Affix reduction in spoken Dutch

Probabilistic effects in production and perception

ISBN: 978-90-76203-28-7
Cover design: Hans Pluymaekers, Peggy Frijns, & Mark Pluymaekers
Printed and bound by Ponsen & Looijen bv, Wageningen

Affix reduction in spoken Dutch

Probabilistic effects in production and perception

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 donderdag 7 juni 2007
om 13.30 uur precies

door

Mark Pluymaekers

geboren op 25 januari 1980 te Meerssen

Promotor:	Prof. dr. R.H. Baayen
Co-promotor:	Dr. M.T.C. Ernestus
Manuscriptcommissie:	Prof. dr. L.W.J. Boves
	Prof. dr. L.C.W. Pols (Universiteit van Amsterdam)
	Prof. dr. V.J.J.P. van Heuven (Universiteit Leiden)
	Prof. dr. I. Plag (Universität Siegen)
	Dr. S. Gahl (University of Chicago)

Het onderzoek beschreven in deze dissertatie werd ondersteund door een subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Contents

Introduction

CHAPTER 1

Speaking is a complex cognitive activity. At the very least, it involves the activation of relevant concepts, the selection of appropriate words and sentence frames, and the transformation of these units into acoustic signals through articulation (Garrett, 1975; Stemberger, 1985; Dell, 1986; Levelt, 1989; Bock, 1995; Levelt, Roelofs, & Meyer, 1999). Speakers taking part in a conversation have even more tasks to perform. For example, they have to keep track of what has already been said. Furthermore, their utterances need to be timed such that they smoothly follow the contributions of other interlocutors (Sacks, Schegloff, & Jefferson, 1974). Since all these tasks require attention, speakers are continuously challenged to make optimal use of their limited cognitive resources.

One activity on which speakers can save considerable effort is articulation. This is due to the characteristics of the speech signal, which contains much more information than is strictly necessary. Miller and Licklider (1950) and Drullman (1995) showed that removing large amounts of acoustic information from the signal does not necessarily hinder comprehension. This does not mean, however, that speakers can randomly select which parts of the speech stream they articulate (Lindblom, 1990). Recently, two hypotheses have been formulated that make predictions about the circumstances in which speakers reduce articulatory effort. The first is the Probabilistic Reduction Hypothesis, introduced by Jurafsky, Bell, Gregory, and Raymond (2001), and the second is the Smooth Signal Redundancy Hypothesis, proposed by Aylett and Turk (2004).

Predicting articulatory reduction

Both the Probabilistic Reduction Hypothesis (PRH) and the Smooth Signal Redundancy Hypothesis (SSRH) depart from the observation that not all elements in an utterance are equally important for conveying meaning. Some words or syllables are highly predictable given their linguistic or situational context. This makes them suitable candidates for articulatory reduction, since listeners have access to other information sources that can help them identify the intended units. In other words, the PRH and the SSRH predict that linguistic units will be more reduced the more predictable they are. It should be noted that in research on reduction, the terms predictability, probability, and redundancy are often used interchangeably. This is understandable, as all three terms refer to the same basic concept. Linguistic units with a high probability of occurrence are predictable, which makes them redundant from an informational point of view.

One important indicator of redundancy is the frequency with which a word occurs in a language. The more often a particular word occurs, the less information it conveys (Shannon & Weaver, 1949). Word frequency is usually estimated by counting the number of occurrences of a word in a large language or speech corpus, although it can also be measured through subjective frequency ratings (Gernsbacher, 1984; Balota, Pilotti, & Cortese, 2001). Word frequency, or, in the terminology of Jurafsky et al. (2001), *prior probability*, has indeed been shown to predict articulatory reduction, in that phonemes occurring in high-frequency words are found to be shorter than the same phonemes occurring in low-frequency words (Umeda, 1977; Van Coile, 1987; Jurafsky et al., 2001).

A word can also be predictable because it has occurred earlier in the conversation. Since conversations tend to center around a single topic, words that have been introduced are often re-used. Fowler and Housum (1987) showed that second occurrences of words in a read-aloud text were shorter and less intelligible in isolation than first occurrences. No such effect was found when two tokens of the same word occurred in a list (Fowler, 1988), or when they were separated by a major episode boundary (Fowler, Levy, & Brown, 1997). Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, and Newlands (2000) observed a clear repetition effect in dialogues, irrespective of whether the speaker or the listener had produced the first token of the word. Importantly, the effect also occurred if the second token of a word was addressed to a different listener. This suggests that the effect of word repetition on articulatory reduction is not strictly driven by the knowledge or needs of the listener. Apart from that, little is known about the psychological mechanism underlying the effect.

A third source of redundancy that has been studied in some detail is the predictability of a word given the surrounding words. If a particular word can easily be predicted given the phrase or utterance it occurs in, speakers can afford to put less effort into its articulation. Several studies have shown that words are more reduced if they occur in a fixed expression than if they occur in a non-predictable environment (Lieberman, 1963; Hunnicutt, 1985; Binnenpoorte, Cucchiarini, Boves, & Strik, 2005). Other research has focused on predictability as a gradient property, which can be estimated by computing co-occurrence statistics from large language corpora. Two measures that have often been used for this purpose are *conditional probability* and *mutual information*, both of which capture the probability of occurrence of a word given one or more of its neighbouring words. More information about these measures will be provided in Chapters 2, 4, and 5. For now, it is sufficient to know that a higher value for conditional probability or mutual information has repeatedly been shown to be correlated with more articulatory reduction (Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Fosler-Lussier & Morgan, 1999; Bush, 2001; Jurafsky et al., 2001; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003).

Aims of the thesis

The main aim of this thesis is to increase our understanding of the circumstances in which speakers reduce articulatory effort. In addition, we aim to gain more insight into how listeners deal with words that have undergone reduction. The guiding hypothesis in this investigation is that linguistic units are more reduced the more predictable they are, as claimed by the PRH and the SSRH. Since these two hypotheses make essentially the same predictions, we do not expect to be able to distinguish between them on the basis of our data.

The research described in this thesis differs from previous research in several ways. First of all, it is exclusively concerned with Dutch, whereas most other studies on the relationship between redundancy and reduction focused on English. Furthermore, different speech styles are investigated, ranging from spontaneous conversations to read-aloud stories and laboratory speech. This allows us to compare the effects of variables like word frequency or contextual predictability in different communicative settings.

However, the most important aspect that sets this thesis apart from previous research is its focus on affixes. Affixes have the interesting characteristic that they are meaningful entities embedded in larger lexical units. As a result, direct comparisons can be made between affixes occurring in different words with different frequencies. Furthermore, affixes in Dutch tend to be unstressed, which provides control over variables such as word stress and sentence accent. Finally, the investigation of affixes enables us to explore previously uncharted sources of redundancy, namely those that are related to morphological structure and morphological predictability.

Measuring articulatory reduction

Reduction of articulatory effort can be measured in many different ways. Some methods, such as electropalatography and ultrasound imaging, allow researchers to directly track the movements of individual articulators. An important disadvantage of such methods is that they tend to be rather costly, both financially and time-wise. Therefore, most researchers restrict themselves to measuring the *acoustic* correlates of reduction, also because the acoustic signal represents the information available to the listener. A wide range of measures can be found in the literature, including mean amplitude (Shields & Balota, 1991), spectral center of gravity (Van Son & Pols, 2003), degree of centralization in vowels (Wright, 1997), and changes in F1 and F2 (Scarborough, 2004). The dependent variable that has been studied most extensively, however, is acoustic duration.

Acoustic duration is also the main dependent variable in the current thesis, for several reasons. First of all, it is intuitively the most straight-forward acoustic correlate of articulatory reduction. According to Browman and Goldstein (1992), speakers can save articulatory effort either by reducing the magnitude of their speech gestures, or by increasing the amount of overlap between gestures. In both cases, durational shortening is expected. Another advantage of duration is that it can be measured fairly reliably in noisy, overlapping speech, allowing us to investigate reduction in spontaneous, face-to-face conversations. Finally, our choice for duration was motivated by the possibility that it offers to use Automatic Speech Recognition technology for acoustic analysis.

Outline of the thesis

The contents of this thesis can be divided into three parts of two chapters each. In Chapters 2 and 3, the focus is on word frequency as a predictor of reduction. Chapter 2 investigates the role of word frequency in spontaneous, everyday conversation, while chapter 3 uses experimental methods to gain more insight into the underlying psychological processes. The central topic in Chapters 4 and 5 is predictability from context. Chapter 4 presents a corpus survey that gauges the effects of repetition and contextual predictability on durational and segmental reduction. In Chapter 5, we investigate whether contextual predictability helps listeners to recognize reduced words. Finally, Chapters 6 and 7 are concerned with the effects exerted by morphological structure and morphological predictability, respectively.

Effects of word frequency: Evidence from corpus data

CHAPTER 2

This chapter has been published as Mark Pluymaekers, Mirjam Ernestus, and R. Harald Baayen (2005a). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118, 2561-2569.

Abstract

This chapter investigated the effects of lexical frequency on the durational reduction of morphologically complex words in spoken Dutch. The hypothesis that high-frequency words are more reduced than low-frequency words was tested by comparing the durations of affixes occurring in different carrier words. Four Dutch affixes were investigated, each occurring in a large number of words with different frequencies. The materials came from a large database of face-to-face conversations. For each word containing a target affix, one token was randomly selected for acoustic analysis. Measurements were made of the duration of the affix as a whole and the durations of the individual segments in the affix. For three of the four affixes, a higher frequency of the carrier word led to shorter realizations of the affix as a whole, individual segments in the affix, or both. Other relevant factors were the sex and age of the speaker, segmental context, and speech rate. To accommodate for these findings, models of speech production should allow word frequency to affect the acoustic realizations of lower-level units, such as individual speech sounds occurring in affixes.

Introduction

In everyday speech, words are often pronounced shorter than their citation forms would suggest. This is not a marginal phenomenon: In Johnson's (2004) study on conversational American English, 25% of the words had one or more segments deleted. Deletion of complete syllables occurred in 6% of the words. Similar observations were made by Ernestus (2000) for Dutch. For example, the word *natuurlijk* ('of course') was sometimes reduced to [tyk]. Despite the frequent nature of these reductions, their presence has not yet been accommodated in any of the main psycholinguistic theories (e.g., Garrett, 1975; Dell, 1986; Levelt, 1989; Bock, 1995).

Reductions have often been linked to word frequency (e.g., Jespersen, 1922; Zipf, 1929), the hypothesis being that high-frequency words are more reduced than low-frequency words. Several explanations have been offered for this relationship, such as the compression of motor routines as a result of practice (Bybee, 2001), or the fact that high-frequency words are more predictable for the listener (e.g., Jurafsky, Bell, Gregory, & Raymond, 2001). Many studies have confirmed the pivotal role of word frequency in predicting diachronic phonetic abbreviations (Zipf, 1929; Bybee, 2001). It has proven more difficult, however, to demonstrate synchronic effects of frequency on acoustic realizations.

The main problem lies in the lack of suitable reference material. Since words generally differ not only in frequency, but also in at least one of their speech sounds, they are bound to differ in duration as well. Therefore, most authors have restricted themselves to comparing instances of the same phoneme occurring in different words. Umeda (1977) found that in American English, word-initial [s]-es were shorter if the frequency of their carrier word was high. Likewise, Cooper and Paccia-Cooper (1980) showed that palatalization of [d] before [j] was more likely in high-frequency than in low-frequency words. Van Coile (1987) used word frequency as a criterion to distinguish between function words and content words in Dutch, and found that vowels occurring in function words were shorter than the same vowels occurring in content words. Finally, Jurafsky et al. (2001) compared tokens of word-final [d] and [t] occurring in English words with different frequencies. In words with a high frequency, the plosive had a greater chance of being deleted, and if it was present in the signal, its duration was significantly shorter.

Over the years, attempts have been made to demonstrate frequency effects on units larger than the phoneme as well. Wright (1979) used pairs of rare and common words matched on length in letters, but, as most speech researchers will agree, this type of matching does not offer enough experimental control for comparing durations. Gregory, Raymond, Bell, Fosler-Lussier, and Jurafsky (1999) measured the durations of a large number of words ending in -*t* or -*d*, and found an effect of word frequency on these durations. However, their target words probably differed on other dimensions as well, such as the number of phonemes and their complexity (Landauer & Streeter, 1973). Therefore, the evidence for effects of word frequency on the durations of larger linguistic units remains inconclusive.

To overcome the difficulties sketched above, we decided to focus on morphemes that can occur in a large number of words with different frequencies: Affixes. This approach is similar to that of Aylett and Turk (2004), who compared syllables occurring in different words. The main difference lies in the fact that affixes by definition carry meaning, while for syllables this is not necessarily the case. An additional advantage of affixes is that most of them never bear stress, providing us with valuable control over factors like word stress and sentence accent.

Most studies on word frequency and reduction used English materials. Recently, a new database of spoken Dutch has become available, containing a large collection of spontaneous, face-to-face conversations. This provided us with an excellent opportunity to investigate the effects of frequency on acoustic reduction in a language other than English.

In summary, this study investigates the effects of word frequency on the durations of Dutch affixes. Durational shortening is of course not the only acoustic correlate of reduction, but the nature of the materials (spontaneous, overlapping speech) precluded us from studying variables such as mean amplitude (e.g., Shields & Balota, 1991) or spectral Center of Gravity (e.g., Van Son & Pols, 2003). In the following section, we describe our materials and method of measurement.

Method

Materials

All materials were drawn from the Corpus of Spoken Dutch (Oostdijk, 2000). This corpus contains approximately 800 hours of speech recordings, of which only the 225 hours of spontaneous, face-to-face conversations were considered for the present study. Orthographic transcriptions are available for the entire corpus. We restricted ourselves to Dutch speakers, since they have been shown to use reduced forms more than speakers from Flanders (Keune, Ernestus, Van Hout, & Baayen, 2005; see Adank, Van Hout, & Smits, 2004, for other acoustic differences between the two varieties of Dutch).

The affixes under investigation were the prefixes *ge-*, *ver-*, and *ont-*, and the suffix *-lijk*. *Ge-* is used mainly to create the perfect participle in Dutch, although it can also function as a nominal or verbal prefix. In this study, we restricted ourselves to the participial use of *ge-*. *Ver-* and *ont-* are verbalizing prefixes expressing states of change (*ver-*) and reversal or inchoation (*ont-*). For example, *ver-* + *plaats* ‘place’ gives *verplaatsen* ‘to move’, and *ont-* + *eigen* ‘own’ gives *onteigenen* ‘to disown’. The suffix *-lijk* can be found in adverbs and adjectives (e.g., *natuurlijk*, ‘natural(ly)’, and *eigenlijk*, ‘actual(ly)’). The citation forms of these four affixes are [xə], [vər], [ɔnt], and [lək].

For each of the affixes, a randomized list was made of all occurrences in the corpus. For each word type containing a target affix, the first token on the list was selected for further analysis. If the quality of the recording was too poor for acoustic analysis, it was replaced with the next token on the list. We considered as word types not only words belonging to different lemmas, but also different word forms of the same lemma. Thus, the sample for *ont-* included both *ontwikkelt* ('develops') and *ontwikkelde* ('developed'). It might be argued that such a sampling procedure leads to an over-representation of high-frequency lemmas in the sample, which could be prevented by selecting only one word form per lemma. An obvious disadvantage of such an approach is that the samples would become considerably smaller, making it more difficult to perform a meaningful statistical analysis. Instead, we checked whether lemma frequency was a better predictor than the frequency of the word form itself, which turned out not to be the case.

Measurements

Acoustic measurements of the target words were made using the software package PRAAT (Boersma, 2001). For all words, we measured the duration of the affix and the durations of the individual segments in the affix (both in milliseconds). Since the amount of background noise differed considerably between tokens, it was hard to establish a general segmentation strategy (see also Vorstermans, Martens, & Van Coile, 1996). Figure 2.1 shows the manual segmentations for the prefix *ont-* in the tokens *ontwaken* 'to wake' (top), and *ontwijken* 'to avoid' (bottom), including the previous word and the first syllable of the stem. *Ontwaken* was relatively easy to segment, since there was hardly any background noise and no overlapping speech around the prefix. The sample for *ontwijken* contained more background noise, resulting in a waveform in which the different segments could not easily be distinguished. In all cases, we placed the segment boundaries where we found clear formant transitions in the spectrogram supported by visible changes in the waveform pattern.

Control variables

Probabilistic measures

Besides frequency, other measures of word probability are known to affect acoustic realizations as well. Fowler and Housum (1987) found that the second realization of a word in a monologue was shorter than the first one. Bard, Anderson, Sotillo, Aylett, Doherty-Snedddon, and Newlands (2000) replicated this effect for dialogues, showing that it was present irrespective of whether the speaker or the listener uttered the first token of the word. To check whether our target words might be subject to repetition effects, we counted how often the target word (or a word from the same inflectional paradigm) had been used in the

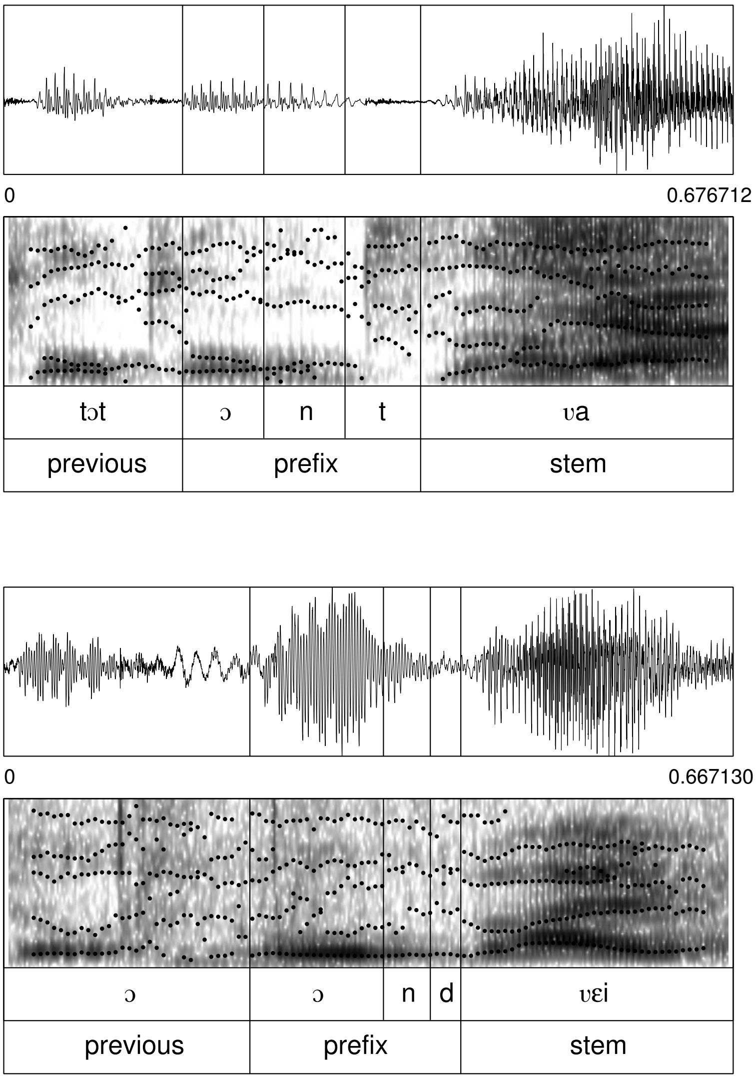


Figure 2.1: Manual segmentations for the tokens *ontwaken* (top) and *ontwijken* (bottom). *Ontwaken* was a more or less ideal case, in which there was no background noise or overlapping speech. For *ontwijken*, the amount of background noise was much greater. In both cases, we placed boundaries where we could see visible changes in the waveform pattern supported by abrupt formant transitions in the spectrogram.

conversation prior to the occurrence of the selected token. Since most tokens turned out to be first occurrences, this factor was not included in the final analyses. In addition, we counted how often the affix under investigation had already occurred. This variable, which varied in value between 0 and 72, turned out to have no effect.

The probability of occurrence of a word also depends on neighbouring words. In recent years, numerous studies have addressed the relationship between predictability from neighbouring words and acoustic reduction (e.g., Hunnicutt, 1985; Fosler-Lussier & Morgan, 1999; Gregory et al., 1999; Bush, 2001; Jurafsky et al., 2001; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003). To determine the predictability of their target words, most authors have used measures like *conditional probability* or *mutual information*, which are computed using frequency estimates from large speech corpora. Both measures capture the likelihood of a certain word occurring given one or more of its neighbouring words. Mutual Information is arguably the most elegant of the two, as it combines conditional probability with the frequency of the word itself. The corresponding equation is as follows (X and Y denote either the previous word and the target word or the target word and the following word):

$$MI(X;Y) = log \frac{(Frequency(XY))}{(Frequency(X)) * (Frequency(Y))}$$

From this equation, it is easy to see why Mutual Information could also be relevant for our purposes. Since word frequency is incorporated in the denominator, any effect observed for frequency might in fact be an effect of Mutual Information in disguise. Our sampling method prevented us from computing Mutual Information values for all of our target words, as some of them were at the beginning or end of utterances. For those words for which Mutual Information could be computed, we checked whether it was a better predictor of duration than word frequency alone. This was never the case.

As we are dealing with morphologically complex words, probabilistic variables affecting morphological processing should also be taken into account. Hay (2003) found that derived words that are more frequent than their stems are judged less morphologically complex by language users. This suggests that speakers may only consider an affix a separate morpheme if the stem is at least as frequent as the combination of stem and affix. If on the other hand the affix-stem combination is more frequent, the word is more likely to be accessed as a whole, implying less psychological reality for the affix. Since Hay showed that this perceived morphological complexity can also affect acoustic realizations, we included the ratio between the frequency of the target word and that of its stem (Word-Stem Ratio) in the analyses.

All frequency estimates, including the ones used to compute Mutual Information and Word-Stem Ratio, were taken from the Corpus of Spoken Dutch and logarithmically transformed.

Other control factors

This section discusses the non-probabilistic variables that were incorporated in this study. When rate of speech is high, words have a higher probability of deviating from the standard pronunciation (Fosler-Lussier & Morgan, 1999). We estimated speech rate by computing the number of syllables per second in the longest stretch of speech containing the target word that did not contain an audible pause. For all words in the utterance except the target word, the number of syllables was determined on the basis of the orthographic transcription. For the target word, we used the information from the manual transcription instead.

Sociolinguistic variables such as sex, age, and regional origin of the speaker also have a considerable impact on pronunciation (e.g., Byrd, 1994; Keune et al., 2005). Information about these three factors was gathered for all speakers. Age was operationalized by subtracting 1900 from the Year of Birth of the speaker.

Words that are positioned at the beginning of an utterance are often acoustically strengthened, while words at the end of an intonational phrase can show durational lengthening (e.g., Fougeron & Keating, 1997; Cambier-Langeveld, 2000; Bell et al., 2003). This was controlled for by coding for each target word whether it was utterance-initial or not, and whether it was utterance-final or not.

Fox Tree and Clark (1997) and Bell et al. (2003) showed that words that occur near disfluencies are lengthened compared to words occurring in fluent contexts. Therefore, the presence of a false start or filled pause directly before or after the target word was also coded.

Not all phonetic environments are equally suitable for reduction. Zsiga (1994) found that word-final consonants are more likely to be reduced if they are followed by another consonant. We determined for each token whether the segment following the affix was a consonant or a vowel. For the prefixed words, we also counted the number of consonants in the onset of the stem (henceforth referred to as Onset Complexity).

Finally, the absence of certain segments in the affix was sometimes included as an extra factor in the analyses. If, for example, the final segment in the affix is absent, this may have implications for the durations of the other segments, as well as for the duration of the affix as a whole. Segments were considered as absent if they could not be isolated in the acoustic signal.

To evaluate the effects of word frequency on duration while controlling for all other possibly relevant factors, we used least squares regression. The application of this method on speech data is described in detail by Bell et al. (2003). The signs of the reported beta coefficients indicate whether there was a positive or a negative correlation between two variables. Since we do not want to report effects that depend crucially on a single data point, we excluded observations that were outliers with regard to leverage or Cook's distance values (Chatterjee, Hadi, & Price, 2000).

Results

Results for ge-

Duration of the prefix as a whole

For *ge-*, the sample consisted of 428 words uttered by 132 different speakers. No speaker contributed more than 12 tokens to the sample. Broad phonetic transcriptions of the encountered realizations are [xə], [x], [χə], and [χ].

We fitted a stepwise multiple regression model to the data with the duration of *ge-* as the response variable. There were four outliers, which were removed. We found main effects of Frequency ($\hat{\beta} = -4.1, t(420) = -3.05, p < 0.005$), Speech Rate ($\hat{\beta} = -8.6, t(420) = -5.62, p < 0.0001$), and Onset Complexity ($\hat{\beta} = -7.3, t(420) = -2.03, p < 0.05$). Words with a higher frequency had shorter realizations of *ge-*. When Speech Rate was high, the prefix was also shorter, as was the case if it was followed by a large number of consonants. The amount of variance explained by this model (also referred to as R^2) was 10%.

Durations of the individual segments

To gain more insight into the articulatory dynamics underlying the above-mentioned effects, separate models were fitted for the two segments in *ge-*.

For the fricative, a model was fitted to the entire data set, including the data points that were outliers in the model for the whole prefix. After the removal of three new outliers, there were main effects of Frequency ($\hat{\beta} = -3.6, t(421) = -3.71, p < 0.0005$), Speech Rate ($\hat{\beta} = -3.5, t(421) = -3.79, p < 0.0005$), and Word-Stem Ratio ($\hat{\beta} = 34.7, t(421) = 2.46, p < 0.05$). This model explained 7% of the variance.

The vowel was present in 414 tokens (97%). Four outliers were removed. Vowel duration was predicted by Frequency ($\hat{\beta} = -2.1, t(406) = -2.31, p < 0.05$), Speech Rate ($\hat{\beta} = -3.0, t(406) = -2.93, p < 0.005$), and Initial Position ($\hat{\beta} = 14.5, t(406) = 2.45, p < 0.05$). The vowel was longer in words that were in Initial Position. The R^2 of this model was 6%.

Discussion of the results for ge-

The results for *ge-* are summarized in Table 2.1. The observed effects of Frequency, Speech Rate, Initial Position, and Onset Complexity all went in the expected direction. The Frequency effect was present for both the fricative and the vowel, as was the effect of Speech Rate.

It might seem counter-intuitive that a higher ratio between the frequency of the word and the frequency of its stem should lead to longer fricatives. After all, a higher value of this ratio is supposed to be associated with less morphological complexity, and hence, less psychological reality for the prefix. We return to this issue in our General Discussion.

Table 2.1: Beta coefficients and significance values of the effects for *ge-*. The beta coefficients indicate the magnitude of the effect in milliseconds. ”–” means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

Predictor	Prefix	Fricative	Vowel
Frequency	-4.1^{**}	-3.6^{***}	-2.1^{*}
Initial Position	–	–	14.5^{*}
Onset Complexity	-7.3^{*}	–	–
Speech Rate	-8.6^{****}	-3.5^{***}	-3.0^{**}
Word-Stem Ratio	–	34.7^{*}	–
Explained variance (R^2)	.10	.07	.06

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

Results for ont-

Duration of the prefix as a whole

There were 102 word types starting with *ont-* in the corpus. The tokens in the sample were uttered by 63 different speakers, who contributed no more than four tokens each to the data set. The realizations we encountered ranged from canonical [ɔnt] to highly reduced [ət].

A model was fitted to the data with duration of *ont-* as the response variable. Three outliers were removed. Prefix duration was predicted by Year of Birth ($\hat{\beta} = -1.4, t(95) = -4.81, p < 0.0001$). Younger speakers produced shorter prefixes. Frequency was not significant as a main effect, but it was in interaction with Speech Rate ($\hat{\beta} = -2.9, t(95) = -3.31, p < 0.005$) and Year of Birth ($\hat{\beta} = 0.2, t(95) = 2.84, p < 0.01$). The interaction between Frequency and Speech Rate is shown in Figure 2.2. Frequency had either a lengthening or no effect when Speech Rate was low (the bottom left and middle panels), a shortening effect when Speech Rate was neither low nor high, and no effect when Speech Rate was extremely high (the top right panel). In Figure 2.3, the interaction between Frequency and Year of Birth is illustrated. For the youngest speakers (the top middle and right panels) the effect of Frequency was absent, whereas for the other age groups a higher Frequency correlated with shorter realizations. All in all, this model accounted for 24% of the variance.

Durations of the individual segments

Since the vowel was present in all 102 tokens, we fitted a model for vowel duration to the entire data set. Three observations were identified as outliers and removed. Younger speakers produced shorter vowels ($\hat{\beta} = -0.3, t(96) = -2.19, p < 0.05$), while women's vowels were longer ($\hat{\beta} = 8.7, t(96) = 2.33, p < 0.05$). The R^2 of this model was 10%.

The nasal was produced in 97 tokens, three of which were outliers. The duration of the nasal

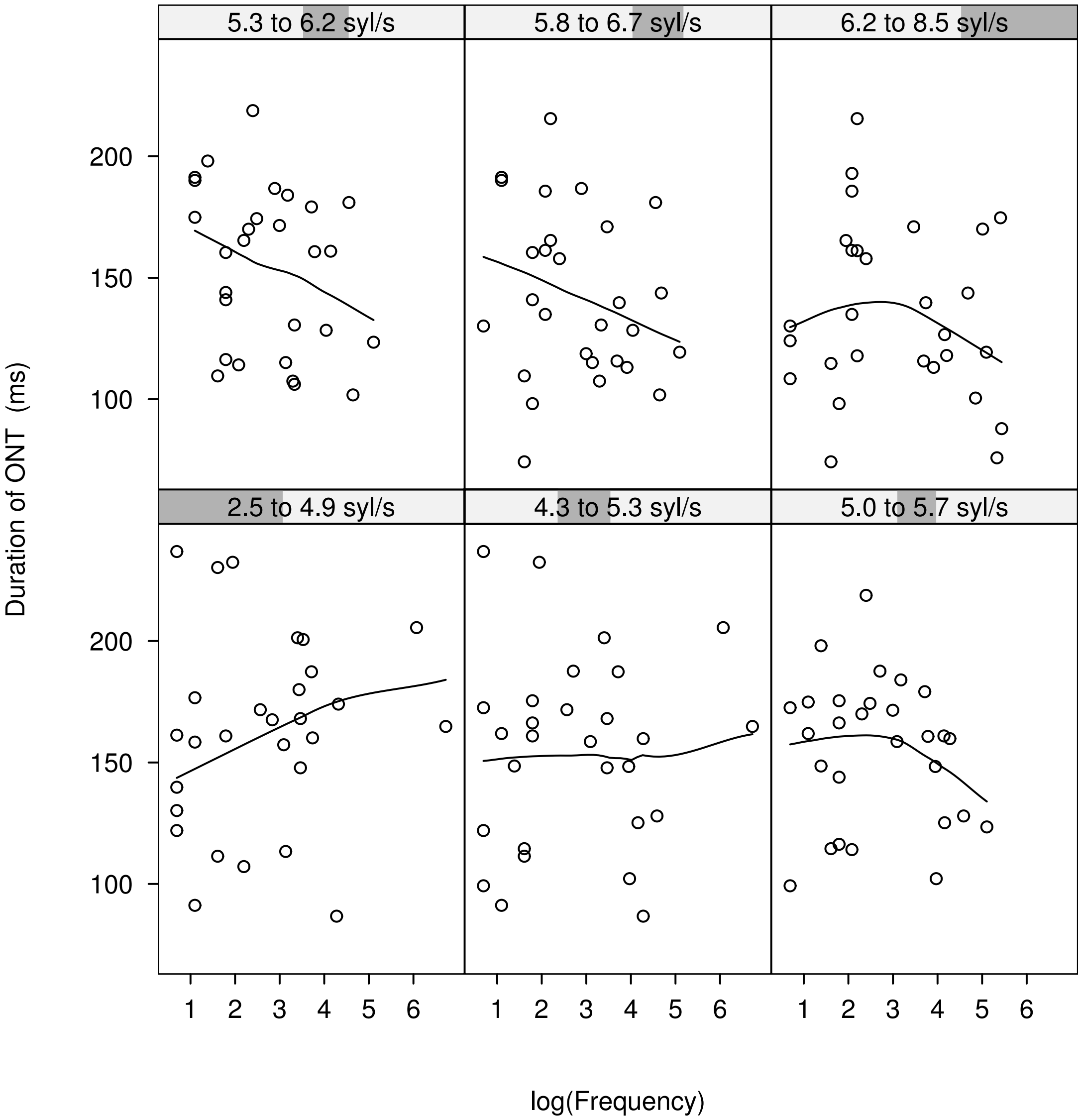


Figure 2.2: Interaction between Frequency and Speech Rate as observed for the duration of *ont-*. The panels should be read from left to right and from bottom to top: Speech Rate is lowest in the bottom left panel and highest in the top right panel. There is no Frequency effect when Speech Rate is low (the bottom left and middle panels) or extremely high (the top right panel).

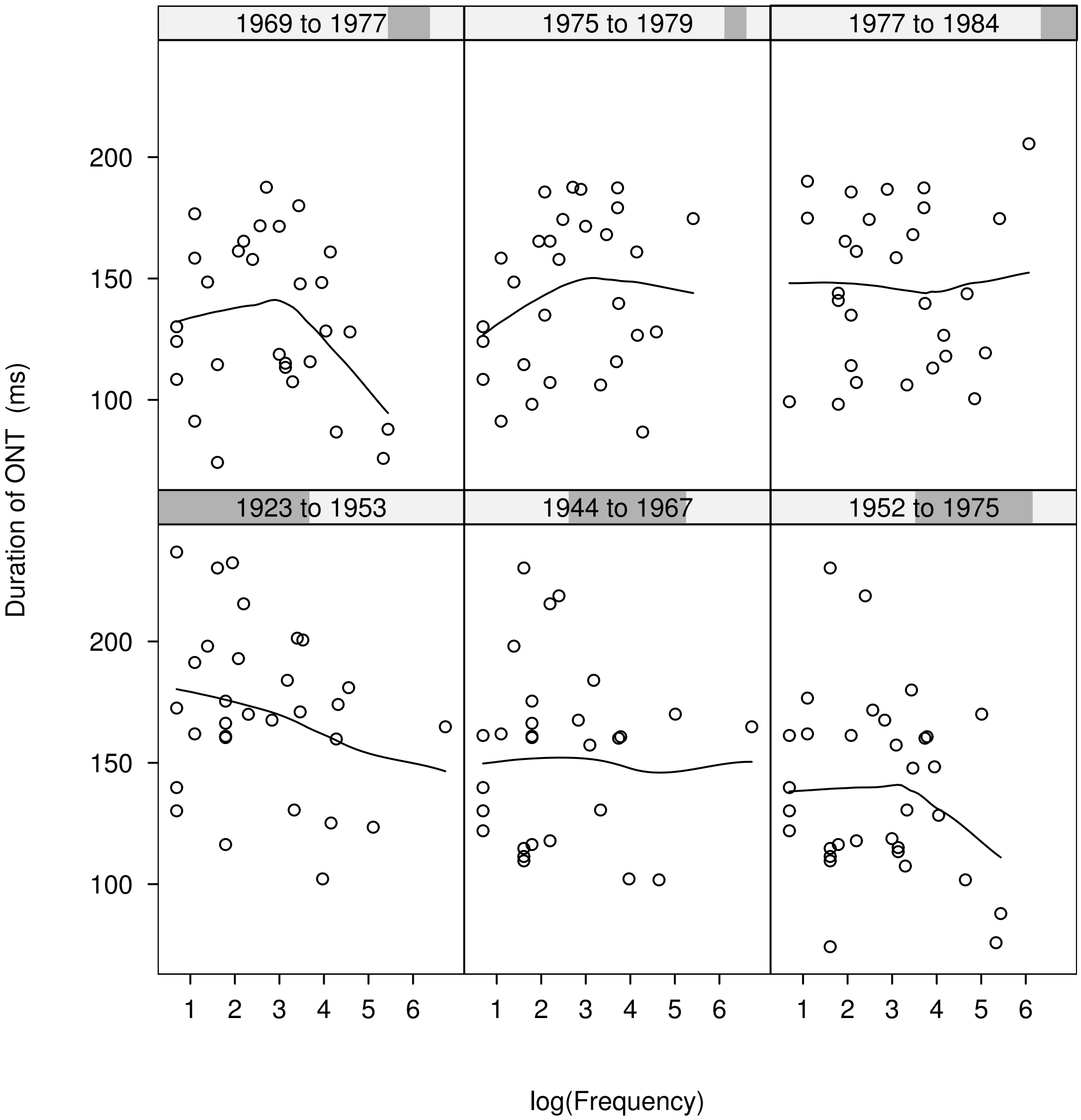


Figure 2.3: Interaction between Frequency and Year of Birth as observed for the duration of *ont-*. The panels should be read from left to right and from bottom to top: The oldest speakers are in the bottom left panel and the youngest speakers in the top right panel. There is no effect of Frequency for the youngest age groups (top middle and right panels).

was affected by the Presence of the Plosive ($\hat{\beta} = -34.2, t(89) = -5.58, p < 0.0001$) and Year of Birth ($\hat{\beta} = -0.2, t(89) = -2.74, p < 0.01$). Younger speakers produced shorter nasals, and if the plosive was absent the nasal was longer. We also found a significant interaction between the Presence of the Plosive and Frequency ($\hat{\beta} = 5.9, t(89) = 3.67, p < 0.0005$). Frequency was only significant if there was no plosive. Together, these three predictors explained 48% of the variance.

Finally, we fitted a model for the duration of the plosive. Three outliers were removed. There were main effects of Frequency ($\hat{\beta} = -17.8, t(66) = -3.50, p < 0.001$), Year of Birth ($\hat{\beta} = -1.0, t(66) = -3.87, p < 0.0005$), and Speech Rate ($\hat{\beta} = -3.6, t(66) = -2.39, p < 0.05$). All effects went in the expected direction. Furthermore, there was an interaction between Frequency and Year of Birth ($\hat{\beta} = 0.3, t(66) = 3.30, p < 0.005$), which was similar to the one observed for the entire prefix (see Figure 2.3). In total, 28% of the variance in the duration of the plosive was explained by this model.

Discussion of the results for ont-

Table 2.2 provides an outline of the results for *ont*-. There were Frequency effects in the expected direction for the duration of the plosive (if it was present) and for the duration of the nasal if the plosive was not present. This did not lead to a main effect of Frequency for the prefix as a whole, but there were two significant interactions.

The interaction with Speech Rate suggested that the effect of Frequency was limited to situations in which Speech Rate was not extremely low or high. The absence of a Frequency effect when Speech Rate is high can be explained by assuming that speakers try to avoid complete deletion of the prefix, since this may hamper communication. This is confirmed by the fact that no tokens were encountered in which *ont*- was completely absent. When Speech Rate is low, on the other hand, there is less need to reduce articulatory effort, which also diminishes the likelihood of finding a Frequency effect.

The interaction with Year of Birth showed that the Frequency effect was not present for the youngest speakers. This can be accounted for by the finding that younger speakers already produced shorter realizations (as evidenced by the negative main effects of Year of Birth observed in all models fitted for *ont*-).

Results for ver-

Duration of the prefix as a whole

The prefix *ver*- occurred in 140 different word types produced by 82 different speakers. The maximum number of tokens uttered by a single speaker was eight. Observed pronunciations included [vər], [və], and [f].

Table 2.2: Beta coefficients and significance values of the effects for *ont-*. The beta coefficients indicate the magnitude of the effect in milliseconds. ”–” means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

Predictor	Prefix	Vowel	Nasal	Plosive
Frequency	–	–	–	-17.8^{***}
Frequency * Speech Rate	-2.9^{**}	–	–	–
Frequency * Year of Birth	0.2^{**}	–	–	0.3^{**}
Frequency * Plosive Present	–	–	5.9^{***}	–
Plosive Present	–	–	-34.2^{****}	–
Sex	–	8.7^{*}	–	–
Speech Rate	–	–	–	-3.6^{*}
Year of Birth	-1.4^{****}	-0.3^{*}	-0.2^{**}	-1.0^{**}
Explained variance (R^2)	.24	.10	.48	.28

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

We fitted a model to predict the duration of *ver-*. After removing three outliers, there were significant main effects of Year of Birth ($\hat{\beta} = -0.5, t(134) = -2.58, p < 0.05$) and Onset Complexity ($\hat{\beta} = -14.7, t(134) = -2.64, p < 0.01$). Younger speakers produced shorter prefixes. If the number of consonants in the onset of the stem was high, the prefix was shorter as well. The R^2 of this model was 11%.

Durations of the individual segments

We fitted separate models only for the fricative and the rime (i.e., the combination of the vowel and [r]), since the vowel and [r] (if present) could not be reliably distinguished.

The fricative was present in all cases. Four outliers were removed. We found main effects of Onset Complexity ($\hat{\beta} = -7.4, t(133) = -2.37, p < 0.05$) and Sex of the speaker ($\hat{\beta} = -12.1, t(133) = -3.43, p < 0.001$). Women produced shorter fricatives. These variables explained 12% of the variance.

For the rime, a model was fitted to the 117 data points for which it was present. Three outliers were removed. There were main effects of Onset Complexity ($\hat{\beta} = -12.8, t(111) = -3.49, p < 0.001$) and Year of Birth ($\hat{\beta} = -0.3, t(111) = -2.52, p < 0.05$), together explaining 17% of the variance.

Discussion of the results for ver-

Beta coefficients and significance values of the effects for *ver-* are given in Table 2.3. For this prefix, there were no effects of Frequency. The effect of Onset Complexity (the higher the number of consonants in the onset of the stem, the shorter the prefix) can be traced back

Table 2.3: Beta coefficients and significance values of the effects for *ver-*. The beta coefficients indicate the magnitude of the effect in milliseconds. ”–” means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

Predictor	Prefix	Fricative	Rime
Onset Complexity	-14.7^{**}	-7.4^{*}	-12.8^{***}
Sex	–	-12.1^{***}	–
Year of Birth	-0.5^{*}	–	-0.3^{*}
Explained variance (R^2)	.09	.12	.17

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

to both the fricative and the rime. As was the case for the other prefixes, younger speakers produced shorter realizations of *ver-*. This effect was mainly due to durational shortening of the rime.

Results for -lijk

Duration of the suffix as a whole

The data set for the suffix *-lijk* consisted of 158 tokens, uttered by 88 different speakers. No speaker contributed more than six tokens to the data set. The realizations we observed ranged from the citation form [lək] to [ə] or [k]. Because of the semantic opacity of many of the words containing *-lijk*, the Word-Stem Ratio we discussed earlier was not included in the analyses.

Again, we first fitted a model to predict the duration of the affix as a whole. Visual inspection of the fitted model revealed that the variance in the residuals was much larger for words in Final Position than for words in Non-Final Position ($F(42, 114) = 4.33, p < 0.0001$). Therefore, separate models were fitted for words in Final and in Non-Final Position. For words in Final Position (43 observations), four outliers were removed. The data set for Non-Final words contained 115 tokens, four of which were outliers.

For words in Non-Final Position, there were main effects of Frequency ($\hat{\beta} = -7.7, t(108) = -3.73, p < 0.0005$) and Year of Birth ($\hat{\beta} = -0.7, t(108) = -3.12, p < 0.005$), explaining 18% of the variance.

If the word was in Final Position, the duration of *-lijk* was affected by Speech Rate ($\hat{\beta} = -36.4, t(36) = -4.24, p < 0.0005$) and the Presence of the Plosive ($\hat{\beta} = 147.5, t(36) = 3.23, p < 0.005$). If the plosive was absent, the suffix was shorter. The R^2 of this model was 47%.

Durations of the individual segments

The [l] was produced in 140 tokens. This time, a split on the basis of Position was not necessary, as visual inspection of the residuals did not reveal any abnormalities. We removed four outliers and found main effects of Frequency ($\hat{\beta} = -2.1, t(130) = -2.95, p < 0.005$), Speech Rate ($\hat{\beta} = -4.2, t(130) = -4.14, p < 0.0001$), Year of Birth ($\hat{\beta} = -0.2, t(130) = -2.96, p < 0.005$), and Final Position ($\hat{\beta} = 10.0, t(130) = 3.66, p < 0.0005$). All effects went in the expected direction. Together, they accounted for 32% of the variance.

The vowel was realized in all but eight tokens. Three outliers were removed. Vowel duration was affected by Speech Rate ($\hat{\beta} = -4.2, t(143) = -2.77, p < 0.01$), Final Position ($\hat{\beta} = 12.2, t(143) = 3.03, p < 0.005$), and the Presence of the Plosive ($\hat{\beta} = -30.6, t(143) = -4.77, p < 0.0001$). If the plosive was absent, the vowel was longer. The R^2 of this model was 23%.

For the plosive, the situation was similar to that of the entire suffix. The variances of the residuals of the initially fitted model differed significantly between tokens that were in Final and Non-Final Position ($F(38, 105) = 7.27, p < 0.0001$). Therefore, we fitted separate models for the 39 Final and 106 Non-Final plosives.

For the Non-Final plosives, three outliers were removed. The only significant effect we found was one of Following Disfluency: if there was no disfluency following the target word, the plosive was shorter ($\hat{\beta} = -66.2, t(101) = -5.73, p < 0.0001$). By itself, this factor accounted for 25% of the variance. For the Final plosives, we removed three outliers and found an effect of Speech Rate ($\hat{\beta} = -22.9, t(34) = -3.28, p < 0.005$) in the expected direction, explaining 24% of the variance.

Discussion of the results for -lijk

An outline of the results for *-lijk* is given in Table 2.4. The duration of *-lijk* was most strongly affected by whether the word was in Final Position or not. For the suffix as a whole and the plosive, this effect was so pervasive that separate models had to be fitted to the Final and the Non-Final data points. For the other two segments, such drastic measures were not necessary, although Final Position remained a significant predictor. The effect always went in the same direction: Final words had longer realizations of *-lijk*. This could be explained by referring to the well-documented phonetic effect of phrase-final lengthening (e.g., Fougeron & Keating, 1997), but we feel that such an explanation would be too restricted. The majority (70%) of the words in Final Position were predicates in an utterance of the type ‘that is [Adjective]’. When saying something like *dat is belachelijk* ‘that is ridiculous’, it would make no sense to reduce articulatory effort on the word *belachelijk*, as it is the only information carrier in the utterance. This might also be reflected in the presence of sentence accent, leading to durational lengthening (e.g., Nooteboom, 1972).

Table 2.4: Beta coefficients and significance values of the effects for *-lijk*. The beta coefficients indicate the magnitude of the effect in milliseconds. ”–” means there was no significant effect. The bottom row shows the amount of variance explained (R^2) by each model.

Predictors	Suffix (Non-Final)	Suffix (Final)	[l]	Vowel	Plosive (Non-Final)	Plosive (Final)
Frequency	-7.7^{***}	–	-2.1^{**}	–	–	–
Final Position	–	–	10.0^{***}	12.2^{**}	–	–
Following Disfluency	–	–	–	–	-66.2^{****}	–
Plosive Present	–	147.5^{**}	–	-30.6^{****}	–	–
Speech Rate	–	-36.4^{***}	-4.3^{****}	-4.2^{**}	–	-22.9^{**}
Year of Birth	-0.7^{**}	–	-0.2^{**}	–	–	–
Explained variance (R^2)	.18	.47	.32	.23	.25	.24

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

Despite the dominance of Final Position as a predictor, we still found effects of Frequency, although these were restricted to the Non-Final suffixes and realizations of [I]. Speech Rate was significant for all durations except that of the Non-Final plosive. The effect of Following Disfluency we found was in line with the earlier findings of Fox Tree and Clark (1997) and Bell et al. (2003).

Interestingly, the two significant effects for Presence of the Plosive went in opposite directions. If the plosive was absent, Final suffixes were shorter, but vowels were longer. It might be the case that speakers more or less compensate for the absence of the plosive by lengthening the vowel. For the subset of Final suffixes, this effect might not have surfaced since the entire suffix was lengthened.

General Discussion

This study provides strong evidence for the relationship between lexical frequency and acoustic reduction. For the Dutch affixes *ge-*, *ont-*, and *-lijk*, we found effects of frequency on the durations of individual segments, the affix as a whole, or both. Apparently, the effect of word frequency in speech production is not restricted to the speed of lexical retrieval; It manifests itself in the subtle acoustic details of the word as well. This lends further support to the probabilistic framework developed by Bybee (2001), Pierrehumbert (2003), and others. They view probabilistic information as an integral part of our linguistic knowledge, exerting its influence at every level of language processing, including articulation.

How can these findings be incorporated in models of speech production? Most current theories are based on either speech errors or reaction time data, and have not been concerned with fine-grained differences in articulation. Nevertheless, models like the one proposed by Levelt, Roelofs, and Meyer (1999) can be modified in such a way that they can account for the results of the current study. One possible modification is the inclusion of reduced word forms in the lexicon, which are selected if the conceptual structure of the message specifies a word as redundant. Although this might work for obviously reduced forms such as [tyk] for *natuurlijk* 'of course', it seems less appropriate to include all possible durational variants of words in the lexicon. A more efficient solution is to pass information about redundancy on to the Articulator, where it could influence the amount of effort put into articulation.

Another possibility is that effects of frequency arise during phonological processing. All words in a speaker's mental lexicon are believed to have a certain resting activation level. When this activation level is high (as is the case for highly frequent words), activation may spread to the constituent speech sounds more quickly than when it is low, resulting in quicker preparations of the speech sounds and thus, shorter articulatory durations (e.g., Balota, Boland, & Shields, 1989). Our current data do not allow us to distinguish between these two hypotheses, as the words under investigation were produced after both conceptual and articulatory preparation had taken place. If, however, frequency effects were to be found in situations where no conceptual preparation was required (e.g., in shadowing or speeded naming tasks), this would suggest that at least part of the effect of frequency on durations arises during the later stages of the speech production process.

One assumption made by many theories of speech production remains problematic, however. They regard the syllable as the principal unit of articulation, as well as the primary locus of frequency effects below the word form level (e.g., Levelt et al. 1999; Sevald, Dell, & Cole, 1995). Given the results of the present study, this view appears to be too restricted. First of all, we find evidence for word-specific frequency effects that operate below the level of the syllable. More importantly, the different segments in a syllable were shown to be subject to different, sometimes even contradictory, forces. This suggests that the motor program that is executed during articulation is very likely not the syllable, since one would expect similar processes to apply to all subcomponents of such a program.

Related to this issue is the question why some segments were affected by frequency while others were not. Although it is very difficult to abstract a general pattern from the data, it is at least clear that durational reduction is not restricted to vowels. Furthermore, we find some evidence that speakers sometimes compensate for the absence of one segment by lengthening another. This presents interesting challenges for frameworks such as Articulatory Phonology (Browman & Goldstein, 1992), since it suggests that reduction of articulatory effort does not necessarily involve either increased overlap or reduced magnitude of speech gestures.

Of course, the morphological status of affixes is relevant as well. Hay (2003) showed that the psychological reality of affixes depends on the ratio between the frequency of the morphologically complex word and the frequency of its stem. Moreover, she found that *t*'s preceding the English suffix *-ly* were more likely to be deleted if this ratio was high. We included the ratio in our analyses and found that a higher value led to longer fricatives in the prefix *ge-*. Apparently, less psychological reality for the affix does not necessarily imply shorter realizations. A possible explanation for this finding is that in most Dutch monomorphemic words, the first syllable receives stress (Booij, 1995). If speakers no longer regard the prefix as a separate morpheme, it will eventually become a ‘normal’ word-initial syllable and therefore more likely to receive stress. The longer duration we observed for the fricative in *ge-* might be the precursor of such a change.

One of the four affixes we investigated, *ver-*, failed to show an effect of frequency. This could be related to the fact that the initial fricative can be devoiced, leading to more variation in the sample. In addition, verbal prefixes like *ver-* and *ont-* contribute much more to the meaning of their carrier words than *ge-* and *-lijk*, which merely signal grammatical function.

Since the main objective of speakers is to get the meaning of their utterances across, it is not inconceivable that relatively meaningful units are less affected by frequency than more or less meaningless ones. Redundancy has many different dimensions, and the role of semantics in defining it should not be underestimated. This is confirmed by other studies, which also report differences between words in their sensitivity to reduction (e.g., Jurafsky et al., 2001; Bell et al., 2003).

A useful step in combining different dimensions of redundancy was taken by Gregory et al. (1999), who incorporated Latent Semantic Analysis in their study. Latent Semantic Analysis (Landauer & Dumais, 1997) is a method to compute semantic relatedness scores from large-scale co-occurrence statistics. Since relatedness scores can be computed between a word and the whole discourse preceding it, this measure effectively combines word repetition, predictability from neighbouring words, and semantic association with other words used in the conversation. Importantly, Gregory et al.'s (1999) results also show that other predictors remain relevant as well. For example, they report effects of mutual information and repetition over and above effects of semantic relatedness. This suggests that this particular measure of semantic relatedness is not the only variable that should be taken into account.

One might ask whether our methodology could also be used to investigate acoustic reduction in stems. Indeed, it would be possible to compare the durations of identical stems that are combined with different affixes. However, in most languages the number of stems that lend themselves to such an approach will be rather restricted, as stems are generally less productive than affixes.

If the aim is to compare reduction in stems to reduction in affixes, word frequency is probably not the best variable to focus on. In non-agglutinative languages, the number of words containing an identical combination of stem and affix will be too small to make a comparison of words with different frequencies possible. For example, the English stem-affix combination *disable* only occurs in *disable*, *disables*, *disabled*, *disabling*, and *disablement*. Therefore, it might be a better idea to study measures like conditional probability or mutual information, and see whether they affect stems and affixes differently (see Chapter 4).

With regard to future research, several issues need to be addressed. First of all, we need to know more about the way the different measures of probability interact. This requires large databases of carefully segmented speech, so that multiple tokens of the same words can be examined in their respective contexts. Second, more attention needs to be paid to semantic variables. Taken together, these lines of research could result in a model in which probabilistic and semantic relationships are exploited to the fullest and play a role at the finest level of acoustic detail.

Effects of word frequency: Experimental evidence

CHAPTER 3

A partial analysis of the data presented in this chapter has been published as Mark Pluymaekers, Mirjam Ernestus, and R. Harald Baayen (2006). Effects of word frequency on the acoustic durations of affixes. *Proceedings of Interspeech 2006*, 953-956.

Abstract

The previous chapter showed that word frequency affects the acoustic durations of affixes in spontaneous speech. In this chapter, we tried to replicate these results in two production experiments. In the first experiment, participants named words containing one of four Dutch affixes. The durations of these affixes were again found to be shorter the higher the frequency of their carrier word. The second experiment was conducted to rule out the possibility that this effect was due to the frequency of the stimulus the participants saw, rather than the frequency of the word the participants had to produce. In a noun/verb categorization experiment with verbal responses, we independently manipulated the frequency of the stimulus word and the frequency of the response word, which always contained the suffix *-lijk*. This suffix was shorter the higher the frequency of the response word, whereas it was not affected by stimulus frequency. These findings support models of speech production that allow cascaded processing all the way down to the articulation phase.

Introduction

One of the most robust findings in the speech production literature is the word frequency effect, first reported by Oldfield and Wingfield (1965). Words with a high frequency of occurrence are produced at shorter latencies than low-frequency words. In the current article, we explore an effect of word frequency that has received considerably less attention in the literature, namely, the effect of frequency on articulatory durations. A growing body of research suggests that high-frequency words have shorter articulatory durations than low-frequency words. To our knowledge, this effect has not yet been incorporated into any of the main speech production theories (e.g., Garrett, 1975; Stemberger, 1985; Dell, 1986; Levelt, 1989; Bock, 1995; Levelt, Roelofs, & Meyer, 1999).

One possible reason for the apparent lack of interest in this effect is that it is notoriously hard to demonstrate. Since words tend to differ not only in frequency, but also in their constituent phonemes, it is difficult to separate effects of word frequency from effects of intrinsic phoneme durations (Landauer & Streeter, 1973). Therefore, most studies on the subject have only compared instances of the same phoneme occurring in different words. For example, Umeda (1977) found that in American English, word-initial [s]-es are shorter if the frequency of their carrier word is high. Van Coile (1987) used word frequency as a criterion to distinguish between function words and content words in Dutch, and found that vowels occurring in function words are shorter than the same vowels occurring in content words. Finally, Jurafsky, Bell, Gregory, and Raymond (2001) investigated the durations of word-final [d] and [t] in American English. Again, a higher frequency of the carrier word was correlated with shorter durations, even after the effects of speech rate, segmental context, and number of syllables had been partialled out.

The three studies mentioned above focused on corpus data. Hence, the speech signal under investigation was the end-product of the complete speech production process, starting with conceptualization and ending in overt articulation. Consequently, the authors could not determine the exact *locus* of the frequency effect. A number of experimental studies have tried to gain more insight into this issue, but, unfortunately, these studies were not always successful in controlling for phonetic and psycholinguistic confounds. Wright (1979) found that lists containing high-frequency words had shorter total reading times than lists containing low-frequency words, but, as Geffen and Luszcz (1983) argued, this total reading time included pauses between words and could therefore not provide conclusive evidence about articulatory durations. A second methodological problem in both of these studies was that high- and low-frequency words were matched on length in letters only, which does not provide enough phonetic control for comparing durations.

A more sophisticated matching procedure was used by Kawamoto, Kello, Higareda, and Vu (1999), who separately compared identical onsets and rimes in high- and low-frequency words. Their main finding was that word frequency affects the duration of onsets, but not the duration of rimes. This result was interpreted as evidence for the Initial Phoneme Criterion, which states that the articulation of a word is initiated as soon as the first phoneme is fully prepared, and that “articulation of each segment continues until the next phoneme in the sequence reaches threshold” (Kawamoto et al., 1999, p. 365). Such an account fails to acknowledge that phoneme durations depend on many different factors, including, but certainly not restricted to, the processing difficulty associated with the next phoneme.

Due to various methodological difficulties, none of the studies discussed so far has been able to draw conclusions about whole words. In Chapter 2 of this thesis, we investigated frequency effects on meaningful units other than the word. From a corpus of face-to-face conversations in Dutch, we selected all occurrences of four unstressed affixes. Subsequently, we measured the durations of these affixes and their constituent segments. Since affixes always consist of the same phonemes, this method provides better phonetic control for comparing durations. Regression techniques were used to bring other phonetic variables under statistical control. For three of the four affixes, the duration of the affix as a whole or the duration of one or more of its segments was found to be shorter the higher the frequency of the word the affix occurred in.

The main aim of the current study was to gain more insight into the psychological processes underlying the effect of frequency on articulatory durations. To this end, we tried to replicate the results of Chapter 2 in a word naming experiment. Word naming was chosen as the experimental task for several reasons. First of all, naming allows us to investigate the role of the listener in the effect. Some accounts of phonetic reduction (e.g., Jurafsky et al., 2001; Aylett & Turk, 2004) implicitly assume that speakers reduce the durations of high-frequency words because listeners need less acoustic evidence to recognize these words. However, if effects of word frequency were to be found in the absence of a listener, this would suggest that at least part of the effect is due to speaker-internal processes alone.

A second advantage of word naming is that it offers the possibility to test whether the frequency effect on durations arises at the conceptual level. Naming is generally assumed to involve very little conceptual preparation on behalf of the speaker (Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Recently, Baayen, Feldman, and Schreuder (2006) tested this assumption by investigating the effects of a large number of form-related and semantic variables in lexical decision and naming. They found that naming latencies were mainly affected by form variables such as word length, neighborhood density, and orthographic consistency, while semantic variables like derivational entropy, word category, and number of synonyms did not play a role. By incorporating a number of semantic variables in our analyses, we can learn more about the nature of the frequency effect. More specifically, if we were to find an effect of frequency on durations, but no effects of any of the semantic variables, we can conclude that the frequency effect on articulatory durations does not arise at the conceptual level.

Finally, the word naming paradigm enables us to investigate the robustness of the frequency effect. By manipulating the presentation rate of the stimuli, we can investigate whether the effect of frequency becomes smaller the faster participants have to articulate. Furthermore, we know from earlier studies that speeding up responses can reduce the magnitude of otherwise well-established effects, such as the effect of frequency on naming latencies (e.g., Kello & Plaut, 2000; Kello, 2004).

Experiment 1

Method

Participants

Twenty-one subjects participated in the experiment in exchange for pay. There were 11 male and 10 female participants, all of whom were native speakers of Dutch. All had normal or corrected-to-normal vision.

Materials

The affixes under investigation were the prefixes *ge-*, *ver-*, and *ont-*, and the suffix *-lijk*. These were the same affixes that were investigated in the previous chapter. For each affix, we selected 60 words containing that affix. These words spanned the entire frequency range and always consisted of three syllables. To draw attention away from the affixes, 210 monomorphemic filler words were included in the experiment. These fillers were mono-, bi-, or trisyllabic, with word stress alternating between the first, second, and third syllable. All words in the experiment had regular spelling.

The 240 targets and 210 fillers were divided into 15 lists of 30 words each. Each list contained 16 randomly selected targets (four for each affix) and 14 randomly selected fillers. The first five words on a list were always fillers, and the remaining words were ordered so that targets containing the same affix were never directly adjacent. All words occurred just once in the experiment.

Design

The main experimental variable was the frequency of the words in which the affixes occurred. Furthermore, we manipulated the presentation rate of the stimuli. Three different rates were employed. In the slow condition, 1500 milliseconds separated the onsets of two subsequent stimuli. In the medium condition, this interval was 1100 milliseconds, while the fast condition allowed participants only 700 milliseconds to give their response.

Participants produced one third of the lists (5 lists, 150 words) in each condition. The lists were divided over the presentation rates so that over the whole experiment (21 participants), each list occurred in each rate exactly seven times. Participants always started with the slow condition, followed by the medium condition and the fast condition. The main reason for keeping the order of conditions constant was our concern that participants who had seen a number of lists in the fast condition might not be able to adopt a slower pace for subsequent lists in a slower condition.

Procedure

Participants were seated in a sound-attenuated booth, behind a Sennheiser microphone. The words were presented on a computer screen, and participants were asked to name them. Participants' responses were recorded on a Tascam DA-20 DAT tape recorder and digitized with a sampling frequency of 16 kHz. In addition, naming latencies were registered by means of a voice key device.

The procedure was the same for all participants. First, five lists were presented in the slow condition, followed by five lists in the medium condition and five lists in the fast condition. Between two lists in the same condition, participants were allowed to pause for a few seconds. Between conditions, they were allowed to leave the booth. During these two longer breaks, participants were told by the experimenter that the presentation rate of the next set of lists would be considerably faster than that of the previous set, but that they should be able to keep up with it if they reacted as quickly as possible. To give participants an idea about the tempo they should adopt, each list was preceded by five fixation points presented at the same rate as the words in that list.

Acoustic measurements

The current experiment generated an enormous amount of speech data. Since hand-measuring the durations of all affixes would be extremely time-costly, we decided to use Automatic Speech Recognition (ASR) technology. Recent research has shown that the reliability of segmentations generated by an ASR system is equal to that of segmentations made by human transcribers, provided that a phonemic transcription of the signal is available to the segmentation algorithm (Vorstermans, Martens, & Van Coile, 1996; Sjölander, 2001).

We trained a Hidden Markov Model (HMM) speech recognizer using the software package HTK (Young et al., 2002). To optimize the ASR's performance on phonemic segmentation, we used context-independent, continuous density HMMs with 32 Gaussians per state (Kessens & Strik, 2004). In total, 37 phone models were trained, representing the 36 phonemes of Dutch and silence. The training material was taken from the phonemically transcribed portion of the subcorpus ‘Library for the Blind’ of the Corpus of Spoken Dutch. In total, the training sample consisted of 13328 read utterances produced by 134 different speakers. The combined duration of these utterances was 6 hours and 39 minutes.

The reliability of the ASR was examined in an independent pre-test. In this test, we compared the positions of phoneme boundaries placed by the ASR to the positions of the same boundaries placed by a trained phonetician. The test materials consisted of 189 randomly selected words from the current experiment. The ASR-generated boundaries were obtained by providing both a parameterized acoustic signal and a phonemic transcription to a Viterbi algorithm, which, using steps of 10 milliseconds, determined the most likely segmentation of the signal given the pre-trained phone models. Comparison between the ASR-generated and hand-made segmentations revealed that 76% of the automatic boundaries was placed within 20 milliseconds of the corresponding hand-coded boundary. The main discrepancies were found in the beginnings of plosives and liquids, which were consistently placed earlier by the ASR than by the phonetician. If the automatic boundaries were shifted 10 and 7 milliseconds to the right, respectively, the percentage of boundaries placed within 20 milliseconds of each other increased to 81%. This level of accuracy is in accordance with international standards (Vorstermans et al., 1996; Sjölander, 2001), and was considered sufficient for the present purposes.

Acoustic analysis of the materials proceeded in three steps. First, the sound file corresponding to a particular target word was parameterized using Mel Frequency Cepstral Coefficients. Then, the parameterized signal was provided to a Viterbi segmentation algorithm, together with a phonemic transcription of the word. This transcription was determined on the basis of two separate reference transcriptions made by two different transcribers. If the transcribers agreed, the consensus transcription was used. If there was disagreement, a third transcriber picked the best transcription of the two. After the segmentation algorithm had segmented the signal on the basis of the phonemic transcription, the beginnings of plosives and liquids were shifted 10 and 7 milliseconds to the right to correct for measurement error.

Statistical analysis

The data were analyzed using linear mixed effect models with subject and item as crossed random effects (Bates & Sarkar, 2005). Separate models were fitted to predict the log-transformed RTs and the log-transformed affix durations. The RT analysis was mainly included to check whether the effects that are normally observed in word naming were also present in the current experiment.

The main predictor variable in the analyses was the log frequency of the word in the

Corpus of Spoken Dutch (Oostdijk, 2000). In addition, several control variables were included. The most important control variable was presentation rate, which had three levels (slow, medium, and fast). Three variables were added to investigate the role of morphological and semantic processing. These were the frequency of the morphological stem in the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995), the number of morphological family members of the word (Schreuder & Baayen, 1997), and the number of synonym sets in which the word participates in WordNet (Miller, 1990; Vossen, Bloksma, & Boersma, 1999). All these values were log-transformed. Other control factors were the length of the word in letters, the affix under investigation, the number of realized segments in the stem (Nooteboom, 1972), the position of the word in the list, the position of the word in the experiment, and the sex of the subject. To control for voice key artefacts, the RT model also included the manner of articulation of the first segment.

The regression models were fitted in multiple cycles. First, all variables were included in the model, as well as their second-order interactions. The variables that showed significant effects were retained and the model was refitted. If any of the variables failed to reach significance in the re-fitted model, it was dropped and the procedure was repeated. The final model contained all variables that showed a significant main effect, as well as all variables that were involved in one or more significant interactions. To eliminate overly influential outliers, all data points with a residual value higher than 2.5 times the standard deviation of residuals were removed from the data set. Then, the model was refitted to the remaining data. This is the model that is reported below.

Results

The reaction time data for two participants were lost as a result of computer error. The model was fitted to all 3983 data points for which the registered RT was longer than 200 milliseconds and for which the participant had produced the correct response (79% of all trials). 136 data points (3.4% of the correct trials) were removed as outliers. The model that was refitted to the remaining data showed that words were named at shorter latencies the higher their frequency of occurrence ($\hat{\beta} = -0.006, t(3864) = -4.57, p < 0.0001$). Presentation rate also had a significant effect ($F(2, 3864) = 290.91, p < 0.0001$). Words were named faster in the medium condition than in the slow condition ($\hat{\beta} = -0.042, t(3864) = -11.07, p < 0.0001$), faster in the fast condition than in the slow condition ($\hat{\beta} = -0.098, t(3864) = -24.05, p < 0.0001$), and faster in the fast condition than in the medium condition ($\hat{\beta} = -0.057, t(3864) = -13.87, p < 0.0001$). Words were named slower the later they occurred in the list ($\hat{\beta} = 0.002, t(3864) = 6.50, p < 0.0001$) and the longer their length in letters ($\hat{\beta} = 0.006, t(3864) = 2.46, p < 0.05$). Furthermore, we found significant effects of manner of articulation ($F(4, 3864) = 6.52, p < 0.0001$) and affix ($F(3, 3864) = 56.05, p < 0.0001$). Since three of the affixes under investigation were prefixes, both of these effects seem to reflect the differential sensitivity of the voice key to different segments. No effects were observed for stem frequency, morphological family size, or number of synonym sets. There were also no significant interactions between variables. The standard deviations of the random effects were 0.09 for subject, 0.03 for item, and 0.10 for the residuals.

The model to predict affix duration was fitted to all 4786 data points for which the duration of the affix was longer than zero and for which the participant had produced the correct response (94% of all trials). 109 data points (2.3% of the correct trials) were removed as outliers. The final model showed that a higher frequency of the word was correlated with a shorter articulatory duration of the affix ($\hat{\beta} = -0.041, t(4664) = -2.82, p < 0.005$). This main effect was modified by an interaction with the number of realized segments in the stem ($F(1, 4664) = 6.68, p < 0.01$): The greater the number of segments in the stem, the smaller the effect of frequency ($\hat{\beta} = 0.007, t(4664) = 2.59, p < 0.01$). There were no interactions between frequency and affix or between frequency and presentation rate. By itself, presentation rate had a significant effect on affix duration ($F(2, 4664) = 223.19, p < 0.0001$). Affixes were shorter in the medium condition than in the slow condition ($\hat{\beta} = -0.033, t(4664) = -4.79, p < 0.0001$), shorter in the fast condition than in the slow condition ($\hat{\beta} = -0.142, t(4664) = -20.20, p < 0.0001$), and shorter in the fast condition than in the medium condition ($\hat{\beta} = -0.109, t(4664) = -15.53, p < 0.0001$). This shows that our manipulation of presentation rate was successful in speeding up articulation. As expected, affixes differed in their intrinsic durations ($F(3, 4664) = 355.38, p < 0.0001$). Furthermore, there was an interaction between affix and the number of realized segments in the stem ($F(3, 4664) = 6.51, p < 0.0005$). Apparently, the number of realized segments in the stem was significant for *ont-*, *ver-*, and *-lijk*, and but not for *ge-*. Finally, a significant main effect was observed for the length of the word in letters ($\hat{\beta} = -0.035, t(4664) = -4.36, p < 0.0001$), showing that affixes were shorter the longer the word they occurred in. No effects were found for stem frequency, morphological family size, or number of synonym sets. The standard deviations of the random effects were 0.12 for subject, 0.09 for item, and 0.19 for the residuals.

Discussion

The main result in Experiment 1 is that affixes were shorter the higher the frequency of the word they occurred in. This finding has several implications. First of all, it shows that at least part of the frequency effect on articulatory durations is purely speaker-based, as there was no listener present in the set-up of the current experiment. Of course, the possibility that our participants were speaking to imaginary listeners cannot be excluded completely, although it seems unlikely given the high cognitive demands associated with the task they had to perform. Second, our results demonstrate that effects of word frequency are not restricted to the initial phoneme of a word, as was suggested by Kawamoto et al. (1999). One of the affixes under investigation was the suffix *-lijk*, and since no interaction was observed between frequency and affix, there is no reason to assume that the effect of frequency was smaller for *-lijk* than for any of the prefixes.

The observation that the effect of frequency was equally strong for all affixes runs counter to the results in Chapter 2, where we failed to find a frequency effect for the prefix *ver-*. A possible explanation for this difference between the two studies is that the speech materials in the current study were less noisy than the spontaneous speech samples investigated in Chapter 2.

We also found that the frequency effect on articulatory durations was quite robust, in that it was present even if the presentation rate of the stimuli was very high. However, we did observe a ceiling effect of a different kind: The frequency effect became smaller as the number of realized segments in the stem of the word increased. This suggests that the amount of durational shortening related to frequency depends directly on the number of segments that speakers are going to produce. A possible explanation for this effect is the use of deadlines in the current experiment; It remains to be seen whether something similar can be observed when participants have more time available to produce their response.

The lack of effects for stem frequency, morphological family size, and number of synonyms indicates that the frequency effect is not likely to be confounded with semantic processing, but rather arises at the form level of the speech production process.

In an additional analysis, we tested whether reaction time was also a significant predictor of affix duration. This turned out not to be the case. Apparently, the articulatory duration of a word is not related to its naming latency. Moreover, we can assume that the frequency effect we observed is not related to the perceptual component of the naming task, which is reflected in the response latencies.

Before we can draw any definitive conclusions, however, we need to address an alternative explanation for the effect of frequency on articulatory durations. In a study on frequency effects in lexical decision, Balota and Abrams (1995) found that responses to high-frequency stimuli tend to be shorter than responses to low-frequency stimuli, regardless of the nature of the response. So, when Balota and Abrams asked subjects to produce the word *normal* every time they saw an existing word, they found that the duration of this word was shorter if the frequency of the stimulus word was high. This finding was replicated in a follow-up experiment with a non-existing response word (*berloe*). Balota and Abrams (1995) explain these results in terms of an *enabled response* model, which states that information accrues faster when subjects react to high-frequency stimuli, allowing responses to be initiated faster and executed more efficiently.

Since the words produced in our experiment were also responses to stimuli with different frequencies, it is not inconceivable that our finding stems from an enabled response mechanism. To make sure that the frequency effect we observed was related to the words the participants had to pronounce, rather than to the words they read, we need to separate the effects of response frequency from the effects of stimulus frequency. This was the main goal of Experiment 2.

Experiment 2

In Experiment 2, we independently manipulated the frequency of the stimulus and the frequency of the verbal response associated with that stimulus. In this way, we could determine whether the effects observed in Experiment 1 were simply due to faster information accrual, or whether they were actually related to the speech production process. Because we wanted each verbal response to be associated with an existing word, we chose noun/verb categorization as the experimental task. Frequency effects have earlier been observed in this task, albeit only on response latencies (e.g., Sereno & Jongman, 1990).

Method

Participants

Twenty subjects (5 male, 15 female) participated in the experiment in exchange for pay. All participants were native speakers of Dutch with normal or corrected-to-normal vision. None of them had participated in Experiment 1.

Materials

The stimulus words were 300 Dutch nouns (200 simplex, 100 complex) and 300 Dutch verbs (200 simplex, 100 complex). All nouns were singular forms, and all verbs were infinitives. The stimuli, which spanned the entire frequency range, were divided into 30 lists so that each list contained 10 nouns and 10 verbs. Furthermore, each stimulus occurred in only one list.

As responses, we selected 60 adjectives ending in the suffix *-lijk*. These adjectives spanned the entire frequency range and were divided into 30 response pairs. The two members of a pair had similar frequencies, the same number of syllables, and the same stress pattern. For each pair, it was randomly decided which member served as response to a noun and which member served as response to a verb. Finally, each response pair was randomly linked to one of the thirty stimulus lists. These associations were not varied between participants.

Design

The main experimental variables were the frequency of the stimulus word and the frequency of the response word. Each subject saw all 30 stimulus lists, albeit in a different random order.

Within a list, a stimulus always occurred in the same position.

Procedure

Participants were seated in a sound-attenuated booth, behind a Sennheiser microphone. The stimuli were presented on a computer screen, and participants were asked to categorize them as nouns or verbs by means of a pre-learned verbal response. Their responses were recorded on a Tascam DA-20 DAT tape recorder and digitized with a sampling frequency of 16 kHz. In addition, naming latencies were registered using a voice key device.

Before the start of each list, participants received a paper card specifying for that particular list which word they were supposed to say if they saw a noun and which word they were supposed to say if they saw a verb. For example, the response words linked to a particular list could be *vreselijk* 'horrible' for a noun and *eindelijk* 'finally' for a verb. Each list was preceded by five practice trials, in which participants were shown either NAAMWOORD (the Dutch equivalent of 'noun') or WERKWOORD (the Dutch equivalent of 'verb') on the computer screen. As soon as the participant had memorized the correct responses for these two categories, the actual stimulus list started. For each stimulus, participants had 2600 milliseconds to complete their response.

Acoustic measurements

The acoustic measurements were performed using the same apparatus and procedure as in Experiment 1.

Statistical analysis

Data analysis proceeded in largely the same way as in Experiment 1. One important difference was that there were now three crossed random effects: Subject, stimulus word, and response word. Furthermore, there were some changes in the control variables. Stem frequency, morphological family size, and number of synonym sets were dropped, as they failed to show an effect in the first experiment. Two new variables were added: The syntactic category of the stimulus (noun or verb) and the number of times a particular response had already been given during the preceding part of the list.

Results

The model to predict RTs was fitted to all 11240 data points for which the RT was longer than 200 milliseconds and for which the participant had produced a correct response (94% of all trials). 258 data points (2.3% of the correct trials) were considered outliers and removed from the data set. In the final model, no effects were observed for either stimulus frequency or response frequency. However, there was an effect of the length of the stimulus in letters: The longer the stimulus, the longer the RT ($\hat{\beta} = 0.029, t(10971) = 8.26, p < 0.0001$). Furthermore, nouns were responded to faster than verbs ($\hat{\beta} = -0.167, t(10971) = 5.58, p < 0.0001$). Participants produced faster responses the later a stimulus occurred in the experiment ($\hat{\beta} = -0.0002, t(10971) = -12.09, p < 0.0001$), but slower responses the later a stimulus occurred in a list ($\hat{\beta} = 0.004, t(10971) = 2.10, p < 0.05$). Apart from these main effects, there were five significant interactions. Three of these involved the syntactic category of the stimulus. The interaction between category and length ($F(1, 10971) = 48.36, p < 0.0001$) showed that the effect of length was virtually absent for verbs ($\hat{\beta} = -0.024, t(10971) = -6.93, p < 0.0001$). Similarly, the effect of position in the experiment was smaller for verbs than for nouns ($F(1, 10971) = 6.09, p < 0.05$; $\hat{\beta} = -0.00003, t(10971) = 2.50, p < 0.05$), as was the effect of position in the list ($F(1, 10971) = 6.53, p < 0.05$; $\hat{\beta} = -0.003, t(10971) = -2.56, p < 0.05$). The interaction between length and position in the list ($F(1, 10971) = 9.49, p < 0.005$) indicated that the effect of length was smaller the later a stimulus occurred in a list ($\hat{\beta} = -0.0007, t(10971) = -2.45, p < 0.05$). Finally, the effect of position in the experiment was smaller for female than for male participants ($F(1, 10971) = 16.31, p < 0.0001$; $\hat{\beta} = 0.00006, t(10971) = 4.03, p < 0.0001$). The standard deviations of the random effects were 0.12 for subject, 0.07 for stimulus word, 0.03 for response word, and 0.18 for the residuals.

The model for duration was fitted to all 11325 data points for which the duration of *-lijk* was longer than zero and for which the participant had produced a correct response (94% of all trials). 220 data points (1.9% of the correct trials) were removed as outliers. Of the two experimentally manipulated variables, only response frequency emerged as a significant predictor: The duration of the suffix *-lijk* was shorter the higher the frequency of the word it occurred in ($\hat{\beta} = -0.013, t(11093) = -3.18, p < 0.005$). This effect was modified by two significant interactions. First of all, the effect was larger for female than for male participants ($F(1, 11093) = 4.26, p < 0.05$; $\hat{\beta} = -0.004, t(11093) = -2.23, p < 0.05$). Second, it was smaller the later a stimulus occurred in a list ($F(1, 11093) = 8.55, p < 0.005$; $\hat{\beta} = 0.004, t(11093) = 2.94, p < 0.005$). There were also significant main effects of sex, stimulus length, position in the list, position in the experiment, and the number of realized segments in the stem. The duration of the suffix was shorter for male participants than for female participants ($\hat{\beta} = -0.181, t(11093) = 2.41, p < 0.05$), shorter for stimulus words consisting of fewer letters ($\hat{\beta} = -0.001, t(11093) = 2.15, p < 0.05$), shorter the later a stimulus occurred in a list ($\hat{\beta} = -0.005, t(11093) = -6.99, p < 0.0001$), shorter the later a stimulus occurred in the experiment ($\hat{\beta} = -0.0002, t(11093) = -9.48p < 0.0001$), and shorter the more segments were realized in the stem ($\hat{\beta} = -0.043, t(11093) = -16.22, p < 0.0001$). Three additional interactions showed that the effect of position in the experiment was smaller for female than for male participants ($F(1, 11093) = 10.41, p < 0.005$; $\hat{\beta} = 0.00003, t(11093) = 3.15, p < 0.005$), that the effect of position in the list was smaller for female than for male participants ($F(1, 11093) = 31.11, p < 0.0001$; $\hat{\beta} = 0.003, t(11093) = 5.56, p < 0.0001$), and that the effect of number of realized segments in the stem was smaller the later a stimulus occurred in the experiment ($F(1, 11093) = 15.05, p < 0.0005$; $\hat{\beta} = 0.00001, t(11093) = 3.88, p < 0.0005$). The standard deviations of the random effects were 0.14 for subject, 0.00 for stimulus word, 0.05 for response word, and 0.12 for the residuals.

Discussion

In Experiment 2, we manipulated both the frequency of the stimulus word and the frequency of the response word. From the results, we can conclude that the response latency was affected by neither of the two variables, while the duration of the suffix was only affected by the frequency of the response word. For our purposes, the second finding is the more important of the two, as it shows that the frequency effect observed in Experiment 1 was related to what the participants had to say, rather than to what they saw. Furthermore, it suggests that the frequency effect on articulatory durations arises very late during the speech production process, given that the responses in the current experiment were well-practiced and therefore did not require extensive conceptual preparation.

It may seem surprising that stimulus frequency did not show an effect on either RTs or durations. This finding might be explained by pointing to the role of stimulus length. In both models, length emerged as a highly significant predictor. Since length and frequency are inversely correlated, the presence of length in the models might preclude stimulus frequency from becoming significant. We explored this possibility by fitting a simple linear model in which we predicted the frequency of a stimulus on the basis of its length. The residuals of this model, which reflect an estimate of frequency that is independent of length, were then entered into the models. The residualized frequency failed to show an effect in either model, while length remained a significant predictor in both models. This shows that the absence of effects for stimulus frequency is not related to the effects we observed for length. Rather, it appears that we simply failed to replicate Balota and Abrams' finding that the frequency of the stimulus affects the duration of the response. This could be due to the task we used, which differed from the task employed by Balota and Abrams. Furthermore, Balota and Abrams used only one response word, which may make it easier to demonstrate effects of stimulus frequency. Finally, there is a possibility that Balota and Abrams' results were also in some way related to stimulus length rather than frequency.

General Discussion

In this section, we review our results and discuss their implications for models of speech production. In the current study, we replicated the results of Chapter 2 experimentally: We found an effect of frequency on duration in two separate experiments using different experimental tasks. This shows that there is indeed an effect of word frequency on articulatory durations. In addition, our findings provide considerable insight into the psychological processes underlying the effect.

First of all, we know now that effects of frequency on articulatory durations can also be observed in the absence of a listener. This finding rules out any account which states that frequency effects occur only because the speaker knows that the listener can easily understand high-frequency words anyway. At least part of the effect is purely speaker-based, in that it arises as an automatic, non-consciously controlled by-product of the speech production process.

We also saw that the effect of frequency on duration is quite robust. In Experiment 1, the presentation rate of the stimuli was varied to investigate whether the frequency effect would become smaller as speech rate increased. This turned out not to be the case: The effect was equally strong for all three presentation rates. However, participants did seem to take the total number of segments they had to produce into account, given that the effect of frequency became smaller as the number of segments increased. No such interaction was observed in Experiment 2, where participants had plenty of time to produce their verbal responses.

What makes the frequency effect on articulatory durations interesting from a modeling perspective, is its implication that the time-course of articulation is not immune to external influences. In most models of speech production, the process of articulation is assumed to be informationally encapsulated. This means that once a particular word has been passed on to the Articulator, articulation is initiated and executed without interference from factors like word frequency or semantic context (e.g., Damian, 2003). Our results, as well as Balota and Abrams' (1995) findings, show that this assumption is no longer warranted. To accommodate our findings, models of speech production should allow cascaded processing all the way down to the articulation phase.

There is another reason why our results cannot easily be explained by the current production models. In a model like that of Levelt et al. (1999), a word that has been selected for production is first morphologically and then phonologically encoded. During phonological encoding, the word is resyllabified in order to fit the syllabic structure of the language. Subsequently, the gestural scores associated with each of the word's syllables are retrieved from the so-called syllabary. In the final stage, these articulatory plans are passed onto the Articulator, which executes them. In such an architecture, it is difficult to see how word frequency can affect articulatory durations. A possible alternative would be to assume that the gestural scores fed to the Articulator are not syllables, but complete words (cf. Johnson, 2006), and that the Articulator executes often-occurring gestural scores more efficiently than rarer ones.

Some of our findings point to interesting directions for future research. For example, the interaction between frequency and sex observed in Experiment 2 suggests that women are more sensitive to frequency information than men. Although there have been other studies reporting such an effect (e.g., Ullman et al., 2002; Lemhöfer et al., submitted), more research is needed to substantiate this claim. Another striking finding in Experiment 2 was that verbs showed smaller effects than nouns for several variables, including length in letters, position in the list, and position in the experiment. Again, similar patterns have been reported before (e.g., Baayen, Dijkstra, & Schreuder, 1997), but additional research remains necessary to shed more light on this issue.

In conclusion, we have provided firm evidence for the existence of a frequency effect on articulatory durations. This effect, which has received relatively little attention in the speech production literature, can be observed in different experimental paradigms, as well as in corpus data. Furthermore, it is demonstrably related to the speech production process. As such, it may be a very useful resource for improving existing models of speech production.

Effects of repetition and contextual predictability: Corpus data

CHAPTER 4

This chapter has been published as Mark Pluymaekers, Mirjam Ernestus, and R. Harald Baayen (2005b). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, 146-159.

Abstract

In this chapter, we shifted our focus from effects of word frequency to effects of predictability from context. Two dimensions of predictability were investigated, namely word repetition and predictability given the neighbouring word. For the seven most frequent words ending in the adjectival suffix *-lijk*, 40 occurrences were randomly selected from a large database of face-to-face conversations. Analysis of the selected tokens showed that the degree of articulatory reduction (as measured by duration and number of realized segments) was affected by repetition, predictability from the previous word, and predictability from the following word. Interestingly, not all of these effects were significant across morphemes and target words. Repetition effects were limited to suffixes, while effects of predictability from the previous word were restricted to the stems of two of the seven target words. Predictability from the following word affected the stems of all target words equally, but not all suffixes. The implications of these findings for models of speech production are discussed.

Introduction

We speak in order to be understood. Nevertheless, the dynamics of conversational interaction may force speakers to reduce articulatory effort on certain words, leading to a temporary decrease in intelligibility. Although reductions occur frequently in spontaneous speech and can be quite extreme (Ernestus, 2000; Kohler, 2000; Johnson, 2004), there is little evidence that their presence actually hinders communication. This has been explained by the hypothesis that speakers only reduce articulatory effort on words that are predictable for the listener, either from the linguistic context or from the situation in which the interlocutors find themselves (e.g., Lindblom, 1990; Jurafsky, Bell, Gregory, & Raymond, 2001). With regard to linguistic context, two factors have received much attention in the literature: *word repetition* and the predictability of a word from its neighbouring words (henceforth, *contextual predictability*). These two factors have in common that they are both concerned with the *informational redundancy* of a word in its context. In this introduction, we review the relevant literature for both variables.

Repetition

Effects of word repetition on reduction were first reported by Fowler and Housum (1987). They found that second mentions of words in monologues were shorter and less intelligible in isolation than first mentions. Bard, Anderson, Sotillo, Aylett, Dohert-Sneddon, and Newlands (2000) replicated this effect for dialogues, showing that it was present irrespective of whether the speaker or the listener had uttered the first token of the word. No repetition effects were found when subjects read words in lists (Fowler, 1988), or when two tokens in a monologue were divided by a major episode boundary (Fowler, Levy, & Brown, 1997). This suggests that it is not so much repetition that matters, but rather whether a word refers to ‘given’ or ‘new’ information.

Hawkins and Warren (1994) argued, however, that first and second occurrences of words differ not only in whether they present ‘given’ or ‘new’ information, but also in their likelihood of carrying sentence accent. First occurrences of content words are more likely to be accented than second occurrences, which could also explain the observed differences in duration and intelligibility. In their study, Hawkins and Warren tried to disentangle repetition effects from effects due to sentence accent and segmental identity. They found no differences in intelligibility between first and second tokens that could not be accounted for by the presence or absence of accent. This led them to conclude that “local phonetic variables, notably sentence accent and the phonetic and phonological properties of individual segments, exert a greater influence on intelligibility than whether or not a word has been used before in the conversation” (Hawkins & Warren, 1994, p. 493).

Does this mean that repetition by itself should no longer be considered a possible predictor of reduction? Recent findings by Gregory, Raymond, Bell, Fosler-Lussier, and Jurafsky (1999) and Aylett and Turk (2004) suggest otherwise. Both studies report effects of the *number* of previous mentions of a word on its duration. This shows that durational differences cannot only be observed between first and second mentions, but also between, for instance, fifth and tenth mentions. Since neither the fifth nor the tenth token of a word in a conversation are likely to be accented, these reductions are probably not due to de-accentuation alone. In other words, there seems to be more to repetition effects than just the presence or absence of sentence accent.

Contextual predictability

Ever since Lieberman (1963), the relationship between contextual predictability and acoustic realizations has captivated researchers in phonetics, linguistics, and psycholinguistics alike. To determine the predictability of their target words, authors have used Cloze tests (e.g., Hunnicutt, 1985) or, as the availability of large speech corpora increased, co-occurrence statistics based on frequency. These statistics can be computed for a wide variety of linguistic units, from syllables (e.g., Aylett & Turk, 2004) to complete syntactic structures (e.g., Gahl & Garnsey, 2004). Most studies, however, focus on words (e.g., Gregory et al., 1999; Fosler-Lussier & Morgan, 1999; Jurafsky et al., 2001; Bush, 2001; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003). Two well-known measures of contextual probability are *conditional probability* and *mutual information*, both of which capture the likelihood of a certain word occurring given one or more of its neighbouring words. We will discuss these measures in more detail below.

Previous studies on the effects of contextual predictability on reduction have produced results that are both consistent and inconsistent. They are consistent in that all significant effects go in the same direction: Words that are more likely to occur are more reduced. They are inconsistent, however, with regard to the relevance of the different measures. Some words are completely unaffected by predictability, while others show effects of two or three probabilistic measures at the same time (Fosler-Lussier & Morgan, 1999; Bell et al., 2003). Furthermore, the results of the various studies are difficult to compare, since all studies used slightly different sets of dependent and independent variables. These methodological differences have directed attention away from other important issues, such as the cognitive and articulatory processes underlying the effects.

Table 4.1: English translations, frequencies (per million and in total) in the Corpus of Spoken Dutch , and citation forms of the seven words investigated in this study.

Word	Eng. translation	Frequency (per million)	Frequency (in total)	Citation form
eigenlijk	actual/actually	1922	18320	/ɛixələk/
natuurlijk	natural/naturally	1440	13602	/natyrlək/
waarschijnlijk	probable/probably	335	3098	/ʊarsxɛinlək/
moeilijk	difficult	320	2736	/mujlək/
duidelijk	clear/clearly	272	2056	/dœydələk/
namelijk	namely	135	1154	/namələk/
makkelijk	easy/easily	96	872	/mɑkələk/

Our approach

It is clear that for both repetition and contextual predictability, several issues remain to be addressed. In this study, we focus on two questions. First, is there an effect of repetition on reduction that is independent of sentence accent and second, what do effects of repetition and contextual predictability reveal about speech production processes? Like most of the previous studies, we use corpus data to investigate these issues. What is new in our approach, is that we focus on words that are morphologically complex. By studying words that have internal structure, we hope to learn more about the effects of repetition and predictability on different parts of the word.

We concentrate on the seven most frequent words ending in the Dutch suffix *-lijk*. These words, which are listed in Table 4.1, are suitable targets for several reasons. First of all, words ending in *-lijk* can be extremely reduced (Ernestus, 2000), and these reductions are at least partly predictable from probabilistic measures such as word frequency (see Chapters 2 and 3) and mutual information (Keune, Ernestus, Van Hout, & Baayen, 2005). Second, being adverbs and adjectives, *-lijk* words are a priori less likely to carry sentence accent. This is especially true for *eigenlijk*, *natuurlijk*, and *namelijk*, which mainly serve as discourse markers. *Duidelijk*, *waarschijnlijk*, *makkelijk*, and *moeilijk* can also function as predicates, presenting new information about the discourse topic. Therefore, the possibility that these words are accented cannot be completely excluded.

As is clear from Table 4.1, the seven target words investigated in this study differ in frequency, phonemic content, meaning, and the number and type of discourse functions they can perform. Since all of these factors can be expected to affect reduction, failure to control for them may limit the possibility of finding effects of repetition or contextual predictability. To overcome this problem, we incorporated the factor ‘word’ as a fixed effect in our analyses.

Of course, this does not control for the discourse function performed by a particular token of a word. There is no reason to assume, however, that discourse function is systematically correlated with either repetition or contextual predictability. It seems very unlikely that, for instance, the 15th occurrence of a word is always associated with one particular discourse function, while the 20th occurrence always performs another. We assume that for our target words, discourse functions are more or less randomly distributed among different tokens, independent of the number of times the word has been mentioned before or how predictable the token is on the basis of neighbouring words. Furthermore, classifying words according to their discourse function is a notoriously difficult activity, which can be regarded as a research topic in itself. Therefore, we considered this beyond the scope of the current chapter.

Still, there were a lot of other variables that had to be controlled. To this end, we used multiple regression analysis. In such an analysis, it is also easy to check whether the effects of two or more variables are additive or interactive. If, for example, repetition effects were to be limited to non-accented words, this would surface in our analyses as a significant interaction between repetition and accent. If, on the other hand, both repetition and accent show significant main effects, this implies that their effects are additive and not confounded. This kind of information is necessary for answering the two research questions formulated above.

Materials and method

The materials were taken from the subcorpus ‘Spontaneous speech’ of the Corpus of Spoken Dutch (Oostdijk, 2000). This subcorpus contains 225 hours of face-to-face conversations, all of which have been orthographically transcribed. We restricted ourselves to speakers from the Netherlands, since they have been shown to use reduced forms more often than speakers from Flanders (Keune et al., 2005). For each of the words in Table 4.1, a randomized list was made of all occurrences in the subcorpus that were not surrounded by pauses or disfluencies. From this randomized list, the first 40 tokens were selected for further analysis. If the recording quality of a selected token was too poor for acoustic measurements, it was replaced with the next token on the list. In total, 280 tokens were analyzed.

The dependent variables in this study were the durations of the stem and the suffix and the number of realized segments in these two morphemes. All acoustic measurements were made by a trained phonetician with the help of the software package PRAAT (Boersma, 2001). Boundaries were placed between the previous word and the stem, between the stem and the suffix, and between the suffix and the following word. If a segment was ambiguous as to whether it belonged to the stem or the suffix (like the [ə] in the realization [namək] for *namelijk* /namələk/), it was considered part of the suffix. In addition, the phonetician determined for each token which segments were realized in the speech signal. This transcription did not start from the citation forms of the target words, but was purely based on the auditory evidence in the signal and the visual information in the waveform. A particular segment was only included in the transcription if there was both visible and auditory evidence for its presence. Finally, the labeller coded for each token whether the stem carried pitch accent or not.

Since most recordings contained at least some background noise, it was hard to establish clear-cut segmentation criteria (see also Vorstermans, Martens, & Van Coile, 1996). Figure 4.1 shows the manual segmentations for two tokens of the word *duidelijk*, including parts of the previous and the following word. The top token was relatively easy to segment, since there was hardly any background noise or overlapping speech. The bottom token was much harder, mainly due to the presence of overlapping speech. In all cases, the phonetician placed boundaries where she could see visible changes in the waveform pattern supported by abrupt formant transitions in the spectrogram.

The fact that all tokens were measured only once may have implications for the generalizability of our results. After all, there is no guarantee that a second measurement of the same tokens, even if performed by the same labeller, would yield exactly the same results. On the other hand, it is unlikely that our results are completely due to the labeller's idiosyncracies, as she was naive with respect to the goals of the study and used similar criteria for all tokens.

To assess the effects of repetition, we determined for each randomly selected token how often the target word had been uttered during the conversation before the selected token occurred. We coded the selected item for the time point during the conversation at which it occurred (for example, after 54 seconds), and counted how often the same word had been uttered before that time point. Given the results of Bard et al. (2000), we did not distinguish between tokens uttered by the same speaker and tokens uttered by other speakers. To reduce the effects of extreme counts, all values were increased by 1 and logarithmically transformed. The original counts varied between 0 and 20, while the transformed values ranged between 0 and 3,04.

As mentioned earlier, contextual predictability can be established in various ways. To avoid the problems associated with testing several probabilistic measures at the same time, we focused on just two variables: Mutual information between the target word and the previous word and mutual information between the target word and the following word. The mutual information between two words is a measure of the reduction in uncertainty about one word due to knowing about the other (e.g., Manning & Schütze, 1999). Therefore, the higher the value for mutual information, the easier one word can be predicted on the basis of the other. To compute mutual information, we used the following equation (X and Y denote either the previous word and the target word, or they denote the target word and the following word; XY denotes the combination of the two words):

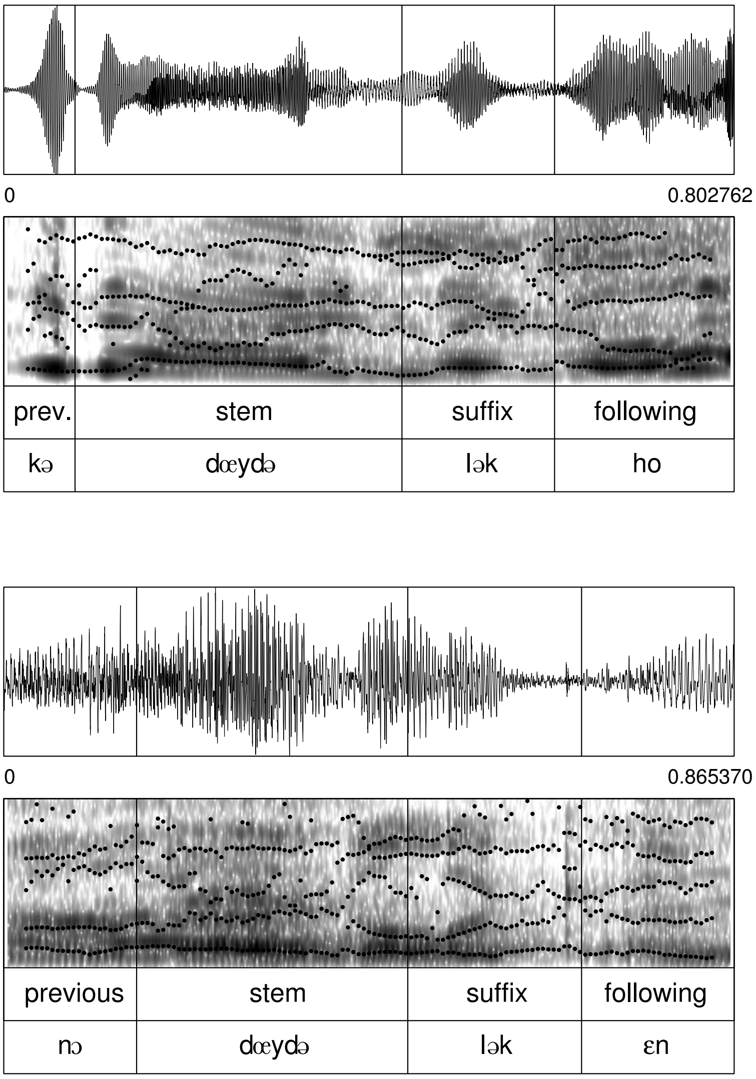


Figure 4.1: Two segmentation examples of the word *duidelijk*. The top token was produced without background noise or overlapping speech, resulting in a waveform that was relatively easy to segment. In the bottom token, it was much harder to determine segment boundaries. In both cases, we placed boundaries where we could see visible changes in the waveform pattern supported by abrupt formant transitions in the spectrogram.

$$MI(X;Y) = log \frac{log\,(Frequency(XY))}{log\,(Frequency(X)) * log\,(Frequency(Y))}$$

The frequency estimates were logarithmically transformed before entering the equation. This was done to minimize the effects of very high frequencies on the outcome of the computation. Furthermore, language users are known to be sensitive to logarithmic values rather than raw frequencies (Rubenstein & Pollack, 1963). All frequency estimates were obtained from the Corpus of Spoken Dutch.

The reliability and stability of the mutual information measure depends crucially on corpus size. If the corpus is too small, the frequency counts for many two-word combinations (the numerator in the equation above) will approach zero, leading to unstable estimates. This was not the case in our sample, as the frequencies of the sampled word combinations ranged from 1 to 1520. As a consequence, the distribution of the log transformed values entering the equation was reasonably symmetric.

Eight other variables known (or expected) to affect reduction were taken into account and designated control variables. First, there were the speaker characteristics Sex, Year of Birth, Education Level, and Region of Secondary Education. Second, Speech Rate (in syllables per second) was computed over the largest chunk of speech containing the target word that did not contain an audible pause. The number of syllables in the chunk was determined on the basis of two sources of information. For all words in the chunk except the target word, we counted the number of vowels in the orthographic transcription. For the target word itself, we counted the number of vowels in the manual segmentation. The total number of vowels was then divided by the overall duration of the chunk. The remaining three control variables were the presence of Pitch Accent on the stem, whether the segment following the target word was a consonant or a vowel (henceforth, Following Segment), and, for reasons explained above, Word. Table 4.2 gives an overview of the most important sample characteristics for each of the seven target words separately.

Results

Analysis

In total, six regression models were fitted: three for the durations of the stem, the suffix, and the word as a whole, and three for the number of realized segments in the stem, the suffix, and the word as a whole. To find the best model in each case, we used a strict model selection procedure. First, we entered the control factors into the model, retaining only those variables that showed a significant effect. Then, the number of previous mentions (Mentions) was added, followed by mutual information with the previous word (MI Previous) and mutual information with the following word (MI Following). If any of these variables failed to show

Table 4.2: Information about the sampled tokens for each of the target words separately (N = 40 for each target word). This information includes the number of tokens carrying pitch accent, the number of different previous and following words observed, the lower and upper values for both Mutual Information measures, and the maximum number of previous mentions observed.

Word	Accented tokens	Previous words	Following words	MI Previous lower	MI Previous upper	MI Following lower	MI Following upper	Max. prev. mentions
eigenlijk	3	30	25	-4.13	-2.80	-4.53	-2.75	20
natuurlijk	0	27	23	-4.51	-2.73	-4.54	-2.25	15
waarschijnlijk	2	28	26	-4.83	-2.78	-4.48	-2.82	4
moeilijk	4	19	24	-3.81	-2.60	-4.28	-2.68	6
duidelijk	4	25	29	-4.68	-2.03	-4.74	-2.84	2
namelijk	2	27	32	-4.62	-2.88	-4.58	-1.95	1
makkelijk	5	21	31	-4.21	-2.65	-4.54	-2.61	6

a significant effect, it was dropped from the equation. The resulting model was checked for interactions between the different variables, which were retained if they added to the predictive power of the model. Subsequently, diagnostic plots were used to identify data points that were outliers with regard to leverage or Cook's distance values. These outliers (usually three or four data points) were removed and the model was re-fitted to the remaining data. If a factor was no longer significant after the removal of outliers, it was dropped and the last two steps of the procedure were repeated. Finally, a bootstrap validation was performed to check for overfitting. During bootstrapping, the proposed model was fitted 200 times to different random selections of our data points. If a particular variable in the model failed to reach significance in too many of these fitting cycles, it was removed from the model. Only those predictor variables that remained significant throughout this whole procedure are reported below.

Regression results

The results of the six regression models are summarized in Table 4.3. It shows for each model which of the predictor variables were significant. The beta coefficients indicating the direction and size of the effects are given in the main text below, as are the corresponding p-values. The factor Word was significant in all analyses, reflecting differences between the target words with respect to meaning, phonemic content and, possibly, word frequency. Since such differences are not the main interest of this study, these effects are not further addressed here. First, we discuss the results for the duration of the stem, followed by the results for the number of realized segments in the stem. These two steps are then repeated for the suffix and the word as a whole.

The stem was longer if it carried Pitch Accent ($\hat{\beta} = 51.0, t(260) = 4.61, p < 0.0001$) and shorter at higher Speech Rates ($\hat{\beta} = -15.3, t(260) = -6.30, p < 0.0001$). There was also an interaction between MI Previous and Word ($F(7, 260) = 6.63, p < 0.0001$), which is illustrated in Figure 4.2. MI Previous was significant for two of the seven target words: *natuurlijk* ($\hat{\beta} = -108.3, t(260) = -5.31, p < 0.0001$) and *eigenlijk* ($\hat{\beta} = -86.2, t(260) = -3.70, p < 0.0005$). In both cases, a higher value for MI Previous correlated with shorter realizations of the stem.

A similar interaction was observed for the number of realized segments in the stem ($F(7, 262) = 4.32, p < 0.0005$). Again, the shortening effect of MI Previous was limited to the words *natuurlijk* ($\hat{\beta} = -1.3, t(262) = -4.66, p < 0.0001$) and *eigenlijk* ($\hat{\beta} = -0.8, t(262) = -2.50, p < 0.05$). There was also a main effect of MI Following ($\hat{\beta} = -0.2, t(262) = -2.14, p < 0.05$), indicating that words that were more predictable from their following words were realized with fewer segments in the stem.

Ten of the 280 tokens in the data set contained no visible or audible trace of the suffix *-lijk* and were therefore excluded from the analyses for the suffix. The duration of the suffix was predicted by Sex ($\hat{\beta} = 11.8, t(248) = 2.30, p < 0.05$), Speech Rate ($\hat{\beta} = -16.4, t(248) =$

Table 4.3: Summary of the regression results for the six models fitted in this study. A star (*) indicates that the variable in question was a significant predictor. The horizontal line in the middle separates the variables of interest (above) from the control variables (below). The bottom row shows the amount of variance explained (R^2) by each model.

	Stem		Suffix		Entire word	
Predictor variable	Duration	Segments	Duration	Segments	Duration	Segments
Mentions			*			
MI Previous						
MI Previous * Word	*	*			*	*
MI Following		*				*
MI Following * Word			*			
Education Level						
Following Segment			*			
Pitch Accent	*				*	*
Region of Sec. Education						
Sex			*			
Speech Rate	*		*		*	
Word	*	*	*	*	*	*
Year of Birth					*	
Explained variance (R^2)	.58	.70	.41	.39	.51	.56

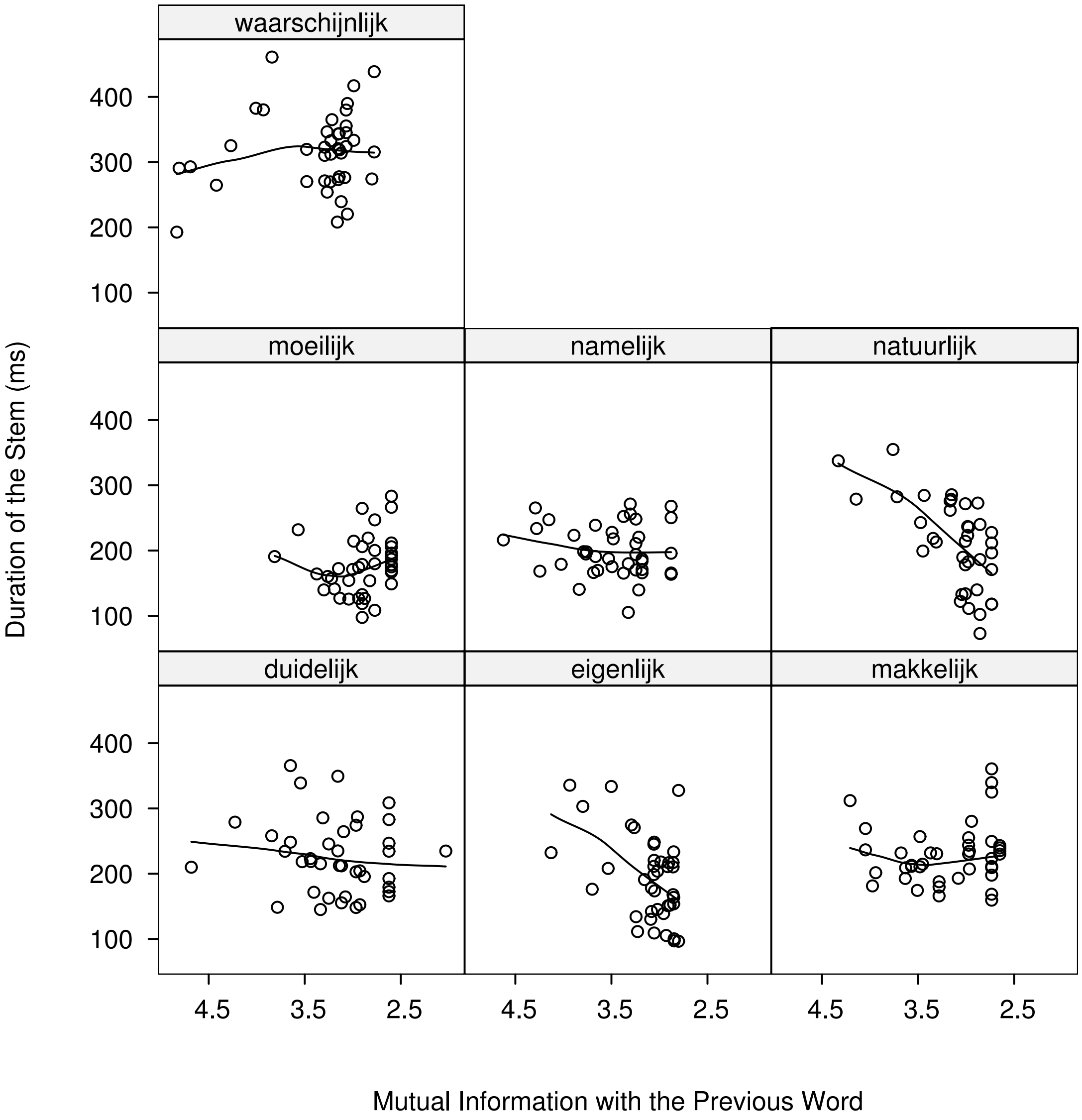


Figure 4.2: Duration of the Stem plotted against Mutual Information with the Previous Word for each of the seven target words separately. The numbers on the x-axis are supposed to have minus signs, but these disappeared due to a printing problem. As can be seen from the descending lines, a higher Mutual Information led to shorter realizations for *natuurlijk* and *eigenlijk*. The lines for *namelijk* and *duidelijk* also seem to fall somewhat, but these effects were not significant.

$-7.79, p < 0.0001$) and Following Segment ($\hat{\beta} = 14.1, t(248) = 2.72, p < 0.01$). Suffixes were longer if they were produced by women, longer if they were followed by a vowel, and shorter at higher Speech Rates. There was also a significant interaction between MI Following and Word ($F(7, 248) = 2.30, p < 0.05$). A higher value for MI Following led to shorter realizations of the suffix, but only for the target words *eigenlijk* ($\hat{\beta} = -41.0, t(248) = -2.30, p < 0.05$) and *namelijk* ($\hat{\beta} = -37.5, t(248) = -3.12, p < 0.005$). Finally, we found an effect of Mentions ($\hat{\beta} = -9.4, t(248) = -2.16, p < 0.05$): The more often the target word had been mentioned in the preceding discourse, the shorter the suffix. The number of realized segments in the suffix was only predicted by the factor Word.

The duration of the word as a whole was predicted by Year of Birth ($\hat{\beta} = -0.9, t(259) = -3.03, p < 0.005$), Speech Rate ($\hat{\beta} = -30.7, t(259) = -8.28, p < 0.0001$), and the presence of Pitch Accent on the stem ($\hat{\beta} = 59.7, t(259) = 3.66, p < 0.0005$). Older speakers produced longer words, words were shorter at higher Speech Rates, and an accented stem led to longer realizations of the word. Again, there was a significant interaction between MI Previous and Word ($F(7, 259) = 6.15, p < 0.0001$), which was very similar to the two interactions mentioned above for the stem. The main difference was that apart from *natuurlijk* ($\hat{\beta} = -147.3, t(259) = -4.90, p < 0.0001$) and *eigenlijk* ($\hat{\beta} = -126.1, t(259) = -3.67, p < 0.0005$), *namelijk* was also significantly shorter if the mutual information with the previous word was higher ($\hat{\beta} = -51.9, t(259) = -2.04, p < 0.05$). This interaction is shown in Figure 4.3.

As expected, words were produced with more segments if the stem carried Pitch Accent ($\hat{\beta} = 0.6, t(260) = 2.25, p < 0.05$). Furthermore, words with high MI Following values contained fewer segments ($\hat{\beta} = -0.3, t(260) = -2.31, p < 0.05$). The interaction between MI Previous and Word was once more significant ($F(7, 260) = 5.47, p < 0.0001$), and again the effect was limited to *natuurlijk* ($\hat{\beta} = -2.2, t(260) = -4.87, p < 0.0001$) and *eigenlijk* ($\hat{\beta} = -1.6, t(260) = -3.01, p < 0.005$).

General Discussion

In this study, we have shown that the durations and number of realized segments of the seven most frequent words ending in the Dutch suffix *-lijk* are affected by word repetition, predictability from the previous word, and predictability from the following word. This section outlines the most important findings and discusses their implications for models of speech production. In addition, we point to directions for future research.

The role of repetition was restricted to a significant effect on the duration of the suffix. It should be noted, though, that this variable approached significance for the durations of the stem and the entire word as well (p-values of .09 and .08, respectively). Apparently, even a crude measure like number of previous mentions, which largely ignores syntactic, prosodic,

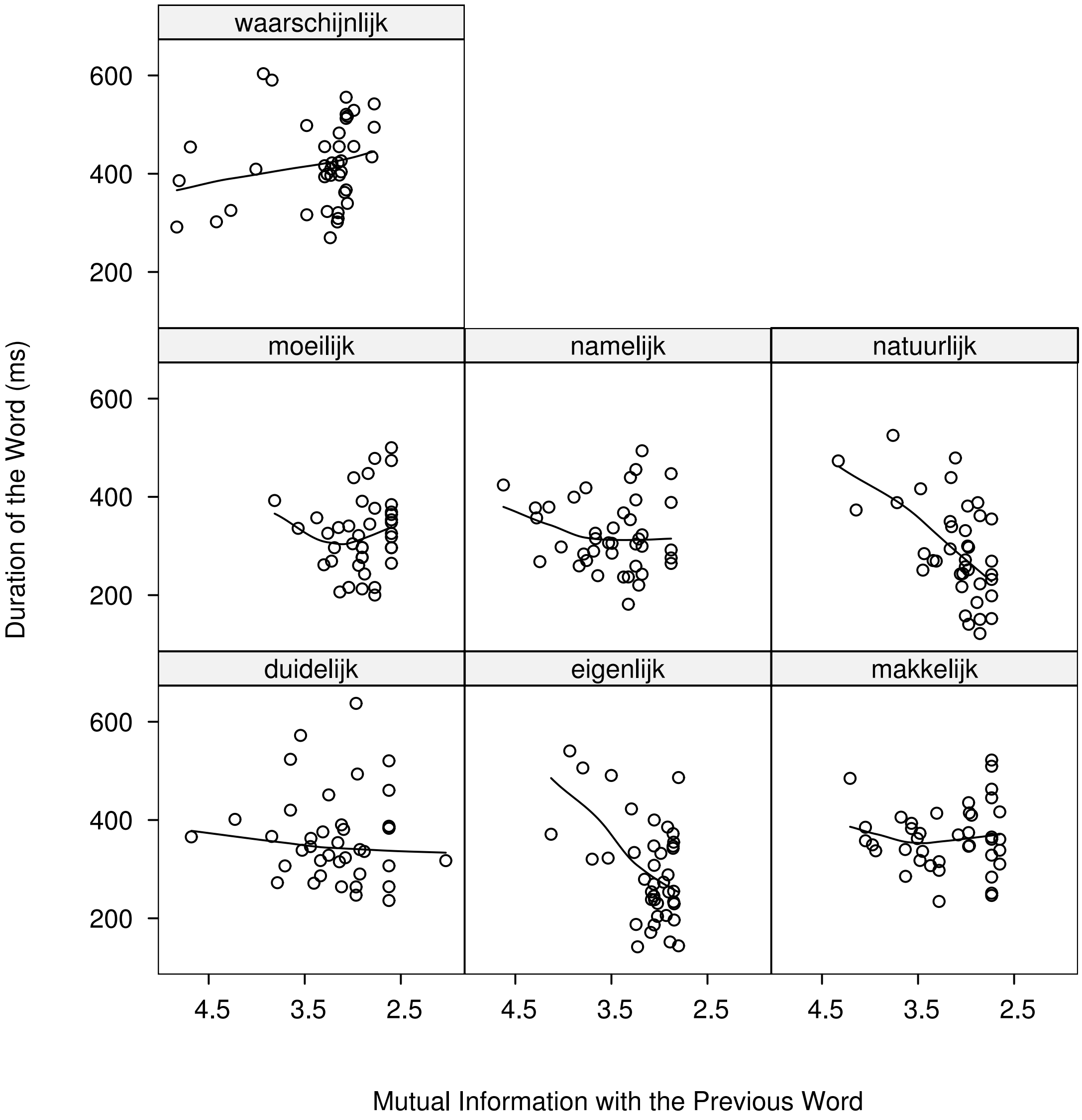


Figure 4.3: Duration of the Word plotted against Mutual Information with the Previous Word for each of the seven target words separately. The numbers on the x-axis are supposed to have minus signs, but these disappeared due to a printing problem. As can be seen from the descending lines, a higher Mutual Information led to shorter realizations for *natuurlijk*, *eigenlijk*, and *namelijk*. The line for *duidelijk* also seems to fall somewhat, but this effect was not significant.

and discourse structure, successfully predicts articulatory durations in spontaneous speech (see also Gregory et al., 1999; Aylett & Turk, 2004).

Furthermore, our results indicate that there is an effect of repetition on reduction that is independent of sentence accent. This is true for several reasons. First of all, we focused on words that are unlikely to be accented, either because they are discourse markers (*eigenlijk*, *natuurlijk*, and *namelijk*), or because they seldom introduce new entities to a discourse (*duidelijk*, *waarschijnlijk*, *moeilijk*, and *makkelijk*). More importantly, we found that even if these words are accented, they still show an effect of the number of previous mentions on the duration of the suffix.

At first glance, these results may appear contrary to the conclusions of Hawkins and Warren (1994) mentioned earlier in this chapter. However, some reservations are in place here. First of all, Hawkins and Warren measured intelligibility, while we were concerned with durations and the number of realized segments. Moreover, the word type used in the current study also differed from the one used by Hawkins and Warren. It may be possible that pure repetition effects are restricted to adverbs and adjectives, while nouns or verbs, which were the focus of Hawkins and Warren's attention, show no such effects. Further research is needed here.

How can our findings be accounted for, then? A possible explanation is offered by Pickering and Garrod (2004), who propose a model of dialogue in which the semantic, syntactic, and phonetic representations of interlocutors become aligned with each other by means of a priming mechanism. As a concomitant result of this priming, the activation of a word at all representational levels increases with each occurrence of that word. This allows speakers to save articulatory effort on words that have been used repeatedly during a conversation, as listeners (whose representations for those words are equally highly activated) require less phonetic evidence to identify them correctly.

In addition to the effect of number of previous mentions, we found several effects of contextual predictability. In this respect, our study adds to the available evidence for the relationship between probability of occurrence and articulatory reduction (e.g., Gregory et al., 1999; Fosler-Lussier & Morgan, 1999; Jurafsky et al., 2001; Bush, 2001; Bell et al., 2003). This is not our only contribution, however. Because we focused on morphologically complex words, we were able to obtain information about the effects of contextual predictability on different morphological parts of our target words. More specifically, our materials allowed us to check whether there were differences between the previous and the following context with respect to the range and the strength of their effects. The picture that emerges from our results is that effects of contextual predictability operate in a way that is not all that simple and straightforward.

Consider the stem, for example. Its duration was only affected by mutual information with the previous word, and this effect was limited to just two of the seven target words: *natuurlijk* and *eigenlijk*. Similar interactions were observed for the duration of the word as a whole and the number of segments in the stem and the word. By themselves, these findings are not too difficult to explain. *Natuurlijk* and *eigenlijk* have far higher frequencies than the other words and, being discourse markers, their semantic contribution to an utterance is relatively small. This makes them highly suitable targets for reduction, especially when their contextual predictability is also high. Further support for this claim comes from Fosler-Lussier and Morgan's (1999) study, in which effects of predictability were also limited to high-frequency words.

For mutual information with the *following* word, the picture was somewhat more complicated. We again found a significant interaction with word: Mutual information with the following word only affected the duration of the suffix in the discourse markers *eigenlijk* and *namelijk*. However, the effects observed for this variable on the number of segments in the stem and the word were main effects, unmediated by the characteristics of the particular target word. Furthermore, unlike effects of previous context, effects of following context operated on both the stem and the suffix. What do these observations tell us about the cognitive processes underlying predictability effects?

First of all, our results cannot be accounted for by simply postulating ready-made motor programs spanning two or more words (e.g., Bybee, 2001; Bush, 2001). Although the 'chunking' of frequently occurring word combinations into multi-word units is cognitively very plausible, such an account fails to explain why effects of previous context were always limited to high-frequency words, while effects of following context affected the stems of all words. Additional evidence against the chunking hypothesis was provided by Gahl and Garnsey (2004), who found correlations between the probability of occurrence of a certain syntactic structure and the durations of words within that structure, regardless of the particular words used. Since it is very unlikely that all different word combinations used in their study were stored as units in the speaker's lexicon, there must be some other explanation for their (and our) findings.

One possibility is that articulation proceeds on a unit-by-unit basis, allowing articulatory effort to be adjusted for each unit on the basis of the informational redundancy of the unit itself (e.g., stem vs. suffix), the word it belongs to (predictable vs. unpredictable), and the syntactic structure it is part of (probable vs. improbable continuations). In fact, most theories of speech production assume that there is a single basic unit of articulation. There has been some debate, however, about which unit is most appropriate for this role.

Given our results, words can be excluded as possible units, since stems and suffixes differed in their sensitivity to different measures. Morphemes, however, do not appear too suitable either, as some of our effects operated across morpheme boundaries while others did not. The syllable, which has been proposed by many researchers (e.g., Levelt & Wheeldon, 1994; Cholin, 2004), faces a similar problem: Some effects were limited to specific syllables, while others affected two or more 'units' at the same time. Furthermore, Chapter 2 has shown that the individual segments in a syllable are all subject to their own specific forces, further challenging the assumed unitary status of the syllable. Segments, on the other hand, have the disadvantage that their corresponding speech gestures often overlap considerably in time. These considerations suggest that the main problem may not lie in our inability to identify the basic unit of articulation, but rather in the assumption that there is one such unit.

As an alternative, we propose that articulatory planning is continuous and not unit-based. To ensure a relatively constant information density, articulatory effort is adjusted throughout the production of the utterance. Parts of the speech stream that carry little information are realized with less articulatory effort than more informative parts. Informativeness is determined on the basis of different dimensions simultaneously: The frequency of the word, the predictability from neighbouring words, the number of times the word has been mentioned, the probability of the syntactic structure it occurs in, and so on. Sometimes these dimensions of informational redundancy interact, while in other cases they exert their influence separately and additively. More research is needed to examine the circumstances and ways in which the different informational measures can interact.

Our results also suggest that there is an asymmetry between predictability effects that arise from planning processes prior to the uttering of a word and predictability effects linked to the preparation of the following context. If speakers are planning the articulation of a word (let us call it ‘target’) that is both highly predictable from the previous context and semantically rather meaningless, they may choose to pronounce it in a highly reduced way. In the meantime, however, the words following the target also need to be planned. During this planning, both the words preceding the target and the target are taken into account, and it is not inconceivable that the target, by virtue of being involved in this subsequent planning, is again subject to articulatory reduction (if the mutual information between the target and the following context is high). These two temporally cascaded planning processes may lead to different degrees of reduction, with the more robust reduction apparently coming from the articulatory planning process in which the target itself is also involved. This “involvement-in-planning” account could explain the differences we observed between effects of previous and following context, although at present it is of course highly tentative and in need of further investigation.

Apart from the points already raised in this discussion, we feel a number of issues need to be addressed in future research. The first issue is the relationship between the activation level of a word and its acoustic realization. There are several indications that an increase in activation leads to acoustic reduction, but little is known about the exact details of this relationship. The second point concerns the balance between speaker-internal and listener-motivated processes in explaining reductions. It is very possible that some reductions are mainly due to cognitive processes on behalf of the speaker, while others occur partly because the speaker actively takes the listener’s knowledge and needs into account. We are convinced that by tackling these issues, speech researchers can finally come to understand the roles of speaker, listener, and context in explaining the enormous phonetic variation inherent in conversational speech.

Effects of contextual predictability in perception

CHAPTER 5

This chapter has been submitted as Mark Pluymaekers, Mirjam Ernestus, and R. Harald Baayen: Recognizing reduced word forms: The role of following context.

Abstract

The previous chapter showed that predictability given the following word is a robust predictor of reduction. In this chapter, we investigated whether this pattern is mirrorred in speech perception. More specifically, we tested the hypothesis that reduced words are recognized faster the more predictable they are given the following word. To this end, subjects were presented with four-word utterances in which the third word was reduced and in which the predictability of the third word given the fourth word was varied. Subjects pressed a button as soon as they were confident about the identity of the third word. Analysis of the response latencies demonstrates that reduced words were recognized faster the more predictable they were given the following word. This shows that listeners, when confronted with reduced word forms, are sensitive to the predictability of a word given its following lexical context.

Introduction

Speech is characterized by immense pronunciation variation. Consequently, models of speech perception should be able to accommodate connected speech processes such as assimilation and reduction, which lead to acoustic realizations that can be quite different from the canonical form. As it turns out, most current models of spoken-word recognition (e.g., McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, & Cutler, 2000) have difficulties dealing with missing or mismatching information in the input. A partial solution to this problem is provided by Scharenborg, Norris, Ten Bosch, and McQueen (2005), who introduce a dynamic programming technique that is more tolerant to mismatches between the acoustic input and canonical lexical presentations. However, the recognition results obtained with this technique are still far from perfect.

Since previous studies have shown that the likelihood of a word being reduced is codetermined by its linguistic context (e.g., Lieberman, 1963; Hunnicutt, 1985), models of speech perception could well benefit from taking contextual information into account. The current study aims to further our knowledge on contextual effects in speech perception by investigating the role of following lexical context in the recognition of reduced word forms.

The recognition of reduced word forms

Even though connected speech processes are ubiquitous in everyday speech, their perceptual consequences have not been investigated in much detail. Furthermore, most studies on the subject have been concerned with relatively small deviations from the standard. The process that has received most attention in recent years is assimilation. In assimilation, a phonological feature (e.g., place of articulation) spreads from one phoneme to a neighboring phoneme, often resulting in a change of phonemic identity (e.g., Lahiri & Marslen-Wilson, 1991; Jongenburger & Van Heuven, 1993). Assimilation can be either progressive or regressive, depending on the position of the assimilated phone relative to the phone it inherits its new feature from. An example of regressive place assimilation in English, borrowed from Gow (2002), is the pronunciation of the phrase *right berries* as [rɑipbɛriz]. Here, the /t/ at the end of *right* inherits the bilabial place of articulation of the following segment, /b/. This example illustrates why assimilation could be problematic for speech recognition: Listeners may interpret the assimilated form as a realization of the equally plausible phrase *ripe berries*.

How do listeners resolve such potential ambiguities? In the literature two different — though not necessarily mutually exclusive — approaches to the problem can be found. The first approach stresses the importance of subtle acoustic cues in the signal, which help listeners recover the original features of the assimilated phone (e.g., Manuel, 1995; Gow, 2002; 2003). The second approach focuses on the viability of assimilation. The pronunciation of /t/ as

[p] is viable in *right berries*, but not in *right ferries*, where /t/ cannot inherit a bilabial place of articulation from the following segment. Assimilated forms have been found to activate the right lexical entries in the mental lexicon only if the assimilation is viable given the phonological context (e.g., Gaskell & Marslen-Wilson, 1996; 1998; Coenen, Zwitserlood, & Bölte, 2001). This can be explained by postulating a so-called phonological inference process, during which phones are evaluated against their phonological context. If listeners hear a [p] immediately before a [b], they consider the possibility that this [p] is underlyingly a /t/. Mitterer and Blomert (2003) showed that such a process is likely to be rooted in early, automatic perceptual processes. Furthermore, it appears to operate without reference to language-specific assimilation rules: Dutch listeners compensate for Hungarian liquid assimilation, even though this type of assimilation does not occur in Dutch (Mitterer, Csépe, & Blomert, 2003).

Taken together, these findings strongly suggest that any ambiguities arising from assimilation are resolved at a pre-lexical level, that is, before contact is made with the lexicon. Recently, Mitterer and Ernestus (2006) investigated whether the same holds if listeners are confronted with a different connected speech process: /t/-lenition in Dutch. Their results suggest that apart from phonological context, the lexical status of the stimulus also plays a role: Listeners are more likely to infer the presence of a /t/ if this enables them to interpret the stimulus as an existing word. Mitterer and Ernestus conclude, therefore, that lexical knowledge plays a greater role in compensating for /t/-lenition than in compensating for assimilation (see also Janse, Nooteboom, & Quené, in press). A possible explanation for this asymmetry is that in /t/-lenition, fewer acoustic cues for the canonical form remain present in the signal (Mitterer & Ernestus, 2006, p. 97).

Of course, /t/-lenition is not by any means the most extreme case of reduction that listeners encounter. In a corpus study on conversational American English, Johnson (2004) found that 25% of the words in the corpus had one or more segments deleted. Deletion of complete syllables occurred in 6% of the words. Similar observations were made by Shockey (2003) for British and American English, Kohler (2000) for German, and Ernestus (2000) for Dutch. For example, Ernestus (2000) observed that the Dutch word *natuurlijk* (/nɑtyrlək/ 'of course') can be reduced to [tyk], and that *eigenlijk* (/ɛixələk/ 'actually') is often pronounced as [ɛik]. In both cases, only the first part of the stressed syllable and the coda of the final syllable are preserved. It is possible that the signal contains subtle acoustic traces of (some of) the deleted segments as well, but it seems unlikely that listeners can restore each of these segments on the basis of the signal alone. Therefore, one would expect lexical context to play an important role in processing extreme reductions such as these.

This intuition is supported by the few studies that have addressed the recognition of extremely reduced words. Ernestus, Baayen, and Schreuder (2002) presented subjects with speech samples from the corpus of conversational speech described in Ernestus (2000).

Subjects heard 54 words, differing in degree of reduction, in one of three conditions: Embedded in a full utterance (Full Context), together with adjacent vowels and intervening consonants (Limited Context), or in isolation. The subjects' task was to provide an accurate orthographic transcription of the speech sample. The results showed that forms with low reduction were almost always recognized correctly, independent of the available context. For words with medium reduction, the Limited Context was usually sufficient. Highly reduced words, however, required more contextual information. In Full Context, these words were almost always recognized correctly, but in Limited Context, the percentage of correctly recognized words dropped to 70%. In isolation, the success rate decreased even further, to no more than 50%. This led the authors to conclude that highly reduced forms are probably not represented by separate, equally accessible entries in the mental lexicon, since this would fail to explain why they are so poorly recognized in isolation.

A slightly different approach was taken by Kemps, Ernestus, Schreuder, and Baayen (2004), who focused on the process by which listeners recover the phonemic information that has been lost as a result of acoustic reduction. It has long been known that listeners are able to perceptually restore speech sounds that have been replaced or masked by extraneous noise (Warren, 1970; Samuel, 1981), and that this process can be influenced by lexical knowledge (Samuel, 1996). In Kemps et al.'s study, subjects were presented with realizations of the Dutch suffix *-lijk* from a corpus of conversational speech, in which the initial consonant /l/ of the suffix was either realized or deleted. Similar to Ernestus et al.'s (2002) study, the stimuli were presented in different contexts. In the Minimal Context condition, only the suffix was presented, whereas in the Full Context condition, the suffix and its carrier word were part of a complete, natural utterance. Subjects had to perform a phoneme monitoring task with /l/ as the target phoneme. In the Minimal Context condition, subjects generally only reported an /l/ if an /l/ was really present in the signal. In the Full Context condition, however, many subjects heard an /l/ even if the presented realization of the suffix did not contain that sound. In these cases, subjects restored the missing phoneme using linguistic information from the utterance. This suggests that context allows listeners to link highly reduced forms to canonical representations in the mental lexicon, and that these latter representations induce restoration processes.

The findings by Ernestus et al. (2002) and Kemps et al. (2004) raise a number of questions. The first question is whether the reported results might have been affected by the type of materials used. Both studies used speech samples from a corpus of spontaneous speech, produced by several different speakers. This may have confused the subjects, who had to adjust to the idiosyncracies of a new speaker for almost every new stimulus. A second and far more interesting question concerns the respective roles of preceding and following context. Neither Ernestus et al. (2002) nor Kemps et al. (2004) investigated the relevance of preceding versus following context. They presented reduced words either in very limited or no context, or in a complete utterance in which the position of the reduced target was not controlled.

Consequently, their experiments provide no information about which specific elements in the context enabled the recognition of the reduced words. In the present study, we aim to gain more insight into this issue by focusing on the role of following lexical context.

The role of following context in speech perception

It has long been known that acoustic information occurring after a particular speech event can alter the conscious perception of that event. For example, Repp, Liberman, Eccardt, and Pesetsky (1978) found that when a /t/ is perceived between the words *gray* and *ship*, whether this segment is perceptually grouped with *gray* or *ship* depends on the duration of the frication noise at the beginning of *ship*. Listeners perceive *great ship* if the frication noise is long, and *gray chip* if the frication noise is short. This shows that later-occurring information can influence perceptual grouping decisions concerning earlier speech sounds.

Following context can also affect the perception of phonemic identity. Mann and Repp (1980) showed that the perceptual boundary between the phonemes /s/ and /ʃ/ shifts as a function of the following vowel: If the following vowel is /u/, listeners are more inclined to perceive /s/ than if the following vowel is /a/. That a similar influence can be exerted by non-speech sounds was demonstrated by Wade and Holt (2005), who found that the frequency of a pure tone following a syllable-onset codetermined whether listeners perceived the onset as /d/ or /g/.

In the above-mentioned studies, the information affecting the perception of earlier speech events was acoustic in nature. However, effects of following context can also be semantically driven. Warren and Sherman (1974) found that noise inserted before the phrase *eel is on the...* was perceived as /p/ if the final word was *orange*, /w/ if the final word was *axle*, /h/ if he final word was *shoe*, and /m/ if the final word was *table*. This also shows that the conscious identification of an ambiguous linguistic unit can be delayed for a relatively long time, until all the information necessary for its disambiguation is available.

Several studies have demonstrated that it is not uncommon for words to be ambiguous until after their acoustic offsets. Using a gating paradigm, Grosjean (1985) found that subjects were sometimes not completely sure about the identity of a target word until no less than three following words had been presented. Bard, Shillcock, and Altmann (1988) argued that Grosjean's results might have overestimated the role of following context, as all target words in his experiment were preceded by the highly uninformative phrase "I saw the". In order to make the test utterances more representative of normal language use, Bard et al. took their stimuli from a corpus of conversational English. Even though the preceding context in these stimuli was much more informative, Bard et al. still observed effects of following context on recognition. No less than 21% of the words in the experiment were not recognized until one or more following words had been presented. Words that were particularly prone to late recognition were function words, words consisting of a small number of phonemes, and words at the beginning of utterances.

The role of following context in recognizing reduced word forms

In the previous section, we saw how following context affects the perception of phonemes and unreduced words. Now, we turn our attention to the role it may play in recognizing reduced words. Previous studies have shown that listeners confronted with assimilation or segment reduction are extremely sensitive to the patterns observed in production (e.g., Coenen, Zwitserlood, & Bölte, 2000; Sumner & Samuel, 2005; Mitterer & Ernestus, 2006). Since it has long been known that words are more likely to be reduced if they are predictable given their linguistic context (Lieberman, 1963; Hunnicutt, 1985), we expect listeners to be sensitive to contextual predictability when processing reduced word forms.

Recently, several studies have shown that words are more reduced the more predictable they are given the directly following word (e.g., Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Fosler-Lussier & Morgan, 1999; Jurafsky, Bell, Gregory, & Raymond, 2001; Bush, 2001; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Keune, Ernestus, Van Hout, & Baayen, 2005; Chapter 4 of this thesis). To estimate this predictability, these studies used measures such as *conditional probability* and *mutual information*, which are computed using frequency counts from large speech corpora. For example, the conditional probability of occurrence of word X given that the following word is Y can be computed as follows (X stands for word X, Y for word Y, and XY for the combination of those two words):

$$P(X|Y) = \frac{Frequency(XY)}{Frequency(Y)}$$

The main aim of the present study is to investigate whether listeners are also sensitive to this conditional probability measure. More specifically, we examine whether listeners recognize reduced words faster the more predictable they are given the word that follows them. In focusing on the role of conditional probability, the current study is the first to investigate the perceptual consequences of a non-categorical, probabilistic variable that has repeatedly been shown to predict reduction in production.

To illustrate our hypothesis, let us consider the phrase *United States*. In this phrase, the conditional probability of the first word given the second word is relatively high: Upon hearing the word *states*, most speakers of English know that the preceding word is quite likely to be *united*. We predict, therefore, that a reduced realization of the word *united* is recognized faster if it is followed by the word *states* than if it is followed by words like *sailors* or *soldiers*, from which it cannot be predicted so easily. This hypothesis is investigated in two experiments.

Experiment 1

Method

Participants

30 subjects from the subject pool of the Max Planck Institute for Psycholinguistics participated in the experiment. All were native speakers of Dutch with no reported speech or hearing disabilities. There were 14 male and 16 female participants. All subjects were paid for participation.

Materials

The experiment was centered around 48 Dutch word pairs (henceforth referred to as bigrams) that were unanimously labeled by the authors as fixed expressions. Half of the bigrams started with a morphologically complex word, while the other half started with a morphologically simplex word. A distinction was made between complex and simplex targets because segments are more often deleted in complex than in simplex words.

Most of the 48 bigrams were idiomatic, in the sense that their meanings could not be derived from the meanings of the individual words (Swinney & Cutler, 1979). The comprehension of idioms has been studied in considerable detail (e.g., Gibbs, 1986; Cacciari & Tabossi, 1988; Colombo, 1993; Titone & Connine, 1994), with most of the research focused on the question which interpretation, the literal or the idiomatic one, is available to the listener first. Although the answer to this question is not directly relevant for our purposes, one of the concepts discussed in these studies proved to be helpful for selecting suitable stimuli. This was the concept of *key*, introduced by Cacciari and Tabossi (1988). The key of an idiom is that word in the expression that allows listeners to recognize it as idiomatic. For example, the key of the English idiom *kick the bucket* is *bucket*, as listeners will not realize that the utterance they are hearing has an idiomatic meaning until they have heard the word *bucket*. To ensure that the first word in a bigram was predictable given the second word but not vice versa, we selected expressions in which the second word was the key. Not coincidentally, this was also the word receiving accent.

For each fixed (F-) bigram, a non-fixed (NF-) counterpart was created. The first word of the NF-bigram was identical to that of the F-bigram, while the second words of the F- and NF-bigrams were matched on initial phoneme, number of syllables and stress position. The NF-bigrams were semantically plausible, but did not form fixed expressions. Examples of F- and NF-bigrams are shown in Table 5.1.

Table 5.1: Examples of fixed (F) and non-fixed (NF) bigrams, including literal English translations.

	Fixed	Translation	Non-Fixed	Translation
Complex	verdacht persoon	‘suspicious person’	verdacht papier	‘suspicious paper’
	bepaalde manier	‘particular way’	bepaalde matras	‘particular matrass’
	vermoorde onschuld	‘murdered innocence’	vermoorde oppas	‘murdered nanny’
	beloofde land	‘promised land’	beloofde licht	‘promised light’
Simplex	droog brood	‘dry bread’	droog bed	‘dry bed’
	blauwe maandag	‘blue monday’	blauwe mantel	‘blue cape’
	oude liefde	‘old love’	oude liedjes	‘old songs’
	witte vlag	‘white flag’	witte vlam	‘white flame’

Pretest

To determine the conditional probability of the first word given the second word for each of our bigrams, we carried out a pretest. 66 subjects, none of whom took part in the main experiments, participated in a rating study conducted via the web page of the Max Planck Institute for Psycholinguistics. Each participant was presented with 48 bigrams, half of which were fixed expressions. Subjects were asked to make three estimates on a 7-point scale: The frequency of the first word, the frequency of the second word, and, crucially, the probability of the first word given the second word. This last estimate will henceforth be referred to as the conditional probability (CP) of the bigram.

The average CP ratings obtained in the pretest ranged from 1.27 to 6.78. To test whether CP ratings differed as a function of fixedness (F vs. NF) or morphological type (complex vs. simplex), we conducted a two-way ANOVA. As expected, F-bigrams generally received higher CP ratings than NF-bigrams (F: 4.31; NF: 2.44; $F(1,92) = 124.85, p < 0.0001$). There was no main effect of morphological type on CP ratings ($p = 0.60$), and no interaction between morphological type and fixedness ($p = 0.59$).

Stimuli

The stimuli were recorded in a sound-proof booth by a female speaker from the West of the Netherlands. The recordings were made on a DAT tape and digitized at a sample frequency of 16 kHz. Two words were added to the beginning of all bigrams to make the stimuli more utterance-like. The added words were identical for the F- and NF-member of a bigram pair. For example, the bigram pair *verdacht persoon* - *verdacht papier* (see Table 5.1) became the stimulus pair *met een verdacht persoon* - *met een verdacht papier*.

Several measures were taken to ensure a degree of reduction that was representative for casual speech. First of all, the utterances were visually presented to the speaker at a very fast presentation rate, forcing her to articulate fast. Furthermore, given that repeated pronunciation is known to lead to more reduction (see Chapter 4), each utterance was recorded seven times.

The stimuli in the current experiment were completely natural, that is, they were presented to the subjects exactly how they had been produced by our speaker. As a consequence, F- and NF-stimuli differed not only in the identity of the fourth word, but also in the acoustic realization of the first three words. However, we did ensure that the degree of segmental and durational reduction in the third word — which was the target for recognition — was matched in the two stimuli. This is illustrated in Figure 5.1, which shows the acoustic realizations of the word *verkeerde* in the F-stimulus *met het verkeerde been* ‘with the wrong leg’ (top) and the NF-stimulus *met het verkeerde beeld* ‘with the wrong image’ (bottom). These two realizations are not only identical in duration, but also spectrally very similar.

To gain insight into the degree of reduction in our target words, several tests were

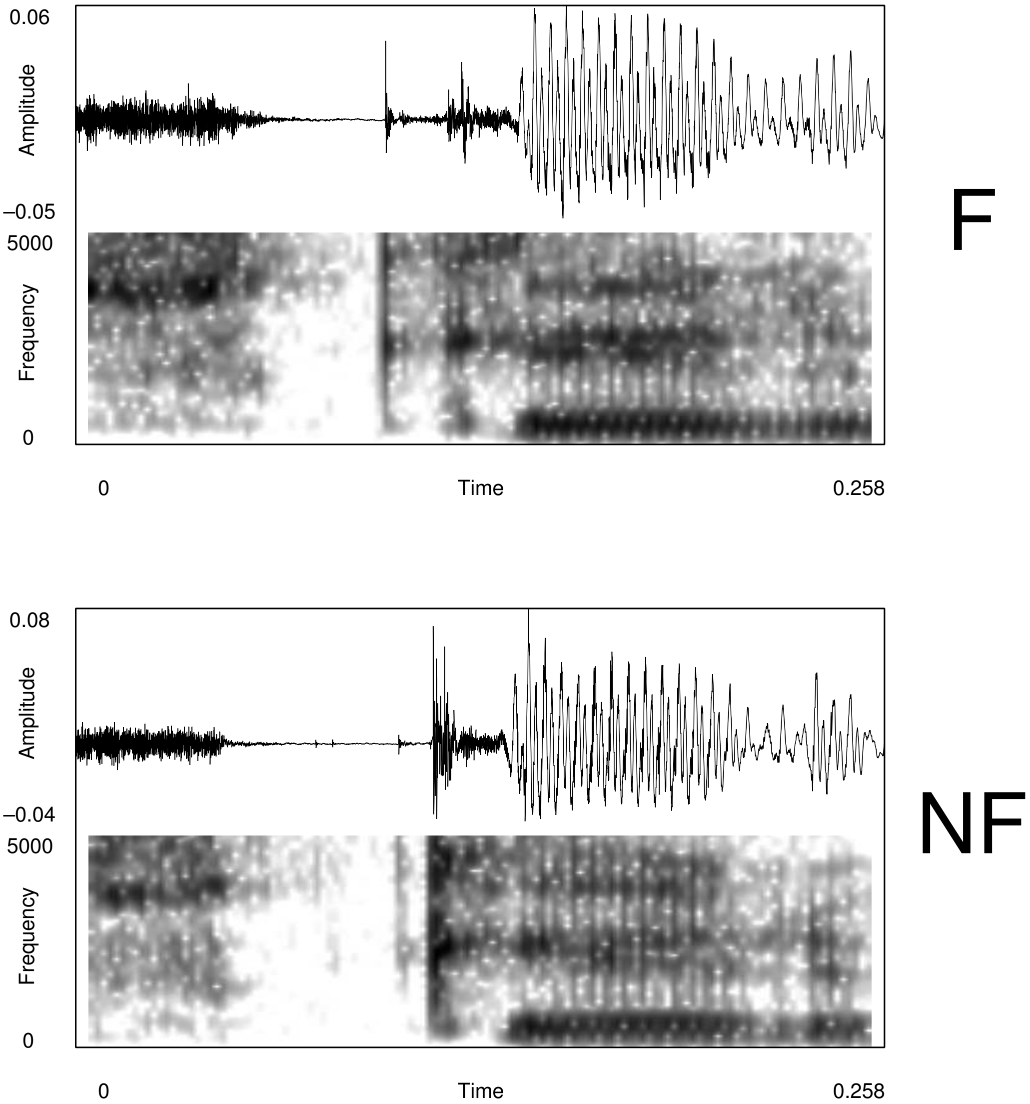


Figure 5.1: Waveforms and spectrograms of the word *verkeerde* in the F-stimulus *met het verkeerde been* (top) and the NF-stimulus *met het verkeerde beeld* (bottom).

performed. First of all, the amount of segmental reduction was assessed using phonemic transcriptions. Two phoneticians independently transcribed each target word, after which the transcription showing the least segmental reduction was chosen as the reference. The difference between the number of segments in this reference and the number of segments in the canonical form of the word was then divided by the number of segments in the canonical form, yielding a measure of segmental reduction that is independent of word length. A t-test revealed that the average amount of segmental reduction was significantly different from zero (mean: 0.20; $t(95) = 11.70, p < 0.0001$), which shows that our target words were indeed segmentally reduced. Furthermore, complex targets were found to be more reduced than simplex targets (Complex: 0.25; Simplex: 0.13; $t(94) = 3.91, p < 0.0005$). Importantly, whether a target word was part of an F- or NF-stimulus did not correlate with the degree of segmental reduction ($p = 0.60$).

The durational reduction in our target words was estimated using canonical realizations of the stimuli. These realizations were recorded in a different recording session by the same speaker. This time, the stimuli were produced only once, with sufficient time available for careful articulation. For each target word, the difference between the duration of its canonical realization and the duration of its realization in the test stimulus was divided by the duration of its canonical realization. This measure of durational reduction turned out to be significantly different from zero, as indicated by a t-test (mean: 0.38; $t(95) = 42.51, p < 0.0001$). Hence, we conclude that our target words were also durationally reduced. Complex and simplex targets did not differ in their degree of durational reduction ($p = 0.83$), and no difference was observed between targets occurring in F-stimuli and targets occurring in NF-stimuli ($p = 0.91$).

In addition to the 96 test stimuli, ten practice stimuli were recorded. These were also four-word utterances, comparable in structure to the test stimuli used in the experiment.

Design

The 96 test stimuli were divided into two lists, both of which contained 24 F-stimuli (12 complex and 12 simplex) and 24 NF-stimuli (12 complex and 12 simplex). If a particular F-stimulus was assigned to one list, the corresponding NF-stimulus was automatically added to the other. Subjects were randomly assigned to a list. In total, they heard the list three times, each time with a different random stimulus order. This repetition allowed us to investigate whether the effect of CP on recognition time changed with repeated exposure.

Procedure

Experiments were run on a standard PC running the NESU package. Subjects were tested individually in a sound-attenuated room. The stimuli were presented over Sennheiser headphones and preceded by a 3.5 kHz tone.

In each trial, subjects heard a four-word stimulus and were required to press a button as soon as they were confident about the identity of the third word (i.e., the target word). They were given 3779 milliseconds (average stimulus duration + 3000 ms) to accomplish this task, starting from the onset of the stimulus.

To check whether participants had recognized the target word correctly, we asked them to name the recognized word into a microphone. Starting from the button press, they had 1500 milliseconds to initiate articulation. Their responses were recorded on DAT tapes and checked for correctness. Asking subjects to give a verbal response had the additional advantage that it provided us with an extra source of data, namely the voice key-registered interval between the button press and the naming of the recognized word.

Before the start of the experiment, subjects performed ten practice trials. Subsequently, the stimulus list was presented for the first time, preceded by three practice trials that had also been part of the practice block. After the first presentation of the stimulus list, subjects were presented with a block of isolated words that was part of a different experiment. The third and the fourth block, which were again part of the current experiment, were identical to the first block, except that a different stimulus order and different practice stimuli from the practice block were used.

Analysis

The data were analyzed using multi-level regression models with subject and item as crossed random effects (Bates & Sarkar, 2005). Two models were fitted: One to predict the logarithmically transformed response latencies for the button press, measured from the onset of the target word, and one to predict the logarithmically transformed naming latencies, measured from the button press. When fitting the models, we followed a two-step procedure. First, the predictivity of a number of control variables was tested. These variables included the sex of the subject (Sex), the duration of the target word (Target Duration), the duration of the utterance (Stimulus Duration), the morphological type of the target word (Type), the number of exposures to the stimulus (Exposures), and the position of the trial in the block (Position).

For the model predicting the naming latency, two additional control variables were used: The place of articulation of the first segment of the target (POA), and the residual values of the model fitted for the button press latency (Button Residuals). The first variable was added to control for voice key artefacts, while the second variable provides a measure of the speed with which the subject is processing that particular stimulus. Only those control variables that showed a significant effect were retained.

In the second step, the predictor variables of interest were added to the model. Three of these variables were computed from the results of the rating study discussed above: The average frequency estimate of the target word (Target Frequency), the average frequency estimate of the following word (Following Frequency), and, above all, the average estimate of the conditional probability of the target word given the following word (CP Rating). In addition, a binary variable was entered which indicated whether the authors considered a particular bigram as fixed (Fixedness). Note that this variable is correlated with CP Rating, in that F-bigrams generally received higher CP ratings than NF-bigrams. Nevertheless, both variables were included in order to determine which of the two was the better predictor. Only the predictor variables that showed an effect over and above the significant control variables were retained.

Finally, interaction terms between predictor and control variables were added to see whether their inclusion improved the predictive power of the model. If this was the case, the particular interaction effect was retained. To eliminate overly influential outliers, data points for which the residual value was more than 2.5 times the standard deviation of the residuals were removed from the data set, and the model was refitted to the remaining data. The resulting model is the one reported below.

Results

The model for the button press RTs was initially fitted to the 3892 data points for which a correct response had been given and for which the button press did not precede the onset of the target word (92% of all trials). 103 data points (2.6% of the correct trials) had residuals larger than 2.5 times the standard deviation and were removed as outliers. The model that was refitted to the remaining data showed that a higher CP Rating correlated with shorter response latencies to the target word ($\hat{\beta} = -0.013, t(3782) = -3.28, p < 0.005$). There was also a significant interaction between CP Rating and Type ($F(1, 3782) = 20.9, p < 0.0001$), suggesting that the effect of CP Rating was larger for simplex than for complex targets ($\hat{\beta} = -0.019, t(3782) = -4.58, p < 0.0001$). A higher Following Frequency was found to correlate with shorter RTs ($\hat{\beta} = -0.042, t(3782) = -3.60, p < 0.0005$), although this effect was modulated by an interaction with Exposures ($F(1, 3782) = 7.03, p < 0.01$): With each repeated exposure, the effect of Following Frequency became smaller ($\hat{\beta} = 0.012, t(3782) = 2.63, p < 0.01$). Furthermore, there were significant main effects for the control variables Exposures ($\hat{\beta} = -0.170, t(3782) = -8.02, p < 0.0001$) and Stimulus Duration ($\hat{\beta} = 0.0004, t(3872) = 6.78, p < 0.0001$). The more often subjects had heard a stimulus, the faster they responded, and the longer the stimulus, the longer the response latency.

The model for the naming latencies was fitted to the 3450 data points for which a correct response had been given and for which a latency had been registered by the voice key (80% of all trials). 78 data points (2.2% of the correct trials) had residuals larger than 2.5 times the standard deviation and were removed from the data set. The model that was refitted to the remaining data showed a significant main effect of Fixedness on naming latencies: If the target and the following word formed a fixed expression, the naming latency was shorter ($\hat{\beta} = -0.087, t(3365) = -3.44, p < 0.001$). There was also a significant interaction between Fixedness and Exposures ($F(2, 3365) = 3.27, p < 0.05$), suggesting that the effect of Fixedness became smaller as the number of exposures increased ($\hat{\beta} = 0.019, t(3365) = 2.36, p < 0.05$). Again, Following Frequency showed a significant main effect, albeit in a different direction than in the model for the button press RTs. The higher the frequency of the following word, the longer it took subjects to name the target word ($\hat{\beta} = 0.018, t(3365) = 3.21, p < 0.005$). This time, there was no interaction with Exposures. Control variables that showed significant effects were POA ($F(4, 3365) = 8.73, p < 0.0001$) and Button Residuals ($\hat{\beta} = -0.254, t(3665) = -12.53, p < 0.0001$).

Discussion

The models for the button press RTs and naming latencies are summarized in Table 5.2. Most importantly, CP Rating had a negative effect on button press RTs. In other words, the higher the conditional probability of the target word given the following word, the faster it was recognized by our subjects. This finding is in accordance with our main hypothesis, which predicts that reduced words are recognized faster the more predictable they are given the following context. It should be noted that in the model for the button press RTs, the continuous CP variable outperformed the binary variable Fixedness as a predictor. This was not the case in the model for the naming latencies, where Fixedness proved to be the better predictor. However, the direction of the effect was basically the same: If the reduced target word was part of a fixed expression, it was named at a shorter latency. All this suggests that reduced words are recognized faster if they are predictable given the following word.

The prominent role of the following word also manifests itself in the effects that were observed for Following Frequency. Target words were recognized faster the higher the frequency of the word that followed them. This can be explained by assuming that the recognition of the target word is facilitated by the processing of the following word, regardless of whether the combination of the two words forms a fixed expression. Since frequent words are processed faster than infrequent words, a higher frequency of the following word will indirectly lead to faster recognition of the preceding target. As soon as the preceding target needs to be produced, however, the frequency of the following word appears to be inhibitory rather than facilitatory. At this point, the lexical entry of the following word is probably activated to such an extent that it hinders the production of other words.

In addition to these main effects, a number of significant interactions were observed. Two of the interactions involved the number of exposures to a particular stimulus. In the current experiment, subjects were exposed to each stimulus three times. It was found that the effect of Following Frequency on button press RTs became smaller the more often a subject had

Table 5.2: Beta coefficients and significance values of the effects observed in Experiment 1. The horizontal line in the middle separates the predictor variables of interest from the control variables. The beta coefficients indicate the magnitude and the direction of the effect. ”–” means there was no significant effect.

	Button press RT	Naming latency
CP Rating	-0.013**	–
CP Rating * Type	-0.019****	–
Fixedness	–	-0.087***
Fixedness * Exposures	–	0.019*
Following Frequency	-0.036***	0.018**
Following Freq. * Exposures	0.012**	–
Button Residuals	–	-0.254****
Exposures	-0.170****	–
Place Of Articulation	–	Several contrasts
Stimulus Duration	0.0004****	–

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

heard a stimulus. The same was true for the effect of Fixedness on naming latencies. These two findings have important methodological implications, since they suggest that the repetition of a stimulus may conceal effects that are present if the stimulus is presented for the first time. The third interaction effect, between CP Rating and Type, showed that the effect of CP Rating was larger for morphologically simplex than for morphologically complex target words. This is unexpected, as the complex words in our experiment showed more segmental reduction than the simplex words. We will return to this issue below.

A possible concern with respect to the results of Experiment 1 is that they might have been caused by subtle acoustic differences between the target words in the F- and NF-stimuli. It is true that the target words were matched on segmental and durational reduction, but this may not have been enough to completely neutralize the impact of stimulus properties on response latencies. Therefore, a second experiment was conducted in which the acoustic realizations of the target and the preceding two words were identical in the two conditions.

Experiment 2

Method

Participants

30 subjects from the subject pool of the Max Planck Institute for Psycholinguistics participated in the experiment. All were native speakers of Dutch with no reported speech or hearing

disabilities. None of them had participated in Experiment 1, or in the pre-test of the materials. There were 8 male and 22 female participants. All subjects were paid for participation.

Materials

The materials were identical to those used in Experiment 1.

Stimuli

New stimuli were created using recordings from the recording session described in the method section of Experiment 1. The realizations of the fourth words were identical to those used in Experiment 1, but the realizations of the preceding three words (including the target) were different. Two splicing conditions were distinguished. In one condition, the following words from the F- and NF-bigrams (e.g., *persoon* and *papier*) were spliced onto the first three words of a realization of an F-utterance that was not used in Experiment 1 (e.g., *met een verdacht* from a realization of *met een verdacht persoon*). In the other condition, the fourth words were spliced onto the first three words of a realization of an NF-utterance that was not used in Experiment 1 (e.g., *met een verdacht* from a realization of *met een verdacht papier*). Bigram pairs were randomly assigned to a splicing condition. The splicing manipulation was performed using the software package PRAAT (Boersma, 2001).

Again, we examined the degree of reduction in our target words, as these were different from the target words in Experiment 1. As in Experiment 1, the average amount of segmental reduction was significantly different from zero (mean: 0.21; $t(47) = 9.07, p < 0.0001$), and complex targets were more reduced than simplex targets (Complex: 0.28; Simplex: 0.15; $t(44) = 3.06, p < 0.005$). The difference between target words in F- and NF-stimuli could not be tested, as the targets in these two stimuli contained exactly the same segments. The average amount of durational reduction also differed significantly from zero (mean: 0.36; $t(95) = 36.88, p < 0.0001$). No differences were observed between complex and simplex targets ($p = 0.15$).

Design

The design was the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1.

Analysis

Statistical analysis was performed according to the same procedure as in Experiment 1. One control variable was added to the models, namely Splicing Condition. Again, this variable was only retained if it showed a significant effect.

Results

The model for the button press RTs was fitted to the 4155 data points for which a correct response had been given and for which the button press did not precede the onset of the target word (96% of all trials). 104 data points (2.5% of the correct trials) had residuals larger than 2.5 times the standard deviation and were removed from the data set. The model that was refitted to the remaining data showed that a higher CP Rating led to shorter response latencies to the target word ($\hat{\beta} = -0.022, t(4043) = -7.83, p < 0.0001$). Following Frequency also showed an effect: The higher the frequency of the following word, the shorter the response latency ($\hat{\beta} = -0.034, t(4043) = -3.54, p < 0.0005$). Again, this main effect was modulated by an interaction with Exposures ($F(1, 4043) = 3.89, p < 0.05$): With each repeated exposure, the effect of Following Frequency became smaller ($\hat{\beta} = 0.008, t(4043) = 1.97, p < 0.05$). Control variables that showed significant effects were Exposures ($\hat{\beta} = -0.157, t(4043) = -7.65, p < 0.0001$), Stimulus Duration ($\hat{\beta} = 0.0005, t(4043) = 9.33, p < 0.0001$), Sex ($\hat{\beta} = -0.258, t(4043) = -2.84, p < 0.005$), and Position ($\hat{\beta} = -0.0005, t(4043) = -2.05, p < 0.05$). Repeated exposure led to shorter reaction times, as did a shorter stimulus duration. Female subjects responded faster than male subjects, and the later a stimulus occurred in a block, the faster it was responded to.

The model for the naming latencies was fitted to the 3397 data points for which a correct response had been given and for which a latency had been registered by the voice key (79% of all trials). 73 data points (2.1% of the correct trials) had residuals larger than 2.5 times the standard deviation and were removed as outliers. The model that was refitted to the remaining data showed that Fixedness was correlated with shorter naming latencies ($\hat{\beta} = -0.062, t(3314) = -2.41, p < 0.05$). The interaction between Fixedness and Exposures just failed to reach significance, but the direction of the effect was the same as in Experiment 1 ($\hat{\beta} = 0.023, t(3314) = 1.91, p = 0.06$). Furthermore, significant main effects were observed for Exposures ($\hat{\beta} = -0.048, t(3314) = -5.73, p < 0.0001$) and Splicing Condition ($\hat{\beta} = 0.044, t(3314) = 2.43, p < 0.05$). Subjects named the target word faster with repeated exposure, as well as if the stimulus was created using the first three words of an F-utterance. As in Experiment 1, Place of Articulation ($F(4, 3314) = 6.14, p < 0.0001$) and Button Residuals ($\hat{\beta} = -0.197, t(3314) = -9.70, p < 0.0001$) were also significant.

Table 5.3: Beta coefficients and significance values of the effects observed in Experiment 2. The horizontal line in the middle separates the predictor variables of interest from the control variables. The beta coefficients indicate the magnitude and the direction of the effect. "–" means there was no significant effect.

	Button press RT	Naming latency
CP Rating	-0.022****	–
Fixedness	–	-0.062*
Following Frequency	-0.034***	–
Following Freq. * Exposures	0.008*	–
Button Residuals	–	-0.197****
Exposures	-0.157****	-0.048****
Place Of Articulation	–	Several contrasts
Position	-0.0005*	–
Sex	-0.258**	–
Splicing Condition	–	0.044*
Stimulus Duration	0.0005****	–

* = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$ **** = $p < 0.0001$

Discussion

The regression results for Experiment 2 are summarized in Table 5.3. Comparison of Tables 5.2 and 5.3 shows that both experiments yielded essentially the same results. Reduced target words were recognized faster the more predictable they were from the following word. Furthermore, naming latencies were shorter for target words belonging to a fixed expression. This replication of the results of Experiment 1 shows that the observed effect of predictability cannot be ascribed to the acoustic properties of the target words or the preceding words.

The results of the two experiments were not completely identical, however. Most importantly, the interaction between CP Rating and Type observed in Experiment 1 was not replicated in Experiment 2. The same was true for the main effect of Following Frequency on naming latencies. This suggests that these effects should be interpreted with caution.

In contrast, the interaction between Following Frequency and Exposures was replicated in Experiment 2, while the interaction between Fixedness and Exposures almost reached significance. Therefore, the methodological issues raised in the discussion of Experiment 1 remain relevant. Furthermore, Splicing Condition turned out to have an effect, a finding to which we will return in the General Discussion below.

General Discussion

This study investigated the role of following lexical context in the recognition of reduced word forms. Since the production literature shows that words are more reduced the more predictable they are given the following word, we expected reduced words to be recognized more easily as their predictability from the following word increased. In two experiments, reduced target words were presented as the third word in a four-word utterance. For each target word, two utterances were created: One in which the target and the following word formed a fixed expression, and one in which the target and the following word formed a semantically plausible combination, but not a fixed expression. In a pre-test, we established that the predictability of the target given the following word was higher in the fixed expressions than in the non-fixed expressions. In the main experiments, we found that subjects recognized reduced words faster the more predictable they were given the following word. Furthermore, the recognized words were subsequently named faster if they were part of a fixed expression. Both effects were found not only when the presented utterances were completely natural, but also when the acoustic properties of the target word were strictly controlled. Therefore, we conclude that listeners, when confronted with reduced word forms, are indeed sensitive to predictability from the following word.

An interesting aspect of this conclusion is its implication that recognition of the target words in our experiments must generally have taken place after the following words had also been (partly) processed. This can also be seen from the timing of the button presses: In both experiments, more than 98.5% of the button presses occurred after the acoustic offset of the target, with more than 78.5% occurring after the acoustic offset of the following word. The average latency between the offset of the target and the button press was 660 milliseconds in Experiment 1 and 552 milliseconds in Experiment 2. Given that the average duration of the following word was 350 milliseconds, it is clear that subjects indeed tended to press the button well after the following word had ended.

What might have caused these relatively long delays? One possibility is that the subjects, who knew that they would be hearing four-word utterances, deliberately waited until they had heard the complete utterance. This would fail to explain, however, why the response latencies could not be completely predicted on the basis of stimulus duration. A more plausible explanation is that the fast and reduced nature of our speech materials more or less forced the subjects to delay their responses until after the final word, simply because there had not been enough time or information available for the identity of the third word to reach awareness.

Our results complement the earlier findings by Ernestus et al. (2002) and Kemps et al. (2004), who found that in the absence of linguistic context, listeners have difficulty in linking extremely reduced word forms to their canonical forms. Since the linguistic context in their experiments included lexico-syntactic as well as prosodic information, it was unclear which elements in the context actually helped listeners recognize the reduced word forms. From the current results, it appears that the lexical identity of the following word is one possible source of information that helps listeners recover the canonical form of reduced words.

This finding is difficult to incorporate into existing models of spoken-word recognition, which are mainly concerned with the recognition of words in isolation. Moreover, these models are often not explicit about the relationship between the recognition of a word and conscious perception. A notable exception in this respect is Adaptive Resonance Theory (ART), developed by Grossberg and colleagues. For the current purposes, we focus on the version of the model outlined in Grossberg and Myers (2000). This version, referred to as ARTWORD by the authors, was used to simulate the results of Repp et al. (1978) (*great ship* vs. *gray chip*). ARTWORD's success in modeling these so-called "backward effects" suggests that it might also be able to account for the results obtained in the present study. Therefore, we now take a closer look at some of the details of the model.

The two most prominent processing components in ARTWORD are a working memory, which stores position-sensitive activity levels of phonemic items, and a list grouping network, which contains representations of linguistic units (e.g., words) referred to as list chunks. The two components are linked by means of bottom-up and top-down weights. During speech perception, those list chunks are activated whose weights best match the activity pattern across the working memory. As soon as the activation reaches a certain critical threshold, the list chunk and the corresponding activity pattern engage in a positive feedback loop. This mutual increase in activation, referred to as resonance, binds the phonemic items into larger linguistic units and raises them into the listener's conscious perception. List chunks can compete with each other for selection, and, in principle, the longest list chunk that is consistent with the bottom-up evidence is selected for resonance.

What enables ARTWORD to accommodate effects of following context, is its ability to temporarily store phonemic information in working memory. As a result, the conscious processing of speech can proceed at a slower rate than the rate with which the speech is coming in. Since resonance usually takes some time to develop, it is very well possible for information occurring after a particular speech event to change the perception of that event. In fact, since conscious perception takes place only after the competition between alternative list chunks has been resolved, it is inevitable that all information necessary for choosing between alternatives will affect the recognition process, even if this information occurs well after the word that has to be recognized.

This characteristic of ARTWORD could also be the key to explaining the findings of the current study. Because of the high speech rate of the stimuli, resonance did generally not develop before the utterance had been presented completely. At this point, listeners had also heard the following word, enabling information about the identity of this word to affect processing of the preceding target. To explain why words that were predictable given the following word were recognized faster, one would need to assume that resonance somehow developed faster for these words. One possible way in which this could be achieved in ARTWORD is by allowing list chunks to not only compete with each other, but also to enhance each other's activation levels. In the case of an expression like *verdacht persoon*, this would mean that bottom-up evidence for *persoon* would lead to an increase in activation for *verdacht*, with the amount of additional activation depending on the strength of the connection between the two words. This could then explain why *verdacht* engages in a resonant feedback loop faster if it is followed by *persoon* than if it is followed by *papier*, from which it receives only little additional activation.

A second, more important adaptation that would need to be made to ARTWORD is that the absence of certain phonemes in the signal should not preclude a list chunk from engaging in a resonant feedback loop. Currently, resonance is only possible if all phonemic items supporting a chunk receive bottom-up activation (Grossberg & Myers, 2000, p. 739). However, most of the target words in our experiments missed at least one phoneme, and nevertheless tended to be recognized correctly. This suggests that either lexical representations do not necessarily contain all phonemes of a word, or that items other than phonemes should be chosen as units in working memory. This last possibility is left wide open by the developers of ARTWORD: "Exactly what the features, and the corresponding levels, represent remains an area of active research. In ARTWORD, these features correspond to standard units of psycholinguistic analysis of English. In general, the psycholinguistic data relevant to a given language determine what units are present in each model level" (Grossberg & Myers, 2000, p. 740-741). This suggests that if it were to be found that phonemes are not the most plausible perceptual units, ARTWORD could be adjusted by changing the type of units present in the working memory.

Given that so many adjustments need to be made to ARTWORD before the model can account for our findings, one might be tempted to ask whether other models might be able to explain our findings more easily. We feel that this is not the case. For example, the Shortlist model (Norris, 1994) assumes that a word is recognized as soon as its activation level reaches a particular critical threshold. The current study shows that the recognition of a word can be affected by, or even depend on, the recognition of the word that follows it. Such an effect cannot be explained in Shortlist, which implicitly assumes that words are recognized in the order in which they are pronounced.

In the remainder of this discussion, we consider some issues about which our results seem to provide valuable information, but which also require further investigation. A first question that remains largely unanswered is to what extent the current findings can be ascribed to the fact that our target words were reduced. In other words, would the effects be similar in size if listeners were presented with unreduced speech? To gain more insight into this issue, we included complex as well as simplex target words in our stimuli. Since the complex words had a higher likelihood of segment deletion than the simplex words, we hypothesized that the complex words might benefit more from a high predictability given the following word. In contrast, we observed in Experiment 1 that the effect of predictability from the following word was smaller for complex than for simplex words. As mentioned earlier, this finding should be interpreted with caution, as it was not replicated in Experiment 2. However, it at least suggests that effects of lexical context are not limited to words that have undergone extreme segmental reduction. This preliminary conclusion receives support from studies on the influence of preceding context on spoken-word processing, such as Zwitserlood (1989) and Van den Brink, Brown, and Hagoort (2001), but a definitive answer to this question can only be given after additional experimentation.

Since many of our experimental stimuli were idiomatic expressions, our data might also provide some insights into the processing of idioms. For example, we observed that the speed of recognition of the target word was affected by the continuous conditional probability measure, while the speed of production of the same word was affected by whether or not the expression was fixed. This distinction is remarkable, as it suggests that the idiomatic status of an expression manifests itself differently during perception and production. During perception, the strength of the connection between the target and the following word (as measured by conditional probability) helps the target word to be recognized faster. However, by the time the target has to be produced, it appears to receive additional activation from a lexical representation of the idiom itself. Such an account would be in accordance with previous work by Sprenger (2003), who claims that idiomatic expressions have separate representations in the mental lexicon. What remains unclear, is why these representations do not openly exert an influence during perception.

Another remarkable finding in the present study was the main effect of Splicing Condition observed in Experiment 2. Target words that originally occurred in a fixed expression were named faster than target words that originally occurred in a non-fixed expression, regardless of the following word in the stimulus. This effect cannot be ascribed to the degree of reduction in the target, which did not differ between the two splicing conditions. An alternative explanation is that the effect is related to differences in coarticulation. It is quite likely that targets from F-utterances were realized with more coarticulation than targets from NF-utterances. Because coarticulation spreads relevant information more evenly across the speech signal, coarticulated words can be easier to recognize (e.g., Scarborough, 2004), which might explain the advantage we observed for targets from F-utterances. Further research is needed to explore the role of such fine phonetic detail in speech perception (e.g., Hawkins, 2003).

Finally, we would like to stress that the present study serves only as a starting point in identifying the contextual elements that help listeners recognize reduced word forms. Besides the directly following word, other information sources are likely to be helpful as well. These include, but are not limited to, preceding linguistic context, prosody, audio-visual cues, and the non-linguistic context provided by the communicative situation. All these information sources are worthy of investigation in their own right. Furthermore, it should be investigated which cues take precedence under which circumstances, and how the different cues interact in situations where multiple cues are available. Only after detailed analysis of the roles of these and other information sources can true progress be made in understanding how listeners recognize reduced word forms.

Effects of morphological structure: The case of -igheid

CHAPTER 6

This chapter has been submitted to Laboratory Phonology 10 as Mark Pluymaekers, Mirjam Ernestus, R. Harald Baayen, and Geert Booij: Morphological effects on fine phonetic detail: The case of Dutch *-igheid*.

Abstract

In the previous chapters, predictor variables were investigated which had already been identified by other researchers as predictors of acoustic reduction. In this chapter, we explored a variable which has received much less attention in the literature, namely morphological structure. The focus was on the Dutch derivational suffix *-igheid* (/əxhɛit/), which occurs in two types of words. In the first type, *-igheid* is analyzed as a single suffix. In the second type, there is a morphological boundary between *-ig* and *-heid*. The main research question was whether this difference is reflected in the duration of the /xh/ cluster. Two hypotheses were distinguished: One based on prosodic structure, which predicts that the cluster is shorter in the first type than in the second type, and one based on the informativeness of the affix given the morphological paradigm, which makes the opposite prediction. All occurrences of *-igheid* in a corpus of read speech were acoustically analyzed using Automatic Speech Recognition technology. The duration of the /xh/ cluster was found to be shorter in words of the second type than in words of the first type, providing support for the hypothesis based on informativeness. This finding suggests that morphological effects on fine phonetic detail cannot always be explained by prosodic structure.

Introduction

The acoustic realization of words and affixes is characterized by immense intra- and inter-speaker variation. Some of this variation is due to noise in the execution of speech motor activities, and therefore lies outside the traditional research domain of (psycho)linguistics. On the other hand, many sources of variation have been uncovered that are directly relevant for (psycho)linguistic theory. These sources include the position of word stress and sentence accent (e.g., Nooteboom, 1972; Van Bergem, 1993; Turk & Sawusch, 1997; Turk & White, 1999), the position of a word within a prosodic domain (e.g., Fougeron & Keating, 1997), whether or not a word is part of a fixed expression (Binnenpoorte et al., 2005), the frequency and predictability of a word (e.g., Lieberman, 1963; Hunnicutt, 1985; Jurafsky et al., 2001; Chapter 1 of this thesis), and speaker characteristics such as sex, age and regional origin (e.g., Labov, 1972; Byrd, 1994; Keune et al., 2005). In the current study, we explore whether over and above these and other relevant factors, morphology also has a role to play in explaining pronunciation variation. We do this by investigating the phonetic implementation of the Dutch derivational suffix *-igheid* (/əxhɛit/).

The morphological structure of -igheid

Strictly speaking, the suffix *-igheid* consists of two separate suffixes: *-ig* and *-heid*. Hence, according to a standard morphological analysis, the noun *groenigheid* 'greenishness' is derived from the adjective *groenig* 'greenish', which is in turn derived from the adjective *groen* 'green'. Insightful as such an analysis may be, it does not necessarily reflect the mental processes underlying the production of such a complex word. As Van Marle (1990) points out, morphological reanalysis may allow speakers to skip the second step and derive *groenigheid* directly from *groen*. A similar observation is made by Haspelmath (1995), who discusses the history of *-igheid*'s German equivalent, the suffix *-igkeit*. Originally, a word like *Müdigkeit* 'tiredness' was derived from *müdig*, which itself was derived from the base word *müde*. When the form ending in *-ig* fell out of use, speakers were more or less forced to analyze words like *Müdigkeit* as a combination of the base word and *-igkeit*, and this is when the new suffix *-igkeit* came into being. Now, *-igkeit* is also applied to adjectives that never had a form ending in *-ig*, such as *gefühllos* 'senseless' (*Gefühllosigkeit* 'senselessness').

For Dutch, the upshot of this reanalysis is that *-igheid* currently occurs in three types of words. In the first type, *-igheid* must be analyzed as a single suffix. For example, the word *vastigheid* 'security' is necessarily derived from *vast* 'solid', as *vastig* does not exist in Dutch. Hence, the morphological boundary in words of this type lies before *-igheid*. In words of the second type, the morphological boundary lies before *-heid*. This is for instance the case in a word like *zuinigheid* 'thriftiness', which has to be derived from *zuinig* 'thrifty' since *zuin* is not a word in Dutch. The third category of *-igheid* words consists of words that could be derived by adding *-igheid* to the base word, as well as by adding *-heid* to an existing form ending in *-ig*. Consider a word like *bazigheid* 'bossiness'. Since *baas* 'boss' and *bazig* 'bossy' are both existing Dutch words, there is ambiguity as to whether *-igheid* functions as a single suffix or not. In the remainder of this chapter, we will refer to words of the first type as +igheid words, to words of the second type as +heid words, and to words of the third type as ambiguous words. The term '*-igheid* words' is used to refer to all words ending in *-igheid*, regardless of their morphological structure.

The phonetic implementation of -igheid

With respect to the phonetic implementation of *-igheid*, one obvious area of interest is the realization of the /xh/ cluster. This cluster, which does not occur morpheme-internally in Dutch, consists of two segments that are hard to distinguish perceptually. Therefore, one might expect the cluster to simplify, for example by deletion of /h/. Schultink (1962, p. 165) claims that /h/-deletion is indeed standard in the pronunciation of *-igheid*. However, whether or not the cluster is simplified might also depend on the morphological structure of the word. In the current study, we examine whether the differences in morphological structure outlined above are reflected in the acoustic duration of the /xh/ cluster.

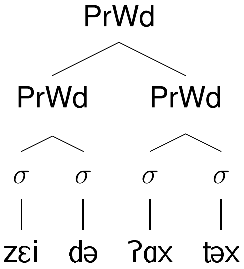
Recently, Hay (2003) demonstrated effects of morphological structure on the acoustic realization of the English phoneme sequence /tl/. Since this sequence does not occur within morphemes, Hay argues that listeners may use it as a cue to a morpheme boundary. In addition, she claims that whether an affixed word is accessed as a whole or through its constituent morphemes is codetermined by the relative frequency of the affixed word compared to the base word. If the affixed word is more frequent than the base, it is more likely to be accessed as a whole. If on the other hand the base word is more frequent, recognition is more likely to take place through the constituent morphemes. This suggests that the phoneme sequence /tl/ functions as a morphological boundary marker mainly in the second case, because it is in these words that morphological decomposition is most likely to take place. Hay predicts, therefore, that the acoustic realization of /t/ is longer in words like *softly*, which is less frequent than its base word *soft*, than in words like *swiftly*, which is more frequent than its base word *swift*. Her prediction is confirmed in a small laboratory experiment: /t/ is indeed longer in words like *softly* than in words like *swiftly*. This shows that it is possible to observe effects of morphological structure on acoustic realizations.

In the remainder of this introduction, we discuss two theoretical accounts that make different predictions about the simplification of the /xh/ cluster in *-igheid*.

A prosodic account

In the generative tradition, effects of morphology on acoustic realizations are believed to be mediated by phonology (e.g., Kenstowicz, 1993, p. 60). In Prosodic Phonology (Nespor & Vogel, 1986), for instance, affixes can form prosodic words of their own, allowing the morphological structure of a complex word to be reflected in its prosodic structure. An example of a Dutch affix that forms its own prosodic word is the suffix *-achtig* (Booij 1995, p. 47). The prosodic structure of the word *zijdeachtig* 'silky' looks as follows:

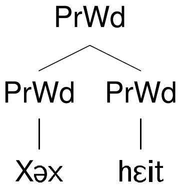
(1)



In (1), there is no resyllabification across constituent boundaries, as evidenced by the absence of Prevocalic Schwa Deletion and the insertion of the glottal stop /ʔ/. This example illustrates the basic assumption in Prosodic Phonology that morphological structure can affect phonetic form through mediation by phonology.

Now, let us consider the prosodic structure of *-igheid* words. According to Booij (1995, p. 47-52), *-heid* forms a prosodic word of its own, while *-ig* prosodifies with the stem. Therefore, it could be argued that all *-igheid* words have the following prosodic structure (X refers to the stem):

(2)



Such a structure works perfectly for words like *zuinig+heid*, in which there is a clear morphological boundary between *-ig* and *-heid*. Since the /xh/ cluster marks a prosodic as well as a morphological boundary in these words, it is unlikely to be simplified.

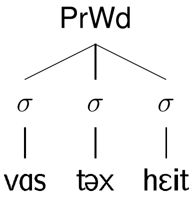
In some *-igheid* words, however, the suffix necessarily functions as a single unit (e.g., *vast+igheid*). For these words the prosodic structure in (2) is less optimal, as it posits a prosodic word boundary between *-ig* and *-heid* even though the two are part of the same linguistic unit. This problem can only be resolved by assuming a different prosodic structure. However, it is not possible to simply assume a prosodic word boundary before *-igheid*, as prosodic words in Dutch cannot start with schwa. Furthermore, there is resyllabification of

Table 6.1: Types of words ending in *-igheid* including examples, morphological structures, and predictions about the phoneme cluster /xh/ based on a prosodic account.

Type	Example	Morphological structure	Prediction for /xh/
+heid	zuinig+heid 'thriftiness'	Xig + heid	not simplified
+igheid	vast+igheid 'security'	X + igheid	simplified
ambiguous	baz+ig+heid 'bossiness'	Mostly X + igheid	simplified

the schwa with the preceding consonant. Therefore, a structure like (3) seems to be the only possible alternative:

(3)



In (3), /h/ no longer occurs at the beginning of a prosodic word, which makes it a likely target for deletion. Consequently, a prosodic account predicts cluster simplification in words in which *-igheid* is a single suffix.

For words in which the morphological boundary can, but need not lie after *-ig*, such as *baz+ig+heid*, it is not self-evident which of the two prosodic structures applies. To determine the most likely morphological parse, Hay's (2003) concept of relative frequency can be used. This means that if the base word (e.g., *baas*) is more frequent than the *-ig* form (e.g., *bazig*), no morphological boundary is assumed after *-ig*, making prosodic structure (3) the most plausible option. If, on the other hand, the *-ig* form is more frequent than the base, structure (2) can be assumed, as there is a morphological boundary between *-ig* and *-heid*. Since base words tend to be more frequent than *-ig* forms, the majority of words in this category will have a prosodic structure like (3). Hence, a prosodic account predicts that the /xh/ cluster will often be simplified in these words.

The predictions that a prosodic account makes about simplification of the /xh/ cluster are summarized in Table 6.1. Henceforth, we will refer to this set of predictions as the Prosodic Structure Hypothesis.

An information-based account

In the previous section, the focus was on the interplay between morphological and prosodic structure. However, morphology also governs the information flow within a word. Affixes can

Table 6.2: Summary of the morphological paradigms typically associated with the three types of *-igheid* words (X refers to the base; + indicates that a particular paradigm member exists).

Paradigm	+heid *zuinig+heid*	+igheid *vast+igheid*	ambiguous *baz+ig+heid*
X		+	+
X -e(n)		+	+
X -ig	+		+
X -ige	+		+
X -igheid	+	+	+
X -igheden	+	+	+
Compounds with X		+	+

change the syntactic category or meaning of a word, which is what makes them informative for the listener. Informativeness has long been recognized as an important predictor of phonetic reduction, in that less informative linguistic units are found to be more reduced than more informative units (e.g., Lieberman 1963; Hunnicutt 1985; Jurafsky et al. 2001; Aylett and Turk 2004). Informativeness can be quantified in several ways, but most recent studies have used probabilistic measures derived from Information Theory (Shannon 1949). For example, Van Son and Pols (2003) developed a measure that estimates the individual contribution of a phoneme to word recognition by computing the reduction in the size of the cohort after that phoneme has been added to the signal.

A related approach can be found in Wright (1997) and Scarborough (2004), among others. Wright (1997) found that words occurring in dense lexical neighborhoods (i.e., words that have a large number of competitor words differing in only one phoneme) were produced with more dispersed vowels than words occurring in sparse lexical neighborhoods. This can be explained by assuming that speakers hyper-articulate words from dense neighborhoods, as these words are easier to confuse with other words. Scarborough (2004) observed that more lexical confusability is also correlated with a higher degree of coarticulation. She explains this finding by pointing to the spreading of information that occurs when segments are coarticulated. By spreading phonemic information more evenly across the signal, speakers increase the likelihood of correct recognition, which is especially relevant if the intended word can easily be confused with other words.

Let us now apply the same principle to words containing *-igheid*. Before the pronunciation of the schwa, the cohort of competitor words will mainly consist of words from the same morphological paradigm. This means that the suffix is more informative the more competitors there are in the paradigm. Table 6.2 gives an overview of the morphological paradigms typically associated with the three types of *-igheid* words.

Table 6.2 shows that ambiguous words generally have the densest paradigms, followed by +igheid words and +heid words. It should be noted, however, that the difference between +heid and +igheid words is larger than the difference between +igheid and ambiguous words. This is because the paradigms of +igheid and ambiguous words include a large number of compounds starting with the base word. Since such compounds do not exist for +heid words, the suffix is much less informative in +heid words than in the other two types.

Of course, one also needs to take the informativeness of the /xh/ cluster itself into account. In +heid words, like *zuinig+heid*, the only element in the cluster that is potentially informative is /h/, as it distinguishes *zuinigheid* ‘thriftiness’ from *zuinige*, an inflectional variant of *zuinig* ‘thrifty’. However, one could argue that /h/ is not suitable for making this distinction, since both /h/ and schwa can be realized as voiceless vowels in Dutch. Therefore, we can conclude that the /xh/ cluster is hardly informative in +heid words and hence, quite likely to be simplified.

In +igheid words, like *vast+igheid*, the cluster is far more informative. First of all, it disambiguates between *vaste* ‘solid’, an inflectional variant of *vast*, and *vastigheid* ‘security’. Second, it distinguishes between *vastigheid* and compounds that start with *vast* followed by a schwa, such as *vasteland* ‘mainland’ en *vastenavond* ‘Mardi Gras’. In some cases, the cluster also signals that the combination of the base word and schwa does not form the onset of a completely different word. This is for example the case in *viezigheid* ‘dirtiness’, which has the same beginning as *vice-decaan* ‘vice-dean’. Since the /xh/ cluster plays such an important role in distinguishing +igheid words from other words in the paradigm, it is not very likely to be simplified.

In ambiguous words like *baz+ig+heid*, the cluster is informative as well. It eliminates *bazen*, the plural form of *baas* ‘boss’, from the cohort, as well as a large number of compounds that start with the combination of the base word and schwa. For a word like *geestigheid* ‘witticism’, these compounds include (but are not limited to) *geestenbanner* ‘exorcist’ , *geesteskind* ‘brainchild’, *geestenleer* ‘spiritualism’, *geestesziek* ‘mentally ill’, and *geestenziener* ‘medium’. This means that the cluster is not very likely to be simplified in ambiguous words either.

Table 6.3 summarizes the predictions an information-based account makes with regard to simplification of the /xh/ cluster. This set of predictions will henceforth be referred to as the Morphological Informativeness Hypothesis.

It could be argued that the information-based account does not really reflect morphological structure. However, the concept of informativeness outlined here is nothing more than the probabilistic consequence of the principle of proportional analogy, which is regarded as pivotal in structuralist as well as word-and-paradigm morphology (e.g., Blevins, 2003). In word-and-paradigm morphology, the morphological unit is not the affix, but the word as it occurs in its morphological paradigm. The informativeness of an affix correlates with the density of the paradigm. As can be seen in Table 6.2, this density is much higher for words like *vast+igheid* and *baz+ig+heid* than for words like *zuinig+heid*. Therefore, it is essential for

Table 6.3: Types of words ending in *-igheid* including examples and predictions about the phoneme cluster /xh/ based on an information-based account.

Type	Example	Morphological paradigm	Prediction for /xh/
+heid	zuinig+heid ‘thriftiness’	sparse	simplified
+igheid	vast+igheid ‘security’	dense	not simplified
ambiguous	baz+ig+heid ‘bossiness’	dense	not simplified

successful communication that *vast+igheid* and *baz+ig+heid* are pronounced more carefully.

In summary, we have presented two hypotheses that make different predictions with respect to the phonetic implementation of the phoneme cluster /xh/ in *-igheid*. According to the Prosodic Structure Hypothesis, cluster simplification is more likely if *-igheid* is analyzed as a single suffix. The Morphological Informativeness Hypothesis, on the other hand, predicts that cluster simplification is less likely if *-igheid* is a single suffix, because the cluster is more informative in such words. These two hypotheses were pitted against each other in a corpus study.

Method

Materials

The materials were taken from the subcorpus ‘Library for the Blind’ of the Corpus of Spoken Dutch (Oostdijk, 2000). This subcorpus comprises 100 hours of recordings of written texts, read aloud by trained speakers from the Netherlands and Flanders. Our main motivation for using read speech rather than spontaneous speech, which is also available in the corpus, was the superior sound quality of the recordings.

All 432 occurrences of *-igheid* in the subcorpus were selected for acoustic analysis. There were 164 different word types in the sample, 100 of which occurred only once. The two most frequent words in the sample were the +heid words *aanwezigheid* ‘presence’, which occurred 52 times, and *nieuwsgierigheid* ‘curiosity’, which occurred 22 times. For each word, the morphological type was determined on the basis of the morphological parse in the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995).

Acoustic analysis

Acoustic analysis of the selected tokens was performed using the ASR device described in Chapter 3. This was done for several reasons. First of all, the ASR was trained such that it bases its decisions purely on the characteristics of the acoustic signal, without reference to linguistic knowledge. This is difficult for phoneticians, who are bound to be influenced by their knowledge of spelling and phonotactics (Vieregge, 1987; Cucchiarini, 1993). Second, ASR devices are perfectly consistent: Multiple analyses of the same acoustic signal will always yield exactly the same result. Finally, recent research has shown that the reliability of segmentations generated by an ASR system is equal to that of segmentations made by human transcribers (Vorstermans, Martens, & Van Coile, 1996; Sjölander, 2001), provided that a phonemic transcription of the signal is available to the ASR algorithm.

For the acoustic analysis of the *-igheid* words, we manually excised the speech signals corresponding to these words from their sentence contexts. Subsequently, the signals were parameterized using Mel Frequency Cepstral Coefficients. Each parameterized signal was provided to the Viterbi algorithm, which automatically segmented the signal into phonemes on the basis of the CELEX transcription of the word. To correct for segmentation error, the beginnings of plosives and liquids were shifted 10 and 7 milliseconds to the right. By following this procedure, we obtained information about the durations of all individual segments in a word.

Statistical analysis and control variables

To see whether morphological type affected the acoustic realization of /xh/ while controlling for other relevant variables, we used multiple regression analysis. Regression analysis is a statistical technique that allows researchers to see whether a particular independent variable has an effect over and above other variables that may be relevant. Therefore, it is an extremely useful tool for analyzing corpus data. Furthermore, prior averaging is not necessary, as regression models are fitted to individual data points and not to means. Finally, it is easy to spot interactions between variables. If, for example, the effect of morphological type was to be limited to words with a high frequency of occurrence, this would surface as a significant interaction of morphological type by frequency in the regression model.

The dependent variable in the analysis was the duration of the /xh/ cluster, as measured by the ASR. Originally, we had planned to also investigate /h/-deletion, but a pre-test established that the presence or absence of /h/ after the fricative /x/ could not be reliably determined, neither by the ASR nor by human transcribers. Although there might be some compensatory lengthening of /x/ if /h/ is deleted, the duration of the cluster as a whole will not become longer as a result of /h/-deletion. Therefore, we feel that the predictions in Tables 6.1 and 6.3 hold for the duration of the /xh/ cluster independently of the presence of /h/.

The main independent variable was morphological type. In the analysis, two types were distinguished: +heid and +igheid. The ambiguous words were classified as +igheid words, for several reasons. First of all, 84% of the ambiguous words was likely to behave morphologically as a +igheid word, as the base word was more frequent than the *-ig* form (Hay 2003). Second, we did not observe significant differences between the +igheid and ambiguous types in any of the analyses we performed. Finally, the predictions concerning cluster simplification are the same for the two types, regardless of the hypothesis under investigation. Therefore, we decided to analyze them as a single category.

It might be argued that the morphological structure of ambiguous words is not exclusively determined by the frequency ratio of the base word and the *-ig* form. For example, semantic structure could also play a role. Consider a word like *bedrijvigheid* ‘industriousness’, which could be derived from *bedrijf* ‘company’ as well as from *bedrijvig* ‘industrious’. From a semantic perspective, the latter derivation seems more plausible. If frequency is taken as the criterion, however, *bedrijvigheid* will be classed as a +igheid word, since *bedrijf* is much more frequent than *bedrijvig*. The reason why we did not take semantics into account in our classification is that determining the semantic relationships between words is a highly intuitive enterprise. The frequency ratio, on the other hand, is based on quantitative estimates, and therefore provides a more objective measure for classifying word forms.

Control variables that were included as covariates were the speaker characteristics sex, age, and country of origin (Netherlands vs. Flanders), the rate of speech, the frequency of the word in the Corpus of Spoken Dutch, and whether the word was in utterance-initial or utterance-final position. Age was operationalized by subtracting 1900 from the year of birth of the speaker. Speech rate was estimated by counting the number of syllables per second in the utterance in which the *-igheid* word occurred. In the Corpus of Spoken Dutch, utterances are defined as stretches of speech that occur between audible pauses. Since in read-aloud speech, segment deletions are relatively rare, the syllable counts were based on the canonical pronunciations of the words in the utterance. Position in the utterance was controlled by means of two binary variables, Initial and Final, which were coded as either true or false for each token. Finally, we analyzed the tokens in terms of their accentual status and found that almost all of them were accented (as evidenced by F_0 movement). This is not surprising, since the *-igheid* words were often the only content words in their respective utterances. Therefore, we felt that it would not be necessary to explicitly control for accentual status in the regression model, as the ratio of accented vs. non-accented tokens would be too uneven to expect any significant effects (or interactions) to emerge.

Results

To predict the duration of the /xh/ cluster, a least squares model was fitted that contained morphological type and all control variables as predictors. To see whether morphological type had an effect over and above the control variables, it was entered into the model last.

In a stepwise model selection procedure, only those variables were retained that showed a significant effect. On the basis of diagnostic plots, we identified three data points that were outliers with respect to leverage or Cooks distance. After removing these outliers, the model was refitted to the remaining data. This analysis showed a significant effect of morphological type on cluster duration($F(1, 423) = 10.49, p < 0.005$). More specifically, the duration of the /xh/ cluster was shorter in words of the +heid type than in words of the +igheid type ($\hat{\beta} = -7.9, t(423) = -3.24, p < 0.005$). This finding, which is illustrated in Figure 6.1, provides support for the Morphological Informativeness Hypothesis.

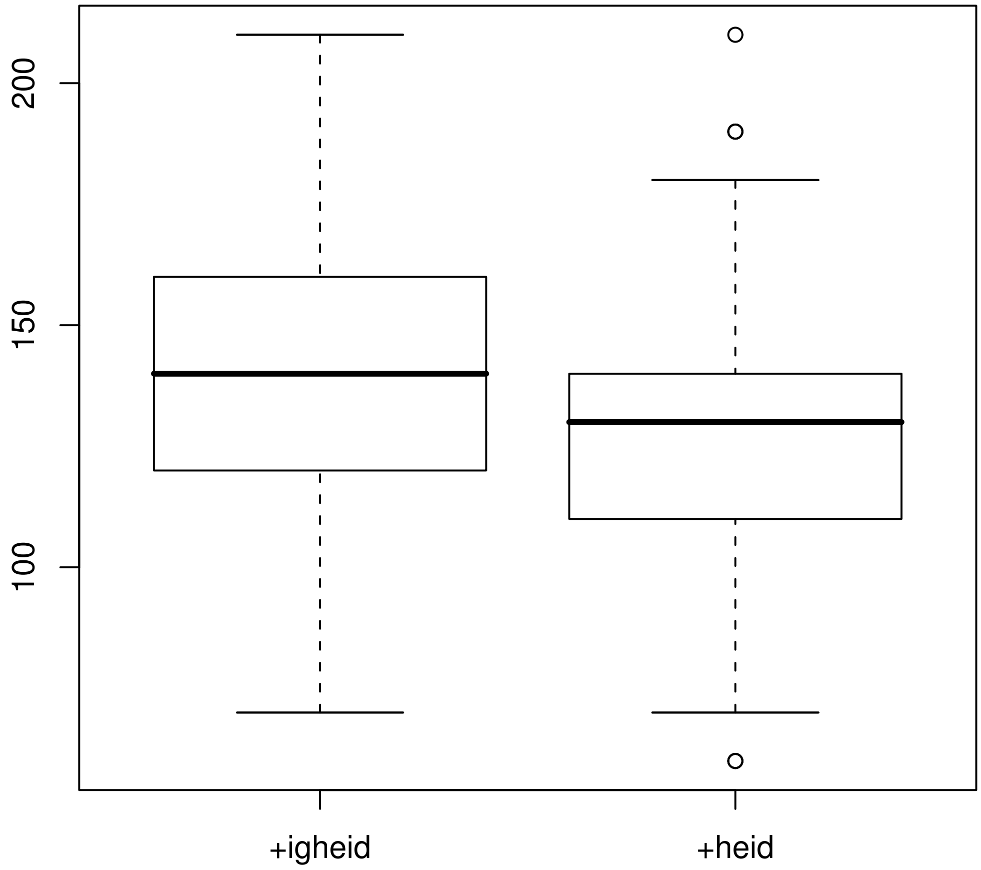


Figure 6.1: Boxplot of the duration of the /xh/ cluster (in milliseconds) as a function of morphological type. The bold lines indicate the medians in both categories.

In addition to morphological type, some of the control variables also showed significant effects. Clusters were longer if the word was in Initial position ($\hat{\beta} = 13.7, t(423) = 2.75, p < 0.01$; mean Initial: 147 ms; mean Non-Initial: 128 ms) or Final position ($\hat{\beta} = 12.3, t(423) = 5.07, p < 0.0001$; mean Final: 137 ms; mean Non-Final: 125 ms). Speakers from Flanders produced shorter clusters ($\hat{\beta} = -20.1, t(423) = -8.07, p < 0.0001$; mean Flanders: 116 ms; mean Netherlands: 139 ms), as did male speakers ($\hat{\beta} = -12.1, t(423) = -4.99, p < 0.0001$; mean male: 122 ms; mean female: 139 ms). All in all, the regression model accounted for 26% of the variance in the duration of the /xh/ cluster.

General Discussion

The current study investigated whether the fine phonetic detail of the Dutch suffix *-igheid*, canonically pronounced as /əxhɛit/, is affected by morphological structure. We conclude that this is indeed the case. The duration of the /xh/ cluster was found to be shorter in words in which *-igheid* is not a single suffix than in words in which it is. This finding lends support to the Morphological Informativeness Hypothesis. According to this hypothesis, the duration of the cluster is affected by its informativeness given the word's paradigmatic neighborhood. For words in which *-igheid* is not a single suffix, such as *zuinig+heid*, the informativeness of the cluster is relatively low, since the possible base word (*zuin*) does not exist in Dutch and its paradigmatic neighborhood is relatively sparse. Thus, it is already clear at the end of [zœyn] that the word to be produced will be *zuinig* or one of its morphological continuation forms. As a result, the /xh/ cluster is relatively uninformative with respect to word identity, which manifests itself in durational shortening.

What makes this finding particularly interesting is that it cannot be explained on the basis of a prosodic account. The Prosodic Structure Hypothesis predicts that the /xh/ cluster would be longer in words like *zuinig+heid*, because it serves as a morphological boundary marker there. Now that the exact opposite has been observed, we can conclude that, contrary to received wisdom, morphological effects on fine phonetic detail cannot always be accounted for by prosodic structure.

Our results also illustrate that intuitions about how a particular word or affix is pronounced can be misleading. Schultink (1962) claims that /h/ is likely to be deleted in all words containing *-igheid*. The observation in the current study that simplification of the /xh/ cluster occurs especially in +heid words is at odds with this intuition. This once more underlines the importance of corpus data in phonological and phonetic research. We believe that corpus data are indispensable for confirming intuitions, or, as we did in the current study, for testing alternative hypotheses concerning the fine phonetic detail of acoustic realizations. However, corpus researchers need to make sure that possibly confounding variables are sufficiently controlled. If this cannot be done by means of experimental design, statistical techniques such as regression analysis should be used.

By saving effort on the articulation of uninformative linguistic units, speakers can free up resources for other cognitive tasks. Simultaneously, listeners might benefit from detailed knowledge about the resulting reduction patterns. Numerous studies have shown that listeners can use fine-grained structural phonetic differences between words to improve word processing (e.g., Davis, Marslen-Wilson, and Gaskell 2002; Hawkins 2003; Salverda, Dahan, and McQueen 2003; Warner et al. 2004; Kemps et al. 2005a; Kemps et al. 2005b; Ernestus and Baayen 2006). Whether they also use the acoustic patterns reported in the current study can only be determined on the basis of a perception study.

Although the current study was exclusively concerned with synchronic reduction, our results may also provide insights about how the pronunciation of *-igheid* will develop diachronically. Since we observed synchronic reduction in the duration of the /xh/ cluster in +heid words, it is not inconceivable that the cluster will also fall subject to diachronic reduction in these words.

With regard to future research, it would be interesting to examine whether patterns similar to the one described in the current paper can be found in other languages. An obvious candidate for such an investigation is the German suffix *-igkeit*, which resembles *-igheid* in terms of morphological structure and has the additional advantage that the two phonemes in the cluster of interest are acoustically well distinguishable. Further exploration of affixes like *-igheid* and *-igkeit* could shed more light on how morphology affects phonetic form without mediation by phonology.

Effects of morphological predictability on interfix duration

CHAPTER 7

This chapter has been published as Victor Kuperman, Mark Pluymaekers, Mirjam Ernestus, and R. Harald Baayen (2007). Morphological predictability and acoustic salience of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121, 2261-2272.

Abstract

This chapter explored the effects of morphological predictability on the acoustic durations of interfixes in Dutch compounds. Two data sets were investigated: One for the interfix *-s-* (1155 tokens) and one for the interfix *-e(n)-* (742 tokens). Both datasets show that the more probable the interfix is given the compound and its constituents, the *longer* it is realized. These findings run counter to the predictions of information-theoretical approaches and can be resolved by the Paradigmatic Signal Enhancement Hypothesis. This hypothesis argues that whenever selection of an element from alternatives is probabilistic, the element's realization is predicted by the amount of paradigmatic support for the element: The most likely alternative in the paradigm of selection is realized with greater acoustic salience.

Introduction

One of the organizing principles of speech production is the trade-off between economy of articulatory effort and discriminability of the speech signal (Lindblom, 1990). Speech communication often takes place in noisy conditions. In order to ensure robust recognition of their acoustic output, speakers need to invest effort in articulation. Yet clear and careful articulation is costly and hence tends to be dispensed efficiently (Aylett & Turk, 2004; Hunnicutt, 1985). As a consequence, elements with low information load (or high predictability) have shorter or otherwise less salient realizations than relatively more informative elements of an utterance.

The informational redundancy of speech elements is often operationalized in terms of the probability (relative frequency of occurrence) of a linguistic unit (e.g., phoneme, syllable, word, or phrase) in its context. High probability has been observed to correlate with acoustic reduction in a large variety of language domains: Syntactic, discourse-related, phonological and prosodic, and lexical (e.g., Aylett & Turk, 2004; Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Fowler & Housum, 1987; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Samuel & Troicki, 1998; Scarborough, 2004; Van Son & Pols, 2003; Van Son & Van Santen, 2005; Chapters 2, 3, and 4 of this thesis). The attested types of reduction include — apart from widely reported durational shortening of syllables and individual phonemes — deletion of phonemes and complete syllables (e.g., Ernestus, 2000; Johnson, 2004), decrease in spectral center of gravity (Van Son & Pols, 2003), decrease in mean amplitude (Shields & Balota, 1991), lower degree of centralization of vowels (Munson & Solomon, 2002), and lower degree of coarticulation (Scarborough, 2004). The informational redundancy associated with a particular unit is a juxtaposition of the unit's probabilities given all relevant contexts. For instance, a word can be predictable because it has a high frequency, but also because it is frequently used with the word that precedes it. Both factors diminish the word's informativeness and both are expected to correlate with durational shortening.

The information-theoretical framework developed by Shannon and Weaver (1949) has been used to explain the association between acoustic salience and informational redundancy. The efficiency of information transmission is optimal if the information in the signal is distributed equally, or smoothly, per time unit (e.g., Aylett & Turk, 2004; Aylett & Turk, 2006). When an important element is transmitted for a longer time, the probability of losing this element to noise decreases and the probability of the element being recognized correctly increases. This theoretical paradigm views acoustic salience (expressed in duration, loudness, etc.) as a means of smoothing the amount of information in the signal over time.

The present chapter shows how the information carried by morphological paradigmatic structure modulates acoustic salience. Previous research (Hay, 2003; Losiewicz, 1992; Chapter 6 of this thesis) reported morphological effects on the acoustic duration of affixes in complex words. The morphological objects that are central in the present study are interfixes in Dutch noun-noun compounds. We will show that the acoustic salience of these interfixes creates an apparent paradox for the proposed information-theoretical principle of “less information, more reduction”, which underlies the Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004), the Probabilistic Redundancy Hypothesis (Jurafsky et al., 2001), and research on speech efficiency (e.g., Van Son & Pols, 2003). In our data, the more predictable the interfix is, the *more salient* its articulation.

The distributional characteristics of the interfixes in Dutch compounds provide a clear-cut example of probabilistic, non-categorical morphological structure. Compounds in Dutch can be realized with the interfix *-s-* (e.g., *oorlog-s-verklaring*, “announcement of war”), or with the interfix *-en-* (or its variant *-e-*) (e.g., *dier-en-arts* “veterinary”). Most compounds in Dutch, however, have no interfix (e.g., *oog-arts* “ophthalmologist”): For ease of exposition, we will henceforth refer to these latter words as compounds with the zero-interfix, or -$\emptyset$-. In the frameworks that adopt deterministic rules, the distribution of interfixes in Dutch is enigmatic and inexplicable. Krott, Baayen, and Schreuder (2001), however, have shown that the distribution of interfixes follows probabilistic principles defined over constituent families. The left (or right) constituent family of a compound is the set of all compounds which share the left (or right) constituent with this compound. For instance, the left constituent family of the compound *banknote* includes *bankbill*, *bankbook*, *bank-draft*, *bank-rate*, and *bankroll*. Krott et al. (2001) and Krott, Schreuder, and Baayen (2002) show that the selection of the interfix is biased towards the interfix that is most commonly used with the given left constituent and, to a lesser extent, with the right constituent. Thus, besides having their own probability of occurrence, interfixes exhibit dependencies on larger morphological units both to the left and to the right (Krott et al., 2002). For this reason, interfixes serve as an appealing testing ground for studying the consequences of morphological predictability for acoustic realization.

The primary focus of the present study is the relationship between the predictability of the interfix given the morphological constituents of the compound, and its duration. We study the information-theoretical approach for two datasets with interfixed compounds and against the backdrop of multiple sources of redundancy, ranging from morphological to phonological and lexical information. Along the way, we replicate findings of laboratory studies of durational reduction for lively read-aloud speech.

Method

Materials

Acoustic materials were obtained from the Read Speech (or the ‘Library for the Blind’) component of the Spoken Dutch Corpus (Oostdijk, 2000). Within this corpus of approximately

800 hours of recorded speech, the Read Speech component comprises 100 hours of recordings of written texts read aloud by speakers of Northern Dutch from the Netherlands and Southern Dutch from the Flanders area of Belgium. In the preparation of the recordings, speakers were pre-screened for the quality of their voice and clarity of pronunciation, and texts were made available to the speakers beforehand for preparatory reading. We chose to concentrate on read speech primarily because of the low level of background noise of the recordings. Quality was essential, since Automatic Speech Recognition (henceforth, ASR) was used for obtaining segmental durations (see below).

Two datasets of Dutch noun-noun compounds were compiled: One with compounds containing the interfix *-s-* and one with compounds containing the interfix *-e(n)-*. Tokens in which the interfix *-s-* was either preceded or followed by the phonemes (s), (z) or (ʃ) were excluded from the dataset, since such an environment makes it difficult to reliably segment the interfix from its neighbor. The final dataset for the interfix *-s-* consisted of 1155 tokens. Similarly, tokens in which the second constituent begins with the segments (n) or (m) were taken out off the dataset of *-e(n)-* interfixes, resulting in a dataset of 742 tokens.

Measurements

Acoustic analysis was performed using the ASR device described in Chapter 3. In the first step of the procedure, the speech signal corresponding to the target compound was manually excised from its utterance context and parameterized using Mel Frequency Cepstral Coefficients. The parameterized signal was then supplied to a Viterbi segmentation algorithm, along with a phonemic transcription of the word. This transcription was taken from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). However, for words with the interfix *-e(n)-*, a cursory inspection of sound files established that many instances of this interfix were not realized as [ə] (the canonical pronunciation in CELEX), but rather as [ən]. An inspection of the sound files from the dataset with the interfix *-s-* revealed cases where the interfix was realized as [s] instead of the CELEX transcription [z]. Therefore, two trained phoneticians independently transcribed the realization of interfixes in both datasets. Initially, they disagreed on 10% of tokens from the *en*-dataset and 13% of tokens from the *s*-dataset. In both cases, they subsequently carried out a joint examination of the problematic tokens and came up with consensus transcriptions. The resulting transcriptions were provided to the segmentation algorithm, which segmented the signal into phonemes. In this way, we obtained information about the durations of all segments for all words.

The acoustic duration of the whole interfix (henceforth, *InterfixDuration*) was taken as the main dependent variable in this study.

Morphological variables

As shown in Krott et al. (2001), the more frequent an interfix is in the left constituent family of a compound, the more biased speakers are to use this interfix in that compound. The measures for this morphologically based bias will be at the center of our interest. They are defined as the ratio of the number of compounds where the left constituent is followed by *-s-*, *-e(n)-*, or *-∅-* respectively, and the total number of compounds with the given left constituent (henceforth, the left family size). To give an example, the Dutch noun *mededeling* "announcement, notice" appears in CELEX as the left constituent in one compound with the interfix *-s-*, *mededeling-s-plicht* "reporting duty", and in two compounds with the interfix *-en-*, *mededeling-en-blad* "newsletter, bulletin", and *mededeling-en-bord* "notice board". The type-based bias of this left constituent family towards the interfix *-s-* is $1/(1+2) = 0.33$. The bias of the interfix *-e(n)-* has the value of $2/(1+2) = 0.66$, while the bias of the zero-interfix is equal to zero. The measures of bias are labeled *TypeSBias*, *TypeEnBias* and *TypeZeroBias*.

Alternative, token-based, estimates of the bias are defined in terms of the frequencies of occurrence, rather than the type count of the compounds. These token-based measures are outperformed in our analyses by the type-based ones and are not reported here. Furthermore, we only consider left constituent families, since the effect of the right bias is reported as either weak or absent (Krott et al., 2002; Krott, Hagoort, & Baayen, 2004).

The predictivity of constituent families for the duration of the interfix may extend beyond the bias measures, which only estimate the ratio of variants in the constituent family, without taking the magnitude (size, frequency, or information load) of the constituent family into account. However, these magnitudes are expected to exhibit effects in our analysis, since they repeatedly emerged as significant predictors in both the comprehension and production of Dutch compounds (e.g., Bien, Levelt & Baayen, 2005; De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002; Krott et al., 2004). To estimate the magnitude of constituent families, we incorporate in our study position-specific measures of entropy proposed by Moscoso del Prado Martín, Kostić, and Baayen (2004). These measures employ the concept of Shannon's entropy (Shannon & Weaver, 1949), which estimates the average amount of information in a system on the basis of the probability distribution of the members of that system. The probability of each member (p_{sys}) is approximated as the frequency of that member divided by the sum of the frequencies of all members. The entropy of a system with *n* members is then the negative weighted sum of log-transformed (base 2) probabilities of individual members:

$$H = -\sum_{i=1}^{n} p_{sys} * \log_2 p_{sys}$$

Note that the entropy increases when the number of paradigm members is high (i.e. family size is large) and/or when the members are equiprobable.

Let us consider the positional entropy measure of the left constituent family of the Dutch noun *mededelingsplicht*. This family consists of three members: *mededelingsplicht* has a lemma frequency of 3, *mededelingenblad* has a lemma frequency of 7, and *mededelingenbord* has a lemma frequency of 15 in the CELEX lexical database, which is based on a corpus of 42 million word forms. The cumulate frequency of this family is $3+5+15 = 23$, and the relative frequencies of these three family members are $3/23 = 0.13$ for *mededelingsplicht*, $7/23 = 0.30$ for *mededelingenblad* and $15/23 = 0.65$ for *mededelingenbord*. The left positional entropy of this constituent family therefore equals $-(0.13 * \log_2 0.13\ + 0.30 * \log_2 0.30\ + 0.65 * \log_2 0.65)\ = 1.30$ bit.

We consider the positional entropy measures for both the left and the right constituent families, henceforth *LeftPositionalEntropy* and *RightPositionalEntropy* as potential predictors of the acoustic duration of the interfix. The informativeness of the right constituent family is meaningful as a measure of the cost of planning the right constituent: Planning upcoming elements with a low information load has been shown in Chapter 4 to predict reduction in the fine phonetic detail of the currently produced elements.

Other variables

Since acoustic duration is known to depend on a wide range of factors, we used stepwise multiple regression to bring these factors under statistical control. Two sets of factors were considered: Lexical frequency-based probabilities, and phonetic, phonological and sociolinguistic variables.

Probabilistic factors

Phrasal level

A higher likelihood of a word given its neighboring words has been shown to correlate with vowel reduction, segmental deletion, and durational shortening (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Jurafsky et al., 2001; Chapter 4 of this thesis). To quantify this likelihood, for each compound token in our data we calculated its mutual information with the preceding and the following word (*BackMutualInfo, FwdMutualInfo*) by using the following equation (X and Y denote either the previous word and the compound, or they denote the compound and the following word; XY denotes the combination of the two words):

$$MI(X;Y) = log\, \frac{(Frequency(XY))}{(Frequency(X)) * (Frequency(Y))}$$

The measures were computed on the basis of the Spoken Dutch Corpus, which contains 9 million word tokens. All frequency measures were (natural) log-transformed. Obviously, the values could not be computed for the instances where the target word was utterance-initial or utterance-final.

For those words for which mutual information with the preceding or the following word could be computed, we checked whether it was a significant predictor of the duration of the interfix over and beyond other factors. Neither *BackMutualInfo* nor *FwdMutualInfo* reached significance in our datasets. This result may originate in the properties of the datasets, which comprise relatively low-frequency compounds. Obviously, these low-frequency compounds have even lower frequencies of co-occurrence with their neighboring words. For instance, for the *s*-dataset the average frequency of co-occurrence of the compounds with the preceding word is a mere 1.63 ($SD = 0.77$), and with the following word a mere 1.20 ($SD = 0.30$). Another explanation may be that effects of contextual predictability do not extend to phonemes in the middle of long compounds. They may only emerge for segments at word boundaries (e.g., Jurafsky et al., 2001; Chapter 4 of this thesis).

Word level

The lexical frequency of a word is known to codetermine articulation (e.g., Jurafsky et al., 2001; Zipf, 1929; Chapters 2 and 3 of this thesis), and therefore we include the natural log-transformed compound frequency (*WordFrequency*) as a control variable in the analyses. Together with the measure of the bias and the left positional entropy, this variable forms a cluster of predictors that capture different aspects of the same phenomenon. The measure of the bias estimates the *proportion* of the positional family of compounds that supports the interfix. The corresponding entropy estimates the number and average information load of the members in this family, i.e., it gauges the reliability of the knowledge base for the bias. Finally, a high compound frequency quantifies the evidence for the co-occurrence of the left and right constituents with the interfix. We expect these variables to behave similarly in predicting the durational characteristics of the interfix.

Segmental level

Another dimension of predictability for segmental duration is the amount of lexical information in an individual segment given the preceding fragment of the word (i.e., given the "word onset"). Following Van Son and Pols (2003), we define an information-theoretic measure that quantifies segmental lexical information (*TokenSegmentalInfo*):

$$I_L = -\log_2 \frac{Frequency((word\,onset)+target\,segment)}{Frequency((word\,onset)+any\,segment)}$$

Van Son and Pols (2003) interpret this measure as estimating the segment's incremental contribution to word recognition. The occurrence of a segment that is improbable given the preceding fragment of the word limits the cohort of matching words substantially and thus facilitates recognition. To give an example, the amount of lexical information of the segment [s] given the preceding English word fragment [kaʊ] is calculated as the negative log-transformed ratio of the cumulate frequency of words that begin with the string [kaʊs] (e.g., *cowskin, cowslip, cowslips*) and the cumulate frequency of the words that begin with the string [kaʊ] plus any segment (e.g., *cows, cowpat, cowshed, cowskin, cowslip, cowslips, etc.*). In the present study, segmental lexical information measures are based on the frequencies of single words, such as made available in CELEX, and do not account for combinations of words, even if those may acoustically be valid matches for the phonetic string. For instance, the combination *cow stopped* is not included in the calculation of the lexical information for the segment [s] in the string [kaʊs].

A positive correlation of this token-based segmental lexical information and segmental duration was reported in Van Son and Pols (2003): They observed r-values in the range of 0.0 to 0.2 for different classes of phonemes grouped by manner of articulation. If segmental lexical information indeed modulates fine phonetic detail, it is a potential predictor of the salience of the interfix.

To this token-based measure of segmental lexical information (*TokenSegmentalInfo*), we add a type-based measure, *TypeSegmentalInfo*, which is based on the *number* of words matching the relevant strings, rather than their cumulated frequencies:

$$S_L = -\log_2 \frac{Number((word\,onset)+target\,segment)}{Number((word\,onset)+any\,segment)}$$

We validated both the token-based and the type-based measures of segmental lexical information against our own dataset to establish how the performance of the type-based estimate S_L compares with that of the token-based measure I_L. Our approach differs from that of Van Son and Pols (2003) in that it considers the divergence of phonemes from their mean durations, rather than the raw durations of these phonemes. Different phonemes, even those that share manner of articulation, intrinsically differ in their durations. Therefore, pooling the durations of large classes of phonemes introduces unnecessary noise in the correlation analyses. We gauged the divergence of each instantiation of every phoneme from the mean duration of this phoneme and tested whether this divergence can be explained by the amount of lexical information carried by the phoneme. Our survey is based on *all* segments in the *s*-dataset and in the compounds of the *en*-dataset in which the interfix is realized as [ə].

We collected the data on durations from the Read Text component of the IFA corpus, a hand-aligned phonemically segmented speech database of Dutch (Van Son, Binnenpoorte, Van den Heuvel, & Pols, 2001). We log-transformed the durations and computed the means and standard deviations for each phoneme. Then, moving phoneme by phoneme through our compound dataset we calculated the z-score for each phoneme, that is, the difference between its actual log-transformed duration and its mean log duration, in units of standard deviations from the mean. The correlation between the observed durational difference and the corresponding amount of type-based segmental lexical information yields an *r*-value of 0.06 ($t(17694) = 7.41, p < 0.0001$). This order of magnitude is comparable with the results that Van Son and Pols (2003) obtained for the token-based measure of lexical information. As the type-based measure influences durations of segments across the dataset, we decided to include it in our analyses of the interfix durations. Thus, we take as control variable the value of *TypeSegmentalInfo* for the (first) segment of the interfix.

Importantly, the durations show a weaker correlation with the token-based segmental lexical information, proposed by Van Son and Pols (2003), than for its type-based counterpart ($r = 0.03, t(17694) = 4.25, p < 0.0001$). This measure also performs worse in the models reported below. Since the token- and type-based measures are highly correlated, we incorporated only *TypeSegmentalInfo* in our analysis.

Phonetic, phonological and sociolinguistic variables

Speech rate is an obvious predictor of acoustic duration (e.g., Crystal & House, 1990; Fosler-Lussier & Morgan, 1999). Two different measures estimating speech rate were included as control variables. First, we defined an utterance-based rate of speech, *SpeechRate*, as the number of syllables in the utterance divided by the acoustic duration of the utterance. Utterance is defined here as the longest stretch of speech containing the compound and not containing an audible pause.

Second, we defined a more local speech rate for the interfix *-s-*. In the *s*-dataset, the interfix *-s-* always belongs to the coda of the preceding syllable. We measured the average segmental duration in the interfix-carrying syllable minus the *-s-* interfix, and considered it as an estimate of the local speed of articulation in the part of the syllable that precedes the interfix *-s-*, henceforth *SyllableSpeed*. The syllable from which the final segment [s] was subtracted is structurally complete, with an onset, a vowel and (in 83% of tokens) a coda of one or more consonants. Note that for words with the interfix *-e(n)-* this measure of local speech rate is not meaningful. It would subtract the complete rhyme of the relevant syllable, leaving only the onset, the duration of which is above all determined by the number and types of its consonants.

Nooteboom (1972) observed that segments are shorter the greater the number of syllables or segments in the word. We considered the total number of segments in the word, *NumberSegments*, and the number of segments following the interfix, *AfterSegments*. We also took into account the sex, age and language variety of the speaker (Keune, Ernestus, Van Hout, & Baayen, 2005). The binary variable *SpeakerLanguage* encodes the speaker's variant as Southern Dutch or Northern Dutch. If the information about age was missing, we filled in the average age of our speakers' population.

Prosody may affect the duration of segments as well. For instance, words at the beginning and the end of utterances show articulatory strengthening (e.g., Bell et al., 2003; Cambier-Langeveld, 2000; Fougeron & Keating, 1997). To control for the word's position in the utterance, we coded each token with two binary variables *UtteranceInitial* and *UtteranceFinal*.

Furthermore, stressed syllables are pronounced longer than unstressed ones (e.g., Ladefoged, 1982). We coded each compound with the interfix *-s-* for whether its interfix-containing syllable carries a (primary or secondary) stress (the binary variable *Stressed*).

The interfix *-e(n)-* is never stressed. The common stress pattern for compounds with the interfix *-e(n)-* is for the primary stress to fall on the syllable immediately preceding the interfix-containing syllable, and the secondary stress on the syllable immediately following the interfix-containing syllable: The insertion of *-e(n)-* prevents a stress clash between constituents. The rhythmic structure of compounds has been proposed as a factor codetermining the selection of the interfix, in addition to lexical constituent families and several other factors (Neijt, Krebbers, & Fikkert, 2003). To test the acoustic consequences of the rhythmic pattern, we coded each compound in the *en*-dataset as to whether the interfix syllable intervenes between two immediately adjacent stressed syllables (the binary variable *Clash*).

Compounds with the interfix *-e(n)-* were coded for the presence or absence of [n] in the acoustic realization of the interfix (*NPresent*), as established by two phoneticians (see Method). Similarly, compounds with the interfix *-s-* were coded for whether the interfix was realized as [z], resulting in the variable *PhonemeZ*.

Finally, the immediate phonetic environment can make a segment more or less prone to reduction. Unstressed vowels in Dutch tend to lengthen before oral stops (Waals, 1999). Therefore, each compound in the dataset with the *-e(n)-* interfix was coded for the manner of articulation of the following segment (binary variable *FollowedbyStop*).

Results

Results for the interfix -s-

The dataset for the interfix *-s-* included 1155 tokens. The number of different word types was 680, and their token frequencies followed a Zipfian distribution ranging from 1 to 19. We fitted a stepwise multiple regression model with the acoustic duration of the interfix as the dependent variable. The values of this variable were (natural) log-transformed to remove skewness of the distribution. The resulting variable *InterfixDuration* has a mean of 4.37 of log units of duration ($SD = 0.35$). We identified 21 data points that fell outside the range of -2.5 to 2.5 units of SD of the residual error, or had Cook's distances exceeding 0.2. These outliers were removed from the dataset and the model was refitted. Below we only report variables that reached significance in the final model.

The strength of the bias for the *-s-* interfix, *TypeSBias*, emerged as a main effect with a positive slope. Surprisingly, the duration of *-s-* was longer for compounds with a greater bias for this interfix ($\hat{\beta} = 0.35, t(1125) = 5.20, p < 0.0001$). A positive correlation with duration was present for the predictor *RightPositionalEntropy* as well ($\hat{\beta} = 0.07, t(1125) = 4.10, p < 0.0001$), indicating that the duration of the interfix increases with the informational complexity of the right constituent. These main effects were modulated by an interaction between *TypeSBias* and *RightPositionalEntropy* ($\hat{\beta} = -0.07, t(1125) = -3.67, p = 0.0003$). Inspection of conditioning plots revealed that the influence of the bias measure was greater when the value of the right positional entropy was low. In addition, *WordFrequency* had an unexpected positive slope that just failed to reach significance: ($\hat{\beta} = 0.01, t(1125) = 1.95, p = 0.0510$). We found no effect of *LeftPositionalEntropy*.

Importantly, the lexical segmental information of the interfix was predictive in the expected direction: Segments conveying more information tended to be longer (*TypeSegmentalInfo*: $\hat{\beta} = 0.12, t(1125) = 3.86, p < 0.0001$).

Among the phonological and phonetic variables, the measure of the speech rate also demonstrated the expected behavior. The greater the local speed of articulation, the shorter the realization of this interfix (*SyllableSpeed*: $\hat{\beta} = -0.51, t(1125) = -5.27, p < 0.0001$). Whether the interfix-carrying syllable was stressed was a significant predictor as well, with stress predicting durational shortening of the interfix (*Stressed*: $\hat{\beta} = -0.09, t(1125) = -3.96, p < 0.0001$). Finally, interfixes realized as [z] were shorter than those realized as [s], as expected given the findings by, for instance, Slis and Cohen (1969) (*PhonemeZ*: $\hat{\beta} = -0.16, t(1125) = -3.17, p = 0.0016$).

All significant predictors were tested for possible non-linearities; none reached significance. The bootstrap validated R^2 of the model was 0.104. The unique contribution of the morpholexical factors *TypeSBias*, *PositionalEntropyRight*, and *WordFrequency* to the explained variance over and above the other predictors was 2.0%, as indicated by the drop in R^2 when these variables were removed from the model.

Discussion of the results for -s-

Three related morpholexical variables emerge as significant predictors of the durational lenghtening of the interfix: *TypeSBias, RightPositionalEntropy* and (marginally) *WordFrequency*. The positive correlations of *TypeSBias* and *WordFrequency* with the duration of the interfix lead to the paradoxical and counterintuitive conclusion that a greater likelihood for a linguistic unit may lead to a longer acoustic realization of that unit, contradicting the information-theoretical approach to the distribution of acoustic salience. We will address this issue in the General Discussion.

The interaction of the right positional entropy with the bias hints at planning processes at work. In Chapter 4, we argued that the planning of upcoming linguistic elements may interfere with the planning and production of preceding elements. We interpret the right positional entropy measure as tapping into the costs of planning the right constituent. The observed interaction indicates that the bias allows greater durational lengthening of the interfix when planning the next constituent is easy.

In accordance with previous reports (e.g., Van Son & Pols, 2003), a high amount of lexical information carried by an individual segment (*TypeSegmentalInfo*) predicts the acoustic lengthening of this segment. In other words, segments with a larger contribution to the word's discriminability are produced with increased articulatory effort, and hence prolonged duration. This highlights the paradox with which we are confronted: Conventional measures, such as the segmental lexical information, behave as expected, while measures for the likelihood of the interfix exhibit exceptional behavior.

The position of the compound in the utterance did not affect the durational characteristics of the interfix significantly, which is in line with observations by Cambier-Langeveld (2000). Cambier-Langeveld argues that final lengthening in Dutch only applies to the last syllable in the word or, if the vowel in this last syllable is [ə], to the penultimate syllable. Thus, the interfix lies beyond the scope of this effect. Similarly, the interfix emerges as outside the domain of influence of initial lengthening.

Results for the interfix -e(n)-

The *en*-dataset contained 742 tokens of compounds. The number of different word types equalled 305, and the Zipfian distribution of tokens per type ranged from 1 to 74. We log-transformed the acoustic durations of the interfixes, which then had a mean of 4.065 log units of duration ($SD = 0.420$). We fitted a stepwise multiple regression model to these durations. This time, 19 data points fell outside the range of -2.5 to 2.5 units of SD of the residual error or had Cook's distances exceeding 0.2. These outliers were removed from the dataset, and the model was refitted. Only predictors that reached significance are reported.

The morpholexical predictors performed as follows: A higher bias for the interfix *-e(n)-*, *TypeEnBias*, correlated with longer interfixes: ($\hat{\beta} = 0.14, t(716) = 5.39, p < 0.0001$). The positional entropy of the right constituent family also had a positive main effect ($\hat{\beta} = 0.08, t(716) = 4.56, p < 0.0001$). The interaction of these two variables was not significant ($p > 0.4$). *LeftPositionalEntropy* and *WordFrequency* did not reach significance either ($p > 0.1$).

As in the model for the interfix *-s-*, a higher amount of lexical information, as attested by *TypeSegmentalInfo* for the first segment of the interfix, correlated with longer articulation ($\hat{\beta} = 0.07, t(716) = 3.09, p = 0.002$). This effect is again in line with predictions of the information-theoretical approach.

The interfixes of 226 tokens (29%) in the dataset were realized as [ən], while 561 tokens were pronounced as [ə]. As expected, the presence of [n] in the interfix implied a substantial increase in the total duration of the interfix. The factor *NPresent* was the most influential predictor ($\hat{\beta} = 0.71, t(716) = 37.80, p < 0.0001$), and its unique contribution to the explained variance of this duration was 55%.

Two phonetic factors contributed to the duration of the interfix. Unsurprisingly, the interfix was shorter when the utterance-based speech rate was higher: (*SpeechRate*: $\hat{\beta} = -0.04, t(716) = -4.17, p < 0.0001$). Factor *FollowedbyStop* also had an effect ($\hat{\beta} = 0.23, t(716) = 13.10, p < 0.0001$), which supports the observation by Waals (1999) that an unstressed vowel is pronounced longer before oral stops. It is noteworthy that Waals' observation, which was made under thoroughly controlled laboratory conditions, is replicated here in more natural read aloud speech.

All significant predictors in the model were checked for non-linearities, none of which reached significance. The bootstrap validated R^2 value for the model was 0.72. The unique contribution of the morphological predictors *TypeEnBias* and *RightPositionalEntropy* to the variance explained by the model was 2.3%, as indicated by the drop in R^2 after the removal of these variables from the model. This contribution is close to that provided by the morpholexical predictors in the *s*-dataset (2.0%).

Discussion of the results for -e(n)-

The analysis of the *en*-dataset replicates the unexpected direction of influence of the morphologically-determined redundancy that we reported for the dataset with the interfix *-s-*. We found again that higher values for the bias estimates correlate with a longer duration of the interfix. We will return to this role of the bias in the General Discussion.

The positive simple main effect of the right positional entropy supports the hypothesis of continuous planning of articulation, according to which the planning complexity of upcoming elements may modulate acoustic characteristics of preceding elements.

Given the dominant contribution of the variable *NPresent* to the explained variance, we set out to establish what factors affected the selection of the variant [ən] versus [ə]. The interfix *-e(n)-* is spelled as either *-e-* or *-en-*, depending on orthographic rules. Compounds spelled just with *-e-* are unlikely to be pronounced with [ən]. The subset of compounds spelled with *-en-* contains 653 tokens. We fitted a logistic regression model that predicted the log odds of the selection of [ən] versus [ə] in this subset. The model uses the binomial link function and considers the presence of [n] in the realization of the interfix as a success, and the absence as a failure. The results demonstrate no effect of *TypeEnBias* on the selection of the phonetic variant ($p > 0.5$). Apparently the realization of an extra phoneme in the interfix is independent of the morphological likelihood of the interfix. The presence of [n] was more likely when *WordFrequency* was high ($\hat{\beta} = 0.63, p < 0.0001$), *RightPositionalEntropy* was high ($\hat{\beta} = 2.11, p < 0.0001$), the speaker's language was Southern Dutch ($\hat{\beta} = 1.37, p < 0.0001$), the number of segments after the interfix was high ($\hat{\beta} = 2.06, p < 0.0001$), and a stress clash was attenuated ($\hat{\beta} = 4.19, p < 0.001$). The likelihood of [n] was lower when *LeftPositionalEntropy* was high ($\hat{\beta} = -0.60, p < 0.0001$).

In a second supplementary analysis, we investigated whether morpholexical factors are better predictors for acoustic duration if we consider the duration of [ə] as the dependent variable, rather than the duration of the whole interfix. In such a model, we expect the presence of [n] to exercise less influence and the morpholexical predictors to have greater explanatory value than in the model for the duration of the interfix as a whole. We fitted a stepwise multiple regression model to the data with the (natural) log-transformed acoustic duration of the phoneme [ə] in the interfix as the dependent variable. After removal of 25 outliers, the model was refitted against the remaining 717 datapoints. In line with our expectations, we observe a decrease in the predictive power of *NPresent* to only 15% of the explained variance, while the share of morphological variables *TypeEnBias* and *RightPositionalEntropy*, which retain significance as predictors of acoustic lenghtening, increases to 4.3% of the explained variance. We conclude that morphological structure codetermines the acoustic characteristics of the interfix *-e(n)-* over and beyond major phonological and phonetic predictors.

General Discussion

According to the information-theoretical approach to acoustic salience, a higher likelihood of a linguistic unit is correlated with more acoustic reduction. The main finding of this study is that the effect of morphologically-determined probability on the duration of interfixes in Dutch compounds runs counter to this prediction. This pattern of results is especially puzzling, since our data also provide evidence *in favor of* the information-theoretical approach in the form of an effect of segmental lexical information. Thus, we do find that a higher probability of a segment given the preceding word fragment leads to more acoustic reduction.

What may be the solution for the problem that the present data appear to pose for the information-theoretical framework? One explanation might be that morphological information has a fundamentally different status from other types of linguistic information, and is typically associated with careful articulation. However, this line of reasoning is refuted by research on prefixes and suffixes in English (e.g., Hay, 2003) and Dutch (Chapters 2, 3, 4, and 6 of this thesis).

Another solution might refer to the fact that interfixes are homophonous with plural markers in Dutch (cf., *boek-en* ”books” and the compound *boek-en-kast* ”bookshelf”). The frequency of the plural word forms might codetermine the duration of the interfix and be confounded with the bias. This explanation, however, can be discarded on the following grounds. First, there was no consistency in the correlation between the frequency of plural nouns and the bias of the interfix across datasets. For the *s*-dataset the correlation was positive ($r = 0.12, t(1154) = 4.24, p < 0.0001$), while for the *-en*-dataset it was negative ($r = -0.28, t(740) = -8.15, p < 0.0001$). Second, the frequency of the plural homophonous forms did not reach significance when included as a covariate in the regression models for both datasets. Finally, previous work on

German compounds by Koester, Gunter, Wagner, and Friederici (2004) has shown that plural suffixes and interfixes may not be perfectly homophonous in terms of systematic fine phonetic detail: Compound constituents followed by an interfix are shorter and have a higher pitch than their stand-alone plural counterparts.

The hypothesis that we would like to offer as a solution for the present paradox is that fine phonetic detail in speech is governed by two orthogonal dimensions, a syntagmatic dimension and a paradigmatic dimension. The information-theoretical approach that underlies the Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004) and the Probabilistic Reduction Hypothesis (Jurafsky et al., 2001), as well as research on speech efficiency (Van Son & Pols, 2003; Van Son & Van Santen, 2005), views information from the syntagmatic perspective by considering the probability of a linguistic unit in its phonetic, lexical, or syntactic context. These syntagmatic relationships are inherently sequential and govern the temporal distribution of information in the speech stream. For instance, the extent to which a segment contributes to the identification of the word *given the preceding word fragment* (Van Son & Pols, 2003) is a syntagmatic measure that is positively correlated with duration: The greater the contribution of the segment, the longer its acoustic implementation.

The syntagmatic measures proceed upon the premise that there is no (probabilistic) variation in the elements forming the word or the syntactic clause to be realized by the speaker. When the speaker wants to express the concept *book*, there is no doubt that the element following [bʊ] is [k].

However, the identity of the elements is not always known with such certainty. The interfix in Dutch compounds is one such example. We label such elements “pockets of indeterminacy”. Paradigmatic relations, here defined over constituent families, provide the probabilistic basis for resolving this indeterminacy. The bias measures quantify the extent of support provided by paradigmatics for the different interfixes available for selection: A greater support increases the likelihood of a given interfix. Our experimental results indicate that such a greater likelihood is paired with a more salient acoustic realization.

Whereas the syntagmatic dynamics of lexical disambiguation are intrinsically temporal, paradigmatic inference is a-temporal in nature. In the a-temporal domain of paradigmatic inference for positions of choice, a greater probability implies a broader empirical basis for selection of a given alternative, and comes with increased acoustic salience.

Importantly, paradigms as a source of support for alternatives for selection are not restricted to morphological structure: We consider paradigms in a general Saussurean sense, as sets of linguistic elements over which the operation of selection is defined (De Saussure, 1966).

The amount of evidence for the alternatives apparently determines the confidence with which an interfix is selected. That a lack of confidence may lead to a decrease in acoustic salience may be illustrated by an analogy: When producing case endings of German nouns, non-native speakers of German may hush up their realizations if they have doubts about the appropriate morpheme, but articulate the endings carefully and clearly if they are certain about which ending to choose. This example serves as an analogy only, and there is no implication that speakers make deliberate, conscious choices based on the morphological bias. The support measured as the bias is rather an estimate of the ”naturalness” of the association between the available interfixes and the constituents of the compound.

Our hypothesis that paradigmatic inference for pockets of indeterminacy leads to salient acoustic detail, henceforth the Paradigmatic Signal Enhancement Hypothesis, offers straightforward, testable predictions at various levels of linguistic structure. Let us first consider the level of morphology. It is well known that English irregular verbs cluster into sets according to the kind of vocalic alternation that they exhibit in the past tense form (*keep/kept*, *run/ran*). The Paradigmatic Signal Enhancement Hypothesis predicts that a past-tense vowel — a pocket of indeterminacy — is realized with increased acoustic salience when the vocalic alternation is supported by a larger set of irregular verbs. Effects of paradigmatic gangs might even be found for the vowels of regular verbs (Albright and Hayes, 2003).

At the interface of morphology and phonology, we call attention to the phenomenon of final devoicing. In German and Dutch, a stem-final obstruent may alternate between voiced and voiceless, compare Dutch [hɔnt] *hond* (’dog’) with [hɔndə] *honden* (’dogs’). Ernestus and Baayen (2003, 2004) have shown that this alternation, traditionally regarded as idiosyncratic, is affected by paradigmatic structures driven by the rhyme of the final syllable. In addition, they have shown that devoiced obstruents (e.g., the [t] of [hɔnt]) may carry residual traces of voicing, and that listeners are sensitive to these residual traces (Ernestus & Baayen, 2006). The Paradigmatic Signal Enhancement Hypothesis builds on these findings by predicting that greater paradigmatic support for voicing will correlate with enhanced acoustic salience of residual voicing in the devoiced obstruent.

Additional evidence for the Paradigmatic Signal Enhancement Hypothesis emerges from research on intrusive /r/ in New Zealand English (Hay & Maclagan, in press): The more likely speakers are to produce intrusive /r/ given a range of linguistic and sociolinguistic factors, the more salient its realization (as reflected in the degree of constriction).

Our speakers read the compounds and thus received unambiguous visual information about the correct interfix. It is therefore remarkable that we nevertheless observed effects of morpholexical factors on the planning and implementation of speech production. We note, however, that the bias of the interfix as determined by the left constituent family has also been observed to affect the speed of reading comprehension of novel and existing compounds. Paradigmatics plays a role even in reading (Krott et al., 2004), whereas we expect the acoustic consequences to have a larger scope when visual cues to the appropriate morphemes are absent, as in spontaneous speech genres. We conclude that competition in morpho-phonological paradigms is intrinsic to language production and comprehension (Ernestus & Baayen, 2006).

Finally, the probabilistic dependencies between morphemes, such as exist between the interfix, the compound's left and right constituents, and the whole compound, challenge the fully decompositional theory of speech production developed by Levelt, Roelofs, and Meyer (1999). According to this model, an abstract lemma representation provides access to a word's individual constituents. The planning for articulation of these individual constituents is fully encapsulated from all other morphemes and their paradigmatic relations. This model is challenged not only by the present findings, but also by those of Van Son and Pols (2003), Hay (2003), Ernestus, Lahey, Verhees, and Baayen (2006), and Chapters 2, 3, and 4 of this thesis. What the present chapter adds to this literature is the surprising observation that fine phonetic detail is not only determined by the properties of the word itself and its nearest phonological neighbors, but also by its morphological paradigmatic structure.

Summary and Conclusions

CHAPTER 8

This dissertation investigated the roles of several probabilistic variables in the production and comprehension of reduced Dutch affixes. The central hypothesis was that linguistic units with a high probability of occurrence are more likely to be reduced (Jurafsky et al., 2001; Aylett & Turk, 2004). We tested this hypothesis by analyzing the acoustic realizations of affixes, which are meaning-carrying elements embedded in larger lexical units. Most of our results proved to be compatible with the main hypothesis, but some seemed to run counter to its predictions. In this chapter, we review our findings and discuss their implications for models of speech production, models of speech perception, and probability-based accounts of reduction.

The role of word frequency

Chapters 2 and 3 investigated whether the acoustic duration of affixes in Dutch is shorter the higher the frequency of their carrier word. Chapter 2 described a study based on conversational speech from the Corpus of Spoken Dutch. Four affixes were investigated: The verbal prefixes *ont-* and *ver-*, the participial prefix *ge-*, and the adjectival suffix *-lijk*. For three of these four affixes (*ge-*, *ont-*, and *-lijk*), we found that the duration of the affix or one or more of its segments was shorter the higher the frequency of the carrier word. As such, this study was the first to present evidence of frequency effects on meaningful linguistic units. Furthermore, it showed that it is possible to observe such effects in spontaneous, everyday speech.

A drawback of looking at conversational speech data is that it precludes us from drawing conclusions about the psychological process underlying the effect. To gain more insight into this issue, we conducted two production experiments, which were presented in Chapter 3. In the first experiment, we observed an effect of word frequency on affix duration in a word naming task. Interestingly, this effect was equally strong for all four affixes studied, which were identical to those investigated in Chapter 2. However, we needed to exclude the possibility that the observed frequency effect was related to perception rather than production (cf. Balota & Abrams, 1995). Therefore, we ran a control experiment in which the stimulus word was not identical to the response word. The frequencies of both words were manipulated independently from each other. The response word always contained the suffix *-lijk*. Only the frequency of the response word affected the duration of this suffix. This shows that the word frequency effect is indeed related to the speech production process, and not just due to a general tendency to produce shorter responses to high-frequency stimuli.

These findings have several implications for models of speech production. First of all, they show that the effect of word frequency in speech production is not limited to the speed of word-form access, as is generally assumed (e.g., Levelt, Roelofs, & Meyer, 1999). Frequency also plays a role during later stages of processing, including the articulation phase. To account for our results, production models need to allow cascading of information, so that information about a word's frequency of occurrence can affect the fine details of its acoustic realization.

Another problematic aspect of speech production models is that the syllable is generally assumed to be the basic unit of articulation (e.g., Levelt & Wheeldon, 1994; Cholin, 2004). In such a model architecture, it is difficult to see how the durational characteristics of identical syllables (all our affixes were also syllables) can differ as a function of the frequency of the word they occur in. Furthermore, we observed in Chapter 2 that different segments belonging to the same syllable could be subject to different, sometimes even contradictory forces. Such differences would not be expected if syllables functioned as psychologically encapsulated articulatory units.

Finally, our results provide evidence against the implicit assumption that frequency effects on articulatory reduction are solely due to the speaker actively taking the listener's knowledge and needs into account. After all, we observed such effects in experimental circumstances where there was no listener present. This suggests that at least part of the effect is purely speaker-internal, in that it arises as a non-consciously controlled by-product of the speech production process.

The role of contextual predictability

In Chapters 4 and 5 of this thesis, the focus was on the predictability of words in their context. Chapter 4 investigated the role of contextual predictability in production by means of a corpus survey of spontaneous speech. For each of the seven most frequent words ending in the suffix *-lijk*, we randomly selected 40 occurrences. Three sources of predictability were explored: The number of times a word had already been mentioned before, the predictability of a word given the previous word, and the predictability of a word given the following word. All three variables showed significant effects on reduction, although they differed in which words and which parts of the words they affected. The number of previous mentions of a word only affected the duration of the suffix. The suffix was shorter the more often the word had already been mentioned in the preceding discourse. Predictability from the previous word only showed an effect on the degree of reduction in the stem, and this effect was limited to just two of the seven words under investigation: *natuurlijk* ‘naturally’ and *eigenlijk* ‘actually’. For these two words, a higher predictability from the previous word was correlated with more durational and segmental reduction of the stem. The most robust effects, however, were observed for predictability from the following word, which affected both the duration of the suffix and the number of segments in the stem. Again, a higher predictability led to shorter acoustic realizations.

Given that predictability from the following word is such a robust predictor of reduction in production, we formulated the hypothesis that reduced words will be recognized faster the more predictable they are given the following word. This hypothesis was tested in Chapter 5. In a perception experiment, we presented subjects with four-word utterances which differed in how well the reduced third word could be predicted from the fourth word. The subjects were asked to press a button as soon as they had recognized the third word. The reaction time for this button press was shorter the higher the predictability of the third word from the fourth word. To gain more control over the acoustic properties of predictable and unpredictable target words, we ran a second experiment using cross-spliced stimuli in which the first three words of an utterance were identical across conditions. Again, subjects recognized reduced targets faster the more predictable they were given the following word. Thus, we can conclude that predictability from the following word plays a role in the recognition of reduced word forms. In determining the identity of a reduced word, listeners take the probability of words given their immediate lexical context into account.

The results in Chapters 4 and 5 shed further light on the structure of the mental lexicon. It appears that words that are mutually predictable can increase each other’s activation levels. In Chapter 4, we found that a word is more reduced the more predictable it is given the following word. According to Balota, Boland, and Shields (1989), more articulatory reduction implies a higher activation level. Hence, a word’s activation level can be increased as a consequence of the planning of an upcoming word with which it frequently co-occurs. In Chapter 5, we observed that processing (or having processed) a following word can speed up the recognition of a preceding word that has been reduced. This is only possible if these two words are somehow linked in the lexicon, for instance by means of a so-called *superlemma* (Sprenger, 2003).

An additional implication of the findings in Chapter 5 is that the speech perception system needs to be able to temporarily store (ambiguous) acoustic information. Otherwise, words could only be processed in the order in which they were pronounced, which in our experiments was clearly not the case. An example of a model that contains such a temporary storage facility is ARTWORD, developed by Grossberg and Myers (2000).

The role of morphology

As remarked in Chapter 1, our focus on affixes allowed us to explore new predictors of articulatory reduction, namely those related to morphology. This was the main objective in Chapters 6 and 7, both of which studied read-aloud speech. In Chapter 6, we investigated whether the acoustic realization of the suffix *-igheid* (/əxhɛit/) is affected by the morphological structure of the word in which it occurs. Basically, *-igheid* occurs in two types of words: Words in which there is only a morphological boundary between *-ig* and *-heid*, such as *zuinigheid* ‘thriftiness’ (*zuin* is not a word in Dutch, while *zuinig* is), and words in which *-igheid* forms a single suffix, such as *vastigheid* ‘security’ (*vastig* is not a word in Dutch, while *vast* is). Two hypotheses were formulated which made conflicting predictions about the realization of the /xh/-cluster in these two word types. According to the Prosodic Structure Hypothesis, the cluster would be more simplified in words in which *-igheid* is analyzed as a single suffix, as the cluster does not span a prosodic boundary there. The Morphological Informativeness Hypothesis, on the other hand, predicted that the cluster would be less simplified in such words, as the informativeness of the suffix given the other words in the morphological paradigm (e.g., *vast*, *vaste*, *vaster*, *vasthouden*, etc.) is relatively high. This second prediction turned out to be correct: The duration of the cluster was longer in words like *vastigheid* than in words like *zuinigheid*. As such, Chapter 6 provides additional evidence for the relationship between the amount of information carried by a linguistic unit and the amount of effort spent on its articulation. Furthermore, it shows that acoustic realizations can be affected by morphological structure (more specifically, paradigmatic structure) without mediation by phonology. This finding provides a considerable challenge for linguistic theory, which generally assumes that there is no direct link between morphology and phonetics.

The research described in Chapter 7 presented an apparent paradox for the theoretical principle of “less information, less articulatory effort”, which was outlined in Chapter 1 of this thesis and worked so well in explaining the data in the other chapters. We observed that the duration of interfixes in Dutch is longer the more predictable these interfixes are given the first constituent of the word. This surprising result can be explained by making a distinction between predictability from a syntagmatic perspective, which is negatively correlated with acoustic salience, and the amount of paradigmatic support for one of a small number of alternatives, which appears to be positively correlated with acoustic salience. This account, which we labelled the Paradigmatic Signal Enhancement Hypothesis, makes testable predictions about the acoustic realizations of other linguistic units that have paradigmatic alternatives.

Topics for further research

It has become clear in the last seven chapters that derivational affixes are fruitful objects for investigation, and that detailed analysis of their acoustic realizations can lead to many new insights. This suggests that it may also be worthwhile to study comparable linguistic units like inflectional morphemes and clitics. Furthermore, it would be useful to replicate the findings of the current thesis in languages other than Dutch, as this would provide more insight into the generalizability of our results and the psychological mechanisms underlying them.

The current thesis mainly investigated the circumstances in which speakers reduce articulatory effort. However, we found in Chapter 5 that at least one of the patterns observed in production is mirrored in perception: Reduced words are recognized faster the more predictable they are given the following word. Listeners could show similar sensitivity to the other production patterns we observed, given that all these patterns are systematically present in the acoustic signal. Whether knowledge of these patterns is actually used during speech perception remains to be investigated in perception experiments.

Finally, we would like to point out that the research described in this thesis was quantitatively oriented, in that relatively large amounts of data were gathered and analyzed. However, this is by no means the only way to gain insight into reduction patterns. In recent studies like Curl (2005) and Plug (2005), the acoustic realizations of a small number of words or phrases were studied as a function of their communicative purpose. Such detailed, qualitative analyses can uncover reduction patterns that may be difficult to extract from large-scale data files. In order to gain a full understanding of the circumstances in which speakers reduce articulatory effort, both types of research need to be combined and integrated.

Concluding remarks

The ultimate aim of the research described in this thesis is to provide data for improving psycholinguistic models of speech production and perception, so that they can account for and cope with the enormous phonetic variation present in the speech signal. It has become clear that most existing theories are not particularly well-equipped to do so. In a production model like that of Levelt et al. (1999), the psychological encapsulation of processes and units precludes higher-level information from affecting articulatory processes. In perception models, on the other hand, the early transformation of the signal into discrete, phoneme-like units does not do justice to the important role played by fine phonetic detail (e.g., Hawkins, 2003). Importantly, more and more researchers are beginning to realize that human speech is far too complex to be explained by theories based on abstract symbols and encapsulated combinatorial rules. This can for instance be seen in probabilistic approaches to linguistics (e.g., Bod, Hay, & Jannedy, 2003), paradigmatic approaches to morphology (e.g., Blevins, 2003), and “phonetically-rich” approaches to speech perception (e.g., Hawkins & Smith, 2001). These promising approaches served as an inspiration during the writing of this thesis, and I sincerely hope that my research can contribute to their further development.

Samenvatting

CHAPTER 9

In alledaags taalgebruik worden woorden vaak korter uitgesproken dan men op basis van hun woordenboekuitspraak zou verwachten. Dit proefschrift onderzocht onder welke omstandigheden sprekers woorden reduceren, en hoe luisteraars met deze gereduceerde woorden omgaan. De voornaamste hypothese hierbij was dat woorden meer gereduceerd zullen worden naarmate ze beter voorspelbaar zijn (Jurafsky, Bell, Gregory, & Raymond, 2001; Aylett & Turk, 2004). Drie bronnen van voorspelbaarheid stonden centraal: de frequentie waarmee een woord voorkomt in de taal, de voorspelbaarheid van een woord gegeven de linguïstische context en voorspelbaarheid die samenhangt met morfologische structuur. Voor elk van deze bronnen werd bekeken hoe zij de akoestische realisatie van affixen beïnvloeden. De belangrijkste bevindingen worden hieronder samengevat.

De rol van woordfrequentie

In de Hoofdstukken 2 en 3 werd onderzocht of de akoestische duur van affixen in het Nederlands korter is naarmate het woord waarvan ze deel uitmaken vaker voorkomt. Hoofdstuk 2 beschrijft een corpusonderzoek gebaseerd op spontane spraak uit het Corpus Gesproken Nederlands. Vier affixen werden onderzocht: de verbale prefixen *ont-* en *ver-*, het participiële prefix *ge-* en het adjectivale suffix *-lijk*. Voor drie van deze vier affixen (*ont-*, *ge-* en *-lijk*) werd gevonden dat de duur van het affix of van één of meerdere segmenten in het affix korter was naarmate de frequentie van het draagwoord hoger was. Als zodanig leverde deze studie als eerste bewijs voor het bestaan van frequentie-effecten op de duur van betekenisvolle eenheden. Daarnaast toonde het onderzoek in Hoofdstuk 2 aan dat dergelijke effecten geobserveerd kunnen worden in spontane spraak.

Een nadeel van het onderzoeken van spontane spraak is dat het weinig informatie oplevert over het psychologisch proces dat aan een bepaald effect ten grondslag ligt. In Hoofdstuk 3 werden twee productie-experimenten beschreven die als doel hadden om meer inzicht te verkrijgen in dit proces. In het eerste experiment, waarin gebruik werd gemaakt van een naming-taak, werd een effect gevonden van woordfrequentie op de duur van affixen.

Opvallend was dat dit effect even groot was voor alle vier de affixen die werden onderzocht (dit waren dezelfde affixen als in Hoofdstuk 2). Voordat uit dit resultaat conclusies konden worden getrokken, moest eerst nog de mogelijkheid worden uitgesloten dat het geobserveerde effect eigenlijk een effect was van stimulusfrequentie, een variabele die gerelateerd is aan perceptie (zie verder Balota & Abrams, 1995). Hiertoe werd een controle-experiment uitgevoerd, waarin de gebruiksfrequenties van het woord dat proefpersonen zagen en het woord dat proefpersonen moesten zeggen onafhankelijk van elkaar werden gemanipuleerd. Uit de resultaten bleek dat het suffix *-lijk*, dat voorkwam in alle woorden die proefpersonen moesten zeggen, korter was naarmate het draagwoord frequenter was. De frequentie van het woord dat proefpersonen zagen had geen enkel effect. Dit laat zien dat het woordfrequentie-effect op akoestische duren inderdaad gerelateerd is aan het spraakproductieproces, en niet alleen kan worden toegeschreven aan de neiging van proefpersonen om bij hoogfrequente stimuli kortere responsen te genereren.

De bevindingen in de Hoofdstukken 2 en 3 hebben enkele belangrijke implicaties voor modellen van spraakproductie. Om te beginnen laten ze zien dat de effecten van woordfrequentie, in tegenstelling tot wat standaard wordt aangenomen (bv. Levelt, Roelofs, & Meyer, 1999), niet beperkt zijn tot die fase in het proces waarin de vormeigenschappen van een woord worden opgehaald. Frequentie speelt ook een rol in latere fases van het proces, inclusief de articulatiefase. Om onze resultaten te kunnen verklaren, moeten modellen van spraakproductie parallelle informatieverwerking toestaan, zodat informatie over de frequentie van een woord een rol kan spelen in de totstandkoming van de precieze akoestische realisatie.

Een ander aspect van spraakproductiemodellen dat problematisch is gezien onze resultaten is de aanname dat de syllabe de basiseenheid van articulatie is (bv. Levelt & Wheeldon, 1994; Cholin, 2004). In een dergelijke modelarchitectuur is moeilijk te verklaren waarom de duren van identieke syllaben (al onze affixen waren ook syllaben) verschillen als een functie van de frequentie van het woord waar ze deel van uitmaken. Daarnaast vonden we in Hoofdstuk 2 dat verschillende segmenten in één en dezelfde syllabe onderhevig waren aan verschillende en in sommige gevallen zelfs tegenstrijdige effecten. Zulke verschillen zouden niet worden verwacht als syllaben zouden fungeren als psychologisch geïsoleerde articulatorische eenheden.

Tenslotte laten onze resultaten zien dat effecten van frequentie op articulatorische reductie niet alleen kunnen worden toegeschreven aan een bewuste keuze van de spreker om rekening te houden met de kennis en wensen van de luisteraar. We observeerden immers een frequentie-effect in experimentele omstandigheden waarin geen luisteraar aanwezig was. Dit suggereert dat tenminste een deel van het effect puur spreker-gebaseerd is, dat wil zeggen, dat het optreedt als een niet bewust gecontroleerd bijproduct van het spraakproductieproces.

De rol van contextuele voorspelbaarheid

In Hoofdstuk 4 en 5 van dit proefschrift stond de voorspelbaarheid van woorden gegeven hun context centraal. Hoofdstuk 4 deed verslag van een corpusonderzoek, waarin de rol van contextuele voorspelbaarheid in spraakproductie werd onderzocht. Voor elk van de zeven meest frequente woorden eindigend op *-lijk* werden veertig willekeurige voorkomens geselecteerd uit het Corpus Gesproken Nederlands. Drie bronnen van voorspelbaarheid werden onderzocht: het aantal keer dat een woord al eerder genoemd was, de voorspelbaarheid van het woord gegeven het voorafgaande woord, en de voorspelbaarheid van het woord gegeven het volgende woord. Alledrie deze variabelen lieten significante effecten zien, al waren deze effecten niet altijd van toepassing op alle woorden, stammen of suffixen. Het aantal keer dat een woord al eerder genoemd was had uitsluitend een effect op de duur van het suffix: het suffix was korter naarmate het woord waar het deel van uitmaakte vaker genoemd was. Voorspelbaarheid gegeven het voorafgaande woord liet alleen effecten zien op de mate van reductie in de stam, en deze effecten waren beperkt tot slechts twee van de zeven woorden: *natuurlijk* en *eigenlijk*. Voor deze twee woorden leidde een hogere voorspelbaarheid gegeven het voorafgaande woord tot meer reductie in de stam, zowel voor wat betreft duur als voor wat betreft het aantal gerealiseerde segmenten. De meest robuuste effecten werden echter gevonden voor voorspelbaarheid gegeven het volgende woord. Een hogere waarde voor deze variabele leidde zowel tot een kortere duur van het suffix als tot minder segmenten in de stam.

Aangezien voorspelbaarheid gegeven het volgende woord een zeer robuuste voorspeller is van reductie, formuleerden we de hypothese dat luisteraars gereduceerde woorden sneller zullen verstaan naarmate ze voorspelbaarder zijn gegeven het volgende woord. Deze hypothese werd getoetst in Hoofdstuk 5. In een perceptie-experiment kregen luisteraars zinnen van vier woorden te horen, die verschilden in de mate waarin het gereduceerde derde woord te voorspellen was gegeven het woord dat erop volgde. De proefpersonen moesten op een knop drukken zodra ze het derde woord verstaan hadden. Analyse van de reactietijden liet zien dat proefpersonen het derde woord sneller verstonden naarmate het beter voorspeld kon worden gegeven het vierde woord. Om meer controle te krijgen over de akoestische eigenschappen van voorspelbare en onvoorspelbare derde woorden, werd een controle-experiment uitgevoerd. Dit experiment maakte gebruik van stimuli waarin de akoestische realisatie van het woord identiek was in de voorspelbare en onvoorspelbare condities. De resultaten waren grotendeels hetzelfde: opnieuw werden gereduceerde woorden sneller verstaan naarmate hun voorspelbaarheid gegeven het volgende woord groter was. We kunnen derhalve concluderen dat voorspelbaarheid gegeven het volgende woord een rol speelt bij het herkennen van gereduceerde woordvormen. Om de identiteit van een gereduceerd woord te achterhalen, maken luisteraars gebruik van informatie over de waarschijnlijkheid van woorden gegeven hun directe context.

De resultaten in Hoofdstuk 4 en 5 bieden nieuwe inzichten in de structuur van het mentale lexicon. Het lijkt erop dat woorden die wederzijds voorspelbaar zijn elkaars activatieniveau kunnen doen stijgen. In Hoofdstuk 4 vonden we dat woorden meer gereduceerd zijn naarmate ze beter voorspeld kunnen worden op basis van het volgende woord. Volgens Balota, Boland, en Shields (1989) is een hogere graad van reductie gecorreleerd met een hoger activatieniveau. Daarom concluderen wij dat de activatie van een woord ook kan stijgen als gevolg van de planning van een volgend woord waar het veel samen mee voorkomt. In Hoofdstuk 5 zagen we dat de verwerking van een volgend woord de herkenning van een voorafgaand woord kan versnellen. Dit is alleen mogelijk als beide woorden op de een of andere manier met elkaar verbonden zijn in het lexicon, bijvoorbeeld door middel van een zogenaamd *superlemma* (Sprenger, 2003).

Een verdere implicatie van de bevindingen in Hoofdstuk 5 is dat het spraakperceptiesysteem de mogelijkheid moet bieden om (ambigue) akoestische informatie tijdelijk op te slaan in een soort buffer; anders zouden woorden altijd alleen maar in de volgorde kunnen worden herkend waarin ze zijn uitgesproken, hetgeen in onze experimenten duidelijk niet het geval was. Een voorbeeld van een model dat beschikt over een dergelijke opslagfaciliteit is ARTWORD, ontwikkeld door Grossberg & Myers (2000).

De rol van morfologie

Omdat affixen het centrale onderzoeksobject van dit proefschrift vormden, konden ook morfologisch georiënteerde predictoren van articulatorische reductie worden onderzocht. Dit was het voornaamste doel in de Hoofdstukken 6 en 7, die beide verslag deden van onderzoek naar voorgelezen spraak uit het Corpus Gesproken Nederlands. In Hoofdstuk 6 onderzochten we of de akoestische realisatie van het suffix *-igheid* (/əxhɛit/) mede afhangt van de morfologische structuur van het woord waar het deel van uitmaakt. Het suffix *-igheid* komt in twee soorten woorden voor: woorden waarin er alleen een morfologische grens is tussen *-ig* en *-heid*, zoals bijvoorbeeld *zuinigheid* (*zuin* is geen bestaand woord, *zuinig* wel), en woorden waarin *-igheid* fungeert als een autonoom suffix, zoals bijvoorbeeld *vastigheid* (*vastig* bestaat niet in het Nederlands, *vast* wel). Er werden twee hypotheses geformuleerd, die verschillende voorspellingen deden over de akoestische realisatie van het /xh/-cluster in deze twee woordsoorten. Volgens de Prosodische Structuurhypothese zou het cluster meer gesimplificeerd worden in woorden waarin *-igheid* een autonoom suffix is, omdat het cluster daar geen prosodische grens overschrijdt. De Morfologische Informativiteitshypothese voorspelt daarentegen dat het cluster juist niet gesimplificeerd wordt in deze woorden, omdat de informativiteit van het suffix gegeven de andere woorden in het paradigma (bv. *vast*, *vaste*, *vaster*, *vasthouden*, etc.) relatief hoog is. Deze tweede voorspelling bleek de juiste te zijn: de duur van het cluster was langer in woorden als *vastigheid* dan in woorden als *zuinigheid*. Dit toont eens te meer aan dat er een relatie is tussen de hoeveelheid informatie die een bepaalde taalkundige eenheid overbrengt en de moeite die wordt gespendeerd aan de uitspraak van die eenheid. Daarnaast laten de resultaten van Hoofdstuk 6 zien dat akoestische realisaties beïnvloed kunnen worden door morfologische (in dit geval, paradigmatische) structuur, zonder dat het betreffende effect gemedieerd wordt door de fonologie. Deze bevinding vormt een behoorlijke uitdaging voor hedendaagse taalkundige theorieën, die ervan uitgaan dat er geen directe verbinding is tussen morfologie en fonetiek.

Het onderzoek dat in Hoofdstuk 7 beschreven is leek in eerste instantie niet te stroken met het principe van “minder informatie, meer reductie”, dat beschreven werd in Hoofdstuk 1 en dat de overige resultaten in dit proefschrift zo mooi kon verklaren. We vonden dat de duren van interfixen in het Nederlands langer waren naarmate deze interfixen beter voorspelbaar waren gegeven de eerste constituent van het woord. Dit opmerkelijke resultaat verklaarden we door een onderscheid aan te brengen tussen voorspelbaarheid uit een syntagmatisch perspectief, die negatief gecorreleerd is met akoestische saillantie, en de hoeveelheid paradigmatische steun voor één bepaald alternatief, die positief gecorreleerd lijkt te zijn met akoestische saillantie. Deze verklaring, door ons de Paradigmatische Signaal Versterkingshypothese genoemd, doet concrete voorspellingen over de akoestische realisaties van andere paradigmatische elementen. Deze voorspellingen kunnen worden getoetst in verder onderzoek.

Ideeën voor verder onderzoek

De zeven hoofdstukken van dit proefschrift hebben laten zien dat derivationale affixen interessante onderzoeksobjecten zijn, en dat een gedetailleerde analyse van hun akoestische realisaties tot vele nieuwe inzichten kan leiden. Dit suggereert dat het wellicht ook nuttig is om onderzoek te verrichten naar vergelijkbare taalkundige eenheden, zoals bijvoorbeeld inflectionele morfemen en clitics. Daarnaast is het belangrijk om de huidige bevindingen te repliceren in andere talen dan het Nederlands, aangezien dit inzicht zou verschaffen in de generaliseerbaarheid van onze resultaten, alsmede in de psychologische mechanismen die eraan ten grondslag liggen.

De nadruk in het huidige proefschrift lag op de omstandigheden waarin sprekers woorden reduceren. In Hoofdstuk 5 vonden we echter dat één van de patronen die we geobserveerd hadden in spraakproductie ook een rol speelde in spraakperceptie: gereduceerde woorden werden sneller herkend naarmate ze beter voorspelbaar waren gegeven het volgende woord. Het is goed mogelijk dat luisteraars ook gevoelig zijn voor de andere patronen die in dit proefschrift beschreven staan, aangezien al deze patronen systematisch aanwezig zijn in het akoestische spraaksignaal. Of luisteraars deze patronen daadwerkelijk gebruiken tijdens spraakperceptie kan verder onderzocht worden met behulp van perceptie-experimenten.

Tenslotte willen we erop wijzen dat het onderzoek uit dit proefschrift kwantitatief georiënteerd was: er werden relatief grote hoeveelheden data verzameld en geanalyseerd. Dit is echter niet de enige manier waarop inzicht verkregen kan worden in reductiepatronen. In recente studies zoals die van Curl (2005) en Plug (2005) werden de akoestische realisaties van een beperkt aantal woorden of zinnen bestudeerd, en werd gekeken hoe deze realisaties varieerden als een functie van het communicatieve doel dat de woorden of zinnen vervulden. Dergelijke gedetailleerde, kwalitatieve analyses kunnen reductiepatronen blootleggen die wellicht moeilijk te abstraheren zijn uit grote dataverzamelingen. Om een volledig beeld te krijgen van de omstandigheden waarin sprekers woorden en zinnen reduceren, zullen beide soorten onderzoek moeten worden gecombineerd en geïntegreerd.

Slotopmerkingen

Het ultieme doel van het onderzoek uit dit proefschrift is om psycholinguïstische theorieën van spraakproductie en perceptie op een dusdanige manier te verbeteren, dat zij de enorme fonetische variatie in het spraaksignaal kunnen verklaren en ermee om kunnen gaan. Het is duidelijk geworden dat de meeste bestaande theorieën niet bijzonder goed hiertoe in staat zijn. In productiemodellen zoals dat van Levelt et al. (1999) zorgt de psychologische isolatie van deelprocessen en eenheden ervoor dat informatie van een hoger niveau geen invloed kan hebben op akoestische realisaties. Het grootste probleem in perceptiemodellen is de vroege transformatie van het spraaksignaal in discrete, foneemachtige symbolen, hetgeen geen recht doet aan de grote rol die subtiele fonetische details spelen bij het verstaan van spraak (bv. Hawkins, 2003). Meer en meer onderzoekers beginnen te beseffen dat menselijke spraak veel te complex is om te verklaren met behulp van abstracte symbolen en psychologisch geïsoleerde regels. Dit is onder meer te zien in probabilistische benaderingen van taalkunde (bv. Bod, Hay, & Jannedy, 2003), paradigmatische benaderingen van morfologie (bv. Blevins, 2003), en fonetisch-rijke benaderingen van spraakperceptie (bv. Hawkins & Smith, 2001). Deze veelbelovende ideeën hebben als inspiratie gediend tijdens het schrijven van dit proefschrift; ik hoop van ganser harte dat mijn onderzoek een bijdrage kan leveren aan hun verdere ontwikkeling.

References

Adank, P., Van Hout, R., and Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America* 116, 1729–1738.

Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90, 119–161.

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.

Aylett, M. and Turk, A. (2005). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119, 3048–3058.

Baayen, R. H., Feldman, L., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 55, 290–313.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 37, 94–117.

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

Balota, D. and Abrams, R. (1995). Mental Chronometry: Beyond Onset Latencies in the Lexical Decision Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 1289–1302.

Balota, D., Boland, J., and Shields, L. (1989). Priming in Pronunciation: Beyond Pattern Recognition and Onset Latency. *Journal of Memory and Language* 28, 14–36.

Balota, D., Pilotti, M., and Cortese, M. (2001). Subjective frequency estimates for 2939 monosyllabic words. *Memory & Cognition* 29, 639–647.

Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language* 42, 1–22.

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

Bates, D. and Sarkar, D. (2005). The lme4 library. *[On-line], Available: http://lib.stat.cmu.edu/R/CRAN/*.

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

Bien, H., Levelt, W., and Baayen, R. H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences* 102, 17876–17881.

Binnenpoorte, D., Cucchiarini, C., Boves, L., and Strik, H. (2005). Multiword expressions in spoken language: An exploratory study on pronunciation variation. *Computer Speech and Language* 19, 433–449.

Blevins, J. (2003). Stems and paradigms. *Language* 79, 737–767.

Bock, J. (1995). Sentence production: From mind to mouth. *In* J. Miller and P. Eimas (Eds), *Handbook of perception and cognition, Vol. 11: Speech, language, and communication*. London: Academic Press, pp. 181–216.

Bod, R., Hay, J., and Jannedy, S. (Eds) (2003). *Probability theory in linguistics*. Cambridge: MIT Press.

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

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

Browman, C. and Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica* 49, 155–180.

Bush, N. (2001). Frequency effects and word-boundary palatalization in English. *In* J. Bybee and P. Hopper (Eds), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, pp. 255–280.

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

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication* 15, 39–54.

Cacciari, C. and Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language* 27, 668–683.

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. Utrecht: LOT.

Chatterjee, S., Hadi, A., and Price, B. (2000). *Regression analysis by example*. New York: John Wiley & Sons.

Cholin, J. (2004). *Syllables in Speech Production: Effects of Syllable Preparation and Syllable Frequency*. PhD thesis, University of Nijmegen, Nijmegen, The Netherlands.

Coenen, E., Zwitserlood, P., and Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* 16, 535–564.

Colombo, L. (1993). The comprehension of ambiguous idioms in context. *In* C. Cacciari and P. Tabossi (Eds), *Idioms, processing, structure, and interpretation*. Hillsdale: Erlbaum, pp. 163–200.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud. *Psychological Review* 108, 204–256.

Cooper, W. and Paccia-Cooper, J. (1980). *Syntax & Speech*. Cambridge: Harvard University Press.

Crystal, T. and House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88, 101–112.

Cucchiarini, C. (1993). *Phonetic transcription: A methodological and empirical approach*. PhD thesis, University of Nijmegen, Nijmegen, The Netherlands.

Curl, T. (2005). Practices in Other-Initiated Repair Resolution: The Phonetic Differentiation of ‘Repetitions’. *Discourse Processes* 39, 1–43.

Damian, M. (2003). Articulatory Duration in Single-Word Speech Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29, 416–431.

Davis, M., Marslen-Wilson, W., and Gaskell, M. (2001). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28, 218–244.

De Jong, N., Feldman, L., Schreuder, R., Pastizzo, M., and Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects. *Brain and Language* 81, 555–567.

De Saussure, F. (1966). *Course in General Linguistics*. New York: McGraw.

Dell, G. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review* 93, 283–321.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *Journal of the Acoustical Society of America* 97, 585–592.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface*. Utrecht: LOT.

Ernestus, M. and Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79, 5–38.

Ernestus, M. and Baayen, R. H (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. *In* L. Goldstein, D. Whalen, and C. Best (Eds), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, pp. 29–51.

Ernestus, M. and Baayen, R. H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics* 42, 873–903.

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

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

Fosler-Lussier, E. and Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication* 29, 137–158.

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

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech* 31, 307–317.

Fowler, C. and Housum, J. (1987). Talkers' Signalling of "New" and "Old" Words in Speech and Listeners' Perception and Use of the Distinction. *Journal of Memory and Language* 26, 489–504.

Fowler, C., Levy, E. and Brown, J. (1997). Reductions of Spoken Words in Certain Discourse Contexts. *Journal of Memory and Language* 37, 24–40.

Fox Tree, J. and Clark, H. (1997). Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition* 62, 151–167.

Gahl, S. and Garnsey, S. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80, 748–774.

Garrett, M. (1975). The analysis of sentence production. *In* G. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, pp. 133–178.

Gaskell, M. and Marslen-Wilson, W. (1996). Phonological Variation and Inference in Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance* 22, 144–158.

Gaskell, M. and Marslen-Wilson, W. (1998). Mechanisms of Phonological Inference in Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance* 24, 380–396.

Geffen, G. and Luszcz, M. (1983). Are the spoken durations of rare words longer than those of common words? *Memory & Cognition* 11, 13–15.

Gernsbacher, M. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General* 113, 256–281.

Gibbs, R. (1986). Skating on thin ice: Literal meaning and understanding idioms in conversation. *Discourse Processes* 9, 17–30.

Gow Jr., D. (2002). Does English Coronal Place Assimilation Create Lexical Ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28, 163–179.

Gow Jr., D. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* 65, 575–590.

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

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

Grossberg, S. and Myers, C. (2000). The Resonant Dynamics of Speech Perception: Interword Integration and Duration-Dependent Backwards Effects. *Psychological Review* 107, 735–767.

Haspelmath, M. (1995). The Growth of Affixes in Morphological Reanalysis. *In* G. Booij and J. Van Marle (Eds), *Yearbook of Morphology 1994*. Dordrecht: Kluwer Academic Publishers, pp. 1–29.

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* 13, 99–188.

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

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

Hay, J. and Maclagan, M. (in press). Social and phonetic conditioners on the frequency and degree of ”intrusive /r/” in New Zealand English. *In* D. Preston and N. Niedzielski (Eds), *Methods in Sociophonetics*.

Hunnicutt, S. (1985). Intelligibility versus redundancy - conditions of dependency. *Language and Speech* 28, 47–56.

Janse, E., Nooteboom, S., and Quené, H. (in press). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. To appear in *Language and Cognitive Processes*.

Jespersen, O. (1922). *Language: Its nature, development and origin*. London: George Allen & Unwin Ltd.

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

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34, 485-499.

Jongenburger, W. and Van Heuven, V. (1993). Sandhi processes in natural and synthetic speech. *In* V. Van Heuven and L. Pols (Eds), *Analysis and synthesis of speech: Strategic research towards high-quality text-to-speech generation*. Berlin: Mouton de Gruyter, pp. 261–278.

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

Kawamoto, A., Kello, C., Higareda, I., and Vu, J. (1999). Parallel Processing and Initial Phoneme Criterion in Naming Words: Evidence from Frequency Effects on Onset and Rime Duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25, 362–381.

Kello, C. (2004). Control Over the Time Course of Cognition in the Tempo-Naming Task. *Journal of Experimental Psychology: Human Perception and Performance* 30, 942–955.

Kello, C. and Plaut, D. (2000). Strategic Control in Word Reading: Evidence From Speeded Responding in the Tempo Naming Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 719–750.

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

Kemps, R., Ernestus, M., Schreuder, R., and Baayen, R. H. (2005). Prosodic cues for morphological complexity: The case of Dutch noun plurals. *Memory & Cognition* 33, 430–446.

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

Kenstowicz, M. (1993). *Phonology in generative grammar*. Oxford: Blackwell Publishers.

Kessens, J. and Strik, H. (2001). Lower WERs do not guarantee better transcriptions. *Proceedings of Eurospeech-2001*, Aalborg, Denmark, pp. 1721–1724.

Keune, K., Ernestus, M., Van Hout, R., and Baayen, R. H. (2005). Social, geographical, and register variation in Dutch: From written ‘mogelijk’ to spoken ‘mok’. *Corpus Linguistics and Linguistic Theory* 1, 183–224.

Koester, D., Gunter, T., Wagner, S., and Friederici, A. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience* 16, 1647–1668.

Kohler, K. (2000). Investigating Unscripted Speech: Implications for Phonetics and Phonology.

Phonetica 57, 85–94.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39, 51–93.

Krott, A., Hagoort, P., and Baayen, R. H. (2004). Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes* 19, 453–471.

Krott, A., Schreuder, R., and Baayen, R. H. (2002). Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies. *Brain and Language* 81, 708–722.

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

Ladefoged, P. (1982). *AUC course in phonetics, 2nd edition*. New York: Hartcourt, Brace, Jovanovich.

Lahiri, A. and Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the lexicon. *Cognition* 38, 245–298.

Landauer, T. and Dumais, S. (1997). 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 Streeter, L. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior* 12, 119–131.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., and Zwitserlood, P. (submitted). Native-language influences on word recognition in a second language: A mega study.

Levelt, W. and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition* 50, 239–269.

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

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

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6, 172–187.

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

Losiewicz, B. (1992). *The effect of frequency on linguistic morphology*. PhD thesis, University of Texas.

Mann, V. and Repp, B. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28, 213–228.

Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Manuel, S. (1995). Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *Journal of Phonetics* 23, 453–476.

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

Miller, G. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography* 3, 235–312.

Miller, G. and Licklider, J. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America* 22, 167–173.

Mitterer, H. and Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics* 65, 956–969.

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

Mitterer, H., Csépe, V., and Blomert, L. (2003). Compensation for phonological assimilation in perception: Evidence from Hungarian liquid assimilation. *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 2321–2324.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94, 1–18.

Munson, B. and Solomon, N. (2004). The Effect of Neighborhood Density on Vowel Articulation. *Journal of Speech, Language, and Hearing Research* 47, 1048–1058.

Neijt, A., Krebbers, R., and Fikkert, P. (2002). Rhythm and semantics in the selection of linking elements. *In* H. Broekhuis and P. Fikkert (Eds), *Linguistics in the Netherlands 2002*. Amsterdam: John Benjamins, pp. 117–127.

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

Nooteboom, S. (1972). *Production and perception of vowel duration: a study of durational properties of vowels in Dutch*. PhD thesis, University of Utrecht, Utrecht, The Netherlands.

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

Norris, D., McQueen, J., and Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioural and Brain Sciences* 23, 299–370.

Oldfield, R. and Wingfield, A. (1965). Response Latencies in Naming Objects. *Quarterly Journal of Experimental Psychology* 17, 273–281.

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

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

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and Robustness. *In* R. Bod, J. Hay, and S. Jannedy (Eds), *Probability theory in linguistics*. Cambridge: MIT Press, pp. 177–228.

Plug, L. (2005). From words to actions: The phonetics of 'eigenlijk' in two communicative contexts. *Phonetica* 62, 131–145.

Repp, B., Liberman, A., Eccardt, T., and Pesetsky, D. (1978). Perceptual Integration of Acoustic Cues for Stop, Fricative, and Affricate Manner. *Journal of Experimental Psychology: Human Perception and Performance* 4, 621–637.

Rubenstein, H. and Pollack, I. (1963). Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior* 2, 147–158.

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.

Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General* 110, 474–494.

Samuel, A. (1996). Does Lexical Information Influence the Perceptual Restoration of Phonemes? *Journal of Experimental Psychology: General* 125, 28–51.

Samuel, A. and Troicki, M. (1998). Articulation Quality Is Inversely Related to Redundancy When Children or Adults Have Verbal Control. *Journal of Memory and Language* 39, 175–194.

Scarborough, R. (2004). Degree of Coarticulation and Lexical Confusability. *In* P. Nowak, C. Yoquelet, and D. Mortensen (Eds), *Proceedings of the 29th Meeting of the Berkeley Linguistics Society*.

Scharenborg, O., Norris, D., Ten Bosch, L., and McQueen, J. (2005). How should a speech recognizer work? *Cognitive Science* 29, 867–918.

Schreuder, R. and Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language* 37, 118–139.

Schultink, H. (1962). *De Morfologische Valentie van het Ongelede Adjectief in Modern Nederlands [The morphological valency of the simplex adjective in modern Dutch]*. Den Haag: Van Goor Zonen.

Seidenberg, M. and McClelland, J. (1989). A Distributed, Developmental Model of Word Recogntion and Naming. *Psychological Review* 96, 523–568.

Sereno, J. and Jongman, A. (1990). Phonological and Form Class Relations in the Lexicon.

Journal of Psycholinguistic Research 19, 387–404.

Sevald, C., Dell, G., and Cole, J. (1995). Syllable Structure in Speech Production: Are Syllables Chunks or Schemas? *Journal of Memory and Language* 34, 807–820.

Shannon, C. (1949). *The Mathematical Theory of Communication*. Chicago: The University of Illinois Press.

Shields, L. and Balota, D. (1991). Repetition and associative context effects in speech production. *Language and Speech* 34, 47–55.

Shockey, L. (2003). *Sound Patterns of Spoken English*. Oxford: Blackwell Publishing.

Sjölander, K. (2001). Automatic alignment of phonetic segments. *Lund University, Dept. of Linguistics Working Papers* 49, 140–143.

Slis, I. and Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction II. *Language and Speech* 12, 80–102; 137–155.

Sprenger, S. (2003). *Fixed expressions and the production of idioms*. PhD thesis, University of Nijmegen, Nijmegen, The Netherlands.

Stemberger, J. (1985). An Interactive Activation Model of Language Production. *In* A. Ellis (Ed.), *Progress in the psychology of language*. London: Lawrence Erlbaum Ass., pp. 143–186.

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.

Swinney, D. and Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18, 522–534.

Titone, D. and Connine, C. (1994). Comprehension of Idiomatic Expressions: Effects of Predictability and Literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 1126–1138.

Turk, A. and Sawusch, J. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics* 25, 25–41.

Turk, A. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics* 27, 171–206.

Ullman, M., Estabrooke, I., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K., and Maki, P. (2002). Sex differences in the neurocognition of language. *Brain and Language* 83, 141–143.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America* 61, 846–858.

Van Bergem, D. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12, 1–23.

Van Coile, B. (1987). A model of phoneme durations based on the analysis of a read Dutch text. *In* J. Laver and M. Jack (Eds), *Proceedings of the European Conference on Speech Technology, vol. 2*, Edinburgh, pp. 233–236.

Van den Brink, D., Brown, C., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 and N400 effects. *Journal of Cognitive Neuroscience* 13, 967–985.

Van Marle, J. (1990). Rule-creating creativity: Analogy as a synchronic morphological process. *In* W. Dressler, H. Luschützky, O. Pfeiffer, and J. Rennison (Eds), *Contemporary Morphology*. Berlin: Mouton de Gruyter, pp. 267–273.

Van Son, R. and Pols, L. (2003). Information Structure and Efficiency in Speech Production. *Proceedings of Eurospeech-2003*, Geneva, Switzerland.

Van Son, R. and Van Santen, J. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication* 47, 100–123.

Van Son, R., Binnenpoorte, D., Van den Heuvel, H., and Pols, L. (2001). The IFA Corpus: a phonemically segmented Dutch Open Source speech database. *Proceedings of Eurospeech-2001*, Aalborg, Denmark.

Vieregge, W. (1987). Basic aspects of phonetic segmental transcription. *In* A. Almeida and A. Braun (Eds), *Probleme der Phonetischen Transkription*. Stuttgart: Franz Steiner Verlag Wiesbaden, pp. 5–55.

Vorstermans, A., Martens, J., and Van Coile, B. (1996). Automatic segmentation and labelling of multi-lingual speech data. *Speech Communication* 19, 271–293.

Vossen, P., Bloksma, L., and Boersma, P. (1999). *The Dutch WordNet (CD-ROM)*, European Language Resources Association (ELRA), Luxembourg.

Waals, J. (1999). *An experimental view of the Dutch syllable*. The Hague: Holland Academic Graphics.

Wade, T. and Holt, L. (2005). Effects of later-occurring nonlinguistic sounds on speech categorization. *Journal of the Acoustical Society of America* 118, 1701–1710.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (2004). Incomplete neutralization and other subphonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32, 251–276.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393.

Warren, R. and Sherman, G. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics* 16, 150–156.

Wright, C. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition* 7, 411–419.

Wright, R. (1997). Lexical Competition and Reduction in Speech: A Preliminary Report. *Research on spoken language processing. Progress Report No. 21, Indiana University*, 471–484.

Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (2002). *The HTK Book 3.2*. Cambridge: Entropic.

Zipf, G. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 15, 1–95.

Zorzi, M., Houghton, G., and Butterworth, B. (1998). Two Routes or One in Reading Aloud? A Connectionist Dual-Process Model. *Journal of Experimental Psychology: Human Perception and Performance* 24, 1131–1161.

Zsiga, E. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics* 22, 121–140.

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

Acknowledgments

When reading a dissertation, the first thing I usually look at is the acknowledgments section. For some reason, I am curious to find out how the author has experienced the process of writing a Ph.D. thesis, and which people played an important role during this process. What I enjoy most about acknowledgments, however, is the obvious sense of relief that can be read between the lines. Now, the time has come for me to write my own acknowledgments. It is something I am very much looking forward to, as I finally get the opportunity to officially express my gratitude to everybody who has helped me during the writing of this thesis.

First and foremost, I would like to thank my supervisors, Mirjam Ernestus and Harald Baayen, for their invaluable support during this whole enterprise. Mirjam acted as my first supervisor and mentor. Whether I was measuring affixes, analyzing data, preparing presentations or writing papers, she was always there to help me. Her critical comments have tremendously improved the quality of my research. Recently, Mirjam has received a lot of praise for her research achievements, all of which is fully justified. Here, I would like to take the opportunity to say that as a supervisor, she is second to none as well. Harald, as promotor, also played an important role in the coming about of this thesis. Besides regularly popping in to chat about my progress, he always managed to find time to help me with statistical analyses and read my papers. Furthermore, his enthusiasm for my research was very contagious. I consider myself lucky to have had Mirjam and Harald as supervisors and I want to thank them from the bottom of my heart.

I would also like to thank the members of the manuscript committee. Lou Boves, Susanne Gahl, Ingo Plag, Louis Pols, and Vincent van Heuven found time to read my manuscript in the hectic period before Christmas, and even managed to come up with lots of useful comments. As a result of their efforts, the quality of my thesis has improved.

The Interfaculty Research Unit for Language and Speech, or IWTS, as it used to be known, proved to be a very inspiring and pleasant working environment. I thoroughly enjoyed all the social activities we organized, but in the end, it was the scientific attitude of the group that made working at the IWTS so special. Furthermore, I was lucky to have great roommates in Laura, Wieke, Victor, and of course Karen, who agreed on becoming my paranimf. A special “thank you” goes out to the student assistants, who have done a great amount of work for my project and without whom I would never have been able to finish in time. So here you go,

in alphabetical order: Anneli, Berit, Femke, Hanke, Hilde, Inge, Jeanne, Luuk, and Mybeth, thanks a lot! Finally, I would like to thank Rob for his intellectual and technical support, Roel for being my (rugby) buddy, Kors for taking me to Café Jos, Marieke, Debby, and Marjolein for the ever-pleasant coffee breaks, Lanneke for her help in practical matters, and Stefan for helping us get rid of any excess pie.

The Comprehension Group of the Max Planck Institute provided a second home for me. Since they were always the first to hear about new results, their useful comments enabled me to prepare myself for tougher audiences. I want to thank Anne Cutler, James McQueen and the other members of the group for their hospitality and support. I would also like to acknowledge Tobias van Valkenhoef, who kept my computer running throughout the project, and the MPI receptionists, who took care of paying the participants of my internet experiment.

When writing a dissertation, it is very important to be able to 'switch off' once in a while. One way of doing that is by working on something completely different. Paul van Tongeren introduced me to the fascinating world of ethics and guided me towards my first philosophical publication. Joost Schilperoord came to Nijmegen to work on our article, but our busy schedules have thus far precluded us from finishing it. Away from work, I found relaxation and friendship with NSRV Obelix and NSDV Double Bull S, and actually managed to win a championship for the first time in my life! In the weekends, I usually went back to Limburg to have a few beers with my closest friends. Somehow, I have never discussed the details of my research with them, which is testament to the great friendship that we have.

The final paragraph of the acknowledgments is usually reserved for family, and this one is not going to be an exception. I would like to thank my parents, Hans, Janneke, and Jeanny for their unconditional love and support. However, there is one person who deserves my gratitude more than anybody. Without her sacrifices, patience, and support, this thesis would never have been written. Peggy, thanks a million for enabling me to do this!

Curriculum Vitae

Mark Pluymaekers was born in Meerssen, The Netherlands, on 25 January 1980. In 1997, he started studying Oriental Languages and Communication at Hogeschool Zuyd in Maastricht, majoring in Japanese. After obtaining his bachelor's degree in 2001, he enrolled in a two-year master's program at Tilburg University. He graduated in 2003 with a thesis on the circumstances in which speakers use fixed expressions. In October 2003, he started a Ph.D. project entitled "Frequency and morphophonological adaptation" at the Interfaculty Research Unit for Language and Speech (IWTS) of the Radboud University Nijmegen. This dissertation describes the research conducted in that project. Currently, Mark works as a lecturer at Hogeschool Zuyd. Furthermore, he is available as a freelance text writer.

Publications

Pluymaekers, M. (2005). Ethische aspecten bij selectieve perceptie in de wetenschap. *Filosofie* 15, p. 61-64. (”Ethical aspects of selective perception in science”).

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

Pluymaekers, M., Ernestus, M., and Baayen, R.H. (2005b). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, p. 146-159.

Pluymaekers, M., Ernestus, M., and Baayen R.H. (2006). Effects of word frequency on the acoustic durations of affixes. *Proceedings of Interspeech 2006*, p. 953-956.

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R.H. (2007). Morphological redundancy and the acoustic salience of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121, p. 2261-2272.

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 categorisation in an Australian language. *Eva Schultze-Berndt*

15. Interpreting indefinites: An experimental study of childrens 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 (a West Chadic language of Nigeria). *Birgit Hellwig*

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

24. Contextual influences on spoken-word processing: An electrophysiological approach. *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: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *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. *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: Probabilistic effects in production and perception. *Mark Pluymaekers*