



Morphological and predictability effects on schwa reduction: The case of Dutch word-initial syllables

Iris Hanique^{1,2}, Barbara Schuppler¹, Mirjam Ernestus^{1,2}

¹Radboud University Nijmegen, The Netherlands

²Max Planck Institute for Psycholinguistics, The Netherlands

Iris.Hanique@mpi.nl, B.Schuppler@let.ru.nl, Mirjam.Ernestus@mpi.nl

Abstract

This corpus-based study shows that the presence and duration of schwa in Dutch word-initial syllables are affected by a word's predictability and its morphological structure. Schwa is less reduced in words that are more predictable given the following word. In addition, schwa may be longer if the syllable forms a prefix, and in prefixes the duration of schwa is positively correlated with the frequency of the word relative to its stem. Our results suggest that the conditions which favor reduced realizations are more complex than one would expect on the basis of the current literature.

Index Terms: pronunciation variation, acoustic reduction, corpus research, morphological structure, word predictability

1. Introduction

In casual speech, words often appear in reduced pronunciation variants [1, 2]. A segment may be relatively short or completely absent. The Dutch word *geleden* 'ago', for example, has the citation form /xələdə/, but the first schwa may be very short and occasionally absent, as in [xlədə]. The present corpus study investigates to what extent the presence and duration of schwa in Dutch word-initial syllables is affected by word predictability and the word's morphological structure.

Insights in reduced speech are useful for the development of both production and comprehension models. In order to create production models that can account for how humans produce reduced speech, we need to know what variables affect their reduced realizations and how. For the development of comprehension models, it is important to know what variation is present in the acoustic input. Moreover, we need information about which variables predict this variation, in order to be able to test whether listeners may use this information as well.

One of the best investigated variables that condition a word's degree of reduction is its predictability. This variable is among others, reflected by the word's frequency of occurrence. Words with a high frequency of occurrence tend to have shorter segments than those with a low frequency [3]. Another measure of a word's predictability is its likelihood given the preceding or following word. Content words (regardless of their frequency) and function words of low or mid frequency that are more predictable given the following word tend to be shorter than those that are less predictable [4, 5]. A role of the word's predictability given the preceding word has only been reported for high frequency function words. Possibly, such an effect has not been reported for content words because so far studies investigating contextual predictability did not focus on initial syllables. Especially that part of the word closest to the preceding or following word may be affected by that context. If so, the final syllable

is likely to be affected by the following word, which is in line with the results presented above [4, 5]. Likewise, the first syllable may mostly be affected by the preceding word, a hypothesis that is tested in the present study.

Whereas the effect of a word's predictability on its degree of reduction has received relatively much attention in the literature, only a couple of studies have addressed the role of a word's morphological structure. Losiewicz [6] showed that English word-final [t] and [d] may be longer if they form affixes of their own (e.g., as in *darted*) than if they are part of the words' stem (e.g., as in *tact*). If this result generalizes to other affixes and other languages, schwas in the initial syllables in our study may be longer if they are part of prefixes, rather than of the words' stems. The current study tests this hypothesis.

More frequent morphologically complex words are claimed to have a higher probability of being retrieved from the mental lexicon as single units (e.g., [7]). If so, highly frequent complex words should pattern with monomorphological words, such that their affixes are more likely to be reduced. This hypothesis is supported by [6] showing that English word final [t] and [d] are shorter both in monomorphemic words (e.g., as in *tact*) and in past-tense morphemes of high frequency verb forms (e.g., as in *started*) than in past-tense morphemes of low frequency verb forms (e.g., as in *darted*). Similarly, the higher the frequency of a morphologically complex word relative to the frequency of its stem, the more likely it is to be retrieved as a whole word from the lexicon and to pattern with monomorphological words [8]. Accordingly, English complex words are more likely to be produced without their final [t]s, the more frequent they are relative to their stem. The present study investigates whether these findings for English suffixes generalize to Dutch prefixes.

In short, this corpus study investigates whether predictability and morphological variables affect schwa reduction in Dutch word-initial syllables. We focus on the presence versus absence of schwa as well as on its duration.

2. Method

2.1. Materials

The present study focusses on the initial syllables *ge-*, *be-*, and *ver-*. Their canonical transcriptions (/xə/, /bə/, and /vət/) contain schwa, the segment under investigation. They occur at the beginning of many word types of various frequencies. Furthermore, these syllables can either be part of a word's stem or be a prefix, which allows for the investigation of the role of morphological structure.

Data were extracted from the ten casual dialogues of the Ernestus Corpus of Spontaneous Dutch [2]. This corpus contains approximately 15 hours of speech, in which 153200 word

tokens were produced, representing 9035 word types. The corpus is very homogeneous with respect to the geographical and social backgrounds of the speakers. All 20 speakers were native speakers of standard Dutch with academic degrees. They were males between 21 and 55 years old, who were born and raised in the West of the Netherlands [2].

For this study, we use the phonemic transcriptions generated automatically in [9]. Automatically generated transcriptions have the advantage of being more consistent than human transcriptions. The automatic generation was based on a forced alignment of the handmade orthographic transcriptions to the speech signal by means of the automatic speech recognition (ASR) toolkit HTK [10]. First, for each word in the orthographic transcriptions, the ASR system retrieved the canonical phonemic transcription and possible pronunciation variants from a pronunciation lexicon. The pronunciation variants in this lexicon were created by the application of 32 reduction rules to the canonical pronunciations of the words. These rules were based on earlier observations of reduced pronunciation variants in casual Dutch [2]. Subsequently, the ASR system selected the pronunciation variant that matched best with the speech signal on the basis of acoustic phone models. These models were 37 32-Gaussian tri-state monophone acoustic models [11], which were trained on the Dutch library of the blind of the Corpus Gesproken Nederlands (CGN, [12]). Since the frame shift was 5 ms and the acoustic models consist of at least three emitting states, segments were assigned a minimum duration of 15 ms (also if they were only 10 ms).

We only considered words for analysis produced without background noise, including speech from other speakers. Furthermore, the second syllable of all words started with a consonant. However, because /h/ is often produced like a voiceless vowel, initial syllables followed by /h/ were omitted from analysis. The words *bedoel* ‘(I) mean’ and *gewoon* ‘just’ and the bigrams *ik geloof* ‘I believe’, *ieder geval* ‘any case’, *geloof ik* ‘believe I’, and *gegeven moment* ‘certain moment’ all occurred very frequently in the data set (143 and 415 times and 32, 45, 61, and 66 times, respectively). Since these words and bigrams may therefore behave differently from the majority, they were discarded as well. Also, the word *genie* ‘genius’ was discarded as it is a loan word with a non-Dutch pronunciation of the initial syllable. Finally, [r] was absent in most tokens of word-initial *ver-*. Therefore, words including the [r] in *ver-* were omitted. In total the final data set consisted of 2292 tokens representing 654 word types from 20 speakers. All words in this data set were content words, except for eight tokens (0.34% of the data).

2.2. Variables

We investigated both the presence versus absence of schwa, and the duration of schwa (in ms). We estimated how these dependent variables are affected by three types of independent variables. The frequencies used to determine some of the independent variables were based on the CGN. This corpus includes the same type of speech as we investigated (i.e., spontaneous speech), which has been shown to be important [13].

The first type of independent variables are word predictability measures. One measure is the word’s frequency of occurrence, which was log transformed with base 2. Another measure is the likelihood of a word given the neighboring words, which was determined by means of the word’s mutual information (MI) with the preceding or following word. This measure is defined as in equation 1, where X and Y denote either the

preceding and target word or the target and following word.

$$MI(X, Y) = -\log_2\left(\frac{Freq(XY)}{Freq(X) \cdot Freq(Y)}\right) \quad (1)$$

Second, we investigated morphological effects by means of two variables. One factor indicated the morphological status of the initial syllable (prefix versus non-prefix). Another variable was the ratio of the frequency of the word’s stem (S) and the frequency of the word itself (W), which we calculated using equation 2 [8].

$$Ratio(S, W) = \log_2\left(\frac{Freq(S)}{Freq(W)}\right) \quad (2)$$

Finally, we included control variables in the statistical models in order to reduce variance in the data. The first control variable identified the initial syllable (*ge-*, *be-*, *ver-*). As a second control variable we included the speech rate, which was defined as the number of syllables per second in the contiguous chunk of speech that contained the target word. The third control variable is a binary factor that indicated whether the word was in utterance final position (final versus non-final). A final variable is the place of articulation of the first consonant of the second syllable. This factor had the levels front ([b, p, m, w, v, f, r]), mid ([d, t, z, ʒ, s, ʃ, l, n]), and back ([x, ɣ, k, ŋ, j]).

3. Results

We investigated the relevance of the various predictors for the presence of schwa and for the duration of schwa by means of mixed effects regression models with speaker and word as crossed random factors. For each dependent variable, we first fitted a model with speech rate, utterance finality, place of articulation of the following consonant, initial syllable, and morphological status as predictors. Subsequently, we added each probabilistic variable separately to the model, using that subset of the data for which the variable could be computed.

3.1. Schwa presence

The significant results for the analyses of the presence versus absence of schwa are summarized in Table 1. The model on the complete data set (M1) showed, as expected, significant main effects of speech rate and utterance finality. Schwa is more likely to be absent at faster speech rates and in non-final utterance positions. In addition, we found a significant main effect of the initial syllable, showing that schwa is more likely to be present in the syllable *be-*, followed by *ver-*, and finally *ge-*. The place of articulation of the following consonant yielded a significant effect as well, which is modulated by an interaction with the morphological status of the initial syllable. Schwa is more likely to be absent if the following consonant is alveolar/palatal (i.e., mid), especially if the initial syllable is not a prefix. Only 146 tokens contained identical consonants before and after the schwa of the initial syllable (irrespective of whether they are voiced). Therefore, the effects of place of articulation on the reduction of schwa cannot be explained by whether the consonants surrounding schwa were identical.

In order to investigate the effects of four probabilistic variables, we created four subsets for which these variables could be computed. The role of each probabilistic variable was analysed by means of a statistical model with as predictors the probabilistic variable and the predictors that were significant in M1 (i.e., speech rate, utterance finality, place of articulation, morphological status, and initial syllable). The first subset (2247

tokens) only included words with a log word frequency greater than zero, in order to have a more or less normal distribution for this variable. Analysis of this subset yielded no statistically significant effect of word frequency. The second subset contained all 1509 tokens directly preceded by another word in the sentence and for which we could calculate the MI with this word. This subset showed no significant effect of this MI.

The third data set contained the 1290 word tokens directly followed by another word in the sentence, which could be used to test the relevance of MI with the following word. Because all tokens have to be utterance non-final, the statistical model did not contain the predictor utterance finality. The model (M2) yielded a significant effect of MI with the following word. Schwa is more likely to be absent if there is less MI between the current and following word. All variables significant in M1 are also significant in this model with MI, except for place of articulation.

Finally, we created a subset to analyse the role of the stem-word ratio. This subset consisted of the 1070 tokens, which contained a prefix and for which the ratio could be calculated. No effects of the stem-word ratio could be detected.

Factor	M	β	F	df_1	df_2	$p <$
MI following	M2	0.06	4.03	1	1280	.05
Morphological status						
× Place of articulation	M1	–	3.87	2	2282	.05
Speech rate	M1	-0.22	25.60	1	2282	.0001
Utterance finality	M1	0.30	4.24	1	2282	.05
Initial syllable	M1	–	13.24	2	2282	.0001
Place of articulation	M1	–	4.93	2	2282	.05

Table 1: Significant results for the presence of schwa. The horizontal line in the middle separates the variables of interest from the control variables. The column M indicates the model described in the text in which the effect was found.

3.2. Schwa duration

We analysed the durations of the 1273 present schwas. In order to have a more or less normal distribution, the durations were log transformed with base 2. Furthermore, schwas longer than 2.5 times the residual standard errors of the values predicted by the statistical models were considered outliers and removed. The significant results can be found in Table 2.

The model on the entire data set for schwa duration (M3) shows significant main effects for speech rate and utterance finality. Schwas tend to be longer at lower speech rates and in utterance final words. Furthermore, there was a main effect of place of articulation, which was modulated by an interaction with initial syllable. Together these effects show that schwa in the syllable *be-* tends to be longer if the following consonant is labial (i.e., front) or /t/. The effect of place of articulation for *ver-* appears comparable to that for *be-*, but does not reach significance. More importantly, the model showed a significant interaction of morphological status with initial syllable. Whether the initial syllable formed a prefix was significant for the syllable *ge-*: Schwa seems to be longer if *ge-* forms a prefix. Although morphological status does not reach significance for *ver-*, the effect appears comparable to that of *ge-*.

As we did for the analyses of the presence of schwa, we created four subsets to test the relevance of the probabilistic variables. Each statistical model testing the relevance of a probabilistic variable contained this probabilistic variable as well

as the predictors that were also significant in M3 (i.e., speech rate, utterance finality, place of articulation, morphological status, and initial syllable). The first subset consisted of the 1255 word tokens with a log word frequency greater than zero. The statistical model for this subset yielded no significant effect of word frequency. The second subset contained all 817 tokens directly preceded by another word in the sentence and for which we could calculate the MI with this word. The model on this subset showed no statistically significant effect of the MI with the preceding word. Similarly, the third subset contained the 689 tokens directly followed by a word in the sentence, which could be used to test the relevance of MI with the following word. This subset yielded no significant effect of this MI.

Finally, the subset for testing the relevance of the stem-word ratio consisted of the 596 tokens starting with a prefix for which the ratio could be calculated. Because all initial syllables in this subset are prefixes, the factor morphological status was not included. The statistical model for this subset (M4) yielded a three-way interaction of the stem-word ratio with the initial syllable and speech rate. Further analysis showed that schwa in the prefix *ge-* tends to be longer if the ratio is higher, especially at lower speech rates. The prefix *ver-* appears to behave similarly to *ge-*, whereas *be-* seems to behave differently. However, both *be-* and *ver-* are represented by a relatively small number of tokens. Further research is therefore necessary to establish the effect of the stem-word ratio on these two prefixes. All variables significant in M3 are also significant in this model, except for place of articulation.

Factor	M	β	F	df_1	df_2	$p <$
Morphological status × Initial syllable	M3	–	3.71	2	1225	.05
Stem-Word ratio × Initial syllable × Speech rate	M4	–	5.67	2	568	.0001
Speech rate	M3	-0.10	31.32	1	1225	.0001
Utterance finality	M3	0.22	23.12	1	1225	.0001
Place of articulation	M3	–	3.46	2	1225	.05
Place of articulation × Initial syllable	M3	–	3.21	4	1225	.05

Table 2: Significant results for schwa duration. The horizontal line in the middle separates the variables of interest from the control variables. The column M indicates the model described in the text in which the effect was found.

4. Discussion and conclusions

The present study investigated which variables affect the presence and duration of schwa in Dutch word-initial syllables. Insights in these variables will inform the development of models of speech production and comprehension. We analysed the presence and duration of schwa in 2292 word tokens representing 654 types, extracted from a corpus of spontaneous Dutch.

We focussed on two types of variables. One is the word's predictability. Several studies have shown that words tend to be more reduced if they are more predictable given the following word (e.g., [4, 5]). In contrast, our data set shows that schwa is more likely to be present if the predictability of the word given the following word is higher. We suggest the following explanation, assuming the finding is valid and not a type 1 error. If the MI between two words is high, the connection between the two words may be so strong that at the moment the first word

is retrieved from the mental lexicon, the following word is retrieved simultaneously. This simultaneous retrieval may interfere with and delay the production of the beginning of the first word. If so, the time necessary to produce the beginning of the first word is longer, which may result in a more clear pronunciation. This explanation also fits with the earlier findings showing that a higher MI is correlated with a higher reduction degree of segments later in the word: An early activation of the following word implies less processing during the production of the end of the first word, which may result in a more reduced pronunciation. This hypothesis predicts that the influence of the predictability of a word given the following word on degree of reduction differs between the beginning and the end of the word. Obviously, this hypothesis needs to be investigated further.

We also investigated the role of the word's predictability given the preceding word. Previous studies documented a role of the preceding word only for function words of high frequencies of occurrence. These studies, however, mainly investigated segments that occur later in the word than the segment that we investigated in the present study. We formulated the hypothesis that the preceding word mainly affects segments early in the word. However, we did not find evidence for this hypothesis: The initial syllables in the content words in our data set do not show any effect of the MI with the preceding word.

Besides MI, we also investigated a word's predictability as indicated by its frequency of occurrence. We did not find any influence of word frequency on the presence of schwa or on its duration. The reason may be that the selected words were all produced within context. Whenever words occur within context, the role of word frequency in processing appears to be less important than their predictability from the context (e.g., [14]).

The second type of variables that we focussed on concerns the word's morphological structure. First, we investigated possible influences of the morphological status of the initial syllable (i.e., whether this syllable forms a prefix). We found that if *ge-* is a prefix, schwa tends to be longer. This effect of morphological status is in line with the finding that English word-final [t] and [d] tend to be longer if they form an affix [6].

Second, we investigated the effect of the ratio of the frequency of the word's stem and the frequency of the word itself. Words with a lower stem-word ratio are more likely to be retrieved from the mental lexicon and consequently may be more likely to show a higher degree of reduction. Our results showed that if the ratio is lower, schwa is shorter in the prefix *ge-*, particularly at lower speech rates. One explanation for why the effect of ratio is less pronounced at higher speech rates may be that at higher speech rates schwas are already reduced. Another explanation may be that at high speech rates words have to be processed quickly, which may favor the direct retrieval of morphologically complex words from the mental lexicon, independently of the word's stem-word ratio.

The reason that ratio affects especially the prefix *ge-* may be that the words starting with *ge-* were mainly past participles. These word forms are semantically transparent, and therefore can both be retrieved from the mental lexicon and be computed on the basis of their morphemes. In contrast, *be* and *ver* mainly form semantically less transparent derivations, which have to be retrieved from the mental lexicon. Their stem-word ratio may therefore be less relevant for processing.

In addition to the effects of predictability and morphology, we found an influence of the place of articulation of the consonant following schwa. Schwa is more likely to be absent if the following consonant has a mid place of articulation. Possibly, schwa is more reduced before such consonants since their

articulation requires less effort than the articulation of back or front consonants after schwa. Interestingly, this effect of place of articulation is smaller if the initial syllable is a prefix. In this latter case, schwa reduction may be (partly) suppressed in favor of a clear production of the prefix.

Furthermore, place of articulation affects the duration of schwa. If *be-* is followed by a front consonant, schwa is longer. After the production of the front consonant [b], the production of a second front consonant may take more effort compared to the production of a back or mid consonant. The production of a relatively long schwa after [b] may facilitate the production of a directly following front consonant.

We conclude that the conditions which favor reduced realizations are more complex than one would expect on the basis of the current literature. A clear picture of these conditions can only be obtained by combining corpus studies with controlled production experiments.

5. Acknowledgments

This work was partly funded by an European Young Investigator Award to the third author and the Marie Curie Research Training Network *Sound To Sense*.

6. References

- [1] K. Johnson, "Massive reduction in conversational American English," in *Proceedings of Spontaneous Speech: Data and Analysis*, Tokyo, Japan, 2004, pp. 29–54.
- [2] M. Ernestus, "Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface," Ph.D. dissertation, Utrecht, 2000.
- [3] M. Pluymaekers, M. Ernestus, and R. Baayen, "Lexical frequency and acoustic reduction in spoken Dutch," *JASA*, pp. 2561–2569, 2005.
- [4] A. Bell, J. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *JML*, pp. 92–111, 2009.
- [5] M. Pluymaekers, M. Ernestus, and R. Baayen, "Articulatory planning is continuous and sensitive to informational redundancy," *Phonetica*, pp. 146–159, 2005.
- [6] B. Losiewicz, "The effect of frequency on linguistic morphology," Ph.D. dissertation, University of Texas, 1992.
- [7] J. Stemberger and B. MacWhinney, "Are lexical forms stored in the lexicon?" in *Theoretical Morphology: Approaches in Modern Linguistics*, M. Hammond and M. Noonan, Eds. San Diego: Academic Press, 1988, pp. 101–116.
- [8] J. Hay, "Causes and consequences of word structure," Ph.D. dissertation, Northwestern University, 2000.
- [9] B. Schuppler, M. Ernestus, O. Scharenborg, and L. Boves, "An automatic method to analyze acoustic reduction in a corpus of conversational Dutch," submitted.
- [10] S. Young, G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book 3.2*. Cambridge: Entropic, 2002.
- [11] A. Hämäläinen, M. Gubian, L. ten Bosch, and L. Boves, "Analysis of acoustic reduction using spectral similarity measures," *JASA*, pp. 3227–3235, 2009.
- [12] N. Oostdijk, "The design of the Spoken Dutch Corpus," in *New Frontiers of Corpus Research*, P. Peters, P. Collins, and A. Smith, Eds. Amsterdam: Rodopi, 2002, pp. 105–112.
- [13] F. Torreira and M. Ernestus, "Probabilistic effects on French [t] duration," in *Proceedings of Interspeech 2009*, Brighton, U.K., 2009, pp. 448–451.
- [14] C. Becker, "Semantic context and word frequency effects in visual word recognition," *Journal of Experimental Psychology: Human Perception and Performance*, pp. 252–259, 1979.