found that listeners were slower to respond to words with misleading acoustic cues, regardless of whether these led to an existing word (i.e. [ˈdʒɒ$_g$b]) or to a nonword (i.e. [ˈdʒɒ$_d$b]), which indicates that spectral detail is used in word recognition.

If listeners are sensitive to subtle phonetic cues in careful laboratory speech, they may also use these cues to predict reduced words in conversational speech. Some researchers even predict that these subtle cues are more relevant in conversational speech than in laboratory speech, since conversational speech is fast, often reduced and accompanied with background noises, and listeners would use all cues available under such adverse listening conditions (Hawkins and Smith, 2001). However, acoustic cues may also be too subtle to help listeners predict words in conversational speech. These cues may be difficult to notice and to process due to the high speed and variability of this speech register.

In the present study, we investigate how semantic/syntactic and acoustic cues in the context help language users predict the identity of acoustically reduced words (or fixed expressions, henceforth "target words", for the sake of convenience), embedded in their natural, reduced contexts. Participants were presented with the contexts of the target words only and the target words themselves were thus missing. Participants were asked to guess the identity of the missing words. This experimental paradigm allowed us to investigate the role of acoustic and semantic/syntactic cues in the context separately from the cues in the reduced word itself.

As reduced pronunciation variants tend to occur in reduced contexts, the contexts that participants were presented with in our auditory experiments contained reduced speech. For instance, Chapters 3 and 4 of this thesis show that reductions may affect participants' processing. We found that it takes more time to activate semantically related words after hearing reduced than unreduced pronunciation variants. If so, the results of our study simply cannot be generalised to the processing of words in unreduced speech.

The present study describes four main experiments investigating how well participants can predict the identity of reduced words if they are only provided with the preceding (Experiments 1 and 2) or the preceding and following context (Experiments 3 and 4) and the reduced words themselves are missing. The context was provided either visually in the form of orthographic transcriptions, such that it only contained semantic/syntactic information (Experiments 1 and 3), or auditorily, such that it contained both semantic/syntactic and acoustic information (Experiments 2 and 4). Participants were presented with the context and were asked to select which out of four options presented on a computer screen was most likely the missing reduced word. We will investigate the contribution of semantic/syntactic cues in the context on the basis of the performance by the participants in Experiments 1 and 3, and by testing the effects of unigram and N-gram frequency. We will focus on the contribution of acoustic cues by comparing Experiments 1 and 2, and Experiments 3 and 4.

In Experiment 1, participants were visually presented with the contexts preceding the target words while these words themselves were missing. In addition, they saw four semantically/syntactically plausible options (as verified by the authors) for each target word and were asked to select the most likely one. The rationale behind this experiment was to establish how well participants could predict reduced words on the basis of only semantic/syntactic cues in the preceding context. If participants just guess, all four options are equally likely. If participants are only sensitive to the lexical frequencies of the four options, they are expected to choose the word with the highest lexical frequency out of the four options. The distractor words were selected such that the target word was the most frequent of the four options in only 5.25% of the trials. Therefore, participants will perform 5.25% correct if they base their decisions purely on word frequency information. Third, if participants show some sensitivity to the preceding context and base their choice on the probability of the four options given the preceding word or the two preceding words (i.e. on bigram or trigram probability), they should also perform below chance level, as we selected the distractor words such that one of them had the highest N-gram (i.e. bigram and trigram) probability in 85.53% of the trials. Finally, only if participants are sensitive to more semantic/syntactic information in the context than is captured by N-gram probability, they may perform above chance level.

2.2 Experiment 1

2.2.1 Participants

Twenty native speakers of Dutch from the pool of participants of the Max Planck Institute for Psycholinguistics (they were nearly all undergraduate students at the Radboud University Nijmegen) were paid to take part in the experiment. None of them reported any hearing loss.

2.2.2 Materials

Speech materials

The materials were taken from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000), consisting of casual conversations between ten pairs of speakers recorded in a soundproof booth. Since high frequency words are more likely to be reduced (e.g. Zipf, 1935; Bybee, 2001), we chose sixteen highly frequent Dutch words or word combinations as target words: *alleen* ‘only’, *allemaal* ‘all’, *altijd* ‘always’, *anders* ‘otherwise’, *bepaalde* ‘certain’, *bijvoorbeeld* ‘for instance’, *eigenlijk* ‘actually’, *gewoon* ‘usual’, *helemaal* ‘totally’, *in ieder geval* ‘in any case’, *misschien* ‘perhaps’, *namelijk* ‘namely’, *natuurlijk* ‘of course’, *op een gegeven moment* ‘at a certain moment’, *over* ‘about’, and *tenminste* ‘at least’. All of these words are adjectives, adverbs, or adverbial phrases and could be left out of their sentences without rendering these semantically incoherent or ungrammatical.

We selected on average five tokens (from different speakers) for each target word type (mean: 4.88 tokens per word, range: 1 to 8 tokens). These tokens had low trigram frequencies with their two preceding words or preceding and following word (2.97 and 1.52 per million respectively in the Spoken Dutch Corpus, Oostdijk, 2002) compared to previous studies. We added 22 filler word tokens which differed from the target tokens only in that they represented different word types, which introduced more variation (and therefore a smaller predictability of the correct option) in the experiment. In the end, the experiment consisted of 78 target word tokens and 22 filler word tokens, produced by eleven speakers.

We extracted the target and filler tokens together with some part of their preceding and following contexts. The amount of context varied for each token, but comprised at least the prosodic phrase in which the token was embedded. On average, the preceding context consisted of 9.63 words (range: 2 to 22, and one token with 29 words), and the following context consisted of 5.05 words (range: 1 to 13 words). The extracted fragments did not contain any overlapping speech or loud background noises. An orthographic transcription of the experimental materials is provided in Appendix A.

The degree of reduction of the target and filler tokens varied from mildly reduced (e.g. [ˈɛiχlək] for *eigenlijk*, with the unreduced pronunciation [ˈɛiχələk]) to highly reduced (e.g. [ˈɛik] for *eigenlijk*). Although the participants thus did not hear or see the target words themselves, a predictor of their performance may be the degree of reduction of this word. Research by, for example, Bell, Brenier, Gregory, Girand, and Jurafsky (2009) shows that words are more reduced if they have a high lexical frequency or if they have a higher probability given the following word. Hence, we expect that words that are more reduced can more easily be predicted on the basis of the context.

We quantified degree of reduction (henceforth *reduction degree in production*) by subtracting the number of segments in the reduced form from the number of segments in the citation form, and dividing its outcome by the number of segments in the citation form. Degree of reduction in our materials varied from 0 to 0.73. If the degree of reduction was equal to 0, there was still some reduction present, for example in the form of consonant or vowel lenition (e.g. full vowels produced as schwas). Target and filler tokens were labelled as highly reduced if *reduction degree in production* was higher than 0.4 (based on the approximate bimodal distribution observed *in reduction degree in production*); otherwise they were labelled as mildly reduced. This resulted in 42 target word tokens classified as mildly reduced and 36 as highly reduced.

In order to verify the intelligibility of the target and filler tokens within their contexts, we carried out a control experiment. In this experiment, twenty native speakers of Dutch, none of whom participated in the main experiments, first heard the reduced word in its original sentence context (e.g. *Ik vertrouw altijd maar op mijn goede geluk* ‘I always rely on good luck.’), and then heard a shorter version of this same fragment, consisting of the same token of the reduced word and its two preceding and following words (e.g. *Ik vertrouw altijd maar op* ‘I always rely on’)[1]. The participants were asked to orthographically transcribe this fragment. They were tested individually in a soundproof booth, on a PC running E-prime 1.2 (Schneider, Eschman, and Zuccolotto, 2002). The materials were grouped in eleven blocks, with each block containing the materials of one of the eleven speakers. Each block was preceded by a familiarisation phase, in which participants were presented with a short (on average 19 second) speech fragment of the speaker to get used to the speaker's (voice) characteristics. Subsequently, participants were presented with two filler tokens, followed by the target tokens. The results showed that three target tokens were difficult to understand (less than 60% correct), and they were not included in the analyses of the subsequent experiments. With respect to the remaining target tokens, participants successfully identified the mildly reduced tokens in 98% (range: 92.5% to 100% correct) and the highly reduced tokens in 94.8% (range: 60% to 100% correct) of the trials. These tokens can thus be well identified in their contexts.

We conducted a second control experiment, in order to investigate whether listeners could understand the target tokens used in our experiments in isolation. Participants listened to the reduced target and filler tokens without their contexts, and were asked to orthographically transcribe the words. The basic experimental procedure was adopted from the previous control experiment. Participants successfully identified the mildly reduced target tokens in 74% (range: 20% to 100% correct) and the highly reduced target tokens in 43.4% (range: 0% to 100% correct) of the trials. These recognition scores are very similar to those obtained by Ernestus et al. (2002). Thus, our two control experiments show that our reduced target tokens are difficult to recognise in isolation but generally easy to understand in context.

We investigated whether the proportion of listeners correctly identifying a target token in isolation correlates with *reduction degree in production* as defined above. We conducted a t-test which showed that *intelligibility in isolation* (the percentage of correct identications in the second control experiment) differs between tokens that were classified as mildly and tokens that were classified as highly reduced (mean for mildly reduced tokens: 68%, mean for highly reduced tokens: 27%, one-tailed t-test: $t(73.6) = 6.27$, $p < .0001$), which indicates that both measures reflect degree of reduction. We decided to use *intelligibility in isolation* as a measure of reduction

[1] In a few cases, the two preceding or following words were inseparable from their neighbouring words because these words had been contracted. In such cases, the context contained one or two (preceding or following) additional words. Conversely, the following context in some cases consisted of only one word because that word was sentence-final.

degree in our analyses of the main experiments, for various reasons. First, *intelligibility in isolation* reflects degree of reduction from a perception perspective, whereas *reduction degree in production* reflects degree of reduction from a production perspective, and this study focuses on perception. For example, it is unclear whether the various types of reduction observable in the signal are equally important in perception (e.g. the absence of [r] versus the absence of [k]), and consequently whether they should be equally important for a measure of degree of reduction. Further, *reduction degree in production* only reflects the relative number of absent segments and does not take lenition (e.g. the pronunciation of full vowels as schwas) into account, but we know that also lenition may affect speech comprehension (e.g. Mitterer and Ernestus, 2006; Warner, Fountain, and Tucker, 2009). Second, it is particularly difficult to determine whether or not a segment has actually been realised (e.g. Ernestus and Baayen, 2011). As a consequence, a measure of reduction based on this information is somewhat unreliable.

We henceforth thus use *intelligibility in isolation* as a continuous measure of a word's degree of reduction although we know that a word's intelligibility is determined not only by its degree of reduction but also by other factors, for example its frequency of occurrence in natural speech contexts (e.g. Howes, 1954, 1957; Newbigging, 1961; Savin, 1963; Soloman and Postman, 1952). We believe that there is a strong relationship between a word's degree of reduction and its intelligibility in isolation. This assumption is supported by the correlation between *intelligibility in isolation* and *reduction degree in production* reported above and by a global inspection of the data which showed a large degree of variability between various tokens of the same word. For example, one token of bijvoorbeeld 'for example' showed a success rate of 90%, whereas a different token of the same word showed a success rate of 27.5%.

In Experiment 1, we investigated the role of semantic and syntactic information in the preceding context by presenting participants visually with this preceding context. The participants were asked to predict the following word on the basis of the preceding context alone, choosing from four semantically and syntactically plausible options presented on the screen.

The options for each trial

The four options included the correct answer, the word that followed the reduced target token in the original sentence, and two other options. An example trial is provided below.

Context:
Het geld is niet van jou en je staat
'The money is not yours and you are'

Options:
1. *altijd* 'always' (correct, reduced word)
2. *rood* 'in debt' (word following reduced word)
3. *bijvoorbeeld* 'for instance'
4. *eigenlijk* 'actually'

The original sentence for this example was *Het geld is niet van jou en je staat altijd rood*, and *altijd* 'always' was thus the target word and *rood* 'red' the following word. We included the following word as one of the four options because we initially wanted to verify whether semantic/syntactic and acoustic contextual information (e.g. prosody or formation transitions) can help listeners predict whether or not an additional word is present in a speech fragment.

The four options did not have identical word-initial sounds, making short-distance co-articulation cues potentially relevant in the auditory version of the experiment. Further, as mentioned above, we made sure that, in most cases, the correct answer was not the option with the highest probability in terms of lexical word frequency and N-gram frequency, for instance word trigram frequency. These frequencies were determined on the basis of the Spoken Dutch Corpus (Oostdijk, 2002). None of the frequency measures was higher for the target word (hence, the correct answer) than for the other options in more than 15% of the trials.

The order of the four options on the screen was randomised (between items and between participants). The order was manually corrected if it formed a semantically and syntactically plausible continuation of the sentence, which could happen for maximally 25 of the 78 trials. To illustrate with the example provided above, the order *bijvoorbeeld eigenlijk altijd rood* 'for instance actually always in debt' would create a plausible continuation of the sentence, possibly leading to more *bijvoorbeeld* responses.

2.2.3 Procedure

The experiment consisted of eleven blocks, and each block contained the materials of one of the eleven speakers. The blocks and the trials within the blocks were randomised across participants. The experiment was self-paced, and was carried out in a soundproof booth. The experiment was programmed in and controlled by E-prime 1.2 (Schneider et al., 2002).

For each trial, the preceding context was presented on the screen for five or eight seconds (depending on the length of the sentence: Eight seconds if it consisted of more than sixteen words, otherwise five seconds), and then the four options appeared on the screen. Participants were asked to guess the following word by pressing one of the four buttons (labelled “1” to “4”) on a response box. The preceding context was then presented a second time so that the experiment was identical to the auditory experiment (see Experiment 2). The four options remained visible so that participants did not have to memorise the four options, which might otherwise interfere with their performance. Then, participants were asked to choose again.

2.2.4 Results and discussion

In all analyses presented in this study we investigated which variables favoured participants’ selection of a given option by means of generalised linear mixed-effects models with the logit link function (see, e.g. Jaeger, 2008) and with random effects for *participant*, *target type* (e.g. *namelijk* or *eigenlijk*), and *target token* (e.g. the third token of *eigenlijk*). We used a backwards stepwise selection procedure, in which predictors were removed if they did not attain significance at the 5% level. The fixed effect factors differed for each model, and will be mentioned for each model separately.

The descriptive statistics for Experiment 1 showed that participants, provided with four plausible options, selected the correct option in 33.27% of the trials, which is above chance (i.e. more than 25%, one-tailed t-test testing whether participants’ performance was significantly higher than 25%; $t(19) = 5.58$, $p < .0001$). This was unexpected, since our target words were discourse markers and adverbs, which could be left out of the sentences. Apparently, listeners can use preceding semantic/syntactic information to also predict these types of words. Further, most target words had lower unigram and N-gram frequencies than at least one of the other three options on the screen. Thus, if participants had used the N-gram probabilities or the lexical frequencies of the four options, or if they had just guessed, we would have seen performance at or below chance level (0.25). The above chance performance therefore indicates that participants used semantic/syntactic information in the wider preceding context to predict the following word.

We analysed participants’ selection of the correct answer versus the other options and included in the statistical model the fixed effects *repetition* (whether the context was presented for the first or second time), position of the correct answer on the screen (henceforth *position correct*), and *intelligibility in isolation*. Since *intelligibility in isolation* was not distributed normally, we converted this variable into an ordinal variable with four levels, representing the four quartiles. The quartile ranges are presented in Table 2.1. This factor was included in all subsequent analyses in this study (instead of the continuous variable).

Level	Intelligibility in isolation
Quartile 1	0% - 22.5%
Quartile 2	22.5% - 55%
Quartile 3	55% - 85%
Quartile 4	85% - 100%

Table 2.1: The quartile ranges for *intelligibility in isolation*.

We only found a main effect of *position correct*, which shows that the participants performed better if the correct answer was the second option and worse if the correct answer was the fourth option on the screen (F (3, 2996) = 32.04, $p < .0001$; 42.19% correct for option 2 and 22.44% correct for option 4). The following experiments also showed effects of *position correct*, suggesting that participants preferred Options 1 and 2 to Options 3 and 4. Since these position effects are not of primary interest for the present study, in the following sections their statistics will be reported in the tables, but they will not be discussed in the text. They are included in the statistical models to reduce the variance.

Importantly, the semantic/syntactic information was clearly insufficient for the participants to predict the target words without errors, as they chose incorrect options in no less than 66.73% of the trials. Given that listeners are sensitive to phonetic detail in laboratory speech, they may use acoustic cues in the preceding context to predict upcoming words in spontaneous speech as well. To investigate this possibility we conducted a second experiment, in which participants were presented with the preceding context of the target words auditorily. If participants use acoustic cues, they should perform better in this experiment than the participants in Experiment 1.

2.3 Experiment 2

2.3.1 Participants

Twenty native speakers of Dutch from the pool of participants of the Max Planck Institute for Psycholinguistics were paid to take part in the experiment. None of the participants had taken part in any of the previous experiments, and none of them reported any hearing loss.

2.3.2 Materials

Experiment 2 is identical to Experiment 1, except that the preceding contexts were presented auditorily. Thus, listeners heard the preceding context of reduced target words, without the target words themselves. The four options were again presented visually, as in Experiment 1. Since truncated speech often sounds very unnatural and may cause listeners to perceive an inserted labial or plosive consonant (e.g. Pols and Schouten, 1978), we added a 500 Hz square wave signal at the end of the preceding context. Square wave signals are not misperceived as speech sounds (Warner, 1998), and are therefore suitable for the current purpose. The signal had a fixed duration (505 ms) and consisted of an onset with gradually increasing amplitude (5 ms), and a 500 ms part with a fixed amplitude of 52 dB. The overall intensity of each sound fragment (excluding the square wave) had an average of 70 dB.

2.3.3 Procedure

The basic procedure was identical to Experiment 1, the only differences being that each speaker block was preceded by a brief familiarisation phase (mean: 19.8 seconds, range: 10.48 seconds to 36.99 seconds) that consisted of a short monologue by the speaker to introduce the participants to that speaker's (voice) characteristics, as in the control experiments. The participants listened to the speech via headphones. They heard each fragment twice (successively), because we wanted to establish whether listeners become more sensitive to subtle acoustic cues in the context when hearing the context a second time.

2.3.4 Results and discussion

The descriptive statistics for Experiment 2 show that participants, provided with four plausible options, selected the correct option in 39.47% of the cases. A regression model was fitted for response accuracy in the combined data set of Experiments 1 and 2. We included the fixed effect factors *repetition*, *position correct*, and *intelligibility in isolation* (with the four quartiles as its levels, see Experiment 1), and, critically, whether the stimuli were presented visually or auditorily (henceforth *presentation mode*).

We found a main effect of *presentation mode* ($F(1,5995) = 10.45$, $p < .01$). Participants performed better if the preceding context was presented auditorily rather than visually (39.47% versus 33.27% correct). In addition to semantic/syntactic cues, listeners apparently used acoustic cues in the preceding context to predict the upcoming word, even though conversational speech is characterised by a high speech rate and contains reduced words, which may obscure acoustic cues that have been proved to be useful in the comprehension of clear speech.

Possibly, participants just relied on acoustic cues since the semantic/syntactic cues in the context were insufficient. We investigated this hypothesis by testing the relevance of acoustic cues in the presence of more semantic/syntactic cues. In Experiments 3 and 4, participants were presented with both the preceding and following context of the target words, again either visually or auditorily. We investigated whether the difference between the visual and auditory modality established in Experiments 1 and 2 also holds when participants are provided with the preceding and following context.

2.4 Experiment 3

2.4.1 Participants

Twenty native speakers of Dutch from the pool of participants of the Max Planck Institute for Psycholinguistics were paid to take part in the experiment. None of the participants had taken part in any of the previous experiments, and none of them reported any hearing loss.

2.4.2 Materials

The context presented in Experiment 1 was extended to include words following the target word. Again, the target words themselves were not presented to the participants. In the presentation to the participants, the preceding and following context were separated by "________", which indicated the position of the target word. The four options presented for each trial were identical to those in Experiment 1, except that the word that followed the target word was replaced by "________", meaning that no word had been left out. See the example provided below.

Context:
Het geld is niet van jou en je staat ________ rood.
'The money is not yours and you are ________ in debt.'

Options:
1. *altijd* 'always' (correct, reduced word)
2. ________ (no words missing)
3. *bijvoorbeeld* 'for instance'
4. *eigenlijk* 'actually'

Since one of the four options was now replaced by "________", Experiments 3 and 4 cannot be well compared to Experiments 1 and 2. Participants dispreferred this option, which may have increased the percentage of correct responses in these experiments.

2.4.3 Procedure

The procedure was identical to that of Experiment 1, except that both the preceding and following context was presented on the screen, for ten seconds if it consisted of more than 25 words, otherwise for seven seconds.

2.4.4 Results and discussion

Participants selected the correct option in 37.27% of the trials. We analysed participants' selection of the correct answer versus the other options and included in the statistical model the fixed effects *repetition*, *position correct*, and *intelligibility in isolation*. We did not find any significant effects.

If participants only used word frequency and N-gram frequency information to predict the missing target words, they would select the correct answer in maximally 33% of the trials, since only in these trials the correct answer had a higher frequency or N-gram frequency with the preceding and following word than the other options. Participants managed to perform significantly better than 33% correct (one-tailed t-test testing whether participants' performance was significantly higher than 33%: $t(19) = 2.82$, $p < .05$), which again suggests that participants are sensitive to semantic/syntactic cues that are not captured by N-gram probabilities.

Since Experiment 2 showed that listeners can use acoustic cues in the preceding context to predict reduced tokens, they may also use acoustic cues if presented with both the preceding and following context. We conducted a fourth experiment, in which participants were presented auditorily with both the preceding and following contexts of the reduced tokens. If listeners also use acoustic cues if they are provided with the full context, they should perform better in this experiment than did the participants in Experiment 3.

2.5 Experiment 4

2.5.1 Participants

Twenty native speakers of Dutch from the same pool of participants as used in the previous experiments participated in the experiment for a small salary. None of these had taken part in these previous experiments, and none of them reported any hearing loss.

2.5.2 Materials

The auditory contexts presented in Experiment 2 were extended to include the words following the target words. Thus, participants were provided with the preceding context, the square wave, and then the following context. The target words themselves were not presented to the participants. The four options were the same as those in Experiment 3 and were presented to the participants visually.

2.5.3 Procedure

The procedure was identical to that of Experiment 2.

2.5.4 Results and discussion

The descriptive statistics for Experiment 4 show that participants selected the correct answer in 48.03% of the trials. We fitted a regression model to compare the effects of the preceding and following auditory context to that of the preceding and following visual context (Experiment 3). We entered the predictors *repetition*, *position correct*, *intelligibility in isolation*, and *presentation mode*. The results are provided in Table 2.2.

Predictor	F value	p value
presentation mode	16.58	$< .01$
repetition	5.70	$< .05$
position correct	4.72	$< .01$
presentation mode : repetition	5.39	$< .05$

Table 2.2: F values and significance values for the model comparing Experiments 3 and 4 (degrees of freedom: 5993).

The participants in the auditory context condition performed significantly better than those in the visual condition (48.03% versus 37.27% correct). Thus, even if provided with semantic/syntactic information in both the preceding and following context of reduced tokens, participants use acoustic cues in the context to improve their performance. Further, participants performed better after having heard the fragment a second time in the auditory presentation mode (50.73% versus 45.33% correct).

In conclusion, our results show that listeners use acoustic cues in the context if they are provided with the preceding semantic/syntactic context, but also if they are provided with the full semantic/syntactic context of reduced tokens. We now investigate, in more detail, which cues increased the predictability of reduced pronunciation variants.

2.6 Further analysis of combined results

In addition to the analyses presented above, we investigated, in more detail, the use of semantic/syntactic and acoustic cues in word identification. With respect to the semantic/syntactic cues, we focused on the question whether N-gram probability effects were equally pervasive in the auditory (Experiments 2 and 4) and visual domain (Experiments 1 and 3). Importantly, in the auditory experiments, the N-gram frequency information is in conflict with the acoustic cues, since we deliberately presented the participants with incorrect options that had higher N-gram frequencies with the words in the context than the correct answers in most trials. Participants presented with the auditory contexts may therefore have focused less on semantic/syntactic cues than the participants who did not have access to the conflicting acoustic information.

With respect to the semantic/syntactic cues, we also investigated whether they were stronger for trials with longer contexts. The amount of context varied considerably between the speech fragments used in this study. The preceding context varied from 2 to 22 words, while the following context varied from 1 to 13 words. Longer contexts probably contain more semantic/syntactic cues, and may hence increase the words' predictability.

With respect to the acoustic cues, we investigated which properties may affect the likelihood that such cues inform listeners. Fast speech demands fast processing, and participants may therefore be less sensitive to acoustic cues in the context, the higher the speed rate (or, on the contrary more sensitive, since these are adverse listening conditions, as hypothesized by Hawkins and Smith, 2001). Further, transitional cues may be more informative and more salient in vowels than in consonants and therefore listeners may find it easier to predict upcoming words that are preceded by words ending in vowels.

We fitted two regression models, comparing Experiments 1 and 2 and Experiments 3 and 4. The experiments with the preceding context could not be compared to the experiments with the preceding and following context, because the latter contained the option "________". Participants dispreferred this option, which could have increased the percentages of correct responses in Experiments 3 and 4. Once again, we included *correctness* as the dependent variable and we used the same predictors as above (*repetition*, *position correct*, *intelligibility in isolation*, and *presentation mode*) in addition to four other types of predictors.

First, we included variables indicating the predictability of the correct answer based on the word's a-priori predictability (its unigram lexical frequency, range: 117.47-2752.85 per million) and its predictability given the two preceding words (word trigram frequency, range: 0-31.98 per million) or preceding and following word (henceforth *surrounding trigram frequency*; only relevant for Experiments 3 and 4, range: 0-15.41 per million), relative to these same frequencies for the three other options. As the frequencies of the different options were not normally distributed, we could not calculate some relative continuous frequency measures and we therefore created three ordinal variables (*word frequency*, *preceding trigram frequency*, and *surrounding trigram frequency*) with three levels: Highest (i.e. the correct answer was the option with the highest frequency), intermediate, and lowest frequency (the correct answer had the lowest frequency). For the variables *word frequency* and *preceding trigram frequency*, the intermediate level included the words with the second and third frequency rank. However, for the variable *surrounding trigram frequency* we excluded the option that no word was missing, since for this option the *surrounding trigram frequency* is the frequency of the preceding word with the two following words, while in some trials there was only a single word following the target. There were no strong correlations between these measures.

Second, we included variables concerning the length of the context. We tested for effects of the length of the preceding context (in both regression models) and the following context (only in the regression model for Experiments 3 and 4). Since the lengths of the preceding and following context did not show normal distributions, we converted these numeric variables into two ordinal variables with four levels, representing their four quartiles (henceforth *length of the preceding context* and *length of the following context*). The quartile ranges for these two variables are presented in Table 2.3.

Level	Range in preceding context	Range in following context	Range in speech rate
Quartile 1	2-6 words	1-2 words	2.50-5.57
Quartile 2	7-9 words	3-5 words	5.57-6.54
Quartile 3	10-12 words	6-7 words	6.54-7.30
Quartile 4	13-29 words	8-13 words	7.30-9.30

Table 2.3: The quartile ranges for *length of the preceding context*, *length of the following context*, and *speech rate*.

Third, we included variables that may provide information about the likelihood that listeners can use acoustic cues given the characteristics of the speech signal. We tested for effects of *speech rate*, defined as the number of syllables of the phrase divided by the duration of the phrase (mean: 6.38 syllables per second). As *speech* *rate* was not distributed normally, we converted also this variable into an ordinal variable with four levels, representing the four quartiles. These quartile ranges are also presented in Table 3. We also incorporated as a predictor the type of the segment (i.e. consonant or vowel) immediately preceding the reduced word (henceforth *preceding sound*). Since both *speech rate* and *preceding sound* were only relevant in the auditory modality, we expect that if these variables have an effect, these effects will interact with *presentation mode*.

Finally, we also included as predictors the variables *trial number*, *block number*, and *block trial number* (i.e. trial number within the given block), which can all capture effects of learning and/or fatigue. These predictors were included mostly in order to reduce the variance in the data.

To begin with, we fitted a regression model for Experiments 1 and 2. The final model is summarised in Table 2.4. We found a main effect of *presentation mode*: Participants gave more incorrect responses in the visual modality (39.47% versus 33.27% correct). Further, we found a main effect of *intelligibility in isolation*: Participants made more errors for words that were difficult to recognise in isolation. This finding is unexpected given a listener driven account of speech reduction, which suggests that speakers reduce especially those words that are highly predictable for the listener (Ernestus and Baayen, 2007). Further, we found an interaction between *presentation mode* and *preceding trigram frequency*: In the visual presentation mode, participants had the tendency to choose an option with a relatively high trigram frequency (25.42% correct for target words with the lowest trigram frequency versus 35% and 33.64% for target words with the intermediate and highest trigram frequencies, respectively). The auditory experiment showed no frequency effects, which suggests that participants relied less on frequency information if they were also provided with acoustic cues in the context. We did not find any effects of *preceding sound*, *length of the preceding context*, *speech rate*, *repetition*, *trial number*, *block number*, or *block trial number*.

Predictor	F value	p value
presentation mode	10.35	< .01
intelligibility in isolation	6.51	< .001
position correct	41.80	< .0001
preceding trigram frequency	0.12	n.s.
presentation mode : preceding trigram frequency	4.33	< .05

Table 2.4: F values and significance values for the model comparing Experiments 1 and 2 (degrees of freedom: 5988).

Subsequently, we fitted a regression model for Experiments 3 and 4, including the same predictors as for the correctness analysis of Experiments 1 and 2. In addition, we included *surrounding trigram frequency*.

Predictor	F value	p value
presentation mode	16.41	< .001
position correct	4.66	< .01
repetition	5.74	< .05
surrounding trigram frequency	4.02	n.s. [1]
presentation mode : repetition	5.32	< .05
presentation mode : surrounding trigram frequency	3.66	< .05

Table 2.5: F values and significance values for the model comparing Experiments 3 and 4 (degrees of freedom: 5989).

The final model is summarised in Table 2.5. As also mentioned in the discussion of Experiment 4, participants gave more correct responses in the auditory than in the visual presentation mode (48.03% versus 37.27% correct). Further, we found again an interaction of *presentation mode* with *repetition*: In the auditory experiment, participants gave more correct responses after repetition (50.73% versus 45.33% correct). Apparently, listeners could make better use of the acoustic cues in the signal after repetition. More importantly for our research question, we found an interaction of *presentation mode* with *surrounding trigram frequency*: In the visual presentation mode, participants gave more correct responses if the target word had the highest trigram frequency rather than the lowest trigram frequency with its preceding and following word (50.21% correct for targets with the highest trigram frequency versus 28.13% for targets with the lowest trigram frequency). No trigram frequency effects were present for the auditory presentation mode. This finding again suggests that participants relied less on these frequency cues if provided with acoustic cues in the context. Interestingly, neither presentation mode showed effects of *preceding trigram frequency*, which suggests that, in Experiments 3 and 4, participants shifted focus from the preceding to the preceding and following context. We did not find effects of *preceding sound*, *length of the preceding context*, *length of the following context*, *speech rate*, *intelligibility in isolation*, *trial number*, *block number*, or *block trial number*.

In summary, comparing the visual and auditory experiments, one can observe a clear difference with respect to the role of trigram frequency: Whereas these measures are highly important in the absence of acoustic cues from the context, they play only a marginal role if these acoustic cues are provided. The role of the acoustic cues seems hardly modulated by speech rate or by the type of the segment preceding the target word.

[1] This effect was only significant in the analysis of variance results; not in the summary of the model.

2.7 General discussion

Ernestus et al. (2002) showed that contextual information is crucial for the understanding of reduced pronunciation variants. The present study investigated which contextual cues listeners can use to predict reduced word tokens or fixed expressions (henceforth "target words", for the sake of convenience) in spontaneous speech. Participants were only presented with contextual information of the target words, and the target words themselves were always completely missing. Previous research has shown that language users benefit from semantic/syntactic and acoustic cues in word recognition. In those studies, language users were placed in highly idealised listening situations, where they were presented with laboratory speech and had to recognise content words which were often almost completely predictable on the basis of their context. In normal listening situations, however, language users have to deal with casual speech, which is characterised by reduced pronunciations of words, high speech rate, hesitations and false starts. Further, listeners not only have to recognise highly predictable content words, but also less predictable words, including discourse markers and adverbs. The present study investigated the roles of semantic/syntactic and acoustic cues under these conditions.

We conducted four main experiments, in which participants were presented with either the preceding (Experiments 1 and 2) or the preceding and following (Experiments 3 and 4) context of reduced target words, either visually (Experiments 1 and 3) or auditorily (Experiments 2 and 4). Importantly, the target words themselves were missing. In contrast to previous studies, we did not use words that were highly predictable on the basis of their context but, instead, words that could be left out of the sentence without significantly changing its meaning. Further, our materials were not recorded in a laboratory setting, but they were extracted from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000). Replicating Ernestus et al. (2002), our control experiments showed that most of the reduced target words were difficult to understand in isolation (52.7% correct on average), but generally easy to understand within context (95.8% correct on average). In our four main experiments, participants were asked to predict the missing target words, choosing from four semantically and syntactically plausible options that were always presented visually.

Since the target words could always be left out of the sentences (without rendering these sentences ungrammatical), one of the four options provided was the absence of that word. In Experiments 1 and 2, in which participants were asked to guess the word following the presented context, this implied that one of the options was the word following that missing word (i.e. the word following the reduced target word). In Experiments 3 and 4, in which participants had to guess the word that was left out from the middle of the presented context, this option was replaced by "________", indicating that no word was missing in the context provided.

We investigated the role of semantic/syntactic cues directly in the two visual experiments, in which the contexts of the reduced target words were presented in the form of orthographic transcriptions (Experiments 1 and 3). First of all, we found that participants predicted the missing target words above chance in both the preceding and full context conditions (pure chance equalled 25%, since there were four options, and we obtained 33.27% and 37.27% correct for the two experiments). This finding is not self-evident, since we used low-predictability words, with low unigram and N-gram probabilities compared to the other three options presented on the screen. Our results therefore suggest that language models completely based on unigram or N-gram probabilities cannot explain language users' sensitivity to semantic/syntactic contextual information. Thus, language users are sensitive to higher-level semantic/syntactic information, for example as captured by Latent Semantic Analysis (Deerwester, Dumais, Furnas, Landauer, and Harshman, 1990), as well.

In fact, we did not find any effects of word frequency on participants' response accuracy at all. This result is in line with previous findings showing that context reduces the effects of word frequency in visual word recognition (e.g. Becker, 1979; Van Petten and Kutas, 1990; Rayner, Ashby, and Pollatsek, 2004). Van Petten and Kutas (1990) recorded ERPs while participants silently read semantically unrelated sentences. They found that low frequency words only yielded larger event-related brain potentials than high frequency words if they appeared early in the sentences: The difference between high and low frequency words disappeared when sufficient context was available, and the words were predictable to some extent given the preceding words in the sentence.

In contrast, there was a reliable effect of trigram frequency. In the preceding context condition, participants were more likely to choose the correct answer if that word formed a relatively frequent word trigram with the two preceding words (i.e. if the word had the highest or an intermediate trigram frequency relative to the other three options). In the full context condition, participants were more likely to choose the correct option if the target formed the most frequent word trigram with the preceding and following word.

Our observation that, when provided with the full context, participants focused on the words' trigram frequency with the preceding *and* following words indicates that in addition to the preceding context, language users can use the following context to recognise words. This conclusion is supported by our finding that participants' accuracy is influenced by how intelligible the missing word is in isolation (established in a control experiment). Previous research has reported a positive correlation between a word's (unigram and N-gram) frequency and its degree of reduction (e.g. Hooper, 1976; Jurafsky et al., 1998; Bell et al., 1999; Jurafsky et al., 2001; Pluymaekers et al., 2005; Bell et al., 2009), which suggests that speakers reduce especially those words that are highly predictable to the listener, as postulated by Aylett and Turk (2004). In contrast to this positive correlation, we found for the experiments with only the preceding context (Experiments 1 and 2, see Table 2.4) that participants made more errors for more reduced target words. This suggests that more reduced words were more difficult rather than easier to predict. This effect was absent when language users were also presented with the following context (Experiments 3 and 4), which shows that more reduced words were as predictable as less reduced words only if also the following context was provided. Participants thus extracted information from the following context in order to predict the missing words.

The effect of the following context on the predictability of reduced words in spontaneous speech is in line with previous research on the comprehension of careful speech (e.g. Warren and Warren, 1970; Warren and Sherman, 1974; Grosjean, 1985). For example, in a study by Warren and Sherman (1974), listeners presented with sentences that contain deliberately mispronounced phonemes (e.g. *George waited for the deli[b]ery of his new color TV*) replaced by noise (leaving only misleading transitional cues) recover from the misleading acoustic cues based on the following context (i.e. listeners heard *deli[v]ery* instead of *deli[b]ery* in the example above). Our results indicate that the role of the following semantic/syntactic context generalises to the processing of words with a low predictability (e.g. discourse markers/adverbs), words that can be left out without significantly changing the sentences' meanings.

The literature contains roughly two accounts for how semantic/syntactic cues might facilitate word recognition. Van Petten and Kutas (1990) claim that a word with a high contextual predictability can be more easily integrated into the preceding context, which facilitates semantic processing. Alternatively, some researchers suggest that language users rely on contextual information to directly predict lexical items and narrow down their lexical search space (e.g. Van Berkum et al., 2005). Both accounts can explain our data.

In the experiments in which the contexts were presented auditorily, listeners could potentially use co-articulation cues because the four words they had to choose from differed in their word-initial sounds. Obviously, other acoustic cues (e.g. prosody) may have been useful as well. We found that participants were better at predicting the reduced words in the auditory experiments than in the visual experiments (43.75% versus 35.27% correct on average). Further, for listeners who heard the full context, these effects were larger when hearing a speech fragment a second time, which may be due to the large amount of information provided in this full context condition.

Participants thus use acoustic cues in the context to their advantage, which is no mean feat, since the provided contexts were extracted from spontaneous conversations and therefore had a high speech rate (mean: 6.38 syllables per second), included other reduced words, and showed high variability (e.g. in speech rate, which varied between 2.5 to 9.3 syllables per second). Apparently, listeners can also use acoustic cues under these adverse listening conditions (as hypothesised by Hawkins and Smith, 2001). In fact, speech rate did not influence how well listeners predicted the missing words. We did not find effects of whether the sound preceding the missing word was a consonant or a vowel either. Further research is required to establish which acoustic cues in the context are particularly useful for the listener and, further, how exactly these cues facilitate the processing of spontaneous speech.

Several speech comprehension models assume a prelexical level of processing in which sounds are converted into abstract categories, such as phonemes (e.g. Shortlist B; Norris and McQueen, 2008). Shortlist B can account for listeners' sensitivity to acoustic cues in the context to the extent that these cues can be captured by the diphone database that is incorporated into the model. This database is based on careful speech and may therefore not contain acoustic transitional cues relevant for spontaneous speech. Further, previous research indicates that listeners are sensitive not only to acoustic cues in sounds directly preceding and following the target sound, but also to those that occur much earlier in the speech stream, for example acoustic traces of [r] in syllables preceding [r] (r-resonance, Kelly and Local, 1986), especially in spontaneous speech (e.g. Heinrich and Hawkins, 2009). Such acoustic cues cannot be captured by a diphone database but are likely to have influenced listeners in our experiments as well. These findings can more easily be explained by speech comprehension models that do not assume any pre-lexical abstraction, but that allow listeners to store all available acoustic cues in the input, for instance in the form of exemplars, to facilitate speech understanding (e.g. Polysp; Hawkins and Smith, 2001).

Importantly, our experiments also provide information on the relative contribution of semantic/syntactic and acoustic cues. These two types of cues can only be teased apart if there is a conflict between these two types of cues. Since the sentences were taken from natural speech, the acoustic properties of the contexts were always congruent with the target words. In most of our stimuli (at least in 65% of the cases), these cues were in conflict with the immediately surrounding context, as the N-gram probabilities were low. We found that, when listeners are presented with acoustic contextual information that (frequently) conflicts with N-gram probability information, N-gram probabilities do not predict participants' choices at all. This suggests that if acoustic and semantic/syntactic cues in the context are in conflict, listeners consider acoustic cues more reliable than the N-gram probability of the words with their surrounding words. This does not mean that semantic/syntactic probabilistic information does not play a role in everyday speech comprehension, but it suggests that whenever there is a conflict, listeners rely more heavily on acoustic cues in the context to predict low predictability words (e.g. discourse markers, adverbs) in casual speech. If these results are replicated in further research, this interaction between the roles of semantic/syntactic and acoustic contextual information should be incorporated into current models of speech comprehension.

Finally, since listeners have great difficulty understanding highly reduced words in isolation, one may hypothesise that they use only context to understand such pronunciation variants. Thus, when English listeners hear "supposed to" pronounced as [səsə], they may deduce its meaning purely on the basis of the context. Our results are in contrast with this hypothesis. Participants performed above chance-level (more than 25% correct) in all four main experiments, but were unsuccessful at predicting the reduced words across the board (less than 50% correct). Nevertheless, one of our control experiments showed that listeners can well recognise the target words when they hear the preceding and following auditory context together with these words (more than 90% correct). We therefore conclude that listeners not only need the context but also the reduced word itself to comprehend this word, which underlines the significance of the (albeit reduced) acoustic properties of highly reduced pronunciation variants.

In conclusion, the present study investigated the contribution of the various types of contextual information available in spontaneous speech. Whereas most studies focused on the role of context in the comprehension of content words in simple sentences in laboratory speech, the present study investigated to what extent listeners can use context to process low predictability words (e.g. discourse markers and adverbs) in natural spontaneous speech. Our results show that listeners use both the preceding and following context to process such words, and that they are sensitive to semantic/syntactic as well as acoustic cues in spontaneous speech contexts, but that they favour acoustic cues in case of a conflict.

2.8 Appendix

Orthographic transcriptions of the materials used in the current study. The reduced target words are underlined. In a few cases the target word will be missing in the English gloss because there is no corresponding English word.

Alleen 'only'

Dat zou kunnen die ken ik <u>alleen</u> van naam inderdaad, ja.
'That is possible that one I <u>only</u> know by name, indeed, yes.'

Ik heb wel mensen gesproken die alleen maar regen gehad hebben.
'I talked to people who only had rain.'

films op Duitsland kijk is inderdaad bijna alleen de
'watch films on the German channel is indeed almost only the'

geld in de archeologie geduwd, maar dan wordt het alleen maar gebruikt om gaten bij de universiteit op te vullen
'money invested in archeology, but then it is only used for closing holes in the university budget.'

ik heb het alleen over dat als ik geld wil sparen
'I'm only talking about if I want to save money'

Ja, het was 2 dagen geloof ik bij mekaar. Ik geloof zaterdag ook, maar ik heb alleen zondag gekeken
'Yes, I believe it was 2 days in total. Saturday as well, I believe, but I only watched on Sunday.'

eh ze werken geloof ik op een tandtechnisch laboratorium ofzo en het is ook alleen 's avonds is vier dagen in de week.
'eh I believe they work at a dental laboratory or something and it is also only in the evening four days a week.'

en dat wordt heel saai natuurlijk als je alleen maar weer die auteur telkens weer ziet.
'and that becomes really boring of course if you only see that author time and time again.'

Allemaal 'all'

o dit is 'm dus en dan stap je allemaal weer in of je neemt nog een patatje ofzo hè.
'oh so this is it and then you all get inside again or take some french fries or something huh.'

Nou, waar we het tijdens het feestje allemaal over gehad hadden.
'Well, all the things that we were talking about at the party.'

daarnaast nog een baantje in Amsterdam, en dan was ik dat allemaal aan het regelen.
'Besides that a job in Amsterdam, and then I was managing everything.'

toen eh heb ik daar een wortelpuntontsteking gehad dus dat moest allemaal weer open enzo
'then eh I had had a root canal infection so they had to open everything up again.'

nou ja je kan ook zeggen van dat doen we allemaal niet meer zo uitgebreid.
'well yes you can also choose to do everything less extensively.'

Als je het zo hebt geregeld dat je zelf in zo'n kamertje als dit zit en het hier vandaan allemaal organiseert
'If you have arranged it such that you have a room like this one and manage everything from here'

dus dat zijn allemaal voordelen.
'so those are all advantages.'

Altijd 'always'

het is in ieder geval leuk dat je op een feestje ziet want dan kan je altijd nog eens eventjes na een kwartiertje
'at least it's fun if you see it at a party because then you can always briefly after 15 minutes'

Ja precies maar daar komen ze altijd als voorbeeld mee
'Yes exactly, but that's the example they always give'

Ik vertrouw altijd maar op mijn goede geluk.
'I always trust my good luck.'

Het geld is niet van jou en je staat altijd rood.
'The money is not yours and you are always in debt.'

nou ja dat hadden we vroeger ook al zo dat is altijd het argument.
'well yes it's the same as it used to be that is always the argument.'

hij geeft me altijd zo het idee dat ik me aanstel.
'he always gives me the impression that I'm exaggerating.'

1 exemplaar plus de overdrukken en daar zijn ze altijd wel heel royaal in dat je minstens 50 overdrukken krijgt.
'1 copy plus the offprints and they are always very generous with these such that you receive at least 50 offprints.'

Anders 'otherwise'
Misschien dat het bij jullie anders ligt, want ik heb het gevoel
'Maybe it's different in your case because I have the feeling'

Ja je kiest er zelf voor als je niet wil moet je wat anders gaan doen.
'Well, it's your own choice; if you don't want it you will have to start doing something else.'

eh bibliografisch nazoek werk erbij wat je allemaal in een keer doet omdat je anders veel meer tijd eraan kwijt bent.
'eh bibliographic research in addition, all at once because otherwise you spend too much time on it.'

kom je op posities waar je anders niet terecht zou komen, nu is het natuurlijk anders omdat het afgeschaft wordt komen er een heleboel plekken vrij.
'you reach positions that would otherwise be unreachable, now it's obviously different because it is being abolished many positions become available.'

En ik denk dat het iedereen gaat ervan uit binnen de bibliotheek dat het allemaal nodig is dat het niet anders kan en dat omdat het zo'n grote ingewikkelde organisatie is
'And I think that everyone at the library assumes that it's all necessary and that there's no other way and that because it's such a large, complex organisation'

vind ik tenminste, de direktie denkt daar anders over
'at least that's my view, the management has a different view'

Bepaalde 'certain'

die worden ook gesponsored door bepaalde firma's die consumentengids volgens mij.
'They are also sponsored by certain cooperations that consumer guide I suspect'

Bijvoorbeeld 'for instance'

als je namelijk buiten Amsterdam woont en je moet Amsterdam door, en dan, als je er langs moet, dan sta je altijd in de file, terwijl ik 's ochtends bijvoorbeeld naar Brabant ga, dan heb ik nooit last van files.
'If you live in the outskirts of Amsterdam, and you have to drive through Amsterdam, and then, if you have to drive though, you are always stuck in traffic, whereas if for instance I go to Brabant I never get stuck in traffic.'

Als ik naar eh als ik bijvoorbeeld geen geld kan pinnen ga ik naar mijn ouders.
'If I go to eh if for instance I cannot withdraw money I go to my parents.'

Maar het is dan zo, ik heb het bijvoorbeeld nu al weer ik geloof
'But it is like this, by now I have for instance I believe'

Eigenlijk 'actually'

omdat ik alleen op de VU die mogelijkheid heb om te doen wat ik nu doe eigenlijk een beetje in een kleine groep mensen
'because only at the VU I got the opportunity to do what I am doing right now actually a bit in a small group of people'

hoeveel locaties nu eigenlijk vervuild zijn.
'how many locations now actually are polluted.'

Hoe dat verder gaat weet niemand eigenlijk, want we zijn net begonnen.
'How this story continues nobody knows actually because we have only started recently.'

nee maar het is toch ook, maar het is toch ook eigenlijk te gek voor woorden.
'No but it is indeed, but it is actually to crazy to be true indeed.'

Dus ja dat is natuurlijk helemaal niet de bedoeling want er is eigenlijk geld te weinig.
'So yes that is of course not the idea because there is actually a lack of money.'

en ja er is zo veel dat vind ik dus eigenlijk zonde van je tijd dan.
'and yes there is so much; so I actually consider that a waste of time then.'

Gewoon 'usual'

nou ik heb het vooral gedaan voor de vakanties dat leek me gewoon heel praktisch
'well I particularly did it for the holidays that simply seemed very practical'

Ja. Daar daar moet je dan die schoenen niet voor aantrekken dan moet je gewoon gympen aantrekken.
'Indeed. For that you should not be wearing those shoes, but you simply should be wearing sneakers.'

Quizzen zeggen me gewoon niets.
'Quizzes simply don't mean anything to me.'

Ja al snel duik je toch weer hetzelfde in, het is eigenlijk gewoon een soort van de draad oppakken.
'Yes in no time you dive immediately in, it is actually simply a matter of acclimatising.'

Maar je houdt toch eh vooral met hardlopen hou ik gewoon last
'But you still eh especially when I'm running it simply remains painful.'

Ja maar er zijn gewoon wel een heleboel bedrijven die eh in zee gaan
'Yes but there are simply enough companies who consolidate with'

Heb je veel hoor mensen die gewoon nooit gaan.
'This happens a lot people who simply never go.'

Helemaal 'totally'

en aan het eind lijkt alsof ie helemaal leeg is
'and in the end it seems like it is totally empty.'

O daar hebben ze het nu helemaal niet over gehad.
'oh they didn't discuss that at all this time.'

Met zeven van die acht bejaarden daarvoor ruzie gehad, en ze kon helemaal niet met die lui opschieten.
'Had recently had fights with seven of the eight elderly, and she could not at all get along with these people.'

Die heeft daar een pak geld onder zijn bed liggen waar je helemaal niet goed van wordt.
'That person has a heap of money underneath his bed, which drives you totally crazy.'

flink met een kwast waardoor het helemaal niet meer zichtbaar was.
'vigorously with a brush as a result of which it was completely invisible.'

Nou dat is niet helemaal waar
'Well that is not entirely true'

precies en ik controleer het ook weer eens drie maanden later van is dat wel gebeurd nou dat wil ik natuurlijk eigenlijk helemaal niet doen.
'exactly and I also check three months later whether that has happened and well I actually really don't want to do that.'

Ik kom net ook aan en het regende helemaal niet, het is net begonnen.
'I also just arrived and it was not raining at all, it just started raining.'

In ieder geval 'in any case'

en het prettige is dan natuurlijk wel dat ze in ieder geval niet examen hoefde te doen.
'and the good thing is of course that she in any case did not have to do an exam.'

zekerheid hebben dat als het fout loopt, dat ik in ieder geval niet verlies
'have the certainty that if things go wrong, I in any case don't lose'

Nou ja je kunt het risico nemen van niet, maar je weet zeker dat er naar gekeken wordt. En in ieder geval binnen de Europese Unie kun je al alle grenzen over.
'Well, you could take the risk, but they will definitely look at it. And at least within the European Union you can cross all borders.'

nu ben ik geloof ik aan het terugwerken naar '87/'88 zijn ze besteld en nooit wat ik heb in ieder geval geen informatie ervan of we ooit wat gehoord hebben.
'Now I am I believe working back to 87/88 they were ordered and never I have in any case no information about whether we ever heard from them.'

Misschien 'perhaps'

terwijl ze zelf ook misschien al een half jaar een baan moeten hebben.
'while they themselves also maybe need a job in half a year.'

wat jij bedoelt is dat je misschien geen kans hebt.
'what you mean is that you might not have a chance.'

en zo zie je dus dat eh je bent misschien wat minder tijd kwijt aan puur aan uitlenen
'and so you see that eh you maybe spend less time on purely on book loans'

hè maar zou zou ik misschien toch koffie kunnen krijgen, of ja?
'hey but could could I maybe have some coffee after all, or yes?'

Namelijk 'namely'

Ja het probleem is namelijk ten eerste
'Yes the problem is namely first of all'

Uiteindelijk duurt zo'n discussie ook altijd heel erg lang want er wordt namelijk volgens, heerlijk, er wordt volgens mij heel veel herhaald.
'Eventually such a discussion always lasts extremely long namely because there is I think, lovely, there is a lot of repetition.'

Natuurlijk 'of course'

in ieder geval bestaat in die vorm want dat weet je natuurlijk ook niet hè, maar toen vond ik het inderdaad plotseling ook leuk worden.
'Even exists in that form because you don't know that of course, but then I indeed also started enjoying it.'

Hm, ja eigenlijk wel alleen werken we natuurlijk ook weer altijd onder een tijdsdruk.
'Hmm, actually yes, except that we of course always work under a certain time pressure.'

En wat en wat een voordeel is van een vaste, voordeel was van een vaste baan is natuurlijk dat je makkelijk leningen kunt krijgen bij de bank.
'and what and what is an advantage of a permanent, was an advantage of a permanent job is of course that you could easily get a loan at the bank.'

bezwaren opzij zettend is natuurlijk dat je je verhandelt hele dure dingen.
'all objections aside, of course you trade very expensive things.'

of dat je ziet van hè dat moeten we eens kijken of daar wat mee is maar dan blijft natuurlijk een aantal dingen zitten.
'or that you see hey that we should take a look at whether there is something wrong with that, but still you of course have some loose ends.'

Nou wat dat betreft is jouw studie natuurlijk ook wel gunstiger, denk ik.
'Well in that respect your study is of course also favourable, I think.'

qua qua prijs is het natuurlijk het voordeligst om een grote partij in te slaan
'as for as for the price it is of course advantageous to buy large amounts'

Op een gegeven moment 'at a certain moment'

Ja maar als je een vast contract hebt kun je natuurlijk ook op een gegeven moment ontslagen worden.
'Yes but if you have a permanent job you can obviously also get fired at a certain moment.'

Nou dan hou je of tenten over of je komt op een gegeven moment slaapzakken tekort.
'Well then either you end up having several tents left, or at a certain moment there is a shortage of tents.'

Over 'about'

terwijl je eigenlijk mag verwachten dat iemand die over een portie ethiek moet
'while you would actually expect that if anyone should have a certain amount of ethics'

Tenminste ‘at least’

Dat is tenminste waar ik, volgens mij hadden we het daar over.
‘that is at least what I, I believe they were talking about that.’

eh maar even voorzichtig zijn met bestellen want dan kan ik tenminste dingen doen.
‘eh be careful right now with placing orders because then I can at least do things.’

zo bekijk ik het wel tenminste anders zou ik dit dus ook niet doen wat ik nu doe.
‘that is at least how I look at it otherwise I would not do what I am doing right now.’

Semantic context effects in the comprehension of reduced pronunciation variants

Chapter 3

This chapter has been reformatted from:
Marco van de Ven, Benjamin V. Tucker, and Mirjam Ernestus (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition*, 39, 1301-1316.

Abstract

Listeners require context to understand the highly reduced words that occur in casual speech. The present study reports four auditory lexical decision experiments, investigating the role of semantic context in the comprehension of reduced versus unreduced speech. Experiments 1 and 2 showed semantic priming for combinations of unreduced but not of reduced primes and low-frequency targets. In Experiment 3, we crossed the reduction of the prime with the reduction of the target. Results showed no semantic priming from reduced primes, regardless of the reduction of the targets. Finally, Experiment 4 showed that reduced and unreduced primes facilitate upcoming low-frequency related words equally if the interstimulus interval is extended. These results suggest that semantically related words need more time to be recognised after reduced primes, but once reduced primes have been fully (semantically) processed, these primes can facilitate the recognition of upcoming words as well as do unreduced primes.

3.1 Introduction

One important characteristic of spontaneous speech is that many word tokens are much shorter than their corresponding citation forms (e.g. Ernestus, 2000; Johnson, 2004). For example, the English words *yesterday* and *ordinary* can be pronounced like [ˈjeʃeɪ] and [ˈɒnrɪ]. Segments may be shorter, completely missing, or realised differently. This may lead to ambiguity, since, for example, the distinction between long and short vowels and between voiced and voiceless stops may be smaller in spontaneous speech than in careful speech. Previous research has shown that listeners need contextual information (e.g. semantic or acoustic information) to understand highly reduced pronunciation variants (e.g. Ernestus, Baayen, and Schreuder, 2002; Kemps, Ernestus, Schreuder, and Baayen, 2004; Chapter 2 of this dissertation), but it is so far unknown what types of contextual information listeners rely on and to what extent. The present study investigates how semantic context contributes to the recognition of mildly reduced and unreduced variants.

Ernestus et al. (2002) extracted pronunciation variants from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000). Tokens were classified as having a low degree of reduction if hardly any or no segments were missing. Tokens were classified as having a medium degree of reduction if they were reduced but consisted of more than the initial, final, and stressed segments. The remaining words were classified as having a high degree of reduction. Participants listened to these variants in isolation, within their phonological context (i.e. together with adjacent vowels and any intervening consonants), or within their sentential context. The participants' task was to orthographically transcribe the speech fragments. The results showed that words of a low degree of reduction were well recognised in all three context conditions. Words of medium or high degree of reduction, however, were only recognised if presented within at least the phonological context (for words with a medium degree of reduction) or the sentential context (for words with a high degree of reduction). Thus, listeners require contextual support to recognise highly reduced pronunciation variants.

A study by Kemps et al. (2004) suggests that listeners reconstruct missing speech sounds when they hear reduced pronunciation variants in context. In a phoneme-monitoring experiment, Dutch participants were presented with target words ending in the derivational suffix *-lijk* [lək] (e.g. *eigenlijk* 'actually', *koninklijk* 'royal') extracted from spontaneous speech. They heard canonical realisations of these words and reduced realisations, in which the suffix [lək] was produced as [k], and were asked to press a button whenever they heard an [l]. If the suffix was presented in isolation, participants correctly pressed a button only for those variants that contained [l] (i.e. only the unreduced variants). However, when the suffix was presented in sentence context, participants also pressed a button for the reduced variants (i.e. without [l]).

It is unclear what type of contextual information is used by listeners to understand reduced pronunciation variants. Reduced speech may be considered as an adverse listening condition. Some studies predict that, in adverse listening conditions, listeners rely more heavily on any available information, including contextual information (e.g. Hawkins and Smith, 2001). However, more recent work indicates that this is not necessarily the case. For example, the role of semantic contextual information appears marginal in listening to low-pass filtered speech, another type of adverse listening condition (Aydelott and Bates, 2004; Aydelott, Dick, and Mills, 2006). Further, research by Andruski, Blumstein, and Burton (1994) suggests that semantic context helps English listeners less in their comprehension of obstruents with reduced compared to unreduced VOT distinctions.

In the present study, we investigated the role of semantic context in the comprehension of reduced pronunciation variants in a series of simple auditory lexical decision experiments with implicit semantic priming. Listeners heard English nouns and pseudonouns, and for each word they had to make a lexical decision. We examined the effects of a target word's semantic relatedness to the preceding word on the recognition of the target word.

All words were thus produced and presented in isolation, rather than in sentence context, in order to isolate semantic effects from other higher level information (e.g. syntax and pragmatics). Follow-up research should indicate whether semantic information in the sentence or discourse context influences the recognition of reduced pronunciation variants in the same way as in our experiments.

Previous research has shown that semantic priming effects are strong for words with a low word frequency, but only marginal (or even absent) for words with a high word frequency (e.g. Becker, 1979; Van Petten and Kutas, 1990; Rayner, Ashby, and Pollatsek, 2004). Therefore, we predicted that semantic priming effects for reduced speech will be largest for words with a low word frequency as well.

Many studies investigating semantic priming use words that on the basis of the preceding context are either highly predictable (e.g. *The opposite of hot is* **cold**.) or unpredictable (e.g. *She read about the* **flower**; e.g. Meyer, 1971; Donnenwerth, Tanenhaus, and Seidenberg, 1981; Bradlow and Alexander, 2007). In everyday listening situations, however, words are seldom highly predictable. Hence, listeners often need to resort to using subtle semantic information in the context. The present study uses a continuous measure of semantic relatedness, and the target words in this study vary from being mildly related to highly related to their preceding words, rather than either semantically highly related or completely unrelated.

We used Latent Semantic Analysis (LSA) to estimate the semantic relatedness of the words (Deerwester, Dumais, Furnas, Landauer, and Harschman, 1990). LSA provides a score, ranging from -1 to 1, that indicates to what extent words are semantically related, where a higher LSA score denotes a stronger semantic relatedness. LSA rests on the assumption that semantically related words tend to occur in similar texts. This computational technique uses deep statistical analysis to infer words' semantic relationships beyond their first-order co-occurrences. On the basis of the distributions of words, these words are placed in a multi-dimensional vector space. LSA scores are obtained by computing the cosine distance between the words' vectors. Previous research has shown that LSA scores can predict human behaviour in psycholinguistic experiments, for example semantic priming in a visual lexical decision task (Landauer and Dumais, 1997a).

In the present study, we investigated the role of semantic contextual information in the processing of unreduced and reduced pronunciation variants. We report four auditory lexical decision experiments, in which participants were presented with unreduced and/or reduced isolated words. In Experiment 1, participants were presented with only unreduced words, in order to establish the baseline effects of semantic context for listeners presented with clear speech in our experiments.

3.2 Experiment 1

3.2.1 Participants

Twenty native speakers of English from the participants pool of the Department of Linguistics, University of Alberta, took part in the experiment, and received course credit for their participation.

3.2.2 Materials

We extracted 154 nouns, with varying word frequencies (range: 40-58322), from the spoken portion of the Corpus of Contemporary American English (85 million word tokens; Davies, 2008). These nouns were used to construct 77 word pairs (see the appendix) that differed in their semantic relatedness. The semantic relatedness of the members of the word pairs ranged from semantically highly related (LSA score: 0.93; e.g. *saddle* - *horse*) to mildly related (LSA score: 0.36; e.g. *snake* - *beak*). We obtained LSA scores for the word pairs by using the Pairwise Comparison interface at the LSA website (Landauer, 1998), where we selected the term to term comparison type, 300 factors, and as the topic space General Reading up to the 1st year of college. Both the LSA scores and the log word frequencies of the second members of the word pairs (i.e. the target words) were normally distributed, and therefore they could be used as numeric variables in regression analyses. The members of a word pair were presented in consecutive trials and we investigated the effect of the semantic relatedness of the word pairs on the recognition of the second members of these pairs.

Further, the experiment contained 174 filler words that were semantically unrelated to their preceding and following word, and 128 pseudowords. The pseudowords were all phonotactically possible words of English, and two Mann-Whitney tests showed that they had the same number of syllables and segments as the existing English words on average ($p > .1$ in both cases; mean number of syllables: 1.5 for the pseudowords versus 1.5 for the existing words; mean number of segments: 5.4 for the pseudowords versus 5.4 for the existing words). Since we included only a limited number of pseudoword fillers we induced a “YES”-response bias, making it particularly difficult to find any priming effects in our data. Further, the many unrelated fillers were included to minimise strategic priming effects. Consequently, any semantic priming effects that show up are robust effects.

The materials were spoken by a male native speaker of Canadian English, who pronounced the words carefully (mean speech rate: 6.79 segments per second). We presented him with the words in a fully randomised order, such that no words were preceded by semantically related words and the speaker’s realisations of the words could not be affected by semantic priming. A different native speaker of Canadian English verified that all the words were pronounced naturally and clearly. The recordings were made in a sound-attenuated booth at the Alberta Phonetics Laboratory, with an Alesis ML-9600 hard disc recorder and a Countryman E6 directional microphone. The sampling rate was 16-bit/44.1 kHz.

After having extracted the individual words from the recordings, we created three lists, in which the 77 semantically related word pairs, the filler words, and the pseudowords were pseudo-randomised so that no more than six existing words or three pseudowords occurred in succession. Further, we avoided rhyme and/or alliteration between words in consecutive trials. There were minimally six participants per list.

3.2.3 Procedure

The participants were tested individually in a sound-attenuated booth, using E-prime 2.0 (Schneider, Eschman, Zuccolotto, 2007), and MBQUART QP805 Demo headphones. They listened to the stimuli over closed headphones and decided as quickly as possible, for each stimulus, whether it was an existing English word. The next stimulus was presented 1000 ms after each button press, or after a time-out of 4500 ms from stimulus onset. We selected these timing parameters on the basis of a pilot experiment, which showed the time participants needed to recognise the unreduced materials and the reduced materials used in Experiment 2, and to get ready for the next stimulus. The materials were presented at a comfortable listening level. The experiment lasted approximately 15 minutes.

3.2.4 Results and discussion

Participants produced 1523 correct responses, 15 incorrect responses, and 2 time outs for the target words (mean RT from the words' uniqueness points excluding the time-outs: 288.04 ms). We analysed participants' RTs for the correct responses, by means of linear mixed-effects models with contrast coding for factors (e.g. Jaeger, 2008). We measured RTs from the words' uniqueness points (UPs), which are the segments in the words at which these words diverge from all other words in the language (Marslen-Wilson, 1987). We measured RTs from the words' uniqueness points rather than from word offsets, because listeners may recognise words before word offset. We preferred the UP over word onset, because the unreduced words in this experiment and the reduced words used in Experiment 2 differ in how quickly they become unique and can be recognised. We determined the words' uniqueness points on the basis of CELEX (Baayen, Piepenbrock, and Gulikers, 1995).

We restricted all the analyses in this paper to the target words rather than including also the primes and the existing filler words, because the LSA scores with the preceding words were distributed normally only for the target words. Moreover, the primes and existing filler words were all preceded by semantically unrelated words, and the present study focuses on listeners' sensitivities to differences in degree of semantic relatedness rather than differences between semantically related and unrelated word pairs.

Furthermore, we restricted all analyses to correct trials directly preceded by other correct trials, excluding those trials for which the data show that the listeners did not recognise the target or the prime (or both). One of the semantically related word pairs (*dance-ballet*) was discarded in our analyses (for all experiments in this study) because the target word (*ballet*) was not recognised by more than 50% of the participants. We also removed those trials for which participants pressed the button prior to the words' uniqueness points, because in these cases participants were probably guessing. In addition, we removed trials for which the RT or the RT for the preceding trial (henceforth *previous RT*) was extremely long (> 1500 ms after stimulus onset), thereby removing trials for which the interstimulus interval was very long (> 2500 ms after stimulus onset). We removed these trials because we wish to investigate the effects of the interstimulus interval and compare our results to those from Experiment 4, in which we used a fixed interstimulus interval of 2500 ms.

The final data set consisted of 1412 trials. We applied a log transformation to the RTs and to the previous RTs in order to obtain normal distributions, and we analysed our data by means of a backwards stepwise selection procedure, in which predictors and interactions were removed if they did not attain significance at the 5% level.

We included the fixed effect variables *lsa* (lsa score) and *target word frequency* (log word frequency). Further, we included five additional variables mainly to reduce variance in the data set. We included the fixed variables *trial number* and *target word duration* (log of the stimulus duration; we took the log of the durations so that the RTs and durations were on the same scale). We also included *previous RT* (log of the RT on the preceding trial), as an indication of the participants' local response speed. Further, we included the random variables *participant* and *word*.

For all the regression models reported in this study, we excluded data points for which the standardised residuals were smaller than -2.5 or larger than 2.5. We then reran the regression models. A summary of the results is provided in Table 3.1.

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	13.294	-	< .0001
trial number	-0.0004	14.53	< .001
previous RT	0.257	79.47	< .0001
target word duration	-1.072	62.05	< .0001
target word frequency	-0.322	2.37	n.s.
lsa	-3.321	2.06	n.s.
target word frequency : lsa	0.432	10.11	< .01
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.053	134.33	< .0001
word	0.039	51.53	< .0001

Table 3.1: Results for the statistical analysis of the logged RTs in Experiment 1.

First of all, there were significant main effects for the control variables *trial number*, *previous RT*, and *target word duration*. Participants responded faster towards the end of the experiment, when the preceding RT was short, and to longer words. As we found similar effects of *target word duration* if we measured the RTs from word onset or word offset, we interpret this target word duration effect as reflecting how much time listeners had to narrow in on the target word and limit the number of competitors prior to the word's uniqueness point.

Most importantly for our research question, we found an interaction between *lsa* and *target word frequency*. This interaction is shown in Figure 3.1. We found semantic priming effects for words with a relatively low word frequency, in line with the literature cited in the Introduction of this paper. In addition, we found a semantic interference effect for words in the highest frequency range. Since this interference effect has not been reported in the literature before, we further investigated this effect. We refitted our regression model to a subset of the data consisting of words

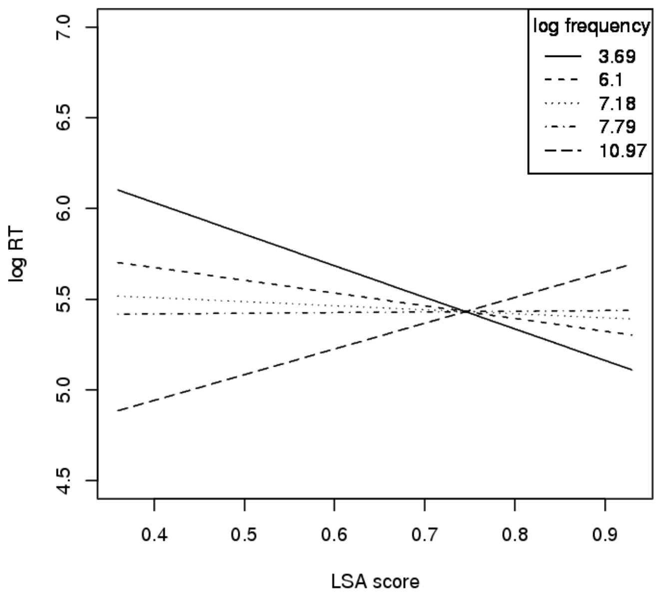

Figure 3.1: The combined effects of *lsa* and *target word frequency* on log RT for the target words in Experiment 1.

in the highest range for lsa and/or word frequency, and found that this interference effects holds for ten word pairs, namely *borders-country*, *chair-room*, *driver-car*, *game-player*, *gold-silver*, *peace-war*, *peaks-mountain*, *saddle-horse*, *team-coach*, and *tooth-dentist*. The negative main effect of *lsa* disappeared if we included additional word pairs. Hence, this semantic interference effect was not restricted to two or three tokens obviously sharing some characteristic in our experiment. We will formulate an explanation for this interference effect in the General discussion of this paper.

The question now arises whether we can find similar semantic context effects for reduced speech, which is characterised by shorter word durations and missing segments. We addressed this issue in Experiment 2, in which we tested reduced pronunciations of the same words as we used in Experiment 1.

3.3 Experiment 2

3.3.1 Participants

Twenty native speakers of English from the same subject pool as used in Experiment 1 received course credit to take part in the experiment. Participants in Experiment 2 had not participated in Experiment 1.

3.3.2 Materials

We created a new set of recordings of the materials used in Experiment 1. For these new recordings, we asked the same speaker of Canadian English to pronounce the same list of words, but now at a faster speaking rate, in order to elicit more reduced speech. Again, another native speaker of Canadian English verified that the words were produced in a natural manner.

The durations of these reduced realisations were significantly shorter than those of the unreduced realisations tested in Experiment 1 ($t(574.20) = -25.47$, $p < .0001$, mean durations: 377.01 ms for the reduced realisations compared to 568.10 ms for the unreduced realisations, mean speech rate: 11.4 and 6.8 segments per second for the reduced and unreduced realisations respectively, see Figure 3.2). We also compared the durations of the unreduced and reduced variants by subtracting the duration of the reduced variant from the duration of the unreduced variant, and dividing its outcome by the duration of the unreduced variant. The descriptive statistics are provided in Table 3.2.

The reduced realisations were not only durationally shorter than the unreduced ones, but also contained fewer segments. The reduced and unreduced realisations of each word were phonetically transcribed and we subtracted the number of segments in the reduced realisation (Experiment 2) from the number of segments in the unreduced realisation (Experiment 1), dividing its outcome by the number of segments in the unreduced realisation. For example, *story* [ˈstɔri] (five segments) was reduced to [ˈstɔi] (four segments), and *player* [ˈpleɪər] (six segments) was reduced to [ˈpleɪr] (five segments), resulting in the scores 0.2 and 0.17, respectively. The descriptive statistics are reported in Table 3.2. In most reduced realisations, no segments were completely missing.

We performed Mann-Whitney tests investigating whether there were differences between the degrees of segmental and durational reduction for the primes, targets, existing filler words, and pseudowords. We did not find any differences ($p > 0.1$ in all cases).

We also tested which segments were typically reduced in our materials. We found that consonants were more frequently missing than vowels (in 18% versus 2.5% of

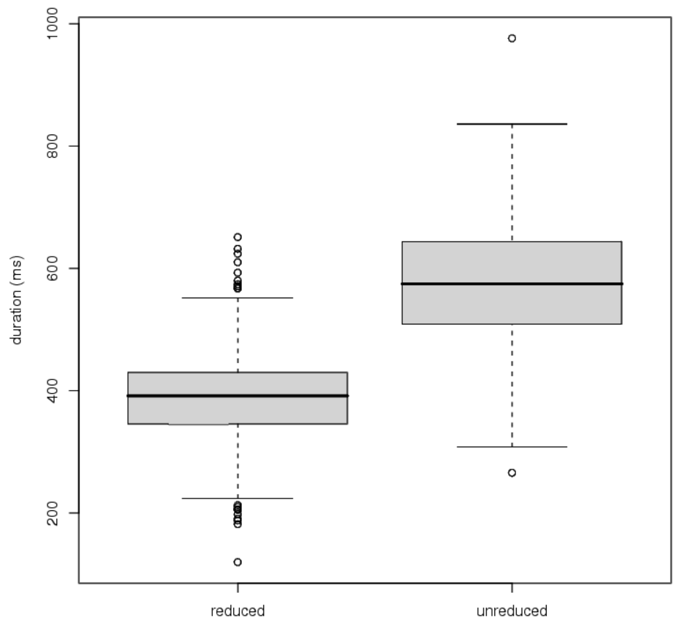

Figure 3.2: Word durations for the reduced and unreduced realisations of the stimuli.

the words). Most of these missing consonants were plosives (84.51%), followed by approximants (12.68%), and fricatives (2.82%), and consonants were missing especially in syllable-final position (85.92%). For example, the word *curtains* was realised like [ˈkɜːrtənz] in the unreduced and like [ˈkɜːrtns] in the reduced condition. Spectrograms and transcriptions of the unreduced and reduced realisation of this word are provided in Figure 3.3. Please note that our materials were only mildly reduced compared to the most extremely reduced pronunciations that can be found in spontaneous speech, and that can only be recognised within their linguistic context. As the reduced words in our study were produced in isolation, they were only mildly reduced and they could be recognised in isolation.

3.3.3 Procedure

The procedure was identical to the one of Experiment 1.

3.3.4 Results and discussion

Participants produced 1477 correct responses, 63 incorrect responses, and no time outs for the target words (mean RT from the words' uniqueness points: 356.59 ms).

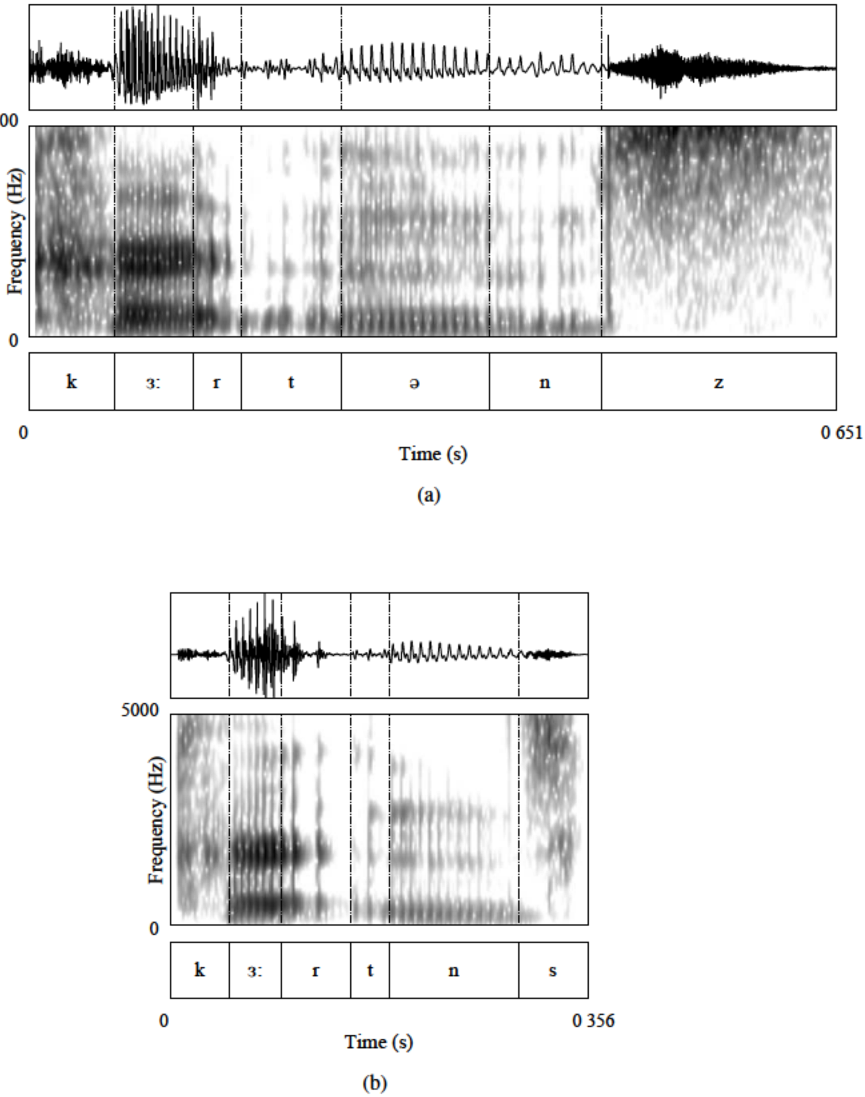

Figure 3.3: Spectrograms and transcriptions for the unreduced (a) and reduced (b) realisation of the word *curtains*.

Stimulus type	Average durational reduction	Range	Average segmental reduction	Range
Target	0.09	0.02-0.18	0.05	0-0.33
Prime	0.09	0.04-0.16	0.03	0-0.40
Filler	0.09	0.02-0.27	0.04	0-0.40
Pseudoword	0.08	0.03-0.16	0.02	0-0.25

Table 3.2: Descriptive statistics for the degree of durational and segmental reduction of the stimuli.

We analysed and compared the number of errors in Experiments 1 and 2 by means of a linear mixed-effects model with the binomial link function (Jaeger, 2008) with the dependent variable *correctness* (correct/incorrect), including the same random and fixed variables as in the analyses for the RTs in Experiment 1, in addition to the fixed variable *register* (indicating whether the primes and targets were reduced or unreduced). We only found an effect for *register*: Participants in Experiment 2 produced more incorrect responses than the participants in Experiment 1 (4% of the reduced target words versus 1% of the unreduced target words; $\beta = -1.255, F(1, 2934) = 18.44, p < .0001$). This finding suggests that although the reduced words in our experiments were produced in isolation and were mildly reduced, they were nevertheless more difficult to understand in isolation than the unreduced variants.

We analysed the RTs for the correct responses, using the same predictors as for Experiment 1, in addition to the fixed variable *register*. After filtering the data, using the procedure described in the Results and discussion section in Experiment 1, the data set consisted of 1359 trials. A summary of the results is provided in Table 3.3.

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	9.44	-	$< .0001$
trial number	-0.0005	40.85	$< .0001$
previous RT	0.224	70.28	$< .0001$
target word duration	-0.793	39.01	$< .0001$
target word frequency	-0.041	4.93	$< .05$
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.060	224.81	$< .0001$
word	0.045	143.69	$< .0001$

Table 3.3: Results for the statistical analysis of the logged RTs in Experiment 2.

The control variables showed the same effects as in Experiment 1. More importantly, we found that, in contrast to Experiment 1, Experiment 2 showed no effects of *lsa*, and only a main effect of *target word frequency*.

The results of these two experiments show interesting differences between the roles of semantic context in the recognition of reduced and unreduced pronunciation variants. In Experiment 1, we found that a higher semantic relatedness with the previous word facilitated the recognition of words with a low frequency, while it hindered the recognition of words with a very high frequency. In Experiment 2, we did not find any effects of semantic relatedness, although the effect of word frequency was omnipresent. In order to test whether the differences between the two experiments attain statistical significance, we fitted a regression model to the combined data sets of Experiments 1 and 2, including the same predictors as for the separate analyses, in addition to the new predictor *register*.

In this analysis, we could not include *target word duration* because this variable now showed a bimodal distribution (i.e. our reduced stimuli were much shorter than our unreduced stimuli). The variables *previous RT* and *register* were highly correlated, and we therefore orthogonalised these two variables. We fitted a simple linear regression model with the dependent variable *previous RT* and the predictor *register* and we used the residuals of this model in our regression analysis, instead of the raw previous RTs. We will only report main effects of and interactions with *register*. First, we found an interaction between *word* and *register* ($\chi^2 = 122.08, p < .0001$), indicating that certain words showed larger effects of reduction than others. Second, we found a three-way interaction between *register*, *lsa*, and *target word frequency* ($\beta = 2.006, F(1, 2588) = 4.03, p < .05$), which confirms the difference in the effects of semantic context in interaction with word frequency in the two experiments.

Experiments 1 and 2 contained either unreduced (Experiment 1) or reduced (Experiment 2) pronunciation variants, which meant that the speech register of the prime and target were always identical. As a consequence, we cannot determine whether the differences in the effects of semantic context between the two experiments are due to the speech register of the prime, the target, or both. We conducted a third auditory lexical decision experiment, in which we crossed the register of the prime (reduced or unreduced) with the register of the target word (reduced or unreduced).

3.4 Experiment 3

3.4.1 Participants

Forty-eight native speakers of English from the same participant pool as used in Experiments 1 and 2 were paid to take part in the experiment. They had not participated in the previous experiments.

3.4.2 Materials

We used the materials of both Experiment 1 (i.e. unreduced speech) and Experiment 2 (i.e. reduced speech). We crossed the reducton of the prime with the reduction of the target. We created four versions of each of the three pseudo-randomised lists, such that we had all possible combinations of the speech register of the prime and target for all prime-target pairs (i.e. a reduced prime followed by a reduced target, a reduced prime followed by an unreduced target, an unreduced prime followed by a reduced target, and an unreduced prime followed by an unreduced target). We made sure that all combinations (reduced-reduced, reduced-unreduced, unreduced-reduced, and unreduced-unreduced) occurred equally often in each randomisation list (19 pairs for each combination). Moreover, the filler words in each list were equally often unreduced as reduced. There were four participants per list.

3.4.3 Procedure

The procedure was identical to those of the previous experiments.

3.4.4 Results and discussion

Participants produced 3600 correct responses, 85 incorrect responses, and 11 time outs for the target words. In order to analyse the errors produced for the target words, we fitted a linear mixed-effects model with the binomial link function (Jaeger, 2008), using the same random and fixed variables as for the combined analysis of Experiments 1 and 2, in addition to the variables that are described more extensively below for the analysis of the RTs. Participants produced more incorrect responses for reduced than for unreduced targets (3% of the reduced target words versus 0.3% for the unreduced target words; $\beta = -2.743, F(1, 3635) = 47.05, p < .0001$).

Reduction of the target	Reduction of the prime	Mean RT (ms)
unreduced	unreduced	270.94
unreduced	reduced	299.52
reduced	unreduced	473.41
reduced	reduced	450.42

Table 3.4: Mean response times measured from the words' uniqueness points (excluding time-outs) for unreduced/reduced target words preceded by unreduced/reduced primes.

The mean RTs for the correct responses are provided in Table 3.4. We again analysed these RTs by means of linear mixed-effects modelling, including the same predictors as for the combined analysis of Experiments 1 and 2, except that the variable

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	6.872	-	< .0001
trial number	-0.001	101.83	< .0001
previous RT	0.268	230.79	< .0001
target word frequency	-0.092	1.97	n.s.
lsa	-1.235	2.23	n.s.
register of the prime (unreduced)	1.400	0.99	n.s.
register of the target (unreduced)	-0.381	188.71	< .0001
target word frequency : lsa	0.125	7.15	n.s. [1]
target word frequency : register of the prime (unreduced)	-0.197	0.87	n.s.
lsa: register of the prime (unreduced)	-1.940	2.11	n.s.
register of the target (unreduced) : register of the prime (unreduced)	-0.229	51.08	< .0001
target word frequency : lsa : register of the prime (unreduced)	0.293	13.22	< .001
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.046	280.51	< .0001
word	0.048	204.29	< .0001
word : register of the target (unreduced)	0.080	131.64	< .0001

Table 3.5: Results for the statistical analysis of the logged RTs in Experiment 3.

register was now replaced by *register of the target* and *register of the prime*, which indicated the speech registers of the target and of the preceding word (reduced or unreduced), respectively. Instead of the raw previous RTs, we used the residuals of a model predicting *previous RT* as a function of *register of the prime* because these variables were highly correlated. After filtering the data, using the procedure described in the Results and discussion section in Experiment 1, the data set consisted of 3060 trials. A summary of the statistical results is provided in Table 3.5.

The two control variables (*trial number* and *previous RT*) showed the same effects as in the previous experiments. More interestingly, participants responded more quickly to unreduced target words, although the exact effects of *register of the target* differed across words. This interaction was also found in subsequent analyses, but we will not mention it again and only list it in the tables. Furthermore, this effect of the *register of the target* was stronger after unreduced primes.

More importantly for our research question, we found a three-way interaction between *register of the prime*, *lsa*, and *target word frequency*. In order to interpret this three-way interaction, we split up the data by *register of the prime*.

[1] This effect was only significant in the analysis of variance results; not in the summary of the model.

We first analysed the target words with unreduced primes. We only included predictors and interactions that were significant in the regression model shown in Table 3.5. A summary of the statistical results is provided in Table 3.6. We found the same main effects and interactions as in Table 3.5, including an interaction between *lsa* and *target word frequency*. This interaction is shown in Figure 3.4. Semantically related primes appear only beneficial for target words with a low word frequency. For target words with a higher frequency, semantic priming effects are marginal, while for the words in the highest frequency range there is a reverse effect of semantic relatedness. These findings are very similar to our findings in Experiment 1, in which all words, including the primes, were unreduced.

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	6.905	-	< .0001
trial number	-0.001	47.53	< .0001
previous RT	0.275	149.18	< .0001
target word frequency	-0.318	2.66	n.s.
lsa	-3.474	0.96	n.s.
register of the target (unreduced)	-0.597	218.09	< .0001
target word frequency : lsa	0.460	13.58	< .001
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.041	86.00	< .0001
word	0.056	77.03	< .0001
word: register of the target (unreduced)	0.083	42.58	< .0001

Table 3.6: Results for the statistical analysis of the logged RTs for target words preceded by unreduced primes in Experiment 3.

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	4.088	-	< .0001
trial number	-0.001	57.55	< .0001
previous RT	0.313	95.63	< .0001
register of the target (unreduced)	-0.372	94.53	< .0001
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.048	125.82	< .0001
word	0.046	92.01	< .0001
word : register of the target (unreduced)	0.072	44.65	< .0001

Table 3.7: Results for the statistical analysis of the logged RTs for target words preceded by reduced primes in Experiment 3.

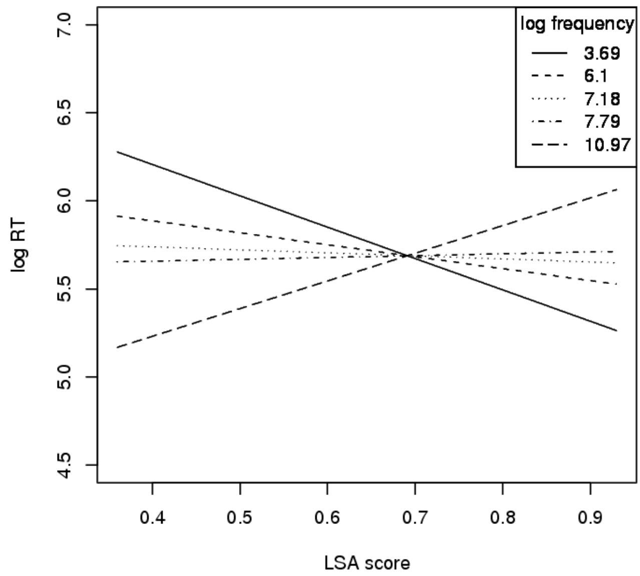

Figure 3.4: The combined effects of *lsa* and *target word frequency* on log RT for the target words preceded by unreduced primes in Experiment 3.

Subsequently, we analysed the target words with reduced primes. A summary of the statistical results is provided in Table 3.7. We did not find an interaction between *register of the target* and *previous RT*. More importantly, we did not find any effects of *lsa* and *target word frequency*, which suggests that, in contrast to unreduced primes, reduced primes hardly influenced the recognition of upcoming semantically related words.

Experiment 3 thus suggests that the absence of semantic context effects in Experiment 2 was due to the reduction of the primes rather than the targets. This finding indicates that the semantically related words are hardly activated by reduced primes. Unreduced primes, on the other hand, can influence the recognition of both unreduced and reduced targets.

Since reduced primes influenced the recognition of upcoming words to a lesser extent than unreduced primes, reduced pronunciation variants appear to be less deeply processed at the point at which participants made their lexical decisions for the following words. Experiment 4 tested whether reduced pronunciation variants are permanently processed less deeply, or whether they are processed as deeply as unreduced variants, but later in time.

3.5 Experiment 4

3.5.1 Participants

Forty-eight native speakers of English, from the same subject pool as those of the previous experiments, received course credits to take part in the experiment. They had not participated in the previous experiments.

3.5.2 Materials

The materials were identical to those of Experiment 3, except that we now used the versions of only two randomisation lists instead of three. As a consequence, there were six participants per list.

3.5.3 Procedure

The procedure was identical to those of the previous experiments, except for that we now used a fixed interstimulus interval of 2500 ms, which is on average 500 to 600 ms longer than the interstimulus interval for the trials analysed from the previous experiments, which was 1000 ms (the time between a button press and the onset of the next stimulus) + 1000 ms (the average RT from stimulus onset for unreduced words) or + 900 ms (the average RT from stimulus onset for reduced words).

3.5.4 Results and discussion

Participants produced 3543 correct responses, 102 incorrect responses, and 51 time outs for the target words. We fitted a linear mixed-effects model with the binomial link function (Jaeger, 2008) to analyse the targets, using the same random and fixed variables as for the analysis of Experiment 3. Again, we only observed a main effect for *register of the target*: Participants produced more incorrect responses for reduced than for unreduced targets (4% of the reduced target words versus 1% of the unreduced target words; $\beta = -1.947, F(1, 3597) = 69.56, p < .0001$).

Reduction of the target	Reduction of the prime	Mean RT (ms)
unreduced	unreduced	201.49
unreduced	reduced	223.13
reduced	unreduced	388.23
reduced	reduced	363.83

Table 3.8: Mean response times measured from the words' uniqueness points (excluding time-outs) for unreduced/reduced target words preceded by unreduced/reduced primes.

The mean RTs for the correct responses are provided in Table 3.8. We again analysed these RTs by means of linear mixed-effects models, including the same predictors as for the analysis of Experiment 3. After filtering the data, using the procedure described in the Results and discussion section in Experiment 1, the data set consisted of 3129 trials. A summary of the statistical results is provided in Table 3.9.

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	7.309	-	< .0001
previous RT	0.199	133.17	< .0001
target word frequency	-0.184	0.50	n.s.
lsa	-2.281	1.23	n.s.
register of the prime (unreduced)	-0.089	1.09	n.s.
register of the target (unreduced)	-0.583	144.43	< .0001
lsa : register of the prime (unreduced)	0.243	1.02	n.s.
lsa : register of the target (unreduced)	0.113	0.84	n.s.
register of the prime (unreduced) : register of the target (unreduced)	0.360	26.59	< .0001
target word frequency : lsa	0.275	4.68	< .05
lsa : register of the prime (unreduced) : register of the target (unreduced)	-0.811	9.95	< .01
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.030	227.30	< .0001
participant : register of the target (unreduced)	0.021	62.75	< .0001
word	0.065	324.26	< .0001
word : register of the target (unreduced)	0.133	195.10	< .0001

Table 3.9: Results for the statistical analysis of the logged RTs in Experiment 4.

Importantly for our research question, we obtained two interactions with *lsa*. We found a two-way interaction between *lsa* and *target word frequency*: While a higher LSA score elicited faster responses to low frequency words, we found inhibition for words in the highest word frequency range (see Figure 3.5). In contrast to Experiment 3, this interaction between semantic relatedness and lexical frequency was significant both for targets preceded by reduced and unreduced primes (rather than only for targets preceded by unreduced primes).

In addition, we found a three-way interaction between *lsa*, *register of the prime*, and *register of the target*. In order to interpret this three-way interaction, we split the data by *register of the prime*.

First, we analysed the targets preceded by unreduced primes. We only included predictors and interactions that were significant in the regression model shown in Ta-

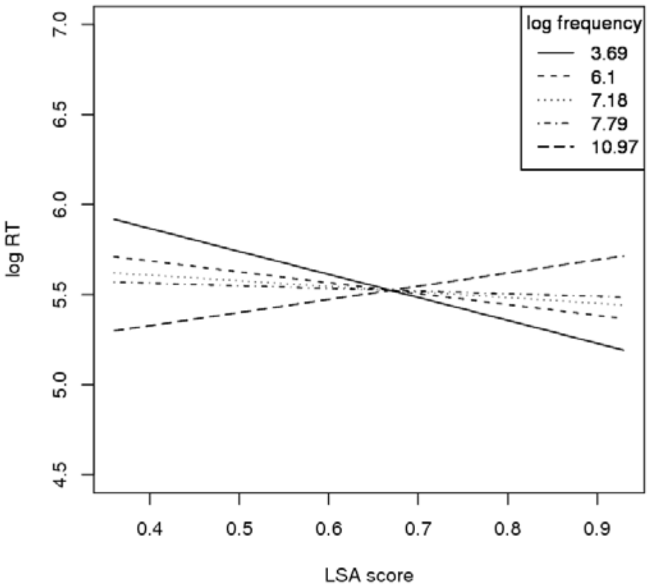

Figure 3.5: The combined effects of *lsa* and *target word frequency* on log RT in Experiment 4.

ble 3.9. A summary of the statistical results is provided in Table 3.10. All the main effects and interactions in Table 3.9 remained significant, including the interaction between *register of the target* and *lsa*, depicted in Figure 3.6. We show in these figures the RTs for words of a low (the minimum word frequency of our target words), intermediate (the mean word frequency of our target words), and high frequency (the maximum word frequency of our target words) separately, since we know that the effect of *lsa* is modulated by *target word frequency*. This figure shows that there is stronger semantic priming for unreduced than for reduced targets with low or intermediate frequencies, while there is weaker semantic interference for unreduced than for reduced targets in the highest word frequency range. We will come back to this interaction in the General discussion.

Subsequently, we analysed the target words preceded by reduced primes. A summary of the statistical results is provided in Table 3.11. The results are very similar to those of the analysis of the targets preceded by unreduced primes, except that we did not find an interaction between *lsa* and *register of the target*. This suggests that after hearing a reduced prime there were similar semantic priming effects for unreduced and reduced targets.

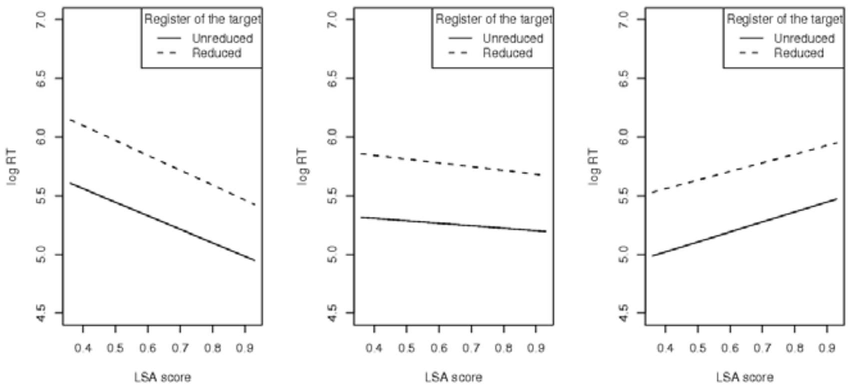

Figure 3.6: The effects of *lsa* on log RT for unreduced and reduced targets with low (left), intermediate (middle), and high (right) word frequencies, preceded by unreduced primes in Experiment 4.

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	6.689	-	< .0001
previous RT	0.231	85.67	< .0001
target word frequency	-0.293	0.63	n.s.
lsa	-3.080	0.85	n.s.
register of the target (unreduced)	-0.327	438.11	< .0001
target word frequency : lsa	0.419	6.59	< .05
register of the target (unreduced) : lsa	-0.477	5.72	< .05
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.023	63.97	< .0001
participant : register of the target (unreduced)	0.010	11.20	< .001
word	0.070	156.43	< .0001

Table 3.10: Results for the statistical analysis of the logged RTs for target words preceded by unreduced primes in Experiment 4.

Experiments 3 and 4 show some important effects of the interstimulus interval on the semantic context effects induced by unreduced and reduced primes: When processing time is limited (Experiment 3), participants show semantic context effects, modulated by *target word frequency*, only for unreduced primes, whereas an extended processing time (Experiment 4) led to such semantic effects for both unreduced and reduced primes. Further, we found that with extended processing time, semantic context effects induced by unreduced primes was smaller for reduced than

Predictor: Fixed effects	$\hat{\beta}$	F	p
Intercept	5.856	-	< .0001
previous RT	0.249	59.83	< .0001
target word frequency	-0.205	0.74	n.s.
lsa	-2.419	0.41	n.s.
register of the target (unreduced)	-0.460	141.32	< .0001
target word frequency : lsa	0.314	4.82	< .05
Predictor: Random effects	**Variance explained**	χ^2	p
participant	0.024	101.50	< .0001
participant : register of the target (unreduced)	0.041	33.80	< .0001
word	0.056	109.84	< .0001

Table 3.11: Results for the statistical analysis of the RTs for target words preceded by reduced primes in Experiment 4.

for unreduced targets. In order to test whether these differences attain statistical significance, we fitted two final regression models, one for targets preceded by unreduced primes and one for targets preceded by reduced primes, to the combined data sets of Experiments 3 and 4. We analysed targets preceded by unreduced and reduced primes separately to simplify the analyses (i.e. to avoid having to test for four-way interactions). We included the same predictors as for the separate analyses, except for *register of the prime*. Further, we added the fixed variable *Experiment* (Experiment 3 or 4). We will only report interactions with *Experiment*.

First of all, we analysed the targets preceded by unreduced primes. We found a main effect of *Experiment* ($\beta = 0.341, F(1, 3066) = 5.59, p < .05$) and an interaction between *Experiment* and *previous RT* ($\beta = -0.077, F(1, 3066) = 9.19, p < .01$), indicating that listeners responded more slowly in Experiment 4, except when the response to the preceding trial was fast. Further, we found an interaction between *Experiment*, *lsa*, and *register of the target*, which indicates that the interaction between *lsa* and *register of the target* was only present in Experiment 4. Subsequently, we analysed the targets preceded by reduced primes. We found a three-way interaction between *Experiment*, *lsa*, and *target word frequency* ($\beta = 0.199, F(1, 2961) = 5.73, p < .05$), indicating that the interaction between *lsa* and *target word frequency* was present only after longer processing time.

In summary, our results indicate that acoustically reduced pronunciation variants can induce semantic priming effects (in interaction with lexical frequency), but only if there is more time available to process these reduced variants.

3.6 General discussion

Previous research has shown that listeners need contextual resources to understand reduced pronunciation variants (Ernestus et al., 2002). It is unclear, however, which resources play a role in the understanding of reduced speech. If we assume that reduced speech is an adverse listening condition, previous research suggests that listeners pay more attention to any type of information, including semantic contextual information, than in the comprehension of clear speech (Hawkins and Smith, 2001). However, research by Aydelott and Bates (2004) and Aydelott et al. (2006) suggests that for one particular adverse listening condition, listening to low-pass filtered speech, the role of semantic contextual information is marginal, although these studies could not ascertain whether listeners actually understood the preceding semantic context. Further, Andruski et al. (1994) have demonstrated that smaller VOT distinctions may reduce semantic priming at short interstimulus intervals (50 ms) but not at longer intervals (250 ms) for English listeners. The role of semantic context in adverse listening conditions is thus unknown, including its role in the processing of reduced speech. The present study directly compared the role of semantic (separately from syntactic) information in the processing of unreduced and reduced speech. As previous research indicates that semantic priming is stronger for words with a lower word frequency (e.g. Becker, 1979; Van Petten and Kutas, 1990; Rayner et al., 2004), we investigated the interaction between the semantic relatedness of a word with its preceding word and word frequency.

We reported four simple auditory lexical decision experiments with implicit semantic priming in English, in which listeners had to make a lexical decision for each word they heard (which were all nouns). The semantic relatedness of the target words with their primes (presented in immediate succession) ranged from mildly related to highly related (only filler pairs were unrelated) as the present study focuses on listeners' sensitivity to differences in the extent to which words are semantically related, instead of differences between semantically highly related and unrelated words. The semantic relatedness of words was estimated by means of Latent Semantic Analysis (LSA; Deerwester et al., 1990). Listeners heard only unreduced pronunciation variants in Experiment 1, only reduced variants in Experiment 2, and both in Experiments 3 and 4. We recorded and presented all words in isolation (instead of in sentences) in order to carefully control the semantic relatedness of each word with its preceding word and to exclude influences from other higher level information (e.g. syntax and pragmatics). As a natural consequence, the stimuli in our experiments were only mildly reduced compared to the highly reduced forms that can occur in sentence medial positions in actual spontaneous speech, and they could be recognised in isolation.

First of all, we investigated, as a baseline, the role of semantic information in the processing of unreduced speech. For words with frequencies ranging from 0.5 to 85 per million (based on Davies, 2008), covering the typical range from low to high frequency words according to Carroll (1967), we found the same semantic effects that have been documented before (e.g. Becker, 1979; Van Petten and Kutas, 1990; Rayner et al., 2004), that is, semantic priming for target words with a low word frequency, and no priming for target words with a high word frequency.

Several accounts have been proposed for the facilitatory effects of semantic information in human speech processing. Some researchers propose that semantic information facilitates word recognition by allowing listeners to reduce lexical search space (e.g. McClelland and Elman, 1986). In these models, the activation of a word spreads to related words. Other researchers suggest that the semantic congruency between a word and its preceding context facilitates semantic integration (e.g. Van Petten and Kutas, 1990).

Surprisingly, we obtained very different results for words with frequencies in our highest word frequency range (from 94 to 686.15 per million). These words show semantic interference, which has not been reported in the literature to our knowledge. This interference effect was found each time we investigated the role of semantic priming after unreduced primes (i.e. in Experiments 1, 3, and 4). Additional analyses showed that these interference effects were not restricted to a couple of atypical word pairs, but they were based on ten different word pairs in our experiments.

The absence of this interference effect in the literature may be explained by the fact that most studies (e.g. Becker, 1979; Van Petten and Kutas, 1990; Rayner et al., 2004) investigated few very high frequency words that are highly related to their preceding words, as they collapsed over various degrees of semantic relatedness. For example, Becker (1979) performed a visual lexical decision experiment with semantically highly related, moderately related, and unrelated word pairs. Importantly, the example provided for the highly related word pairs in this study (*freezing-cold*) is only mildly related (lsa score: 0.43) compared to the highly related target word pairs in our study. Hence, that study collapsed over words with mildly related to highly related preceding words, which explains why this study, and most other studies reported in the literature, could not detect the semantic interference effect found in the present study. Studies are necessary that explicitly investigate the comprehension of high frequency words preceded by semantically highly related words to see whether our results can be replicated using different word pairs, preferably also in different languages.

The question arises how these interference effects can be integrated into existing psycholinguistic models of speech comprehension. The interference can be explained in connectionist models of speech comprehension like TRACE (McClelland and Elman, 1986). In these models, high frequency words have relatively high resting activations and (as mentioned above) semantic priming is explained as the spreading of activation of words to semantically related words, which consequently can be recognised more quickly. This implies that, before the recognition of the prime words, semantically related words with very high resting activation levels may get activation levels that are higher than the activation levels of the prime words due to this spreading of activation, which would inhibit the recognition of these prime words. These highly frequent words therefore need to be suppressed before listeners can recognise the prime words. Apparently, this suppression can lead to activation levels that are lower than the words' resting activation levels.

Subsequently, we investigated semantic context effects in the processing of reduced speech. In Experiment 2, where listeners were presented with only reduced words, we did not find any semantic context effects. In Experiment 3, we investigated whether these marginalised semantic context effects for reduced speech were due the reduction of the primes, the reduction of the targets, or both. In this experiment, listeners heard both unreduced and reduced words. We found semantic effects after unreduced, but not after reduced primes, regardless of whether the target was reduced. These results suggest that semantic information in reduced words plays a smaller role, compared to unreduced words, in the recognition of upcoming related words. Hence, it seems that semantic effects are attenuated if the preceding semantic context is more difficult to process.

These findings raise the question why semantic priming effects are smaller for reduced primes. On the one hand, reduced pronunciation variants may not activate semantically related words as well as do unreduced pronunciation variants. Alternatively, it may take longer before reduced pronunciation variants activate semantically related words and semantic effects only emerge if there is ample time to process the reduced variants.

In Experiment 4, listeners had more time (500 or 600 ms longer on average, depending on the speech register of the prime) to process the prime before we presented the target. We found semantic effects (in interaction with word frequency) from unreduced as well as reduced primes. This suggests that the activation of semantically related words takes more time for reduced speech, but eventually semantic effects from reduced speech have a similar magnitude as from unreduced speech.

Interestingly, the results of Experiment 4 suggest that, after a longer interstimulus interval, unreduced primes show stronger semantic context effects for unreduced targets than for reduced targets. A possible explanation may be that a reduced pronunciation is somewhat unexpected after an unreduced word, since in natural situations completely unreduced words, like the ones in our experiment, tend to be surrounded by other clearly pronounced words (e.g. in formal conversations). Reduced words, in contrast, tend to be surrounded by words of all types of reduction degrees. This difference may only appear after longer processing time, when listeners have had more time to fully process the primes.

What do these observations reveal about the role of semantic information in the processing of spontaneous speech? They suggest that semantic information plays a role in the processing of both unreduced and reduced speech if there is sufficient time to fully process the words that contain this semantic information, including access to the semantic entries of these words in the lexicon. Since semantically related words (in particular nouns) are often separated by other words (e.g. function words) in many languages, there is probably sufficient time to fully process reduced variants in most cases. These findings predict that whenever semantically related words occur in immediate succession, these words are less reduced, or listeners resort to different contextual resources. For example, listeners may rely more heavily on acoustic information in the context of reduced words.

The present study serves only as a starting point for establishing the role of semantic context in the processing of spontaneous speech. We tested the recognition of isolated words in the context of other isolated words. Follow-up research is required to establish whether semantic information also influences the processing of reduced words embedded in natural sentences, since it has been suggested that, at least in cross-modal priming, associative priming effects may be smaller under such conditions (e.g. Norris, Cutler, McQueen, and Butterfield, 2006).

Further, the present study includes various types of pronunciation variation, including durational reduction and segmental reduction. Previous research by Aydelott-Utman, Blumstein, and Burton (2000) suggests that, at least in identity priming, durational reduction inhibits priming, while syllable structure variation (e.g. [pəˈlis] versus [ˈplis] for *police*) may actually enhance priming. More research is required to investigate whether all types of variation influence semantic priming similarly.

In summary, the present study illustrates that the effects of a word's semantic relatedness with its preceding word occur later in the processing of reduced speech than in the processing of unreduced speech. This finding suggests that semantic information in reduced words only facilitates the recognition of upcoming words if there is sufficient processing time. Since reduced words are often separated by other words, listeners probably benefit as much from semantic information in their processing of reduced speech as they do in their processing of unreduced speech.

3.7 Appendix

The semantically related word pairs tested in this study and their LSA scores:

1. axe-knife 0.52
2. balloon-airplane 0.53
3. basement-stairs 0.59
4. beard-eyebrows 0.65
5. bees-honey 0.81
6. blanket-pillow 0.54
7. boat-dock 0.77
8. book-pages 0.62
9. borders-country 0.51
10. bottle-glass 0.55
11. candy-chocolate 0.62
12. cat-mouse 0.72
13. chair-room 0.75
14. church-pope 0.75
15. cloud-drops 0.65
16. coasts-sea 0.66
17. collar-dress 0.53
18. cousin-aunt 0.68
19. curtains-window 0.76
20. dance-ballet 0.81
21. dishes-kitchen 0.75
22. driver-car 0.79
23. farm-crop 0.72
24. feather-eagle 0.66
25. fire-chimney 0.63
26. game-player 0.88
27. gold-silver 0.88
28. hook-fish 0.66
29. horn-bull 0.4
30. household-wives 0.57
31. house-porch 0.65
32. huts-village 0.72
33. ice-glaciers 0.84
34. idea-topic 0.64
35. island-mainland 0.77
36. jacket-shirt 0.72
37. king-princess 0.64
38. knee-elbow 0.74
39. lid-box 0.57
40. lock-door 0.7
41. milk-cheese 0.82
42. month-receipts 0.63
43. morning-breakfast 0.79
44. music-concert 0.8
45. nurses-patients 0.76
46. ocean-shelf 0.55
47. paper-ink 0.72

48. parade-crowds 0.58
49. party-birthday 0.43
50. peace-war 0.76
51. peaks-mountain 0.9
52. pie-butter 0.66
53. pollen-flower 0.88
54. presents-christmas 0.45
55. rabbit-fox 0.64
56. river-flood 0.67
57. saddle-horse 0.93
58. school-junior 0.59
59. sister-baby 0.37
60. sky-mist 0.57
61. snake-beak 0.36
62. spoon-bowl 0.6
63. storm-lightning 0.69
64. story-plot 0.78
65. sunrise-noon 0.51
66. symbols-map 0.58
67. teacher-classroom 0.9
68. team-coach 0.89
69. tooth-dentist 0.89
70. tower-castle 0.5
71. town-streets 0.58
72. voice-tone 0.61
73. water-bucket 0.48
74. wax-plaster 0.52
75. web-spider 0.77
76. wheel-gears 0.75
77. wine-beer 0.85

Semantic facilitation in bilingual everyday speech comprehension

Chapter 4

This chapter has been reformatted from:
Marco van de Ven, Benjamin V. Tucker, and Mirjam Ernestus. Semantic facilitation in bilingual everyday speech comprehension. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, pages 1245-1248.

Abstract

Previous research suggests that bilinguals presented with low and high predictability sentences benefit from semantics in clear but not in conversational speech (Bradlow and Alexander, 2007). In everyday speech, however, many words are not highly predictable. Previous research has shown that native listeners can use also more subtle semantic contextual information (Van de Ven, Tucker, and Ernestus, 2009). The present study reports two auditory lexical decision experiments investigating to what extent late Asian-English bilinguals benefit from subtle semantic cues in their processing of English unreduced and reduced speech. Our results indicate that these bilinguals are less sensitive to semantic cues than native listeners for both speech registers.

4.1 Introduction

In casual speech, words are often much shorter and have fewer segments than in clear speech (Johnson, 2004; Ernestus, 2000). For example, the English word *apparently* and the Dutch word *eigenlijk* 'actually' can be pronounced like [ˈpʰeɾɪ̃] and [ˈɛik], respectively. These types of acoustically reduced pronunciation variants are highly frequent in casual speech. To illustrate, in an earlier version of the Buckeye Corpus of American English conversational speech (88,000 word tokens), complete syllables are absent in six percent of the words (Johnson, 2004), and in the Ernestus corpus of casual Dutch (91,718 word tokens; Ernestus, 2000), ca. 19.4% of the syllables are missing (Schuppler, Ernestus, Scharenborg, and Boves, to appear).

Native listeners have difficulty understanding these highly reduced pronunciation variants in isolation (Ernestus, Baayen, and Schreuder, 2002). Previous research indicates that they can use both acoustic and semantic/syntactic cues in the context to deduce reduced words, but they favour acoustic cues whenever these two are in conflict (Van de Ven, Tucker, and Ernestus, 2009; see also Chapter 2 of this dissertation). Further, they benefit more from semantic contextual information when listening to unreduced speech compared to reduced speech (Van de Ven et al., 2009).

Whereas several studies have thus shed some light on how native listeners deal with reduced pronunciations, it is unclear how non-native listeners do so. The present study investigates whether late Asian-English bilinguals (native speakers of an Asian language who acquired English after childhood) can make use of subtle semantic cues in their processing of unreduced and reduced pronunciation variants. We selected Asian-English bilinguals because Asian languages and the English language are typologically highly distinct and share few cognates, which implies that semantic priming only occurs if the speakers know the English words (instead of just the cognates in their native language).

Previous psycholinguistic research suggests that the role of semantic contextual information is highly similar in the processing of bilinguals' native (L1) and non-native (L2) languages (Hahne and Friederici, 2001).

However, other studies have shown that there are also crucial differences between native and bilingual listeners in their use of semantic contextual information. For example, when listening to speech in noise, non-native listeners benefit less from semantic/syntactic contextual information (Mayo, Florentine, and Buus, 1997). Similarly, bilinguals only benefit from semantic cues in the context when presented with clear speech (rather than conversational speech; Bradlow and Alexander, 2007).

Importantly, these studies are based on sentences, which means that the role of semantic contextual information cannot be separated from other types of higher level information (e.g. syntax and pragmatics). Further, these studies tested bilinguals' processing of high predictability (e.g. *A bicycle has two wheels.*) versus low predictability (e.g. *Dad looked at the pork.*) sentences. However, in everyday speech comprehension, many words are not highly predictable, and listeners often have to resort to much more subtle semantic contextual information.

As mentioned above, previous research has shown that native listeners use subtle semantic contextual information if they are presented with unreduced pronunciation variants, whereas they use this information to a smaller extent if presented with reduced variants (Van de Ven et al., 2009). The present study investigates whether bilinguals are as sensitive to subtle semantic information as native listeners in their processing of unreduced and reduced pronunciation variants. On the basis of previous findings (e.g. Bradlow and Alexander, 2007; Van de Ven et al., 2009; Mayo et al., 1997) we predict that monolingual and bilingual listeners show similar sensitivity to semantic cues in the processing of unreduced speech, but no sensitivity to semantic cues in the processing of reduced speech.

We report two auditory lexical decision experiments with semantic priming, in which bilinguals were presented with unreduced (Experiment 1) or reduced (Experiment 2) pronunciation variants. Similar to Van de Ven et al. (2009), listeners were presented with isolated words, rather than with sentences, in order to isolate the effects of semantic facilitation from the influences of other sources of higher level information.

Since this study focuses on the role of subtle semantic cues, we needed a fine-grained measure of words' semantic relatedness. For this purpose we used Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnas, Landauer, and Harshman, 1990), which is a computational technique that estimates words' semantic relatedness on the basis of the numbers of direct and indirect co-occurrences of words in paragraphs or texts, extracted from a large collection of written corpora. LSA scores range from -1 to 1, where values closer to -1 indicate a low and values closer to 1 indicate a high semantic relatedness between words. Previous research has shown that LSA scores predict semantic priming effects obtained in auditory lexical decision tasks for native listeners (Van de Ven et al., 2009; Landauer and Dumais, 1997; Landauer, Foltz, and Laham, 1998).

4.2 Experiment 1

4.2.1 Introduction

We investigated whether a word's semantic relatedness to its preceding stimulus influences lexical decision times differently for native listeners (i.e. the control group) than for bilinguals, by replicating Experiment 1 of Van de Ven et al. (2009), identical to Experiment 1 in Chapter 3 of this dissertation, with bilingual speakers. This experiment contains word pairs with LSA scores ranging from 0.36 to 0.93, which includes word pairs with members that are mildly, moderately, and highly related. The members of each word pair were presented in consecutive trials.

4.2.2 Participants

Twenty undergraduate students from the University of Alberta, Department of Linguistics participant pool took part in the experiment, and received course credit for their participation. They were all late Asian-English bilinguals (i.e. they acquired English after childhood) whose native languages were Mandarin (n = 11), Cantonese (n = 7), Korean (n = 1), and Japanese (n = 1). They had passed a TOEFL (Test of English as a Foreign Language) examination with a minimum score of 580 (or equivalent). The data from these twenty bilinguals are compared to the data from the twenty native speakers of Canadian English who were tested in Experiment 1 in Van de Ven et al. (2009).

4.2.3 Materials

The stimuli are those of Experiment 1 in Van de Ven et al. (2009). The 154 target nouns have widely varying word frequencies (range: 40-58322), which were estimated on the basis of frequency counts for the spoken section of the Corpus of Contemporary American English (385 million word tokens; Davies, 2008). These target nouns form 77 word pairs with LSA scores ranging from mildly related (e.g. *snake* - *beak*, LSA score: 0.36) to highly related (e.g. *gold* - *silver*, LSA score: 0.88) word pairs. Thus, if any effects of semantic similarity show up for the target stimuli in our experiment, these effects are the result of listeners' sensitivity to subtle distinctions in semantic relatedness. In addition, the experiment contained 87 semantically unrelated filler pairs and 128 pseudowords. The limited number of pseudowords induces a "YES"-response bias, which makes it difficult to find any semantic priming effects. Hence, if any semantic priming effects show up in the data, these are robust effects.

Existing words and pseudowords were presented in a pseudorandomised order. We used the same three randomisation lists as in Van de Ven et al. (2009), in which the order of the stimuli was manually corrected in case more than six existing words or three pseudowords occurred in succession. Further, Van de Ven et al. (2009) avoided rhyme and alliteration between words in consecutive trials.

The materials were recorded in a sound-attenuated booth by Van de Ven et al. (2009). They asked a male speaker of Canadian English to produce the words in clear citation speech, and the words were presented in a randomised order. The members of the target word pairs never occurred in immediate succession to prevent effects of words' semantic relatedness on their realisations. A different native speaker of Canadian English verified that all words were produced in a natural fashion.

4.2.4 Procedure

The procedure was identical to that of Experiment 1 in Van de Ven et al. (2009). The experiment was a self-paced auditory lexical decision task with implicit semantic priming, in which participants had to make a lexical decision, by means of a button press, for each stimulus presented. The next trial was presented 1000 ms after each button press, or after a time-out of 3500 ms. The experiment was conducted in a sound-attenuated booth at the Alberta Phonetics Laboratory. The materials were presented over closed headphones at a comfortable listening level, and the experiment lasted approximately 15 minutes.

4.2.5 Results

Together, the native speakers and bilinguals produced 12495 correct responses, 384 incorrect responses, and 241 time outs. We analysed participants' RTs (from stimulus offset) for the correct responses by means of linear mixed-effects models. As we are primarily interested in the effects of subtle semantic cues, we analysed only the RTs to the target words. All the regression models reported in this study were fitted to the subset of trials for which the response to the preceding trial was also correct. In addition, we excluded outliers for the RTs on the preceding trial to ascertain correct processing of the prime. Thus, our results cannot be explained by listeners' failure to understand the prime or by guessing strategies. Further, we removed data points for which the standardised residuals of the final model were smaller than -2.5 or larger than 2.5. We took the log of the RTs in order for them to show a normal distribution, and we used a backwards stepwise selection procedure, in which predictors and their interactions were removed if they did not attain significance at the $\alpha < .05$ level. We included the main predictors *native language* (monolingual or bilingual), *lsa* (LSA score of the word with its preceding word), *target word frequency* (log of the word frequency), *prime word frequency* (log word frequency of the preceding word), and *previous RT* (log of the previous RT, residualised for *native language*). Further, we included the fixed variables *trial number* and *word duration* (log of the stimulus duration), and the random variables *participant* and *word*, in order to reduce the variance in the data. A summary of the results is provided in Table 4.1.

To begin with, we found an interaction between target word frequency and lsa. This interaction indicates that a stronger semantic relatedness of a target word with its preceding word facilitates the recognition of this target word (e.g. *balloon* in the word pair *airplane - balloon*), but only if this target word has an intermediate or low

Predictor	$\hat{\beta}$	F	p
Intercept	12.873	-	< .0001
trial number	-0.002	22.09	< .01
target word frequency	-0.212	17.64	< .01
lsa	-2.284	1.5	n.s.
previous RT	0.278	210.28	< .0001
word duration	-0.789	39.81	< .0001
native language	-0.751	25.25	< .01
trial number : lsa	0.003	1.7	n.s.
trial number : native language	0.002	1.13	n.s.
lsa : native language	0.666	0.15	n.s.
target word frequency : lsa	0.229	4.68	< .05
trial number : lsa : native language	-0.003	5.9	< .05

Table 4.1: Results for the statistical analysis of the RTs for Experiment 1.

word frequency. For words with a very high word frequency, a stronger semantic relatedness leads to inhibition. This interaction holds for the native and bilingual listeners (for a possible explanation of this interaction, we refer to Van de Ven et al., 2009). Hence, the bilinguals show similar patterns to those of the native listeners.

Interestingly, we also found a three-way interaction between *native language*, *lsa* and *trial number*. In order to interpret this interaction, we split the data by *native language*. We found that bilingual listeners benefited less from semantic cues towards the end of the experiment, whereas there was no such attenuation for the monolingual control group.

In addition, we found shorter RTs if the response for the preceding trial was also fast, for longer words, and for words presented later in the experiment.

These results indicate that Asian-English bilinguals are highly sensitive to subtle semantic cues in English, similar to native listeners, and they use this sensitivity especially in their recognition of low frequency words. However, bilinguals apparently also had more difficulty employing these semantic cues, since they showed smaller semantic priming effects towards the end of the experiment.

This raises the question to what extent bilingual listeners can use semantic cues to process reduced speech, characterised by shorter word durations and missing segments. Previous research has shown that native listeners have difficulty using semantic cues in such adverse listening conditions (Van de Ven et al., 2009). We investigated this issue by replicating Experiment 2 of Van de Ven et al. (2009), identical to Experiment 2 in Chapter 3 of this dissertation, in which participants performed lexical decisions for acoustically reduced words, with bilingual listeners.

4.3 Experiment 2

4.3.1 Participants

Twenty undergraduate students from the University of Alberta, Department of Linguistics participant pool, who had not participated in the previous experiment, received course credit for their participation. The native languages of these bilinguals were Mandarin (n = 7), Cantonese (n = 7), Korean (n = 5), and Japanese (n = 1). They were compared to the twenty native speakers of Canadian English tested in Experiment 2 in Van de Ven et al. (2009).

4.3.2 Materials

The materials are those of Experiment 2 in Van de Ven et al. (2009). The speaker of Experiment 1 was instructed to produce the word list of Experiment 1 as quickly as possible. Once more, a native speaker of Canadian English verified that all words were produced in a natural fashion, particularly focusing on the way the speaker reduced these words. This recording procedure resulted in reduced words that were shorter than the unreduced materials used in Experiment 1 (673.71 ms versus 388.52 ms on average, two-tailed t-test: $t(864.8) = 46.58, p < .0001$). In addition, the materials contained some mild segmental reductions: On average, 0.19 segments were deleted per word compared to the realisations of the words in Experiment 1 (range: 0-2 segments per word).

4.3.3 Procedure

The procedure was identical to those of Experiment 1 in the present study and the experiments in Van de Ven et al. (2009).

4.3.4 Results and discussion

Together, the native and bilingual participants produced 12317 correct responses, 630 incorrect responses, and 173 time outs. We again analysed the participants' RTs by means of linear mixed-effects models, using the same procedure and including the same predictors as for Experiment 1. A summary of the results is provided in Table 4.2.

The results showed faster responses towards the end of the experiment, for more frequent words, if the previous response was fast, and for longer words. Further, we found a two-way interaction between *trial number* and *native language*, which indicates that bilinguals show a larger decrease in reaction times towards the end of the experiment than native listeners.

Predictor	$\hat{\beta}$	F	p
Intercept	10.196	-	< .0001
trial number	-0.0007	137.5	< .0001
target word frequency	-0.052	11.41	< .01
lsa	0.085	0.127	n.s.
previous RT	0.326	371.44	< .0001
word duration	-0.662	31.13	< .0001
native language	-0.02	7.18	n.s.
lsa : native language	-0.328	7.24	< .05
trial number : native language	0.0003	5.33	< .05

Table 4.2: Results for the statistical analysis of the RTs for Experiment 2.

More importantly, we found an interaction between *native language* and *lsa*. This interaction indicates that bilingual listeners show smaller semantic facilitation than native listeners, although this facilitation was not statistically significant for native listeners either.

These two experiments show striking differences between the processing of unreduced and reduced pronunciation variants, both by native and Asian-English bilingual listeners. We tested whether these differences attain statistical significance by fitting a regression model to the combined data set of Experiments 1 and 2, using the predictors and their interactions that were present in the final models of Experiments 1 and 2. In addition, we tested for main effects of and interactions with *experiment*. A summary of the results is provided in Table 4.3.

First of all, we found a main effect of *experiment*, which shows that listeners respond faster to unreduced than to reduced words. Interestingly, we found an interaction between *target word frequency* and *lsa*, which suggests that this interaction was also present in the data from Experiment 2, although this did not appear from the analysis of just the data for Experiment 2. Apparently, it only surfaces after increasing the sample size. More importantly for our research question, we found a two-way interaction between *native language* and *lsa*, rather than a three-way interaction with *experiment*. This finding indicates that, after combining the data sets of Experiments 1 and 2 (thereby increasing the statistical power), bilingual listeners overall show significantly smaller semantic facilitation compared to the native listeners. Finally, we found an interaction between *trial number*, *experiment*, and *native language*, which indicates that monolingual listeners especially show faster RTs compared to bilingual listeners towards the end of Experiment 1. These results indicate that bilingual listeners are generally less sensitive to subtle semantic cues than native listeners, even though they clearly use them in the processing of unreduced pronunciation variants.

Predictor	$\hat{\beta}$	F	p
Intercept	7.177	-	< .0001
trial number	-0.0007	115.52	< .0001
target word frequency	-0.165	17.59	< .01
lsa	-1.238	0.81	n.s.
previous RT	0.294	527.95	< .0001
word duration	-0.692	148.85	< .0001
experiment	-0.23	18.17	< .01
native language	-0.086	28.64	n.s.
trial number : native language	0.0003	0.006	n.s.
trial number : experiment	0.001	6.58	< .05
experiment : native language	-0.078	3.57	n.s.
lsa : native language	-0.21	4.58	< .05
target word frequency : lsa	0.177	3.91	< .05
trial number : experiment : native language	-0.0005	5.31	< .05

Table 4.3: Results for the statistical analysis of the RTs for the combined data set of Experiments 1 and 2.

Finally, our results show similar word frequency effects in L1 and L2 speech processing, in line with Imai, Walley, and Flege (2005). Our study extends the finding reported in Imai et al. (2005), by showing that these frequency effects are not only similar for unreduced, but also for reduced speech.

4.4 General discussion

The present study investigated Asian-English bilinguals' sensitivity to subtle semantic cues in the processing of unreduced and reduced pronunciation variants. We selected Asian-English bilinguals because these listeners' native and non-native languages are typologically distinct. Moreover, these languages share few cognates, which implies that semantic priming for the bilinguals will show up only if the speakers know the English words (rather than just their native language counterparts). We conducted two auditory lexical decision experiments with implicit semantic priming, in which listeners had to make a lexical decision for every word presented, and consecutive words differed in their semantic relatedness from mildly related to highly related.

First of all, bilingual listeners showed patterns that are very similar to those of native listeners. Bilinguals are sensitive to subtle semantic cues, and they use these cues especially to recognise words with a lower frequency of occurrence. Hence, bilinguals' sensitivity to semantic cues in the non-native language is not restricted to distinctions between semantically related and unrelated words, but it also includes more fine-grained information about the extent to which words in a non-native language are semantically related. Moreover, our results indicate that this sensitivity generalises to bilinguals whose L1 and L2 language are typologically distinct and share few cognates. Further research has to show whether this fine-grained information is due to direct L1-L2 transfer, or whether bilinguals maintain a different semantic network for the non-native language.

Importantly, however, we also found that bilinguals in general show smaller semantic facilitation effects than native listeners. This finding indicates that bilingual listeners have more difficulty than native listeners in using subtle semantic cues in their L2, not only when listening to conversational speech, as may be hypothesised on the basis of Bradlow and Alexander (2007), but also in their processing of clear speech. Further research is required to test whether bilinguals can benefit from semantic cues in actual face-to-face conversations, in which they can use more sources of information (e.g. the larger discourse, visual information, etc.), or whether listeners' use of such higher level information remains less effective in bilingual speech comprehension.

To conclude, our results show that bilingual listeners are sensitive to subtle semantic cues in their processing of unreduced speech in a typologically distinct non-native language. However, compared to native listeners, they have more difficulty using these cues in the processing of both reduced and clear speech, which possibly affects the extent to which they benefit from such cues in everyday listening situations.

The role of the acoustic properties of reduced words in the processing of spontaneous speech

Chapter 5

Marco van de Ven and Mirjam Ernestus. The role of the acoustic properties of reduced words in the processing of spontaneous speech.

Abstract

In casual speech, many words are pronounced with fewer segments than suggested by their citation forms. Listeners use context to recognise reduced pronunciation variants, yet many reduced variants cannot be recognised based on context alone. This study investigates the roles of the acoustic information from the first few segments and the durations of reduced pronunciation variants in the recognition of these variants.

We report three auditory gating experiments, in which participants heard the context and some segmental and/or durational information from reduced target words with unstressed initial syllables. Listeners gave too short answers if they were provided with durational information from the target words. Apparently, listeners are unaware of the reductions that can occur in spontaneous speech. More importantly, listeners required fewer segments to recognise target words if the first syllable was strongly reduced and the unstressed vowel was absent. This result shows that the word recognition process need not be hampered by the reduced pronunciation variants that occur in casual speech.

5.1 Introduction

In casual speech, words are generally realised much shorter and with less articulatory effort than is suggested by their citation forms (e.g. Ernestus, 2000; Johnson, 2004). For example, the English word *ordinary* can be pronounced like [ˈɔnri] and, likewise, the Dutch word *natuurlijk* ‘of course’ may be reduced to [ˈtyk]. These types of reduced pronunciations are ubiquitous in spontaneous speech. To illustrate, Johnson (2004) found that, in American English, segments are changed or missing in 25% and syllables are missing in 6% of the word tokens. Similarly, in Dutch, segments are changed or missing in 48% of the word tokens and syllables are missing in approximately 19% of the word tokens (Schuppler, Ernestus, Scharenborg, and Boves, 2011). The present study investigates how listeners understand these reduced pronunciation variants, focusing on the roles of different types of acoustic information in the variants themselves.

Several studies have investigated how listeners recognise reduced pronunciation variants. First of all, Ernestus, Baayen, and Schreuder (2002) tested the contribution of context to the understanding of highly reduced pronunciation variants. They found that listeners have difficulty recognising highly reduced pronunciation variants out of context (ca. 50% correct). Identification problems remained when these variants were presented together with their phonological contexts (the neighbouring vowels and intervening consonants; ca. 70% correct), but within sentence context, listeners did not have any difficulty recognising these reduced variants (more than 90% correct). These findings indicate that listeners require context to recognise highly reduced pronunciation variants.

Chapter 2 of this dissertation investigated the contribution of semantic/syntactic and acoustic information from the context to the recognition of these reduced variants. They conducted four cloze tasks, in which listeners were presented with the context of reduced pronunciation variants without these variants themselves, and they were asked to guess these missing words, choosing from four semantically and syntactically plausible options. Listeners did not choose the reduced target words in more than 50% of the trials, despite the fact that their lexical search space had been reduced to four words (the four words they had to choose from). This finding suggests that the role of context alone (i.e. without the reduced pronunciation variants themselves) in the processing of these reduced variants is quite limited.

This hypothesis was confirmed by Janse and Ernestus (2011), who reported an experiment in which participants were only presented with orthographic transcriptions of the preceding and following context of reduced pronunciation variants, and an experiment in which the participants also heard the reduced variants (but the context was again presented visually). The results showed that listeners could not identify most target words on the basis of their written contexts alone (53% correct), but the auditory presentation of the target words significantly increased participants' performance (90% correct). These findings indicate that contextual information only becomes highly informative once listeners have heard the reduced pronunciation variants.

Thus, the literature shows that listeners require both the context and acoustic information from reduced pronunciation variants to recognise these variants. The present study investigates how much and what types of acoustic information from the reduced variants (e.g. the first consonants/vowels, or perhaps word duration) are required.

We focus on the recognition of words with unstressed initial syllables, since these syllables tend to be acoustically reduced with missing segments (e.g. the Dutch verb form *verkoopt* [vər'kopt] 'sells' may be realised like [f'kopt]). Such reductions are likely to influence word recognition, since the reduction is located (far) before the word's uniqueness point (henceforth UP). The reduction may therefore increase uncertainty about the word's identity during the word recognition process. To give an example, the Dutch words *verlaten* 'leave' and *flater* 'blunder' with the citation forms [vər'latən] and ['flatər] may be pronounced like ['flatə] and ['flatər]), and this may lead to increased temporary ambiguity.

Previous research suggests that reductions do indeed hinder word recognition. Nearly all this research (see the General discussion) test listeners' comprehension of reduced words in isolation or simple constructed sentences (e.g. Ernestus and Baayen, 2007; Ranbom and Connine, 2007; Tucker, in press; Tucker and Warner, 2007; Chapter 3 of this dissertation). For example, Tucker (in press) conducted an auditory lexical decision task with American English words containing either unreduced or reduced flaps. Most of these flaps occurred in word-medial position, before the words' UPs. Tucker found longer recognition times from word offset for words with reduced rather than unreduced flaps, which shows that reductions before the UP can delay the word recognition process.

These results suggesting that reductions hamper speech comprehension are surprising given that reductions are highly frequent in spontaneous speech and given that there is no clear evidence that listeners have difficulty processing everyday conversational speech. Importantly, these studies compared the processing of reduced and unreduced words presented in isolation, or in simple short sentences, whereas in everyday situations listeners only hear reduced pronunciation variants within sentence context. Moreover, these studies investigated read speech (in which reductions normally do not occur), rather than spontaneous speech materials. This raises the question whether there are also processing disadvantages for reduced words embedded in their natural sentence context.

The various psycholinguistic models of speech comprehension propose different accounts for the processing of reduced pronunciation variants. These models can be placed on a continuum with purely abstractionist models on one end, and purely exemplar-based models on the other end of the continuum.

The finding that reduced words are more difficult to process than unreduced words is as expected by abstractionist models of speech comprehension. These models generally assume that only a single lexical representation, the word's citation form, is stored in the mental lexicon. Listeners activate a set of lexical candidates by matching the sounds that they have recognised to those in the words' citation forms, and these candidates then compete for recognition (e.g. Marslen-Wilson and Welsh, 1978; Norris, 1994). Thus, if listeners hear the first two segments of the unreduced realisation of the English word *ordinary* (i.e. [ˈɔr]), listeners activate *ordinary*, as well as the lexical competitors, such as *order* and *orca*. Abstractionist models predict that reductions inhibit word recognition especially if the first realised sounds do not match the first sounds of the word's citation form. For example, if listeners hear *ordinary* realised like [ˈɔnri], abstractionist models predict that, after hearing its first two segments ([ˈɔn]), listeners activate mostly the lexical candidates *honour* and *on*, rather than the target word *ordinary*. This leads to a lexical garden path, unless, on the basis of traces in the acoustic signal or phonotactic constraints, the missing sounds can be reconstructed (e.g. Ernestus, in press).

Exemplar-based models account for the recognition of reduced variants in a very different way. These models assume that listeners store exemplars for all pronunciation variants (e.g. Goldinger, 1998). Consequently, listeners may activate both *ordinary* and *honour* after hearing the first two segments of the reduced realisation [ˈɔnri], and reductions may actually facilitate word recognition if, as a result of these reductions, words become unique more quickly. For example, the Dutch word *verschillende* [vərˈsχɪləndə] 'different' may be realised like [fˈsχɪlnə], and therefore becomes unique after the fourth rather than the sixth segment (i.e. [ɪ]). Exemplar-based models account for the finding that listeners cannot recognise highly reduced pronunciation variants in isolation or simple constructed sentences by assuming that exemplars contain information about the contexts the pronunciation variants occurred in (Hawkins and Smith, 2001), and listeners can only recognise reduced pronunciation variants if they are embedded in a viable context. Hence, they assume that only in their natural contexts reduced words are recognised at least as easily as unreduced words.

In order to further specify both abstractionist and exemplar models, we need to know which types of information in reduced pronunciation variants listeners rely on in particular. The present study reports three auditory gating experiments, in Dutch, designed to gain more insight into this issue. In these experiments, listeners were presented with the natural (preceding and following) context of reduced pronunciation variants (henceforth target words), and some acoustic information from these variants themselves (except for the baseline condition, see below). These materials were extracted from a corpus of spontaneous speech. Each target word had an unstressed initial syllable, and the second syllable carried main stress.

In Experiment 1, we investigated the role of the first realised consonant or consonant cluster (henceforth "consonant cluster", for the sake of convenience) and the importance of the presence of the first vowel in the recognition of reduced pronunciation variants. The experiment consisted of two parts. In part one (*gate 1*), for half of the trials, participants heard the preceding and following context, separated by a square wave, without the reduced pronunciation variants themselves. In the second part of the experiment (*gate 2*), for the second half of the trials, participants heard the preceding context and the initial consonant cluster of the target word, followed by a square wave and the following context.

This initial consonant cluster consisted of only the onset consonants from the citation form if the first vowel was present in the acoustic signal, whereas it consisted also of consonants from the coda and/or the onset of the stressed syllable from the citation form if the first vowel was absent. For example, the Dutch word *principe* 'principle' with the citation form [prɪnˈsipə] was realised like [pˈsipə] in one token and like [pəˈsipə] in a different token from the experiment, and the participants in the second half of Experiment 1 heard the segments [pˈs] and [p] of these target words respectively. Initial consonant clusters contained more than the word-initial onset consonants in the citation form in almost half of the trials.

This experiment allowed us to make two important comparisons. First of all, we could compare the conditions with and without the initial consonant cluster, and this comparison shows the contribution of the initial consonant cluster to the recognition of the reduced target words. Second, we could compare tokens with initial consonant clusters consisting of only word-initial onset consonants (e.g. [p] for [pəˈsipə]) to tokens with initial consonant clusters consisting of also consonants from the coda and/or the onset of the second, stressed syllable of the target word (e.g. [pˈs] for [pˈsipə]). This variation allowed us to investigate the effects of missing vowels on the word recognition process. Since vowels may play a less important role in word recognition than consonants (e.g. Cutler, Sebastiań-Galleś, Soler-Vilageliu, and Van Ooijen, 2000; Bonatti, Peña, Nespor, and Mehler, 2005; Mehler, Peña, Nespor, and Bonatti, 2006), the word recognition process may hardly be hindered by these missing vowels. In fact, listeners may benefit significantly from hearing more consonants than those from the original word onset.

In addition to segmental cues, listeners may use prosodic information to recognise reduced variants. Hence, listeners may use the duration of a reduced pronunciation variant, combined with the durations of the (segments in) surrounding words, which indicate speech rate (Nooteboom, 1980), to deduce its number of syllables/segments, which may facilitate the recognition of this variant. We tested this hypothesis in Experiment 2, which was identical to Experiment 1, except that the duration of the square wave (combined with the duration of the initial consonants in the second half of the experiment) now equalled the duration of the reduced target word. Listeners may better recognise the target word when provided with this durational information, because they may be able to guess its number of syllables/segments.

Finally, Experiment 3 investigated the role of the consonants and vowels from the second, stressed syllable in the recognition of the reduced target words. Previous research has suggested that when speakers reduce, they tend to maintain the onsets and nuclei of the stressed syllables (Ernestus, 2000). Further, stressed syllables play a more important role in lexical selection than unstressed syllables (e.g. Norris and Cutler, 1985; Grosjean and Gee, 1987). Experiment 3 again consisted of two parts, and the duration of the square wave was fixed.

In part one (*gate 3*), participants heard the context, and the reduced target words up to and including their first vowel. This vowel was either the vowel from the first, unstressed syllable (e.g. the first schwa in [pəˈsipə]) or the vowel from the second, stressed syllable (e.g. [i] in [pˈsipə]). This part allows us to compare the contribution of the vowel and consonants from the stressed syllable, in the absence of the first unstressed vowel, with the contribution of the first unstressed vowel in the absence of the second stressed syllable to the recognition of the target words. On the one hand, listeners may be better at recognising target words with unstressed vowels, because no vowel is missing. On the other hand, if the first vowel *is* missing, listeners hear more consonantal information and they hear the vowel from the stressed syllable, and listeners may benefit from hearing this information.

In part two (*gate 4*), listeners heard the context and the target words up to and including the consonant cluster immediately following the first vowel. For example, for the Dutch word *principe* 'principle' listeners heard [pəˈs] and [pˈsip] for the realisations [pəˈsipə] and [pˈsipə], respectively. This part shows to what extent hearing these additional consonants influences participants' performance.

In Experiments 1-3, we will not only investigate the role of the acoustic properties of the target words, but we will also investigate how this acoustic information interacts with the information from the target word's predictability given the context. Previous research has shown that participants predict words using the frequencies of the candidates given the preceding or following words if the context is provided visually, in the form of orthographic transcriptions (Chapter 2 of this dissertation). This probabilistic effect appears to be smaller if participants are provided with context in acoustic (i.e. if they hear the context) rather than orthographic form (Chapter

2 of this dissertation).

In conclusion, this series of experiments will shed some light on which types of acoustic information from reduced pronunciation variants contribute to the recognition of these variants in the processing of spontaneous speech (rather than clearly articulated laboratory speech), and to what extent. The segmental information and the durations of the segment sequences provided in gates 1-4 for tokens with or without the first vowel are summarised in Table 5.1.

Stimulus type	Gate 1 Baseline	Gate 2 C_i	Gate 3 C_iV	Gate 4 C_iVC_j
Vowel present	Ø	76.00 ms ([p])	125.62 ms ([pə])	194.90 ms ([pəˈs])
Vowel absent	Ø	141.20 ms ([pˈs])	218.49 ms ([pˈsi])	277.10 ms ([pˈsip])

Table 5.1: An overview of the segments provided in gates 1-4 (and their average durations) for tokens in which the first vowel was acoustically present or absent (exemplified by two tokens for the target word *principe*: [pəˈsipə] and [pˈsipə]). C = consonant cluster, V = vowel.

5.2 Experiment 1

5.2.1 Participants

Twenty native speakers of Dutch from the Max Planck Institute for Psycholinguistics participant pool were paid to take part in the experiment. None of them reported any hearing loss.

5.2.2 Materials

The materials were extracted from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000), which consists of casual conversations between ten pairs of Dutch native speakers, recorded in a sound-attenuated booth. We selected 38 high-frequency multisyllabic Dutch word types with unstressed initial syllables, always starting with a consonant in their citation form, as our target stimuli, provided in Appendix A. Many of these word types were content words, or at least they contributed substantially to the meaning of the utterance. In addition, we selected 20 different Dutch word types, including words with word-initial stress and monosyllabic words, as filler items, to introduce more variation in the experiment.

For each target word type, we selected on approximately two tokens (one token for 23 word types, two tokens for nine word types, three tokens for two word types, and four tokens for four word types), and in all tokens there were consonants and/or vowels missing in the first syllable. We tried to select as many tokens in which the vowel was present as tokens in which the vowel was absent in the first, unstressed syllable. Since, for most word types, we could not find a token with and a token without the first vowel, we selected tokens with and without the first vowel of different word types. Further, we selected 1.5 tokens for each filler word type on average.

We extracted these tokens together with the prosodic phrase containing the target or filler token (mean preceding context: 5.46 words, range: 2 to 18 words; mean following context: 4.12 words, range: 1 to 15 words). None of the extracted speech fragments contained overlapping speech or loud background noises.

We started by verifying the intelligibility of 73 possible target tokens and 30 filler tokens, embedded in their contexts, by means of a control experiment, because we only wanted to include tokens that could easily be recognised in context. Twenty native speakers of Dutch heard the full sentence fragment (e.g. *Kan je op verschillende [*fˈsχɪln*] manieren doen.* 'You can do that in various ways.'), followed by the reduced target word and its two preceding and following words (e.g. *je op verschillende [*fˈsχɪln*] manieren doen.* 'You do that in various ways.'). The participants were instructed to orthographically transcribe this shorter fragment (i.e. consisting of five words in total). This experiment (as well as all subsequent experiments reported in the present study) was carried out in a sound-attenuated booth, with E-prime 1.2 (Schneider, Eschman, Zuccolotto, 2002). The experiment consisted of twenty blocks, and each block contained the materials of one of the twenty speakers from the corpus. The blocks and trials within blocks were randomised across participants. In each block, listeners first heard a familiarisation phase, consisting of a short monologue (on average 21.46 seconds) by the speaker, which allowed the participants to get used to the speaker's (voice) characteristics. Further, two filler tokens preceded the target tokens in each block. We found that most, but not all, our stimuli were relatively easy to understand in their contexts (93.72% correct, range: 16.67% - 100% correct).

For the main experiments, we selected those stimuli that were easy to understand in their contexts (more than 75% correct in the control experiment). In total, the main experiments contained 63 experimental tokens and 30 fillers, produced by twenty speakers. We included an orthographic transcription of the 63 experimental tokens as an appendix (Appendix A).

Subsequently, we carried out a second control experiment to assess how easily the filler and target tokens could be recognised in isolation. This experiment was identical to Control Experiment 1, except that the words were presented in isolation. Twenty participants took part in this experiment, and these participants had not participated in Control Experiment 1. Participants recognised the target tokens in 69.24% of the trials on average (range: 0% - 100%), which indicates that listeners require context to consistently recognise these reduced pronunciation variants, in line with previous research (e.g. Ernestus, Baayen, and Schreuder, 2002; Chapter 2 of this dissertation).

We evaluated the degree of reduction of the target and filler tokens by means of a phonetic transcription of the materials (see again Appendix A). Two transcribers determined which segments were present in the speech signal, and the locations of the segment boundaries, and these boundaries formed the basis for the gates used in our main experiments. The average distance between the segment boundaries placed by the two transcribers equalled 2.11 ms. There was disagreement on the presence/absence of consonants in the first syllable in 12.7% and on the presence/absence of vowels in the first syllable in 15.87% of the target tokens. Whenever there was a difference between the two transcriptions, a third transcriber (the first author) determined the correct transcription.

The descriptive statistics for the reduction in the initial consonant cluster are shown in Table 5.2. The transcriptions showed that vowels were missing in the initial syllable (and thus in the initial consonant cluster) in 31 target tokens (49.21%), for instance, in [f**a**ˈmili] realised like [fˈmili]. Spectrograms and transcriptions of two tokens of *principe*, one realised with and one without the first unstressed vowel, are provided in Figure 5.1. Missing vowels lead to phonotactically illegal consonant clusters in 15 target tokens (23.81%, e.g. [vərˈkopt] realised like [**f**ˈ**k**opt]). We included more tokens with legal than with illegal initial consonant clusters because legal initial consonant clusters may be more difficult for the listener if we assume that listeners use phonotactic cues to reconstruct the citation forms of reduced pronunciation variants. Further, at least one consonant was missing in the initial consonant cluster of 30 target tokens (47.62%, e.g. [p**r**oˈjɛkt] realised like [pəˈjɛkt], and hence [r] was missing). As shown in Table 5.2, there were no segments missing in the initial consonant cluster in approximately 40% of the target tokens, which meant that, in these tokens, the first vowel was always present and there were segments missing in the coda rather than the onset of the first syllable.

In *gate 1*, participants heard only the preceding and following context of the reduced target words, separated by a square wave. In *gate 2*, participants also heard the initial consonant cluster. This cluster comprised only the onset consonants from the word's citation form if the first vowel from the citation form was present in the acoustic signal (e.g. [p] for the token [pəˈsipə]). This was the case for 32 target tokens, forming 50.79% of all target tokens. The cluster comprised also consonants from the coda and/or from the onset of the second, stressed syllable from the word's citation form if the first vowel from the citation form was absent (e.g. [pˈs] for the token [pˈsipə]): participants heard two consonants in nine target tokens (14.29% of all target tokens), three consonants in fourteen target tokens (22.22% of all target tokens), and four consonants in eight target tokens (12.70% of all target tokens).

Missing segments	Phonotactically legal	Phonotactically illegal
None	27 (42.86%)	-
Vowel only	-	6 (9.52%)
Consonants only	5 (7.94%)	-
Vowel + consonants	16 (25.40%)	9 (14.29%)

Table 5.2: Absolute numbers (and percentages) of target words with different types of reduction in the initial consonant cluster, broken for the phonotactic well-formedness of this cluster.

Previous research has shown that truncated speech sounds highly unnatural and may lead listeners to perceive an inserted labial or plosive consonant (Pols and Schouten, 1978). This problem can be overcome by adding a square wave to the speech fragment (Warner, 1998). Therefore, we used a square wave to indicate the original location of the target word. We used a 500 Hz square wave, which consisted of an onset of five ms with gradually increasing amplitude and a part of 500 ms with a fixed amplitude of 52 dB. The intensity of the sound fragments (without the square wave) was normalised to 70 dB.

We created twenty different randomisation lists for the experiment, one for each participant. In each list, we randomised the speaker blocks across the parts and the trials within the blocks. Participants heard the materials of a particular speaker in either *gate 1* or *gate 2*. At the end of the first half of the experiment (i.e. after the 47th trial, given that the total number of trials equalled 93), the current speaker block was completed before *gate 2* began, and *gate 2* started with a novel speaker. As a consequence, part one contained more target tokens than part two (33 versus 30 target tokens on average).

5.2.3 Procedure

The procedure in Experiment 1 was identical to those of the control experiments, except that it consisted of two parts: One in which participants heard only the preceding and following context of the reduced target and filler tokens, and another in which they heard the context in addition to the initial consonant cluster of the reduced target and filler token. Thus, in *gate 1* participants heard the preceding and following context separated by a square wave. In *gate 2*, participants heard the preceding context and the initial consonant cluster, followed by a square wave and the following context. In both parts, participants were instructed to orthographically transcribe the target words.

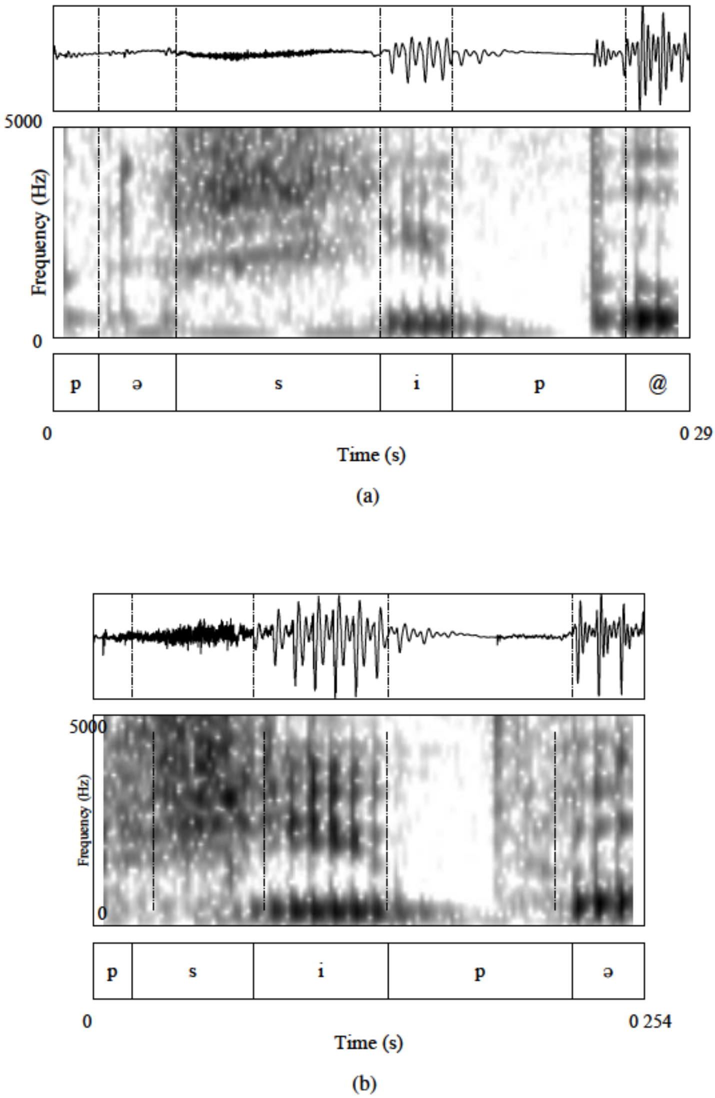

Figure 5.1: Spectrograms and transcriptions for two tokens of the word *principe* ‘principle’.

5.2.4 Results and discussion

Two trained transcribers labelled the correctness of the participants' responses (correct/incorrect). Participants produced 430 correct responses and 830 incorrect responses for the target words. We analysed the correctness of participants' responses (in the analysis of the present but also subsequent experiments) by means of generalised linear mixed-effects regression with the logit link function (Jaeger, 2008). Predictors were removed if they did not attain significance at the 5% level. We included random effects for *participant*, *target type* (e.g. the Dutch word *programma* or *verdieping*), and *target token* (e.g. the first or second token of the Dutch word *programma*).

First of all, we entered the fixed predictors *gate* (*gate 1* versus *gate 2*) and *first vowel deletion* (whether the vowel was missing in the initial consonant cluster). The variable *first vowel deletion* indicated whether we only presented consonants from the onset (e.g. only [f] for the realisation [fəˈkopt]), or also from the coda and/or the onset of the second, stressed syllable (e.g. [fˈk] for the realisation [fˈkopt]). The presence of the first vowel correlated with the presence of all consonants in the initial consonant cluster. Detailed analyses (for all main experiments reported in this study) showed, however, that the presence of the first vowel was the best predictor of the correctness of the responses.

We also incorporated as predictors the likelihoods of the target words given the preceding and following words in the sentences. We determined the words' bigram frequencies with their preceding or following words in the sentences, based on the Spoken Dutch Corpus (Oostdijk, 2000). Since not the full range was covered in either of these frequency measures, we converted these variables into the factors *preceding bigram frequency* and *following bigram frequency*, consisting of four levels, representing their four quartiles. The quartile ranges are provided in Table 5.3.

Level	Preceding bigram frequency	Following bigram frequency
Quartile 1	0-3	0
Quartile 2	4-29	1-5
Quartile 3	30-127	6-41
Quartile 4	128-2482	42-1275

Table 5.3: The quartile ranges for *preceding bigram frequency* and *following bigram frequency*.

Finally, we included the control variable *trial number* (trial number within each gate), in order to capture effects due to learning or fatigue. The results of the mixed effects model are shown in Table 5.4.

Predictor	F	p
gate	83.28	< .01
first vowel deletion	1.46	n.s.
gate : first vowel deletion	29.26	< .0001

Table 5.4: F values and significance values for the model for Experiment 1 (degrees of freedom: 1256).

We found a main effect of *gate* and an interaction between *gate* and *first vowel deletion*, indicating that participants performed better when hearing the initial consonant cluster, and this effect was stronger if the first vowel was missing and the consonant cluster consisted of more than just the consonants from word onset (see Table 5.5). This finding shows that listeners could use these additional consonants from the reduced target words to recognise these target words, despite the missing vowel.

The fact that the statistics show a main effect of *gate* but no main effect of *first vowel deletion* is as expected, given that tokens with the initial unstressed vowel were recognised better than tokens without the initial unstressed vowel in *gate 1*, but worse in *gate 2* (see Table 5.5). The numerically present trend for *gate 1* that less reduced words were easier to guess is reversed in *gate 2*, where the listeners hear more segmental information from more reduced words. Further, note that the tokens with or without vowels largely represent different word types and differences between word types were at least partly accounted for by the random factor *target type*, which showed a high standard error (1.664). As a consequence, the statistical model was conservative with respect to a potential effect of first vowel deletion.

Gate	Vowel realised	Vowel absent
gate 1	31.25%	20.67%
gate 2	32.89%	53.95%

Table 5.5: The percentages correct for tokens with or without the vowel from the initial unstressed syllable in Experiment 1.

We also investigated participants' responses in the incorrect trials. For this purpose, the same two transcribers first of all marked whether the incorrect response was contextually appropriate. Responses were labelled as appropriate if they could fit within the syntactic structure of the sentence, and the resulting sentence made any sense. For example, the response *papier* 'paper' was labelled contextually inappropriate for the sentence *Met die slaapzakken ook in het verleden wel problemen gehad* *eigenlijk.* 'With those sleeping bags in the past also had problems actually.', because if we replace the target word *problemen* with *papier*, then the sentence becomes uninterpretable.

Further, the transcribers marked the correctness of the word's first segment, second segment (if the first segment was correct), third segment (if the first and second segment were correct), the word-final segment, and the number of syllables. A segment was labelled as correct if the sound corresponding to the grapheme provided by the participant matched that of the word's citation form. For example, in certain regions of the Netherlands voiced fricatives are always pronounced as voiceless, and therefore if a participant's answer for the target word [vərˈkopt] *verkoopt* started with an 'f', then the first segment of this answer was labelled as "correct".

For all measures, and for all the experiments reported in this study, there was high agreement among the transcribers (more than 95%). The descriptive statistics for the incorrect trials for tokens with or without the first unstressed vowel are provided in Table 5.6 (the total number of incorrect trials equals 100%). These descriptives suggest that, in case of an error, participants could better identify the first segments of the word's citation form in *gate 2* than in *gate 1*, which is as expected, since participants heard these segments in *gate 2*, whereas they did not in *gate 1*. Moreover, participants could better identify the first segments of the word's citation form in *gate 2* if the first vowel was missing compared to when it was present. In both cases, however, listeners could not always recognise the initial consonants or did not always use this segmental information.

Further, participants provided more contextually appropriate responses for tokens without first vowel deletion than for tokens with first vowel deletion, and this difference was smaller in *gate 2* than in *gate 1*. Possibly, participants had more difficulties understanding the contexts of target tokens with absent first vowels, since highly reduced word tokens tend to occur in acoustically reduced contexts.

Finally, in 22.49% of the incorrect trials, participants provided answers that were semantically and syntactically possible and that shared their first segments with those of the reduced target word. Apparently, Dutch allows multiple word candidates on the basis of the context and the word's first segment. For example, for the sentence *Ik had vandaag weer een auto geleend.* 'Today I borrowed a car again.', two participants answered *vanochtend* 'this morning', and this response shared the initial consonant with the target word and was contextually appropriate. For these utterances, participants have to hear more segments than the initial ones to guess the words.

To summarise, our results show that listeners had difficulties guessing the target word on the basis of the context alone or on the basis of the context and the initial consonant cluster. Importantly, performance was better if the consonant cluster re-

Gate	Vowel realised	Contextually appropriate	1st segment	1st & 2nd segment	1st - 3rd segment	Final segment	Number of syllables
gate 1	No	32.95%	7.66%	0.77%	0.38%	16.48%	10.34%
gate 1	Yes	57.14%	7.36%	2.16%	0.00%	25.11%	15.58%
gate 2	No	32.09%	39.55%	14.18%	8.21%	14.93%	8.21%
gate 2	Yes	41.67%	32.35%	6.86%	2.45%	24.51%	22.55%

Table 5.6: Percentage of contextually appropriate responses, and percentages of correct first segments, final segments, and number of syllables for the incorrect trials in Experiment 1, broken down by the gate and whether the initial unstressed vowel was realised.

sulted from a missing vowel and consequently consisted of more than the original word onset consonants. This finding indicates that hearing additional consonants outweighs missing the first unstressed vowel in word recognition.

Listeners apparently need more information from reduced pronunciation variants than the initial consonants, which could be more segmental information, or perhaps the durations of the variants, as discussed in the Introduction. Listeners may be able to use word duration as a cue to predict the words' number of syllables, and therefore they may be able to recognise words more easily if durational information is provided. We investigated this hypothesis in Experiment 2.

5.3 Experiment 2

5.3.1 Participants

Twenty native speakers of Dutch from the Max Planck Institute for Psycholinguistics participant pool, none of whom participated in the previous experiments, were paid to take part in the experiment. None of them reported any hearing loss.

5.3.2 Materials

The materials were identical to those of Experiment 1, except that the duration of the square wave now equalled the duration of the reduced word token (in *gate 1*) or the duration of the square wave combined with that of the initial consonant cluster equalled the duration of the reduced word token (in *gate 2*). We used a minimum duration of 20 ms because a pilot experiment indicated that for shorter durations participants had difficulty locating the square wave. The minimum duration of 20 ms meant that, in *gate 2*, the combined duration of the square wave and the initial consonant cluster for three filler tokens was longer than these reduced filler tokens themselves.

5.3.3 Procedure

The experimental procedure was identical to that of Experiment 1, except that participants were now told that the duration of the square wave equalled the duration of the missing word in *gate 1*, and of the missing part of the word in *gate 2*.

5.3.4 Results and discussion

Two trained transcribers labelled the responses, using the same criteria as for Experiment 1. Participants produced 335 correct responses and 925 incorrect responses for the target words. The percentages correct for tokens with and without the first vowel are provided in Table 5.7.

Gate	Vowel realised	Vowel absent
gate 1	22.32%	15.50%
gate 2	28.95%	41.58%

Table 5.7: The percentages correct for tokens in Experiment 2, broken down by whether the initial unstressed vowel was realised.

We fitted a regression model for the combined data set of Experiments 1 and 2, because we wanted to investigate whether listeners performed better in Experiment 2 than in Experiment 1, since they were now provided with duration information from the reduced target words. We also wanted to test whether this durational information influenced how much listeners relied on other sources of information, for example bigram frequencies. We included the same random and fixed variables as for Experiment 1, in addition to *experiment* (Experiment 1 versus Experiment 2). The results are provided in Table 5.8.

Predictor	F	p
experiment	13.93	< .001
gate	156.18	< .0001
following bigram frequency	0.002	n.s.
preceding bigram frequency	3.19	< .05
first vowel deletion	2.27	n.s.
experiment : following bigram frequency	4.02	< .05
gate : first vowel deletion	35.96	< .0001

Table 5.8: F values and significance values for the model comparing Experiments 1 and 2 (degrees of freedom: 2506).

We found a main effect of *experiment*, and an interaction between *experiment* and *following bigram frequency*. Surprisingly, participants gave more correct responses in Experiment 1 than in Experiment 2. Durational information thus did not make the task easier, but appeared misleading. This was true especially for target words with low bigram frequencies with their following word (37.37% versus 37.31% correct for Quartile 1 and Quartile 4 in Experiment 1, but 23.68% versus 33.46% correct for Quartile 1 and Quartile 4 in Experiment 2). Possibly, participants relied more heavily on the following bigram frequency in Experiment 2 (and therefore produced more errors for the low predictability words in Experiment 2), because they found the durational information misleading and therefore used more different resources to predict the reduced target words.

In addition, our results showed a main effect of *preceding bigram frequency*, indicating that, in both Experiments 1 and 2, listeners better predicted words with higher bigram frequencies with their preceding words (13.19% correct for Quartile 1 versus 38.15% correct for Quartiles 3-4 on average). This effect only emerged in this combined analysis of Experiments 1 and 2 probably due to a lack of statistical power in the separate analyses of these experiments.

Further, we again found a two-way interaction between *gate* and *first vowel deletion*, showing that listeners better predicted target words in *gate 2*, especially if these target words had word-initial consonant clusters that comprised more than the onset consonants of the word-initial syllable.

We then investigated participants' responses in the incorrect trials in Experiment 2. The descriptive statistics for the incorrect trials for tokens with or without the first unstressed vowel are provided in Table 5.9. Similar to Experiment 1, participants more frequently recognised the first segments of the word's citation form in *gate 2* than in *gate 1*, and they produced fewer contextually appropriate responses if the first unstressed vowel was missing. There is also an important difference between the errors in Experiment 1 and Experiment 2: we found much higher percentages of contextually inappropriate responses in Experiment 2 than in Experiment 1 (63.50% versus 51.29%). Apparently, listeners paid less attention to context if the durational information was misleading.

Why did participants perform worse if they were provided with additional, durational information to rely on? Listeners appear generally to be unaware of the reductions that occur in spontaneous speech (e.g. Kemps, Ernestus, Schreuder, and Baayen, 2004) and consequently participants may have tried to match the duration of the square wave to the durations of the words' citation forms. Since the target words were all segmentally and durationally reduced, participants may consequently have preferred candidates that are shorter than the citation forms of the target words. If so, this may be reflected in the lengths of the responses; that is, participants may have provided shorter responses in Experiment 2 than in Experiment 1. We tested this hypothesis by first converting participants' responses into phoneme sequences, and by subsequently comparing the lengths of the responses in phonemes (henceforth *response length*) in Experiments 1 and 2. Since these lengths were not distributed normally, we converted *response length* into a binary variable by applying a median split so that we could analyse our results by means of binomial regression: We labelled responses as "long" if they contained more than five phonemes; otherwise we labelled responses as "short". We then fitted a generalised linear mixed-effects regression model with the logit link function for the dependent variable *response length*, including the same random variables as for the analysis of the correctness of the responses. We found a significant main effect of *experiment* ($\beta = 0.477, F(1, 2518) = 8.20, p < .01$), indicating that participants provided shorter responses in Experiment 2 than in Experiment 1 (4.92 versus 5.31 phonemes on average).

Gate	Vowel realised	Contextually appropriate	1st segment	1st & 2nd segment	1st - 3rd segment	Final segment	Number of syllables
gate 1	No	25.90%	7.19%	2.88%	2.16%	18.71%	8.99%
gate 1	Yes	39.85%	8.81%	2.30%	0.00%	16.86%	17.24%
gate 2	No	28.24%	30.59%	12.94%	11.18%	19.41%	8.82%
gate 2	Yes	43.98%	27.78%	6.48%	1.85%	23.15%	15.28%

Table 5.9: Percentage of contextually appropriate responses, and percentages of correct first segments, final segments, and number of syllables for the incorrect trials in Experiment 2, broken down by the gate and whether the initial unstressed vowel was realised.

To conclude, listeners are misled by the durational information from reduced pronunciation variants if this durational information is provided separately from other acoustic information. These results support the hypothesis that listeners are unaware of the reductions in spontaneous speech, and therefore they cannot use word duration by itself to recognise these reduced pronunciation variants.

So far, we have established the contribution of the initial consonant cluster and of word duration to the recognition of reduced pronunciation variants. In Experiment 3, we investigated the contributions of the first realised vowel and of the subsequent consonant or consonant cluster (henceforth the "second consonant cluster", for the sake of convenience) to the recognition of these variants.

5.4 Experiment 3

5.4.1 Participants

Twenty native speakers of Dutch from the same pool of participants as used in the previous experiments participated in the experiment. They received a small salary for participating in the experiment, and they had not participated in any of the previous experiments. None of the participants reported any hearing loss.

5.4.2 Materials

Each stimulus used in Experiment 1 was extended to include the first realised vowel of the reduced word for *gate 3* and this vowel as well as the second consonant cluster for *gate 4*. For example, for the target word *principe* 'principle' pronounced like [pəˈsipə], participants heard [pə] in *gate 3* and [pəˈs] in *gate 4*. Conversely, for a different token of this target word, pronounced like [pˈsipə], participants heard [pˈsi] in *gate 3* and [pˈsip] in *gate 4*. This meant that two tokens of the target word *manier* (both realised like [mˈni]) were presented in full in *gate 3* (3.17% of the trials), while 24 target words (38.10%) were presented in full in *gate 4* (e.g. [ˈmir] for the target word *manier* 'manner' and [ˈχɑt] for the target word *gehad* 'had').

Phonetic transcriptions showed that in 13 target stimuli the added consonant cluster consisted of consonants separated by an unstressed vowel in the words' citation forms (20.63%). For example, the Dutch word *verschillende* [vərˈsχɪləndə] 'different' was realised like [fˈsχɪln] and participants also heard the consonants immediately following this second unstressed vowel in the word's citation form (e.g. participants heard [fˈsχɪln] rather than [fˈsχɪl]). Moreover, transcriptions showed that in 40 target stimuli there were consonants missing in the added consonant cluster (63.49%). For example, the Dutch word *vanzelf* 'by itself' [vɑ**n**ˈzɛlf] was realised like [vəˈzɛlf] and [n] was missing, and participants heard [və] in *gate 3* and [vəˈz] in *gate 4*. Since participants were misled by durational information in Experiment 2, we used the same square wave with a fixed duration as in Experiment 1.

5.4.3 Procedure

The experimental procedure was identical to the procedure of Experiment 1.

5.4.4 Results and discussion

Participants produced 745 correct responses and 515 incorrect responses for the target words. The percentages correct for tokens with and without the first unstressed vowel are provided in Table 5.10. First of all, we investigated the contribution of the first vowel, in addition to the initial consonant cluster, to the recognition of reduced pronunciation variants. That is, we combined the data for *gate 2* from Experiment 1 and *gate 3* from Experiment 3, and we fitted a regression model, including the same fixed and random effects as for Experiment 1. The results are provided in Table 5.11.

Gate	Unstressed vowel	Stressed vowel
gate 3	38.10%	64.44%
gate 4	54.93%	81.79%

Table 5.10: The percentages correct for tokens in which listeners either heard the unstressed or stressed vowel from the target words in gates 3 and 4.

Predictor	*F*	*p*
gate	4.02	n.s. [1]
preceding bigram frequency	1.73	n.s.
first vowel deletion	6.56	n.s. [1]
gate : preceding bigram frequency	10.05	< .0001
gate : first vowel deletion	6.67	< .05

Table 5.11: F values and significance values for the model for gate 2 and gate 3 in Experiments 1 and 3 respectively (degrees of freedom: 1250).

We found two interactions with *gate*. The interaction with *first vowel deletion* indicates that the addition of the first vowel had a stronger effect if this vowel was the (second) stressed vowel than if it was the (first) unstressed vowel. The interaction with *preceding bigram frequency* demonstrates that bigram frequency effects were restricted to *gate 2*, where participants received the least segmental information.

Subsequently, we fitted a regression model for the complete data set of Experiment 3 in order to determine the effect of the consonant cluster following the first vowel in the stimuli. We included the same random and fixed variables as for Experiment 1, in addition to the fixed variable *complete auditory form* (whether the reduced target word was presented in full). The presence of the first vowel correlated with the presence of other unstressed vowels and with the presence of all consonants in the second consonant cluster. Detailed analyses showed that the presence of the first, unstressed vowel was the best predictor of the correctness of the responses. The results are provided in Table 5.12. We found main effects of *gate* and *first vowel deletion*, indicating that listeners better recognised words in *gate 4* and if listeners heard the vowel from the stressed rather than the unstressed syllable (see Table 5.10).

[1] This effect was only significant in the analysis of variance results; not in the summary of the model.

Predictor	*F*	*p*
gate	68.51	< .0001
first vowel deletion	12.11	< .001

Table 5.12: F values and significance values for the model for Experiment 3 (degrees of freedom: 1257).

Finally, we investigated participants' responses in the incorrect trials in Experiment 3. The descriptive statistics are provided in Table 5.13. The results for *gate 4* show that in most incorrect trials, listeners did not recognise the first three segments of the reduced target word at all. Further, like the incorrect answers for Experiments 1 and 2, also those for Experiment 3 suggest that if the first unstressed vowel was present, the surrounding segments, including the first consonants, were pronounced more clearly. This appears from three observations. First, the presence of the first, unstressed vowel facilitated the recognition of preceding consonants and, second, led to more contextually appropriate responses. Third, listeners provided more contextually appropriate responses if they could identify the word's first segment (61.49% compared to 25.89% of all incorrect trials). For example, for the sentence *Nou hij verzamelt al heel lang kinderboeken.* 'Well he has been collecting children's books for a very long time.', participants answered *verkoopt* 'sells' and *vertaalde* 'translated', and these responses share the onset consonant of the target word and are contextually appropriate.

Gate	Vowel realised	Contextually appropriate	1st segment	1st & 2nd segment	1st - 3rd segment	Final segment	Number of syllables
gate 3	No	14.53%	20.51%	9.40%	4.27%	21.37%	13.68%
gate 3	Yes	48.08%	35.10%	11.54%	7.21%	21.15%	22.12%
gate 4	No	18.87%	13.21%	13.21%	11.32%	30.19%	15.09%
gate 4	Yes	43.07%	32.12%	14.60%	8.76%	25.55%	18.98%

Table 5.13: Percentage of contextually appropriate responses, and percentages of correct first segments, final segments, and number of syllables for the incorrect trials in Experiment 3, broken down by the gate and whether the initial unstressed vowel was realised.

To summarise, our results show a significant contribution of the first vowel as well as the second consonant cluster to the recognition of reduced pronunciation variants. Moreover, listeners better recognised target words if they heard the vowel from the stressed syllable and the first, unstressed vowel was missing than if they heard the vowel from the initial, unstressed syllable. This suggests again that the possibly disturbing absence of a vowel may be compensated for by the more useful information from a stressed vowel and additional consonants. Finally, we found that the role of bigram frequency information decreased once listeners heard the first realised vowel.

5.5 General discussion

Previous research has shown that listeners require both the context and acoustic information from reduced pronunciation variants to recognise these variants (Ernestus, Baayen, and Schreuder, 2002; Janse and Ernestus, to appear; Chapter 2 of this dissertation). The present study investigated which types of acoustic information from reduced variants themselves are necessary for the recognition of these variants, focusing on the recognition of variants with unstressed initial syllables. (Ernestus et al., 2002; Janse and Ernestus, 2011)

In these unstressed initial syllables, segments are frequently missing, and since these reductions occur before the words' uniqueness points, they are likely to influence the word recognition process. Intuitively, missing segments in the initial syllable are likely to create ambiguity and increase uncertainty during the recognition process (e.g. the Dutch words *verlaten* 'leave' and *flater* 'blunder' with the citation forms [vərˈlatən] and [ˈflatər] may sound like [ˈflatə] and [ˈflatər]), especially if we assume that listeners only store the words' citation forms in their mental lexicons.

The present study investigated the role of the first realised consonant or consonant cluster (henceforth "initial consonant cluster"), the first realised vowel, and the subsequent realised consonant or consonant cluster (henceforth "second consonant cluster"). Importantly, the initial consonant cluster consisted of more than the onset consonants from the citation form if the vowel from the initial unstressed syllable was missing, which was the case in approximately fifty percent of the tokens (see Table 5.1 in the Introduction). This variation allowed us to investigate the effects of the absence of the first vowel on word recognition. Similarly, the second consonant cluster consisted of more consonants if the second unstressed vowel was missing. In addition to the role of segmental information, we investigated to what extent listeners rely on the durations of reduced pronunciation variants to recognise these variants.

We carried out three auditory gating experiments, in which participants heard fragments of spontaneous speech always consisting of the preceding context, some segments of the reduced target words (except for the baseline condition, *gate 1*, in which listeners heard only the context), a square wave, and the following context. Participants were instructed to orthographically transcibe the target word. They heard the initial consonant cluster (*gate 2*), the initial consonant cluster and the first realised vowel (*gate 3*), or the initial consonant cluster, the first realised vowel, and the second consonant cluster (*gate 4*) of the target words. The only difference between Experiments 1 and 2 was that the duration of the square wave in Experiment 2 equalled the duration of the missing part of the reduced target word, and this experiment allowed us to investigate the role of word duration by itself in the recognition of reduced pronunciation variants.

We found that participants' performance improved with every gate (percentages correct for *gate 1*: 26.02%; *gate 2*: 43.19%; *gate 3*: 51.13%; *gate 4*: 68.07%). Importantly, the presentation of the first consonant cluster improved their performance more if this cluster resulted from the absence of an unstressed vowel (percentage correct for tokens realised with the first unstressed vowel: 32.89%, and for tokens realised without this vowel: 53.95%). The absence of the first vowel improved participants' performance not only when they heard just the first consonant cluster but also when they additionally heard the following stressed vowel and second consonant cluster (i.e. in *gates 2-4*).

These findings suggest that listeners require less segmental information to recognise tokens realised without their first unstressed vowels. However, the sequences that we presented from tokens realised with and without their first unstressed vowels differed in their durations. Participants may therefore have recognised tokens realised without their first unstressed vowels better just because their sequences were longer. Interestingly, the stimuli for tokens with first vowel deletion in *gate 2* and for tokens without first vowel deletion in *gate 3* (e.g. [pˈs] versus [pə] for the Dutch word *principe* 'principle') were approximately equally long (i.e. we did not find a significant difference in duration, $p > 0.1$), as were the stimuli for tokens with first vowel deletion in *gate 3* and those for tokens without first vowel deletion in *gate 4* (e.g. [pˈsi] versus [pəˈs]; we found only a marginal difference in duration, $p = 0.09$). By numerically comparing these gates we can evaluate differences in participants' performance for stimuli of approximately the same duration. The data suggest again that participants performed better for tokens with than for tokens without first vowel deletion (*gate 2* versus *gate 3*: 53.95% versus 38.10% correct, and *gate 3* versus *gate 4*: 64.44% versus 54.93% correct).

We conclude that listeners may require less segmental information to recognise words if the vowel is missing in the word-initial syllable. The absence of the unstressed vowel appears less important than the presence of additional consonants. This result corroborates previous findings (e.g. Cutler et al., 2000; Bonatti et al., 2005; Mehler et al., 2006) suggesting that consonants play a larger role in lexical access than vowels.

The result contrasts with previous findings suggesting that reductions inhibit word recognition. These previous findings nearly all come from experiments testing listeners' comprehension of reduction in read-aloud isolated words or words embedded in simple short sentences (e.g. Ernestus and Baayen, 2007; Ranbom and Connine, 2007; Tucker, in press; Tucker and Warner, 2007; Chapter 3 of this dissertation). Only Brouwer, Mitterer, and Huettig (in press) tested the comprehension of reduced words in their natural contexts, like we did. Brouwer and colleagues conducted eye-tracking experiments showing the orthographic representations of the words on a computer screen, which may have activated the words' citation forms. Possibly, these orthographic representations are responsible for the inhibition that these authors found for reduced forms. Future eye-tracking studies should indicate whether the differences between the findings presented by Brouwer and colleagues and the present study are indeed due to these ortographic representations.

How can our findings be accounted for by current models of speech comprehension? Abstractionist models of speech comprehension (e.g. Marslen-Wilson and Welsh, 1978; Norris, 1994) can account for the finding that listeners only recognise highly reduced pronunciation variants within a viable context by positing context-dependent reconstruction mechanisms for missing segments. These mechanisms should be able to operate without phonotactic cues indicating that segments are missing, since, in many of our target words, missing vowels did not lead to phonotactically illegal consonant clusters. Possibly, these context-dependent mechanisms may use subphonemic cues, signalling missing segments, that may have been present in our materials.

Exemplar-based accounts of speech processing (e.g. Goldinger, 1998) can also account for our findings. First, these models assume that reduced as well as unreduced variants are stored in the mental lexicon, and therefore listeners do not need to reconstruct missing segments and can easily recognise reduced variants even if they are phonotactically completely legal. Second, the storage of reduced variants explains why listeners require less segmental information to recognise reduced variants than unreduced variants, if the reduced variants have earlier uniqueness points, as shown in our study. Finally, these models assume that listeners store exemplars of pronunciation variants together with their contexts, and hence these contexts are required to recognise these variants.

The present study also investigated the effects of durational information in reduced words on their recognition. In Experiment 2, the duration of the square wave, or its duration combined with the duration of the initial consonant cluster, equalled that of the reduced target word. Surprisingly, listeners found this durational information misleading, and they provided more errors and significantly shorter responses in Experiment 2 compared to Experiment 1. In line with Kemps et al. (2004), this finding shows that listeners are unaware of the reductions that occur in spontaneous speech, and on the basis of the durations of the target words they therefore expected these target words to contain fewer segments in their citation forms.

In all three experiments, we tested the contribution of contextual information to the recognition of reduced target words as a function of the amount and type of acoustic information listeners heard from these target words. We found clear preceding bigram frequency effects for participants who heard only the context or the context and the initial consonant cluster of the target word, but no frequency effects for participants who also heard the first vowel or the first vowel and the second consonant cluster. This finding shows that the segments from reduced pronunciation variants, despite their reduced nature (e.g. shorter segment durations, spectral reduction), play a more important role in the processing of these pronunciation variants than probabilistic information based on context, in line with Chapter 2 of this dissertation. Further, we found stronger following bigram frequency effects for participants who heard square waves with durations that equalled the durations of the missing parts of the target words (i.e. in Experiment 2). These square waves were shorter than the square wave used in Experiment 1. As a consequence of this shorter square wave, there was a shorter period between the preceding context or the initial consonant cluster and the following context, which may have led listeners to rely more heavily on the words' probability given the following context. On the whole, listeners seem to rely more heavily on contextual information if the acoustic information is insufficient or misleading.

In addition to investigating which acoustic information improved listeners' performance, we investigated the characteristics of their incorrect responses. These incorrect responses mainly suggest that listeners often had difficulties identifying the words in the contexts and the segments in the target words, which is unsurprising since the stretches of speech were spliced from casual speech. For instance, we found that participants frequently provided answers with incorrect consonants, which suggests that there was often uncertainty about the identity of the individual speech sounds. Moreover, if the first, unstressed vowel was absent and the whole stretch of speech was probably more reduced, listeners more often provided answers with incorrect initial consonants and contextually inappropriate answers.

The errors also showed that participants gave more contextually inappropriate answers if provided with durational information about the target words, without corresponding spectral information, which they considered misleading. Furthermore, approximately 20% of the errors were contextually appropriate and shared the initial segment(s) of the reduced target word, which shows that for many target words the first few segments were insufficient to rule out all but the target words.

Finally, this study demonstrates that the gating paradigm (Grosjean, 1980), originally designed for studying the comprehension of laboratory speech, can also be used for studying the comprehension of highly reduced pronunciation variants in conversational speech. By placing gates at the end of segment boundaries (rather than after intervals of fixed durations), we controlled for the amount of segmental information provided. As shown in our study, this version of the gating paradigm can provide useful insights into the effects of reduction on word recognition.

To conclude, the present study has provided evidence that if listeners are presented with reduced pronunciation variants embedded in their natural contexts, they may require less segmental information to recognise these variants if, as a result of these reductions, these reduced variants become unique more quickly.

5.6 Appendix

Orthographic transcriptions of the materials and phonetic transcriptions of the target words used in the present study. We have underlined the target words in the orthographic transcriptions.

Daarna [darˈna] ‘subsequently’

*Ze hadden ons gevraagd of wij de allerlaatste keer in die boot wilden roeien en daarna [*nəˈna*] zou die in stukken gehakt worden.*
‘They had asked us to row that boat for the last time and subsequently it would be cut up in pieces.’

*En een jaar daarna [*nəˈna*] ben jij erbij gekomen.*
‘And the year after that you joined us.’

*Hij heeft daarna [*təˈna*] helemaal opnieuw leren praten.*
‘After that he had to learn to talk again from square one.’

dezelfde [dəˈzɛlvdə] ‘the same’

*Het was precies dezelfde [*tˈsɛlə*] tijd.*
‘It was exactly the same time.’

Familie [fɑˈmili] ‘family’

*En daar zit nu ook de hele familie [*fˈmili*] weer bij, of niet?*
‘And the whole family will join once again, right?’

*Ja, bij jullie familie [*fˈmili*] zijn jullie echt snel.*
‘Yes, in your family they are really quick.’

*Het ging dan meer om familie [*fˈmili*] bezoek dus het hoefde niet.*
‘It was then more like a family visit, so it did not have to.’

Gegevens [χəˈχevəns] ‘data’

*En dan de gegevens [*ˈχevəs*] aan te vullen.*
‘And then update the data.’

Gehad [χəˈhɑt] ‘had’

*Ik heb een tijd gehad [*ˈχɑt*] dat ik veel naar eh naar Derrick keek.*
‘I have had a period in which I frequently watched eh Derrick.’

*Of heb jij ook te maken gehad [*ˈχɑd*] met eh ambtelijke teksten zeg maar?*
‘Or have you also had to deal with eh so called official texts?’

*Ik heb ook een periode van een jaar ofzo gehad [*ˈχɑt*] dat ik één keer gereden had.*
‘I have also had a period of one year or so, in which I drove only once.’

*Want ik heb nooit het idee gehad [*ˈχɑt*] dat de organisatie een probleem was.*
‘Because I have never had the feeling that the organisation was a problem.’

Gesproken [χəˈsprokən] ‘speaking’

*De mensen van wie je normaal gesproken [*ˈsprokə*] veel vuurwerk ziet.*
‘The people that normally speaking show fireworks.’

Goedkoop [χutˈkop] ‘cheap’

*Een grote partij in te slaan en dan heel goedkoop [*χəˈkop*] aan te bieden.*
‘Stock a large amount and then offer them at a very low price.’

Goedkope [χutˈkopə] ‘cheap’

*Straks staan ze allemaal tegen die goedkope [*χəˈkop*] tenten aan te loeren.*
‘Soon they will all be looking at those cheap tents.’

Hetzelfde [hɛtˈzɛlvdə] ‘the same’

*Dat was de tweede keer dat we op hetzelfde [*ˈsɛldə*] instituut zaten.*
‘That was the second time that we were at the same institute.’

*Je betaald exact hetzelfde [*ˈsɛldə*] bedrag als vorig jaar.*
‘You pay exactly the same amount as last year.’

Kunstmatige [kʏnstˈmatɪχə] ‘artificial’

*Waarom we voor een kunstmatige [*kəsˈmatχ*] taal hebben gekozen.*
‘Why we opted for an artificial language.’

Manier [mɑˈnir] ‘manner’

*Nee, maar het is toch de manier [*ˈmir*] waarop het gebouw gemaakt is.*
‘No, but still it is the way the building was constructed.’

*Maar dat was al op die manier [*mˈni*] gegarandeerd.*
‘But that was already guaranteed in that way.’

*Maar goed, dan wordt het toch op een of andere manier [*mˈni*] vastgelegd.*
‘But well, that will still be recorded in some way.’

*Moet natuurlijk dat geld op de een of andere manier [*ˈmir*] beheren.*
‘Of course has to administer that money in a certain way.’

Moment [moˈmɛnt] ‘moment’

*Maar hij leest op dit moment [*ˈmɛn*] meer kinderboeken dan ik.*
‘But at this moment he reads more children’s books than me.’

*In Amsterdam duurt het op dit moment [*ˈmɛt*] heel lang.*
‘In Amsterdam it takes very long at this moment.’

Normaal [nɔrˈmal] ‘normally’

*Maar waar wordt dit normaal [*nəˈmal*] voor gebruikt?*
‘But what is this normally used for?’

Partij [pɑrˈtɛi] ‘batch’

*Als ik iets koop dan moet het maximaal een partij [*pəˈtɛi*] van 75 stuks zijn.*
‘If I buy anything then it has to be maximally a batch of 75 pieces.’

Partijen [pɑrˈtɛiən] ‘batches’

*In het verleden heb je vrij forse partijen [*pəˈtɛi*] afgenomen.*
‘In the past you bought quite large batches.’

Principe [prɪnˈsipə] 'principle'

*Ik kan eh in principe [*pəˈsipə*] gaan wanneer ik wil.*
'I can eh in principle go whenever I want to.'

*De dingen die je meet zijn in principe [*pəˈsipə*] makkelijker.*
'The things that you measure are in principle easier.'

*Ik voel me daar in principe [*pˈsip*] ook helemaal niet bij thuis.*
'In principle I really do not feel comfortable there.'

*Boeken die er in principe [*pˈsipə*] hadden kunnen zijn.*
'Books that in principle could have been there.'

Problemen [proˈblemən] 'problems'

*Met die slaapzakken ook in het verleden wel problemen [*ˈplemə*] gehad eigenlijk.*
'With those sleeping bags in the past also had problems actually.'

Procent [proˈsɛnt] 'percent'

*Nee dan wil ik toch echt 25 procent [*pˈsɛnt*] korting op die eerste prijs van je hebben.*
'No then I really want to have a 25 percent discount on that first price of yours.'

*Dan kan ik daar wel eh twintig procent [*pəˈsɛnt*] afkrijgen denk ik.*
'I think I can get eh a twenty percent discount.'

Programma [proˈχrɑmɑ] 'programme'

*Maar dat vind ik een slecht programma [*pəˈχɑmɑ*] eigenlijk.*
'But I consider that a bad programme actually.'

Project [proˈjɛkt] 'project'

*Hij is weer met een ander Europees project [*pəˈjɛk*] bezig.*
'He is working on a different European project again.'

Vakantie [vɑˈkɑnsi] 'holiday'

*Echt het idee van op vakantie [*fˈkɑnt*] misschien een auto huren ofzo.*
'Really the idea of maybe renting a car during the holidays or something.'

Vandaag [vɑnˈdaχ] ‘today’

Ik had *vandaag* *[*fəˈda*] weer een auto geleend.*
‘Today I borrowed a car again.’

Vanzelf [vɑnˈzɛlf] ‘by itself’

Dat het opeens *vanzelf* *[*vəˈzɛlf*] gaat.*
‘That suddenly it goes automatically.’

Verdieping [vərˈdipɪŋ] ‘floor’

Op die *verdieping* *[*fəˈnipɪŋ*] ergens op de Keizersgracht.*
‘On that floor somewhere along the Keizersgracht.’

Verhaal [vərˈhal] ‘story’

Ik zal het *verhaal* *[*ˈfal*] vertellen ja.*
‘I will tell the story, yes.’

Het moraal van het *verhaal* *[*ˈfal*] kwam er voor mij op neer van*
‘The moral of the story to me was that’

Verjaardag [vərˈjardɑχ] ‘birthday’

Jij was niet op de *verjaardag* *[*fəˈjad*] van Jet, toch?*
‘You were not present at Jet’s birthday, were you?’

Ik vind een *verjaardag* *[*fəˈjar*] is nog wel leuk om te doen.*
‘I think a birthday is still enjoyable to do.’

Verkeerd [vərˈkert] ‘wrong’

Hij had toch wel eh een *verkeerd* *[*fˈkɪt*] tentje of iets dergelijks.*
‘He did have a eh wrong tent or something.’

Verkeerde [vərˈkerdə] ‘wrong’

Dat ze een grote kans hebben om eh het *verkeerde* *[*fəˈkɪdə*] pad op te gaan.*
‘That they run a larger risk to eh go off the track.’

Verkoopt [vərˈkopt] ‘sell’

*Want jij verkoopt [*fˈkopt*] er tenslotte meer.*
‘Because after all you sell more.’

Verleden [vərˈledən] ‘past’

*Mijn oma heeft verleden [*ˈflej*] jaar voor het eerst in januari haar verjaardag gevierd.*
‘Last year, my grandmother celebrated her birthday in January for the first time.’

*Maar dit jaar ga ik niet het risico lopen, want verleden [*ˈfled*] jaar ben ik het schip in gegaan.*
‘However, this year I will not run that risk, because last year I was financially disadvantaged.’

Verloopt [vərˈlopt] ‘elapse’

*Van hoe hoe dat afscheid verloopt [*fˈlopt*] van een vakgroep.*
‘Of how how one takes leave of a research group.’

Verschillende [vərˈsχɪləndə] ‘different’

*Kan je op verschillende [*fˈsχɪln*] manieren doen.*
‘You can do that in various ways.’

*Corpus dat bestaat uit materiaal van verschillende [*fˈsχɪlə*] taalfasen, toch?*
‘Corpus that consists of materials from various phases of language development, right?’

*Ik vind wel een heleboel verschillende [*fˈsχɪləndə*] dingen leuk wat dat betreft.*
‘I like a lot of different things as far as that is concerned.’

*Steekproef te nemen van verschillende [*fˈsχɪlnə*] vakgebieden.*
‘To take a sample of different research fields.’

Vertellen [vərˈtɛlən] ‘tell’

*Dus ik kan meer vertellen [*fˈtɛlə*] wat ik wel leuk vind.*
‘So I can better tell you what I do like.’

Vervelend [vərˈvelənt] ‘annoying’

*Gewoon het idee dat je niet af en toe even kan praten over je werk vond ik heel vervelend [*ˈvelənt*], want dat had ik dus erg weinig vond ik zelf.*
‘Simply the thought that you cannot occasionally talk about your work, I considered very annoying, because I thought I had very little opportunity to do that.’

*Ik moet ze ook netjes houden, want anders is het voor jou vervelend [*ˈvent*] als ik ze*
‘I also need to keep them tidy, because otherwise it is very annoying for you if I’

Verzamelt [vərˈzamәlt] ‘collects’

*Nou hij verzamelt [*fəˈzamɔt*] al heel lang kinderboeken.*
‘Well he has been collecting children’s books for a very long time.’

Voornamelijk [vorˈnamələk] ‘mainly’

*Het ging die ene persoon dan ook voornamelijk [*vˈnamək*] om het programma.*
‘It concerned that one person who mainly for the programme.’

*Je komt weleens langs en voornamelijk [*vˈnamə*] zit je in de kroeg.*
‘You occassionally pass by and you are mainly spending time in the pub.’

Waarschijnlijk [warˈsχɛinlək] ‘probably’

*En jullie hebben waarschijnlijk [*wəˈsχɛik*] alleen al het oud-Engelse deel eruit gevist.*
‘And you have probably only extracted the Old English part.’

Zoals [zoˈwɑls] ‘such as’

*Net zoals [*ˈzəz*] wat ik een keertje bij de Albert Heijn had.*
‘Just like what I once had at the Albert Heijn.’

*Te vieren zoals [*ˈzɔz*] dat gebruikelijk is.*
‘To celebrate it as usual.’

Summary and conclusions

Chapter 6

This dissertation investigated the contribution of the semantic/syntactic and acoustic information from the context and the acoustic information from reduced pronunciation variants to the comprehension of these variants. In this chapter, we review the findings presented in this dissertation, and draw conclusions on the basis of these findings. We will not discuss our results in chronological order (by chapter), but we will categorise our results by topic.

6.1 The acoustic context

Chapter 2 investigated the role of acoustic cues in the context in the processing of reduced pronunciation variants. We conducted four cloze tasks, in which participants were presented with either the preceding or the preceding and following context of reduced target words. This contextual information was presented either visually, in the form of orthographic transcriptions, or auditorily (i.e. the actual speech materials were played). Importantly, participants never heard or saw the reduced target words themselves (i.e. listeners were only presented with their contexts), and we investigated to what extent participants could use the context to predict these target words. Participants' lexical search space was reduced to four lexical items (they had four options to choose from). These four options had different initial sounds, and therefore transitional cues in surrounding segments may be useful to predict the reduced target words. Participants' performance increased by approximately six percent if they heard (rather than read) the preceding context, and by approximately eleven percent if they heard (rather than read) the preceding and following context. These results show that listeners can use acoustic cues from the context in conversational speech, in addition to semantic/syntactic cues, despite the high speech rate and the fact that segments are realised with less articulatory effort than in laboratory speech, or they are not realised at all.

6.2 The semantic/syntactic context

We investigated the role of semantic/syntactic information in the context in Chapters 2-4. In Chapter 2, we found that listeners can use semantic/syntactic contextual information to predict adjectives and discourse markers on the basis of their sentential context. Participants who were provided with four options to choose from could predict the reduced target words on the basis of the preceding semantic/syntactic context in 33% of the trials, and on the basis of the full semantic/syntactic context in 37% of the trials.

Only a small proportion of these context effects can be ascribed to just the probability of a word given its directly surrounding words (e.g. word bigram or trigram probability). This shows that we need more sophisticated measures of contextual probability, for example based on Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnas, Landauer, and Harshman, 1990), to capture listeners' sensitivity to semantic/syntactic contextual information. Hence, in Chapters 3-4, we measured the semantic relatedness of words with their preceding words by means of Latent Semantic Analysis.

In Chapter 3, we used the auditory lexical decision task with implicit semantic priming. Native listeners heard unreduced and reduced words that varied in their semantic relatedness with their preceding words, and we investigated semantic effects during the processing of these unreduced and reduced words. Our study focused on subtle differences in words' semantic relatedness (i.e. all prime-target pairs were semantically related, rather than related or unrelated), and we estimated the extent to which words were related by means of Latent Semantic Analysis.

We found that listeners require more time to use the meanings of acoustically reduced words (compared to unreduced words) to recognise upcoming words. Once reduced words have been fully processed, their (subtle) semantic information can facilitate the recognition of upcoming (reduced) words to the same extent as the semantic information from unreduced words. Further, we found that if listeners had sufficient time to fully process the words, unreduced words facilitated the processing of upcoming related unreduced words more strongly than upcoming related reduced words. Possibly, listeners have more difficulty recognising reduced words following unreduced words because, unlike reduced words, completely unreduced words are typically followed by other completely unreduced words. Most importantly, our results suggest that the semantic information from reduced words can facilitate the recognition of upcoming related reduced words if these are separated by other words (e.g. discourse markers) in the sentence.

In Chapter 4, we presented the same unreduced and reduced materials to non-native listeners. Previous research suggests that non-native listeners have difficulties using semantic information to enhance their processing of upcoming words, but only if they are presented with fast, conversational speech (Bradlow and Alexander, 2007). We found that non-native listeners in general have more difficulty using semantic information than native listeners, regardless of whether words are reduced or unreduced. These divergent findings may be explained by the fact that our study focused on subtle differences in words' semantic relatedness, rather than contrasting words that are and that are not semantically related.

6.3 The acoustic properties of reduced words

Chapter 5 tested the contribution of the acoustic properties of reduced pronunciation variants with unstressed initial syllables to the recognition of these variants. Participants heard the preceding context, some segmental information from a reduced target word, a square wave (indicating the location of the missing part of the target word), and the following context. They were instructed to write down the target word. The segmental information that participants heard consisted of the initial consonant cluster, the initial consonant cluster and the first vowel, or the initial consonant cluster, the first vowel, and the subsequent consonant cluster.

Not surprisingly, participants better recognised the reduced target words, the more segmental information they heard. Moreover, participants benefited more from hearing consonants than vowels of the reduced target words, suggesting that consonants play a larger role in lexical access, in line with previous research (e.g. Cutler, Sebastiań-Galleś, Soler-Vilageliu, and Van Ooijen, 2000; Bonatti, Peña, Nespor, Mehler, 2005; Mehler, Peña, Nespor, and Bonatti, 2006). Interestingly, we found that listeners required less segmental information to recognise target words with first vowel deletion than target words without first vowel deletion. Apparently, the absence of the first, unstressed vowel is less important than the presence of additional consonants.

This finding suggests that reductions need not always hinder the word recognition process, contrary to previous research that nearly all used read-aloud words in isolation or simple sentences (e.g. Ernestus and Baayen, 2007; Ranbom and Connine, 2007; Tucker and Warner, 2007; Tucker, in press). In fact, listeners may sometimes benefit from these reductions if, as a consequence, words become unique more rapidly.

Chapter 5 also investigated the role of the duration of a reduced word in its recognition. In one experiment, participants heard a square wave with the duration of the missing part of the reduced word. Surprisingly, listeners found this durational information misleading, since they performed worse and provided shorter answers than when the square wave had a fixed duration. This result indicates that the duration of the square wave made listeners expect target words with fewer segments in their citation forms than these words actually had. Kemps, Ernestus, and Schreuder (2004) found that listeners unconsciously restore missing segments in their processing of reduced words. In line with Kemps and colleagues, our findings show that listeners are unaware of the reductions that occur in spontaneous speech.

6.4 The interaction between semantic/syntactic and acoustic information

We investigated the interaction between the roles of semantic/syntactic and acoustic information during the processing of reduced pronunciation variants in Chapters 2 and 5. Chapter 2 investigated the relationship between a word's degree of reduction and how easily listeners predict the word on the basis of its context. Surprisingly, this chapter showed that listeners have more difficulty predicting words on the basis of their preceding context alone, the higher their degree of reduction. This finding is unexpected given a listener-driven account of speech reduction suggesting that speakers reduce especially those words that are highly predictable for the listener (Ernestus and Baayen, 2007). Apparently, certain word types (e.g. adverbs and discourse markers) may be reduced more strongly, the less predictable they are. It seems that these types of words are not so much reduced due to their high predictability, but in order to decrease the articulatory effort for the speaker.

Chapter 2 also tested the interaction between semantic/syntactic information (based on a word's co-occurrence frequency with its surrounding words) and acoustic contextual information in the recognition of reduced pronunciation variants. We found that, when the semantic/syntactic and acoustic cues in the context provide conflicting information regarding the identity of upcoming reduced words, listeners rely more heavily on acoustic cues than on semantic/syntactic cues to predict these words. This is a remarkable finding, given the tremendous amount of acoustic variability in conversational speech. Apparently, listeners pay more attention to acoustic information than to semantic/syntactic information in the context, despite this large amount of acoustic variability.

Chapter 5 tested the contribution of the semantic/syntactic information to the recognition of reduced words as a function of how much acoustic information listeners have heard from these reduced words themselves. This chapter showed that listeners only relied on contextual probabilities (measured by word co-occurrence frequencies) if they heard little or no acoustic information from the reduced target words, or if they found the acoustic information confusing. Similar to Chapter 2, this finding demonstrates that listeners rely more heavily on acoustic cues than on probabilistic cues from the context when trying to recognise reduced pronunciation variants. Apparently, listeners especially rely on words' bigram and trigram probabilities if they have little or no acoustic information about words to rely on.

6.5 Implications for models of speech comprehension

Psycholinguistic models of speech comprehension need to account for the effects found in this dissertation. At one end of a continuum, traditional abstractionist models (e.g. Marslen-Wilson and Welsh, 1978; Norris, 1994) assume that, for each word, the mental lexicon contains only a single lexical representation (i.e. the word's citation form). At the other end, exemplar-based models (e.g. Goldinger, 1998) assume that the mental lexicon contains multiple representations for each word, including representations for unreduced as well as reduced pronunciation variants.

First, these models should explain the effects of the acoustic information in the context on the recognition of reduced pronunciation variants. Chapter 2 of this dissertation showed that listeners can rely on acoustic cues in the context to facilitate their recognition of reduced variants. These acoustic cues probably include information from the "local" acoustic context, such as progressive and regressive assimilation, but also acoustic cues in the broader context, such as R-resonance, since the latter appears to be especially important in the processing of conversational speech (Heinrich and Hawkins, 2009). Speech comprehension models that assume a prelexical level of processing (e.g. Shortlist B; Norris and McQueen, 2008) typically use a diphone database to capture the transitions between speech sounds, and these models therefore can only account for "local" acoustic context effects. These models thus cannot explain more widespread acoustic context effects. These effects can be captured by exemplar-based models if they assume that the mental lexicon contains information about the contexts that lexical representations occurred in (Hawkins and Smith, 2001).

Second, psycholinguistic models need to explain the contribution of the segmental information from the reduced variants themselves to their recognition. Abstractionist models explain the finding that listeners cannot recognise reduced pronunciation variants in isolation by assuming that these variants are recognised by means of context-dependent reconstruction mechanisms, and these mechanisms are possibly aided by subphonemic cues signalling the missing segments. If we assume that listeners reconstruct missing segments, these missing segments may enhance word recognition if, as a consequence, words become unique more quickly.

Exemplar-based models assume that the different pronunciation variants are stored in the mental lexicon, which explains why listeners require less segmental information if reduced variants have earlier uniqueness points. These models account for the finding that listeners require context to recognise reduced pronunciation variants by assuming that lexical representations are stored together with the contexts they occurred in (Hawkins and Smith, 2001).

Finally, psycholinguistic models need to account for the role of the semantic/syntactic information in the context in the recognition of reduced variants. Roughly two accounts have been proposed for how semantic/syntactic information facilitates the word recognition process, both of which can be integrated into either abstractionist or exemplar-based models of speech comprehension. On the one hand, Van Petten and Kutas (1990) claim that predictable words are more easily integrated into their preceding context. Listeners may be able to predict a word based on its context by comparing its ease of integration to that of other lexical candidates. On the other hand, Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) claim that listeners use contextual information to directly narrow down their lexical search space. Both accounts can explain the semantic/syntactic context effects found in Chapters 2 and 5 of this dissertation, and our semantic priming effects (in interaction with word frequency) found in Chapters 3. It is less clear how these accounts can explain our finding that acoustic information plays a larger role in speech comprehension than semantic/syntactic information (Chapters 2 and 5 of this dissertation). Further research is required to investigate how this interaction can be incorporated into these models.

Interestingly, Chapter 3 found semantic interference effects for words with very high word frequencies: Listeners processed words with very high frequencies more slowly if they were more strongly semantically related to their preceding words. This interference effect can easily be accounted for by connectionist models of speech comprehension like TRACE (McClelland and Elman, 1986). In these models, words with very high frequencies have very high resting activation levels, and these levels may become even higher through the spreading of activation from semantically related words. These activation levels may then become higher than those of the prime words, inhibiting the recognition of these prime words. Consequently, listeners need to suppress these highly frequent, semantically related words before they can recognise the prime words. This suppression can lead to activation levels lower than the words' resting activation levels, as a result of which these words become more difficult to recognise.

On the whole, the results of this dissertation suggest that abstractionist models need to be adjusted before they can deal with conversational speech. Exemplar-based models, on the other hand, still need to specify how much information is stored in the mental lexicon, and which types of information listeners can use to recognise words. Future research should show what adjustments are necessary for abstractionist and exemplar-based models, in order to deal with the comprehension of conversational speech.

6.6 Methodology

As mentioned in the Introduction, as a byproduct of its main research question, this dissertation investigated to what extent the comprehension of reduced pronunciation variants in conversational speech can be studied by means of existing psycholinguistic paradigms.

First of all, we used an unaltered version of the lexical decision task, with implicit semantic priming, to get a first impression of the influence of reduction on the role of semantic information during word recognition. We used this paradigm in Chapters 3 and 4 of this dissertation, investigating semantic priming in native and non-native speech processing. This paradigm can provide interesting insights into the effects of reduction on the time-course of lexical activation and the activation of a word's semantic information (as also shown by, for example, Sumner and Samuel, 2005). Since listeners were presented with isolated words, additional research should indicate whether these findings hold for reduced words embedded in their sentence contexts.

Second, we modified the cloze task, first described by Taylor (1953). In a typical cloze task, participants are presented with text fragments in which one or several words are missing, and they are instructed to fill in the missing word(s). In our auditory cloze tasks, participants heard a short familiarisation phase for each speaker to get used to the speaker's (voice) characteristics and speaking style. We then presented participants auditorily with the context of reduced target words, and a square wave indicated the location of these target words. We used a square wave (rather than just silence) to indicate the target word's location, because truncated speech sounds highly unnatural and may lead listeners to perceive an inserted labial or plosive consonant (e.g. Pols and Schouten, 1978), whereas square waves are less often misperceived as speech sounds (Warner, 1998). Since the reduced words were low-predictability words (e.g. adjectives and adverbs), participants were presented with four plausible options on a computer screen and were asked to select the target word that was missing. This method allowed us to investigate to what extent semantic/syntactic and acoustic information from the context can contribute to the recognition of reduced words.

Third, we adapted the classic auditory gating task. In this task, introduced by Grosjean (1980), listeners are presented with incremental portions of target words (i.e. the gates) embedded in their contexts. In most gating studies, the gates have a constant size (e.g. 50 ms). In conversational speech, segments may be shorter or completely missing, and gates with fixed durations may contain more or less segmental information depending on the word's degree of reduction. Hence, we used a modified version of the auditory gating paradigm in Chapter 5 of this thesis. In our version of this paradigm, we placed the gates at the end of the first realised consonant (cluster), the first realised vowel, and the second realised consonant (cluster), thereby controlling the amount of segmental information provided in each gate (see Cutler and Otake, 1999, for a similar approach, using the gating paradigm to study the role of pitch-accent information in spoken word recognition). This method allowed us to compare listeners' recognition of words with and without first vowel deletion, providing useful insights into the effects of reduction on the word recognition process.

This dissertation investigated listeners' processing of reduced pronunciation variants by means of one online experimental paradigm (i.e. auditory lexical decision) with isolated words and two offline paradigms (i.e. gating and the cloze task) with materials extracted from corpora of conversational speech. We need sophisticated studies using (novel) online techniques that allow researchers to present reduced pronunciation variants within their natural contexts. Such studies can provide more insights into the way listeners process conversational speech.

6.7 Topics for further research

The chapters in this dissertation have illustrated that listeners combine a variety of resources to recognise reduced pronunciation variants.

In the current dissertation, we investigated the comprehension of reduced variants in settings where listeners do not have to formulate responses while processing conversational speech. It would be interesting to investigate how listeners process this type of speech in more interactive settings. Such demanding listening situations may lead listeners to rely more heavily on contextual information, since this information may allow listeners to predict certain words (e.g. nouns) already prior to hearing their first sounds.

Moreover, studies are necessary that investigate how listeners deal with reduced pronunciation variants under noisy conditions, given that conversational speech is often accompanied by background noise (e.g. other people talking in the background), and listeners are nevertheless able to cope with reductions in these adverse listening conditions. Listeners may turn out not only to rely on information from the auditory modality, but perhaps also the visual modality (e.g. lip movements), during their processing of conversational speech.

Finally, this dissertation has demonstrated that established psycholinguistic experimental methods can be adjusted for studying at least some aspects of the recognition of reduced pronunciation variants. Future studies should try to adapt also online experimental paradigms for the purpose of studying the comprehension of reduced pronunciation variants. Only by observing the comprehension of reduced variants from different viewpoints can we obtain a better picture of how listeners deal with these variants.

6.8 Concluding remarks

The research presented in this dissertation aimed to provide insights into the processes underlying the comprehension of reduced words, and, in this way, to contribute to the further development of speech comprehension models. It has become clear that the comprehension of reduced words involves a complex interplay between the semantic/syntactic cues from the context and acoustic cues from the context as well as the reduced words themselves. Listeners can rely on both semantic/syntactic and acoustic information from the context to predict upcoming reduced words, and once segmental information from reduced words becomes available, listeners can integrate this segmental information, and reduction does not appear to hinder the word recognition process.

Bibliography

Andruski, J., Blumstein, S., and Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52:163–187.

Aydelott, J. and Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, 19:29–56.

Aydelott, J., Dick, F., and Mills, D. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, 43:454–464.

Aydelott-Utman, J., Blumstein, S., and Burton, M. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, 62:1297–1311.

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

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

Becker, C. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 5:252–259.

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

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., and Gildea, D. (1999). Forms of English function words – effects of disfluencies, turn position, age and sex, and predictability. In *Proceedings of the 14th International Congress of Phonetic Sciences*, pages 395–398, San Francisco, California.

Blank, M. and Foss, D. (1978). Semantic facilitation and lexical access during sentence processing. *Memory and Cognition*, 6:644–652.

Bonatti, L., Peña, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16:451–459.

Bradlow, A. R. and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121(4):2339–2349.

Brouwer, S. (2010). *Processing strongly reduced forms in casual speech.* PhD dissertation, Radboud University Nijmegen, Nijmegen.

Brouwer, S., Mitterer, H., and Huettig, F. (in press). Speech reductions change the dynamics in competition during spoken word recognition. *Language and Cognitive Processes*.

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

Carroll, J. B. (1967). On sampling from a lognormal model of word frequency distribution. In Kučera, H. and Francis, W. N., editors, *Computational Analysis of Present-Day American English*, pages 406–424. Brown University Press, Providence.

Cutler, A. and Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, 105:1877–1888.

Cutler, A., Sebastiań-Galleś, N., Soler-Vilageliu, O., and Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory and Cognition*, 28:746–755.

Davies, M. (2008). The corpus of Contemporary American English (coca): 385 million words, 1990-present. *[On-line], Available: http://www.americancorpus.org.*

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28:218–244.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society For Information Science*, 41:391–407.

Donnenwerth-Nolan, S., Tanenhaus, M., and Seidenberg, M. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 7:170–180.

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

Ernestus, M. (in press). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*.

Ernestus, M. and Baayen, R. (2007). The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration and frequency of occurrence. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 773–776, Saarbruecken, Germany.

Ernestus, M. and Baayen, R. (2011). Corpora and exemplars in phonology. In Goldsmith, J., Riggle, J., and Yu, A., editors, *The Handbook of Phonological Theory (2nd ed.)*. Wiley-Blackwell, Oxford.

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

Gaskell, G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31:447–463.

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

Gow, D. (2002). Does English coronal place assimilation create lexical amibiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28:163–179.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28:267–283.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, 38(4):299–310.

Grosjean, F. and Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25:135–155.

Hahne, A. and Friederici, A. D. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4:123–141.

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

Hawkins, S. and Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics - Rivista di Linguistica*, 13:99–188.

Hawkins, S. and Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, 22:493–511.

Heinrich, A. and Hawkins, S. (2009). Effect of r-resonance information on intelligibility. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association*, pages 804–807, Brighton.

Hooper, J. B. (1976). *An introduction to natural generative phonology*. Academic Press, INC, New York.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speech of recognition. *Journal of Experimental Psychology*, 48:106–112.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 29:296–305.

Imai, S., Walley, A. C., and Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America*, 117(2):896–907.

Jaeger, F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59:434–446.

Janse, E. and Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, 39:330–343.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, pages 29–54, Tokyo, Japan. The National International Institute for Japanese Language.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., and Raymon, W. (1998). Reduction of English function words in Switchboard. In *Proceedings of ICSLP-98*, pages 3111–3114.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. John Benjamins, Amsterdam.

Kelly, J. and Local, J. (1986). Long-domain resonance patterns in English. In *Proceedings of IEE International Conference on Speech Input/Output: Techniques and Applications*, pages 304–309, London.

Kemps, R., Ernestus, M., Schreuder, R., and Baayen, R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 19:117–127.

Kemps, R., Wurm, L., Ernestus, M., Schreuder, R., and Baayen, R. (2005). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, 20:43–73.

Landauer, T. (1998). Lsa @ cu boulder. *[On-line], Available: http://lsa.colorado.edu.*

Landauer, T. and Dumais, S. (1997a). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104:211–240.

Landauer, T. and Dumais, S. (1997b). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25:259–284.

Marslen-Wilson, W. and Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, 8:1–71.

Marslen-Wilson, W. and Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101:653–675.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25:71–102.

Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10:29–63.

Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40:686–693.

McClelland, J. L. and Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.

McDonald, S. and Shillcock, R. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6):648–652.

McLennan, C. T. (2006). The Time Course of Variability Effects in the Perception of Spoken Language: Changes Across the Lifespan. *Language and Speech*, 49:113–125.

Mehler, J., Peña, M., Nespor, M., and Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, 42:846–854.

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

Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90:227–234.

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

Morton, J. and Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, 15:43–51.

Newbigging, P. (1961). The perceptual reintegration of frequent and infrequent words. *Canadian Journal of Psychology*, 15:123–132.

Nooteboom, S. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, 67:276–287.

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

Norris, D. and Cutler, A. (1985). Juncture detection. *Linguistics*, 23:689–705.

Norris, D., Cutler, A., McQueen, J., and Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, 53(2):146–193.

Norris, D. and McQueen, J. (2008). Shortlist B: A bayesian model of continuous speech recognition. *Psychological review*, 115:357–395.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52:189–234.

Oostdijk, N. (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter*, 5:4–8.

Oostdijk, N. (2002). The design of the Spoken Dutch Corpus. In Peters, P., Collins, P., and Smith, A., editors, *New Frontiers of Corpus Research*, pages 105–112. Rodopi, Amsterdam.

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

Pols, L. and Schouten, M. (1978). Identification of deleted consonants. *Journal of the Acoustical Society of America*, 64:1333–1337.

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

Rayner, K., Ashby, J., and Pollatsek, A. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30:720–732.

Salverda, A., Dahan, D., and McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90:51–89.

Savin, H. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35:200–206.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-prime user's guide. Psychology software tools.

Schneider, W., Eschman, A., and Zuccolotto, A. (2007). E-prime getting started guide. Psychology software tools.

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

Shockey, L. (2003). *Sound patterns of spoken English.* Oxford: Blackwell Publishing, Malden.

Soloman, R. and Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, 43:195–201.

Sumner, M. and Samuel, A. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52:322–338.

Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30:415–433.

Tucker, B. (in press). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*.

Tucker, B. V. and Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 1949–1952, Saarbrücken, Germany.

Van Berkum, J., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31:443–467.

Van de Ven, M., Tucker, B., and Ernestus, M. (2009). Semantic context effects in the recognition of acoustically unreduced and reduced words. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association*, pages 1867–1870, Brighton.

Van den Brink, D., Brown, C., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influence during spoken-word recognition. *Journal of Cognitive Neuroscience*, 13:967–985.

Van den Brink, D., Brown, C., and Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2):364–372.

Van Petten, C. and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18:380–393.

Warner, N. (1998). *The Role of Dynamic Cues in Speech Perception, Spoken Word Recognition, and Phonological Universals*. PhD. dissertation University of California, Berkeley.

Warner, N. (to appear). Methods for studying spontaneous speech. In Cohn, A., Fougeron, C., and Huffman, M., editors, *Handbook of Laboratory Phonology*.

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

Warren, R. and Sherman, G. (1974). Phonemic restoration based on subsequent context. *Perception and psychophysics*, 16:150–156.

Warren, R. and Warren, R. (1970). Auditory illusions and confusions. *Scientific American*, 223:30–36.

Zipf, G. K. (1935). *The Psycho-Biology of Language*. Houghton Mifflin, Boston.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32:25–64.

Summary

Introduction

The majority of research on human speech processing focuses on the processing of carefully pronounced isolated words or sentences. This type of speech differs fundamentally from the type of speech listeners encounter in everyday listening situations. During informal conversations (casual speech), speech rate is much higher than in "careful" speech, resulting in hesitations, and in absent sounds or complete syllables. For example, the English words *yesterday* and *probably* may sound like *yesyay* and *proly*. It is unclear how listeners deal with these reduced pronunciation variants. Previous research indicates that listeners require contextual information to recognise reduced pronunciation variants (Ernestus, Baayen, and Schreuder, 2002). This dissertation aims to increase our understanding of how listeners process spoken language, by investigating which types of information contribute to the recognition of these reduced variants, and to what extent. The findings presented in this dissertation can be used to improve current models of speech recognition.

In this dissertation, we investigated the roles of three types of information during the recognition of casual speech:

1. The role of sentence structure and/or lexical information. Previous research on the processing of careful speech has shown that listeners recognise words more quickly, the more predictable they are on the basis of their preceding context. This dissertation investigated whether listeners can use such cues during their processing of casual speech, since this type of speech is characterised by a high speech rate and words (also in the context) are frequently pronounced less clearly (for example *see* preceded by *supposed to* pronounced like *susu*, where *u* stands for the schwa, the first vowel in *supposed*).

2. The role of the acoustic information in the context. For example, listeners could use prosodic information and assimilation (e.g. *good paper* pronounced like *goob paper*) to predict upcoming reduced pronunciation variants.

3. The role of acoustic information in the reduced variant itself. Although listeners require context to recognise reduced pronunciation variants, context alone may not be sufficient to recognise these variants in many cases. We therefore investigated the

contribution of the acoustic cues in reduced pronunciation variants (e.g. the initial consonants/vowels) presented together with the context they occurred in.

Results of this dissertation

The three different topics are discussed in the four experimental chapters of this dissertation. We now provide a short summary of the most important results per topic.

First of all, we expected that listeners could use sentence structure and/or lexical information from the context to predict reduced pronunciation variants. Our results confirm that these types of contextual information contribute to the recognition of these reduced variants, for words that do not contribute significantly to the meaning of the sentence in which they occur (e.g. the Dutch discourse markers *eigenlijk* ‘actually’ and *natuurlijk* ‘of course’, which may be pronounced like *eik* and *tuuk*). Listeners do not merely use information from the neighbouring words (i.e. the “local context”), but especially also from the broader, sentence context.

Nevertheless, these context effects appear to operate differently during the processing of casual speech. Previous research using “careful” speech has shown that listeners more quickly recognise a word if it is preceded by a word that is related in meaning. For example, the noun *cat* is recognised more quickly after *mouse* than after *house*. Our results suggest that listeners require more time to use preceding related words in casual speech to facilitate the recognition of an upcoming word if the preceding word is acoustically reduced. If listeners have sufficient time to process reduced pronunciation variants, reduced variants can facilitate the recognition of upcoming words to a similar extent as unreduced variants. Moreover, listeners appear to have more difficulty using information from unreduced pronunciation variants to recognise upcoming reduced variants.

We have also investigated to what extent non-native listeners can use lexical information to recognise reduced pronunciation variants. Listeners probably rely more heavily on contextual information while listening to a non-native language, especially if they are presented with fast, casual speech. Compared to native listeners, non-native listeners appear to have more difficulty using the preceding context to recognise an upcoming word, both for unreduced and reduced words. This result indicates that listeners are less sensitive to the meaning relationships between words in a non-native language, although these listeners had acquired strong communicative skills in the non-native language. This finding suggests that non-native listeners have more difficulties understanding casual conversations.

The effects of lexical information during the processing of casual speech do not only arrive later in time, but they are also relatively weak. Our results illustrate that listeners rely more heavily on acoustic information than on lexical information and/or information conveyed by sentence structure if these two sources of information are contradictory. This is an unexpected finding, given that the context is pronounced quickly and less clearly. Casual speech is also characterised by a large amount of variation, both in speech rate and the sounds that are being produced. Listeners nevertheless seem to rely heavily on the acoustic information when they try to predict the upcoming word in a sentence.

Our results indicate that contextual information plays an important role during the recognition of the reduced pronunciation variants that occur in casual speech. One may therefore assume that the role of acoustic information from reduced variants themselves in the recognition of these variants is only trivial. Nothing is farther from the truth: Our results show that listeners can predict reduced pronunciation variants in no more than 30% of the cases. In the remaining cases, listeners apparently also require the acoustic information from the reduced pronunciation variants themselves. Listeners use especially the consonants (rather than the vowels) from reduced pronunciation variants to recognise these variants. In fact, listeners more quickly recognise reduced variants if the first, unstressed vowel is missing. Apparently, listeners can compensate these missing vowels by means of the following consonants that become available more quickly.

In one of the experiments participants were provided with durational information about the reduced pronunciation variants. Surprisingly, participants in this experiment performed much worse and gave shorter responses than in the other experiments. Participants apparently expected relatively short words on the basis of the durational information that was presented to them. This result shows that listeners are not aware of the reduced pronunciation variants that occur in casual speech.

Conclusion

The research presented in this dissertation has provided new insights into the way listeners recognise reduced pronunciation variants in everyday listening situations. Our results have shown that listeners use different types of information from the context and from the reduced variants themselves, although lexical information from the context seems to become available later in time if the context is acoustically reduced.

Samenvatting

Introductie

Veruit het meeste onderzoek naar de verwerking van gesproken taal richt zich op de verwerking van duidelijk uitgesproken woorden of zinnen. Dit type spraak verschilt echter fundamenteel van het type spraak dat luisteraars doorgaans voorgeschoteld krijgen. Tijdens informele conversaties (spontane spraak) is de spreeksnelheid over het algemeen veel hoger dan in “nette” spraak, wat zorgt voor weggelaten klanken en zelfs volledig weggelaten lettergrepen. Zo kan het fragment *blijf ik ongeveer op hetzelfde* klinken als *blijfk ogfe opt selde* (Ernestus, 2000). Het is vooralsnog onduidelijk hoe luisteraars omgaan met dit soort verkorte uitspraakvarianten. Eerder onderzoek laat zien dat luisteraars de zinscontext nodig hebben om verkorte uitspraaksvarianten te herkennen (Ernestus, Baayen en Schreuder, 2002). In dit proefschrift hebben we onderzocht welke informatie bijdraagt aan de herkenning van deze uitspraakvarianten, en in welke mate. Deze kennis is van belang om beter te kunnen begrijpen hoe mensen gesproken taal verwerken en voor het verbeteren van huidige spraakherkenningsmodellen.

In dit proefschrift hebben we onderzoek gedaan naar de rol van drie typen informatie tijdens de herkenning van spontane spraak:

1. De rol van zinsstructuur en/of lexicale informatie. Eerder onderzoek heeft laten zien dat luisteraars duidelijk uitgesproken woorden die voorspelbaar zijn op basis van de voorafgaande context sneller kunnen herkennen. Dit proefschrift heeft onderzocht of luisteraars dergelijke informatie kunnen gebruiken tijdens de verwerking van spontane spraak, ondanks het feit dat veel woorden (en dus ook relevante woorden in de context) snel en minder duidelijk zijn uitgesproken (bijvoorbeeld *project* voorafgegaan door *Europees* uitgesproken als *upees*).

2. De rol van akoestische informatie in de context. Luisteraars zouden bijvoorbeeld intonatie en assimilatie (bijvoorbeeld *geen bal* uitgesproken als *geem bal*) kunnen gebruiken om verkorte uitspraakvarianten te voorspellen.

3. De rol van akoestische informatie in het woord zelf. Hoewel luisteraars context nodig hebben om verkorte uitspraakvarianten te herkennen, is context alleen in veel gevallen waarschijnlijk niet genoeg. We kijken daarom naar de bijdrage van de

akoestische informatie in de verkorte uitspraakvarianten zelf (bijvoorbeeld de eerste medeklinker en klinker in *selde* voor *hetzelfde*), wanneer deze in context worden aangeboden.

Resultaten in dit proefschrift

De drie verschillende onderwerpen kwamen in de vier experimentele hoofdstukken van dit proefschrift aan bod. Nu volgt een beknopte samenvatting van de belangrijkste resultaten per onderwerp.

Allereerst verwachtten we dat luisteraars zinsstructuur en/of lexicale informatie uit de context kunnen gebruiken om verkorte uitspraakvarianten te voorspellen. Onze resultaten bevestigen dat deze contextuele informatie bijdraagt aan de herkenning van deze verkorte uitspraakvarianten, zelfs wanneer deze woorden relatief weinig toevoegen aan de betekenis van de zin waarin zij voorkomen (bijvoorbeeld *eigenlijk* en *natuurlijk*, uitgesproken als *eik* en *tuuk*). Luisteraars gebruiken niet alleen de direct aangrenzende woorden (de "lokale context"), maar vooral ook de bredere zinscontext.

Toch lijken deze contexteffecten anders te werken tijdens de verwerking van spontane spraak. Eerder onderzoek met "nette spraak" laat zien dat luisteraars een woord sneller herkennen als het wordt voorafgegaan door een woord met een verwante betekenis. Het zelfstandig naamwoord *kat* wordt bijvoorbeeld sneller herkend na *muis* dan na *ruis*. Onze resultaten suggereren dat luisteraars meer tijd nodig hebben om voorafgaande betekenisverwante woorden in spontane spraak te gebruiken om een volgend woord te herkennen, indien het voorafgaande woord verkort is uitgesproken. Als luisteraars voldoende tijd hebben om verkort uitgesproken woorden te verwerken, dan kunnen zij deze woorden even goed gebruiken bij het herkennen van volgende woorden als duidelijk uitgesproken woorden. Bovendien kunnen luisteraars context minder goed gebruiken als de context duidelijk en het woord zelf verkort is uitgesproken.

Verder hebben we onderzocht in hoeverre niet-moedertaalsprekers lexicale informatie kunnen gebruiken bij het herkennen van verkorte uitspraakvarianten. De verwachting is dat luisteraars contextuele informatie harder nodig hebben tijdens het luisteren naar een tweede taal, vooral wanneer zij snelle, spontane spraak krijgen aangeboden. Vergeleken met moedertaalsprekers blijken niet-moedertaalsprekers minder goed in staat om de voorafgaande context te gebruiken bij het herkennen van een volgend woord, zowel voor verkort als voor duidelijk uitgesproken woorden. Dit resultaat laat zien dat luisteraars minder gevoelig zijn voor de betekenisrelaties tussen woorden in een vreemde taal, ook als deze niet-moedertaalsprekers een sterke taalvaardigheid in de tweede taal hebben verworven. Niet-moedertaalsprekers zullen hierdoor waarschijnlijk meer moeite hebben om spontane conversaties te kunnen volgen.

De effecten van lexicale informatie tijdens de verwerking van spontane spraak komen niet alleen later, maar zijn ook relatief zwak. Onze resultaten laten zien dat luisteraars vertrouwen op de akoestische informatie uit de context als deze informatie in strijd is met de zinsstructuur en lexicale informatie uit de context. Dit is onverwacht, omdat de context veelal bestaat uit snel en verkort uitgesproken woorden. Hierdoor is contextuele informatie in het algemeen minder duidelijk aanwezig. Bovendien is er veel variatie in het spraaksignaal, zowel in spreeksnelheid als in de klanken die worden geproduceerd. Niettemin lijken luisteraars sterk op deze akoestische informatie te vertrouwen bij het voorspellen van het volgende woord in de zin.

Onze resultaten laten zien dat contextuele informatie een belangrijke rol speelt tijdens de herkenning van de verkorte uitspraakvarianten die in spontane spraak voorkomen. Hierdoor zou je kunnen denken dat de akoestische eigenschappen van het woord zelf niet van belang zijn voor het herkennen van verkorte uitspraakvarianten. Niets blijkt minder waar: Onze resultaten laten zien dat luisteraars verkorte uitspraakvarianten in slechts ongeveer 30% van de gevallen kunnen voorspellen op basis van hun context. In de overige gevallen hebben luisteraars tevens akoestische informatie uit de woorden zelf nodig. Luisteraars blijken niet zozeer klinkers maar vooral medeklinkers uit verkorte uitspraakvarianten te gebruiken om deze te herkennen. Sterker nog, luisteraars herkennen verkorte uitspraakvarianten eerder als de eerste, onbeklemtoonde klinker is weggelaten door de spreker (bijvoorbeeld *flaten* in plaats van *verlaten*). Blijkbaar kunnen luisteraars deze weggelaten klinker compenseren met de daarop volgende medeklinkers die eerder beschikbaar komen.

In een van de experimenten werd de proefpersonen informatie over woordduur aangeboden. In dit experiment presteerden proefpersonen verrassend genoeg veel slechter dan in de overige experimenten. Bovendien gaven zij veel kortere antwoorden dan de proefpersonen in de andere experimenten. Blijkbaar verwachtten proefpersonen relatief korte woorden op basis van de woordduur die hen werd aangeboden. Dit laat zien dat luisteraars zich niet bewust zijn van de verkorte uitspraakvarianten die in spontane spraak voorkomen.

Conclusie

Het onderzoek in dit proefschrift heeft meer inzicht gegeven in de manier waarop luisteraars in het dagelijks leven verkort uitgesproken woorden herkennen. Onze resultaten hebben laten zien dat luisteraars hierbij gebruik maken van verschillende typen informatie uit de context en uit de verkort uitgesproken woorden zelf, hoewel betekenisinformatie uit de context later beschikbaar lijkt te zijn indien de context verkort is uitgesproken.

Acknowledgments

This dissertation is the result of research that I carried out at the Comprehension Group of the Max Planck Institute for Psycholinguistics. This dissertation could not have been completed without the support of various individuals who are gratefully acknowledged here.

First and foremost, I am profoundly indebted to my primary supervisor, Mirjam Ernestus, for teaching me all the necessary steps for becoming an independent researcher and providing invaluable support throughout my PhD. She always made time to hear the latest results of my experiments, and created a positive working environment. I am also grateful to her for allowing me to travel around the globe to attend conferences in Banff, Krakow, Brighton, Windsor (Canada), Albuquerque, and Tokyo to boot! Second, I wish to thank my promotors, Rob Schreuder and Anne Cutler. I am indebted to Rob for many stimulating discussions related and unrelated to my research, and for always being the first “participant" in all of my experiments. I am thankful to Anne Cutler for allowing me to join her research group, for her enthusiasm whenever I presented my work during the Comprehension Group meetings, and for providing insightful comments on this dissertation.

I would like to show my appreciation to the members of the reading committee, Roeland van Hout, Mark Pitt, and Rob van Son, for providing constructive comments that helped to improve this dissertation.

I also wish to express my gratitude to Francisco Torreira and Iris Hanique, my room mates during the largest part of my PhD, who agreed to be my paranymphs. It has been a real pleasure to share an office with both of you! I extend my appreciation to those people who frequently joined the lunches at the MPI, Nicole Altvater, Arina Banga, Daniel Brenner, Helen Buckler, Nienke Dijkstra, Esther Hanssen, Loes van Horen, Karen Keune, Lucinka Kočková-Amortová, Anneke Neijt, Kevin Russell, Jessamyn Schertz, Wieke Tabak, Sho Tsuji, and Laura de Vaan; these lunched formed a nice break from the everyday routine.

I am grateful to Benjamin V. Tucker for allowing me to pay several visits to his lab, and for collaborating with me on a study that resulted in Chapters 3 and 4 of this dissertation. Further, I am grateful to Ross Munro, a former undergraduate student in the Alberta Phonetics Lab, for being the “voice" for my English experiments. Thanks also to Ann Bradlow for allowing me to visit her lab in between two conference visits.

I would also like to thank my colleagues at the Max Planck Institute for providing a very comfortable and highly inspiring working environment. Specifically, I would like to thank the members of the Comprehension Group for giving valuable feedback on various presentations of my work, and the members of the Technical Group for keeping my CPU running, and for preparing the equipment for various trips.

I also wish to express my utmost appreciation to the student assistants who have done a tremendous amount of work for my PhD: Corine Baayen, Jens Frielink, Merel van Goch, Iris Hanique, Tijl Grootswagers, Luuk van der Knaap, Femke de Kruiff, David Peeters, Vitória Magalhães Piai, Henriette Raudszus, Timo Veldt, and Marieke Vermeulen.

I very much enjoyed those moments when I played table tennis in the basement of the MPI, regardless of whether I won or lost, with Victor Kuperman, Christophe Dos Santos, Raymond Bertram, and Koji Miwa. Thank you for providing these pleasant distractions!

I sincerely thank my parents, Harrie and Francien, for the opportunities they gave me, and for always being there for me. Finally, I am very grateful to Marieke, for all her patience, love, and care.

Curriculum Vitae

Marco van de Ven was born in Oss, The Netherlands, on 19 January 1982. He studied English language and culture at the Radboud University Nijmegen and the research master Linguistics at Utrecht University, both of which he completed in 2007. In 2007, he started a PhD project entitled “The role of acoustic detail and context in acoustic reductions in Dutch and English” at the Language Comprehension Group of the Max Planck Institute in Nijmegen, after being awarded a PhD stipend from the Max-Planck-Gesellschaft. In 2009-2010, he also worked as a junior lecturer at the Dutch department of the Radboud University Nijmegen. Currently, he is a lecturer at the Radboud University in Nijmegen, and at the University of Amsterdam.

Publications

Van de Ven, M. and Ernestus, M.: The role of the acoustic properties of reduced words in the processing of spontaneous speech.

Van de Ven, M., Ernestus, M., & Schreuder, R. (submitted): Predicting Acoustically Reduced words in Spontaneous speech: The role of Semantic/Syntactic and Acoustic cues in the context.

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

Van de Ven, M. & Gussenhoven, C. (2011): On the timing of the final rise in Dutch falling-rising intonation contours. *Journal of Phonetics*, 39, 225-236.

Van de Ven, M., V. Tucker, B., & Ernestus, M. (2010). Semantic facilitation in bilingual everyday speech comprehension. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (pp. 1245-1248). Causal Productions Pty Ltd.

Van de Ven, M., V. Tucker, B., & Ernestus, M. (2009). Semantic context effects in the recognition of acoustically unreduced and reduced words. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (pp. 1867-1870). Causal Productions Pty Ltd.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking. Investigations on the time course of semantic, syntactic and phonological processing. Miranda van Turennout
2. The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming and electromagnetic midsaggital 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 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 Janssen
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. Anna Margetts
13. From speech to words. Arie van der Lugt
14. Simple and complex verbs in Jaminjung: A study of event categorization 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 Weber
17. Moving eyes and naming objects. Femke 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 H. de Jong
21. Fixed expressions and the production of idioms. Simone A. Sprenger
22. The grammatical coding of postural semantics in Goemai. Birgit Hellwig
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. Fermin Moscoso del Prado Martin
24. Contextual influences on spoken-word processing. Daniëlle van den Brink
25. Perceptual relevance of prevoicing in Dutch. Petra M. 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 Meeuwissen
28. Morphology in auditory lexical processing. Rachèl J.J.K. Kemps
29. At the 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. Marlies 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 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 B. 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-contrastive variation. Christiane Dietrich
41. Cognitive cladistics and the relativity of spatial cognition. Daniel B.M. Haun
42. The acquisition of auditory categories. Martijn Goudbeek
43. Affix reduction in spoken Dutch. Mark Pluymaekers
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. Valesca Kooijman
45. Space and iconicity in German Sign Language (DGS). Pamela 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 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 Wegener
52. Prosodic structure in speech perception and production. 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 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