How Robust Are Exemplar Effects in Word Comprehension?

Iris Hanique[1,2], Ellen Aalders[1], Mirjam Ernestus[1,2]

[1]Centre for Language Studies, Radboud University Nijmegen, the Netherlands

[2]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Abstract

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

Keywords: speech comprehension, exemplar effects, pronunciation variation, acoustic reduction

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

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

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

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

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

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

Our targets were Dutch verbs that start with the unstressed prefixes *be-* or *ver-*. Words of this type are common in Dutch and their prefixes often contain reduced schwas in casual speech (Hanique, Ernestus, and Schuppler, 2013; Pluymaekers, Ernestus, and Baayen, 2005). In all four experiments, primes and targets could differ in their degree of reduction: Segments in the reduced tokens were shorter than in the unreduced tokens, and some segments were completely missing. Our unreduced tokens therefore represent tokens that are typically found in slow speech, while our reduced tokens represent tokens that can be found at a high speech rate in casual speech. We hypothesized that if participants react more quickly to targets showing the same degree of reduction as their primes, participants must have accessed the exemplars of these primes.

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

Second, we investigated whether exemplar effects also occur if the repetition of words is less clear for participants in our experiment than in experiments reported in the literature. In the experiments in the literature, the number of trials varied from 48 (McLennan and Luce, 2005) to 436 (Craik and Kirsner, 1974) and between 33% and 50% of the trials formed word repetitions. Furthermore, the majority of these experiments used an explicit memory task (old-new judgments). Since it was clear to the participants that many words were repeated, they may have used a strategy in which they directly accessed exemplars. In Experiments 1 and 3 of our study, participants listened to 288 trials in which 34% of the trials formed word repetitions (similar to McLennan and Luce, 2005). In Experiments 2 and 4, we increased the number of trials to 800 and 864 respectively, and decreased the percentage of trials constituting word repetitions to less than 20% (almost 16% in Experiment 2 and 18% in Experiment 4).

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

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

Experiment 1

Method

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

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

Table 1 presents an overview of the number of the different types of stimuli. The experiment contained 48 pairs of primes and targets. We wished to keep the number of trials intervening between primes and targets small so that, at least in this respect, our experiment resembled the experiments in the literature that showed exemplar effects. These 48 prime-target pairs were therefore divided over two parts. Each part consisted of two blocks: the first block had 24 primes and 48 foils and the second block contained the corresponding 24 targets and 48 foils. Each word type occurred in only one part of the experiment. In order to better hide the aim of the experiment, in the second block of each part, we repeated existing words (the targets) as well as 24 pseudo-infinitives (foils)[i].

INSERT TABLE 1 ABOUT HERE

We used two pronunciation variants for the primes, targets, and foils: an unreduced one, carefully articulated at a slow speech rate, and a reduced one, with shorter and possibly absent segments. A prime and target represented either the same or a different pronunciation variant. All stimuli were recorded by two Dutch native speakers: one male (henceforth *Speaker A*) and one female (*Speaker B*). Stimuli were recorded over the course of multiple recording sessions. Since speakers typically do not produce casual speech in front of a microphone, we had to tell our speakers that the reduced stimuli had to sound as if uttered in casual speech. The instructions given to the speakers determined whether tokens were categorized as reduced or unreduced. For each word type that occurred as prime and target or as repeated foil, each speaker recorded several unreduced and reduced tokens (see Figure 1 for an example). From these tokens we selected the two best tokens for each pronunciation variant for a given speaker, so that primes and targets (and repeated foils) were always different tokens. For the remaining foils, we recorded either reduced or unreduced variants and selected the best token for a given speaker.

INSERT FIGURE 1 ABOUT HERE

We analyzed the recordings to examine whether the reduced and unreduced stimuli differed in degree of reduction, and whether the speakers varied in the degree of the pronunciation difference between reduced and unreduced stimuli. For the 384 recorded primes and target tokens, we created broad phonetic transcriptions using the forced alignment procedure described by Schuppler, Ernestus, Scharenborg, and Boves (2011). From these transcriptions we extracted the duration of the whole word and determined whether schwa was present. The averages are presented in Table 2. Subsequently, we analyzed these two measures as dependent variable. For the presence of schwa, we fitted logistic mixed effects regression models and for word duration we fitted mixed effect regression models, with *prefix* (*be*- vs. *ver*-), *speaker* (Speaker A vs. B), and *variant* (reduced vs. unreduced) as fixed effects and *word type* (e.g., *vertolken* or *beschrijven*) as random effect. Table 3 shows the resulting models. As shown in Figure 2, reduced stimuli were significantly shorter than unreduced stimuli. This difference was larger for stimuli produced by Speaker B. Similarly, schwa was more often absent in reduced stimuli and in stimuli produced by Speaker B. The automatically generated transcriptions suggest that schwa was even frequently absent in Speaker B's unreduced realizations (10.4%). In general, our analyses clearly demonstrate that the reduced tokens are more reduced than the unreduced tokens. In addition, Speaker B shows a larger difference between the reduced and unreduced tokens than Speaker A.

INSERT TABLES 2 AND 3 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

To test whether the differences between the unreduced and reduced tokens and between two speakers could also be perceived by naïve listeners, we conducted a rating experiment. We asked 50 participants aged between 18 and 29 (mean 21) to rate 60 foils and all primes and targets on a 6-point scale ranging from very unintelligible (rating score 1) to very intelligible (rating score 6). We created eight different pseudo-randomized orders of the stimuli, so that together the eight lists contained every token of each word (i.e., two primes and two targets produced by each speaker). Each participant heard one list in which each word type occurred once and which contained both reduced and unreduced stimuli produced by Speaker A as well as Speaker B. Since the rating scores were not normally distributed, we converted the scores to a factor in which scores 1, 2, and 3 were treated as unintelligible and scores 4, 5, and 6 as intelligible. We then fitted a logistic linear mixed effects model (Faraway, 2006) based on all primes and targets with *word type* and *participant* as crossed random effects. This model yielded significant effects of *speaker* ($\beta = 2.89$, $z = 10.51$, $p < .0001$) and *pronunciation variant* ($\beta = 3.86$, $z = 9.46$, $p < .0001$) and their interaction ($\beta = -2.39$, $z = -3.51$, $p < .001$), indicating that reduced items were less intelligible than unreduced items, especially if the items were produced by Speaker B (Speaker A: reduced 96.7% intelligible and unreduced 99.2%; Speaker B: reduced 72.5% and unreduced 98.7%). The two speakers clearly differed in their pronunciation and intelligibility of the reduced variants and it is therefore interesting to compare exemplar effects for these two speakers.

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

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

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

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

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

Results and Discussion

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

As none of the participants made errors in more than 20% of the trials, none were excluded from our analyses of the response times. We restricted our analyses to those target words for which more than 80% of the responses were correct, which led to the exclusion of the word *bekransen* 'to garland'. Table 4 shows the statistical model based on the remaining 1980 trials (85.9% of all trials). Response times measured from word onset were 943 ms on average and ranged from 522 to 2375 ms. The effects of our control predictors showed that responses were faster to words carrying the prefix *be-* (mean: 913 ms) than *ver-* (971 ms); to words produced by Speaker B (mean: 879 ms) compared to Speaker A (1003 ms); and to words with a higher frequency of occurrence. In addition, responses were faster if the word itself or its prime was shorter. Finally, responses were faster the faster the response to the prime or the preceding trial.

INSERT TABLE 4 ABOUT HERE

Importantly, we found a significant main effect of *variant match*, which indicated that responses were faster if the prime and the target represented the same pronunciation variant (mean: 933 ms) compared to different variants (952 ms). *Variant match* did not significantly interact with the random effects *word type* or *participant*, suggesting that the effect does not depend on a subset of word types or participants. Furthermore, the interaction between *variant match* and *speaker* was not significant, which suggests that the effect of *variant match* does not differ for the two speakers. We examined whether the effect was also significant for the two speakers separately and it was (Speaker A: $\beta = 0.018$, $t = 2.29$, $p < .05$; Speaker B: $\beta = 0.021$, $t = 2.61$, $p < .05$). Although our phonetic analyses and the rating study clearly showed differences between the stimuli produced by the two speakers, the effect of *variant match* is thus similar for both speakers. A possible explanation is that each participant heard only one speaker. As listeners typically adapt very rapidly to a new speaker (e.g., Dahan, Drucker, and Scarborough, 2008), participants had probably already adapted to the speaker during the first block. Consequently, the differences between the speakers did not play a substantial role.

To further investigate the robustness of exemplars, in Experiment 2 we increased the number of non-repeated foils. This experiment consisted of 800 trials. As only 16% of the trials formed word repetitions, this setup closely approximates natural conversations, in which speakers avoid repetition by often replacing content words by pronouns. As we were not able to create large numbers of stimuli using the prefixes *be*- and *ver*- only, Experiments 2 also contained foils with the prefixes *in*-, *aan*-, and *ont*-. In addition, we increased the average number of trials between primes and targets. Since we thought these manipulations would make it harder to find exemplar effects, we tested more participants.

Experiment 2

Method

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

Materials. We used the same stimuli as in Experiment 1 plus additional foils. The Dutch lexicon contains approximately 500 tri-syllabic infinitives with the prefix *be*- or *ver*- and a unique stem (Celex; Baayen, Piepenbrock, and Gulikers, 1995), including very low frequency infinitives (e.g., *verzoeten* ‘to sweeten’, *bewolken* ‘to cloud over’, and *verzagen* ‘to saw up’*)*. The additional foils therefore also represented three other prefixes: *aan*-, *in*-, and *ont*-. In order to ensure that each of the five prefixes was presented 160 times in the entire experiment, we added 32 infinitives starting with *be*- and *ver*- and 480 infinitives carrying either the prefix *aan*-, *in*-, or *ont*- (see Table 5). The number of existing and pseudo-infinitives starting with *in*-, *aan*-, or *ont*- was unequal, as only a limited number of existing *ont*-infinitives are available. To avoid repetition of the prefixes *be*- and *ver*- only, 30 foils with the prefix *in*-, *aan*-, and *ont*- were also presented twice.

INSERT TABLE 5 ABOUT HERE

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

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

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

Results and Discussion

We analyzed the response times of Experiment 2 with the same method and predictors[ii] as used in Experiment 1 except for the predictor *experiment part*. Three participants and the word *bekransen* 'to garland' were excluded from analyses as their error rates were above 20%. Table 4 shows the statistical model based on the remaining 5111 trials (81.9% of all trials). The average response time was 981 ms (range: 525 to 2108 ms). All control predictors that were significant in Experiment 1 were also significant in this experiment.

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

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

Experiment 3

Method

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

Materials and procedure. We used the same stimuli and recordings as in Experiment 1. In contrast to the previous experiments, half of the trials in a stimulus list were produced by Speaker A and the other half by Speaker B. Furthermore, whereas primes were either reduced or unreduced, targets were always reduced. We created three different pseudo-randomizations of the trials with the same restrictions as Experiment 1. We created eight lists for each randomization by varying the variant and speaker of the prime and the speaker of the target, which resulted in 24 different stimulus lists. A prime and target were again separated by 67 trials on average. The procedure and duration of a session were identical to those of Experiment 1.

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

Results and Discussion

As the error rates of all participants were lower than 20%, no participants were excluded from analyses. The word *bekransen* 'to garland' had an error rate above 20% and was again omitted from further analyses. The statistical model of Experiment 3 was based on the remaining 2004 trials (85.2% of all trials) and is shown in Table 6. Response times were, on average, 943 ms (range: 540 to 1943 ms). The control predictors that were significant in the preceding experiments were again significant and showed similar effects. In addition, we found a difference between the two experiment parts, indicating that responses obtained in the second part (mean: 921 ms) were faster than in the first part (966 ms).

INSERT TABLE 6 ABOUT HERE

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

The difference found between Experiments 1 and 3 was supported by an analysis of the combined dataset of the responses to Speaker A in Experiment 1 and the responses in Experiment 3, which shows an interaction between *experiment* and *variant match* ($\beta = -0.031$, $t = -2.7$, $p < .05$): Whereas *variant match* explains variance in Speaker A's part of Experiment 1, it does not in Experiment 3. An analysis of the combined dataset of the responses to Speaker B in Experiment 1 and the responses in Experiment 3 gave the same result ($\beta = -0.023$, $t = -2.1$, $p < .05$). These results suggest that exemplar effects were greater in Experiment 1, if they were present in Experiment 3 at all.

Neither Experiment 2 nor Experiment 3 showed exemplar effects. Nevertheless, we decided to conduct Experiment 4, which is a combination of Experiments 2 and 3: Participants heard the targets in the same four conditions as in Experiment 3, while the experiment was identical to Experiment 2 in the number and diversity of the foils and in the distances between primes and targets. If the null results in Experiments 2 and 3 were due to Type II errors, we would expect to find exemplar effects in Experiment 4. Moreover, we can combine the results from Experiment 4 with those from Experiments 2 and 3 to see whether these increased datasets present evidence for exemplar effects.

Experiment 4

Method

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

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

Results and Discussion

To analyze the response times, we used the predictors[iv] from Experiment 3, except for *experiment part*. All participants and target words were included in the analyses, except for the target words *bekransen* 'to garland' and *beschaven* 'to civilize', as they had error rates higher than 20%. The statistical model of Experiment 4, based on the remaining 2459 trials (75.3% of all trials), is presented in Table 6. The average response time was 956 ms (range: 549 to 2444 ms). The same significant control effects were found as in Experiment 3, with the exception of *word frequency*.

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

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

Additional Analysis of All Experimental Data

So far, we investigated the presence of exemplar effects by analyzing the datasets with two categorical predictors (*variant match* and *speaker match*). The variation between a reduced and an unreduced realization differs between speakers (see Figure 2), word types, and word tokens. We therefore also analyzed all datasets with a continuous predictor indicating the similarity in reduction between the prime and target, namely the absolute difference between the log-transformed duration of the prime and the log-transformed duration of the target. Only the analysis of Experiment 1 showed a significant main effect of this continuous predictor ($\beta = 0.17$, $t = 2.2$, $p < .05$), indicating that responses were faster if the duration difference between prime and target was smaller. In addition, none of the experiments showed an interaction between this predictor and *speaker*. These results indicate that even a more sensitive predictor shows no exemplar effects in Experiments 2, 3, and 4, and thus confirm the results obtained with the categorical predictors *variant match* and *speaker match*.

General Discussion

In this paper, we investigated exemplar effects in a series of priming experiments with lexical decision tasks, in which primes and targets represented the same or a different pronunciation variant. We examined the robustness of exemplar effects under more natural conditions than in experiments reported in the literature so far (e.g., Craik and Kirsner, 1974; Goldinger, 1996; Palmeri et al., 1993) and did so in four ways. First, we studied the generalizability of exemplar effects over two very different speakers. Second, we investigated whether exemplar effects arise if the repetition of words is less clear for participants than in experiments showing exemplar effects reported in the literature. Third, we investigated if exemplar effects arise when listeners are exposed to not one but two types of pronunciation variation in the experiment (i.e., degree of reduction and speaker voice). Finally, in contrast to earlier studies, primes and targets were never completely identical.

In Experiment 1, 34% of the 288 trials formed word repetitions and each participant only listened to one of the two speakers. This experiment showed a clear exemplar effect: responses were faster to targets that represented the same pronunciation variant as their primes. In contrast to earlier studies (e.g., Mattys and Liss, 2008; McLennan et al., 2003; McLennan and Luce, 2005; Palmeri et al., 1993), in our experiments, primes and targets were always different recordings, even when they represented the same pronunciation variant produced by the same speaker. The results of Experiment 1 thus show that even if the target is not completely identical to the prime, its processing can be facilitated by the exemplar formed by its prime.

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

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

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

In Experiment 2, we made the repetition of words less obvious by simultaneously increasing the number of trials between a prime and target from 67 to 405 on average and reducing the proportion of trials forming word repetitions to 16%. This more closely approximates natural conversations, in which the frequent replacement of content words by pronouns decreases the repetition of content words. Although the statistical power of Experiment 2 was greater than Experiment 1 (due to the larger number of participants), Experiment 2 showed no exemplar effects. This indicates that exemplars effects are negligible when the repetition of words is less clear for participants.

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

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

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

Another possible explanation for the absence of exemplar effects has to do with the difference in reduction patterns between our speakers. As illustrated in Figure 2, a reduced pronunciation produced by Speaker A may be very similar in word duration to an unreduced pronunciation produced by

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

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

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

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

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

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

References

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

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

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

Dahan, D., Drucker, S., & Scarborough, R. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108*(3), 710-718.

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

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

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

Goldinger, S. (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 49-54). Saarbrücken, Germany.

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

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

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

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

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

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

McLennan, C., Luce, P., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 539-553.Oostdijk, N. (2002). The design of the Spoken Dutch Corpus. In P. Peters, P. Collins, & A. Smith, *New Frontiers of Corpus Research* (pp. 105-112). Amsterdam: Rodopi.

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

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

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

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

Appendix

Stimuli in All Experiments

Targets

begieten begluren begraven begrijpen begroeten bekladden beklimmen bekransen

bekrassen beschaven beschermen beschrijven besmeren bestelen bestoken bestraffen

bestralen besturen betasten betrappen bevriezen bevruchten bezingen bezorgen

verbannen verbranden vergeven vergrijzen verkiezen verklappen verkleumen

verkreuken vermoeien verprutsen verslapen verslikken verspelen versperren

verspreiden verstijven verstoten vertellen verteren vertolken vertragen vertrappen

vertrekken verzachten

Repeated Pseudo-Word Foils

bedangen bedinken begannen begoeren begranzen begruien beklegen bekonnen

bekrapen bekrempen benotten bepleuten beplonten beporken beschakken beschoeten

besmotten bestermen bestraaien betaffen betroeren bevichten bevrammen bezeiten

verbloffen verbrissen verdechten verdilgen verdoepen verdrooien verfalmen

verfrinsen vergippen vergoeten vergreuzen verguilen verklenen verknillen verkoezen

verscharpen versnallen verstoemen verteuven vertiemen vertilmen verwilken

verzekken verzwukken

Additional Foils in Experiments 1 and 3

Existing Words

bedanken bedaren begeren beginnen begroten beheksen beheren bejagen bekeren

bekronen beleggen belonen bemerken bereiden besmetten besmetten bestraten

betreffen betwisten bevolken bevrijden beweren bewerken bezetten verbergen verbouwen verdampen verdenken verduren verdwalen vergoeden vergokken vergroeien verkleuren verknallen verlangen vermaken vermengen verplaatsen verplichten verrekken verstikken verstoppen vertakken vertalen vertoeven verzenden verzinnen

Pseudo-Words

bedelken bedirven bedoeren begennen begrooien bekliegen bekreipen belamen bemonnen benoeten benuiden bepelen bepraven beristen beschekken beslatten bespraaien bestroeien betreuden bevengen bewirken bezekken bezieten bezoelen verbliffen verblijmen verbrussen verdetsen verdirven vergroemen vergussen verkirsen verknaren verloenken verlunken verniemen verpatten verpippen verrosten versmeuden versmieden verspallen vertoelen verwalken verwijpen verzoepen verzwekken verzweugen

Additional Foils in Experiments 2 and 4

Existing Words

aanbellen aanblazen aanblijven aanboren aanbraden aanbreken aanbrengen aandienen aandikken aandragen aandraven aandrijven aandringen aandrukken aanduiden aandurven aanduwen aangrijpen aanhaken aanhalen aanhangen aanharken aanhebben aanhechten aanheffen aanhoren aanhouden aankaarten aankijken aanklagen aanklampen aankleden aankloppen aanknopen aankomen aankopen aankunnen aanladen aanleggen aanleren aanliggen aanmaken aanmanen aanmelden aanmeren aanmerken aanmeten aannaaien aannemen aanpakken aanpappen aanpassen aanplakken aanplanten aanpoten aanpraten aanprijzen aanraden aanraken aanreiken aanrichten aanrijden aanroepen aanroeren aanrukken aanschaffen aanscherpen aansluiten aansnellen aansnijden aanspannen aanspoelen aansporen aanspreken

aanstampen aantreffen bedanken bedaren begeren beginnen begrenzen begroten

beheksen beheren bejagen bekeren bekronen beleggen belichten belijden belonen

bemerken beschimpen besluipen besluiten besmetten bespatten bestraten betrachten

betreffen betrekken betwisten bevallen bevolken bevrijden beweren bezatten

bezetten inbedden inbeelden inbeuken inblazen inblikken inbreken inbrengen

inchecken indammen indekken indelen indenken indeuken indienen indikken

indraaien indringen indrinken indrogen indruisen indrukken induiken indutten

induwen inenten ingieten ingooien ingrijpen inhouden inhuren inkerven inkijken

inklemmen inkleuren inkomen inkopen inkorten inkrimpen inladen inlassen

inleiden inleven inlezen inlichten inlijsten inlijven inlopen inlossen inluiden

inpakken inpassen inperken inplannen inprenten inprikken inrichten inroepen

inruilen inschenken inscheuren inschikken inseinen insluipen insluiten insneeuwen

insnoeren inspannen inspringen inspuiten instappen insteken instemmen instoppen

instromen intoetsen invallen ontbieden ontbijten ontbinden ontbloten ontbossen

ontbreken ontdekken ontdooien ontduiken onteren onterven ontfermen ontgelden

ontglippen ontgroeien ontgroenen onthalen ontharen ontheffen onthullen onthutsen

ontkennen ontkiemen ontkleuren ontkomen ontkrachten ontkurken ontladen ontlasten

ontlopen ontluiken ontnemen ontpitten ontplooien ontpoppen ontroeren ontroven

ontruimen ontschepen ontschieten ontsluiten ontsmetten ontsnappen ontspannen

ontsporen ontspringen ontstemmen ontstijgen ontstoppen ontvallen ontvellen

ontvetten ontvlammen ontvluchten ontvolken ontwaken ontzeggen ontzuilen

verbergen verbluffen verdampen verdenken verdoffen verdoven verduren verdwalen

vergelden vergoeden vergokken vergroeien vergulden verkleuren verklikken

verknallen verkopen verkroppen vermaken verpakken verplaatsen verplichten verrekken versjouwen verstikken verstoppen vertalen vertikken vertoeven vertonen verzenden verzinnen

Pseudo-Words

aandonnen aandriegen aandrieven aandrikken aangoeven aanhachten aanhiffen aanhoffen aankliegen aanklijzen aankloeden aankluigen aanknoepen aankuimen aanmekken aanmetten aanmoenen aanmolden aanmonnen aanmorken aanpatten aanpemmen aanpepsen aanpeuten aanplenten aanplikken aanploeken aanpretten aanproeten aanproetsen aanproezen aanprossen aanpruiten aanpuiten aanrakken aanrappen aanrijtsen aanriksen aanroeden aanruiden aanschieven aanschorpen aansnieden aansnoelen aansoeren aanspallen aansprokken aanstatten aanstempen aanstijken aanstoempen aanstopen aanstrappen aantekken aantisten aantoeken aantuiken aanvollen aanvregen aanwannen aanwienen aanwoeken aanwoezen aanwooien bededden bedirven bedoeren bedrakken bedwilmen begennen begronzen begrooien bekliegen bekreipen bemonnen benoeten bepelen bepraven beproeten beristen beschekken beslatten besnuien bespodden bespraaien bestarmen bestirmen bestroeien bestroeken betreuden bewirken bezanken bezekken bezeuken bezieten bezoelen inbeuten inbingen inboerken inbriksen inbummen indieken indinnen indoenen indommen indonken indremmen indrokken indrouwen indruigen indutsen ingitten ingoetsen ingoeven ingrannen inkarven inkeeuwen inkimmen inkloepen inkluipen inkummen inlijgen inlissen inloepen inloesten inloezen inmekken inmonten inplinnen inpranten inprekken inproetsen inraksen inralen inrannen inrensen inrutten inschekken inschoezen insienen inslappen insnieren inspienen insprangen instenden instijken instramen instreppen instreumen instrupsen intaben intaren intetsen intotsen intrieken invollen invruizen invuigen inweggen inwoetsen ontbatten ontblammen ontblitten ontboeden ontboesen ontbrikken ontbrinden ontbrotsen ontbunnen ontfilken ontfirmen ontflepsen ontgalden ontgirten ontglappen ontgleppen ontgratsen ontgreben ontgreeuwen ontgrennen ontgreunen ontgrooien ontgruinen onthatsen onthuitsen ontkarken ontkemmen ontkidden ontkloeren ontkoerken ontkrechten ontkrichten ontkrippen ontladden ontledden ontleppen ontlietsen ontlodden ontluitsen ontpatten ontpeuten ontplaaien ontpuiten ontschuipen ontslatten ontslitten ontsmatten ontsmoeren ontsmouten ontsmudden ontsmuiten ontsnieken ontsnippen ontsprengen ontsteugen ontstiepen ontstimmen ontstotsen ontstuigen ontveeuwen ontvelken ontveuzen ontvieten ontvilken ontvlachten ontvlichten ontvlimmen ontvloemen ontvollen ontvuiten ontwannen ontwienen ontwietsen ontwirren ontwoenen ontwoetsen ontworren ontwotsen ontwuinen ontzallen ontzelen ontzuiken verbiegen verbleuven verbliffen verbrussen verdappen verdetsen verdirven verdotsen verdwiezen vergellen vergeppen vergetten vergroemen vergussen verkessen verkirsen verkruggen verloenken verlunken verniemen verpatten verpippen verschorpen versloenzen versluppen versmeuden versmieden verspallen vertoelen verwalken verzwekken verzweugen

Repeated Existing Words

aanbakken aanbieden aanstellen aansterken aanstichten inbakken inbinden instellen instinken instorten ontaarden ontberen ontvouwen ontwarren ontwennen

Repeated Pseudo-Words

aanbreuden aanbrinden aandinnen aanstotten aanstruipen inbetsen inbieben inweuden inzanken inzuien ontbaaien ontbetten ontbissen ontziggen ontzoegen

Additional Foils in Experiment 4

Repeated Existing Words

Stimuli are the same as in Experiment 2 with the following additions:

bedwingen bereiden beschijnen beslissen betuigen bevinden bewerken bezuren verbouwen verdwijnen verlangen vermengen verplegen verschuilen vertakken verzuimen

Repeated Pseudo-Words

Stimuli are the same as in Experiment 2 with the following additions:

bedelken bekeugen belamen benuiden beseppen besmieren bevengen bezwuren verblijmen verknaren verlienen verrosten verslaten verwijpen verzanen verzoepen

Corresponding Address

Mirjam Ernestus

Radboud University Nijmegen & Max planck Institute for Psycholinguistics

P.O. Box 310

6500 AH Nijmegen

The Netherlands

m.ernestus@let.ru.nl

Acknowledgments

The authors would like to thank Lou Boves and Louis ten Bosch for their useful comments. This work was partly funded by a European Young Investigator Award and by an ERC consolidator grant (284108) to the third author.

Footnotes

[i] By accident, we included two identical existing words as foils (i.e. *besmetten*). We therefore also repeated one existing word foil.

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

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

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

Table 1

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

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

Table 2

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

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

Table 3

Statistical models of the phonetic analysis of the recorded primes and targets.

Fixed effects	Word duration			Presence of schwa		
	β	*t*	*p* <	β	*z*	*p* <
Prefix (ver)	56.78	4.01	.0001	0.75	1.97	.05
Speaker (Speaker A)	101.51	22.14	.0001	2.33	6.21	.0001
Variant (unreduced)	131.21	28.96	.0001	4.69	10.11	.0001
Speaker × variant	-53.20	-8.23	.0001	-	-	*n.s.*

Table 4

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

Fixed effects	*Experiment 1*			*Experiment 2*		
	β	*t*	*p* <	β	*t*	*p* <
Prefix (*ver*-)	0.06	5.9	.0001	0.08	7.23	.0001
Speaker (Speaker A)	0.13	4.9	.0001	0.08	4.91	.0001
Word frequency	-0.01	-2.3	.001	-0.02	-3.63	.0001
Target duration	0.39	14.2	.0001	0.40	19.26	.0001
RT prime	0.16	8.5	.0001	0.07	6.61	.0001
RT preceding trial	0.21	10.2	.0001	0.12	12.37	.0001
Variant match (mismatch)	0.02	3.5	.0001	-	-	*n.s.*

Random effects		*sd*		*sd*
Word type	intercept	0.03	intercept	0.03
Word type	RT preceding trial	0.08		
Participant	intercept	0.09	intercept	0.10
Participant			target duration	0.09
Residual		0.15		0.14

Table 5

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

Prefix	Primes	Targets	Repeated foils				Foils		Total	
	Existing		Existing		Pseudo		Existing	Pseudo	Existing	Pseudo
	B1	B2	B1	B2	B1	B2	B1/B2	B1/B2	B1/B2	B1/B2
be-	24	24			24	24	32	32	80	80
ver-	24	24			24	24	32	32	80	80
in-			5	5	5	5	76	64	86	74
aan-			5	5	5	5	76	64	86	74
ont-			5	5	5	5	58	82	68	92
Total	48	48	15	15	63	63	274	274	400	400

Table 6

Statistical models for the response times of Experiment 3 and 4. Estimated standard deviation is indicated by sd.

Fixed effects	*Experiment 3*			*Experiment 4*		
	β	*t*	*p* <	β	*t*	*p* <
Prefix (*ver-*)	0.05	3.9	.0001	0.05	4.07	.0001
Speaker (Speaker A)	0.08	7.9	.0001	0.08	11.09	.0001
Word frequency	-0.02	-2.3	.05	-	-	*n.s.*
Target duration	0.40	6.3	.0001	0.29	5.27	.0001
RT prime	0.12	6.2	.0001	0.11	3.76	.001
RT preceding trial	0.14	8.4	.0001	0.22	13.27	.0001
Experiment part (part 2)	-0.05	-4.6	.0001	-	-	*n.s.*

Random effects		*sd*		*sd*
Word type	intercept	0.05	intercept	0.04
Word type	speaker	0.05	RT prime	0.11
Participant	intercept	0.09	intercept	0.08
Participant	experiment part	0.06	RT prime	0.14
Residual		0.14		0.17

List of Figures

Figure 1. Examples of recorded stimuli: two unreduced and two reduced variants of vertolken /vərtɔlkə/ ‘to interpret’, produced by Speaker B.

Figure 2. Boxplot of word duration split to speaker and pronunciation variant.

Figure 1. Examples of recorded stimuli: two unreduced and two reduced variants of *vertolken* /vərtɔlkə/ 'to interpret', produced by Speaker B.

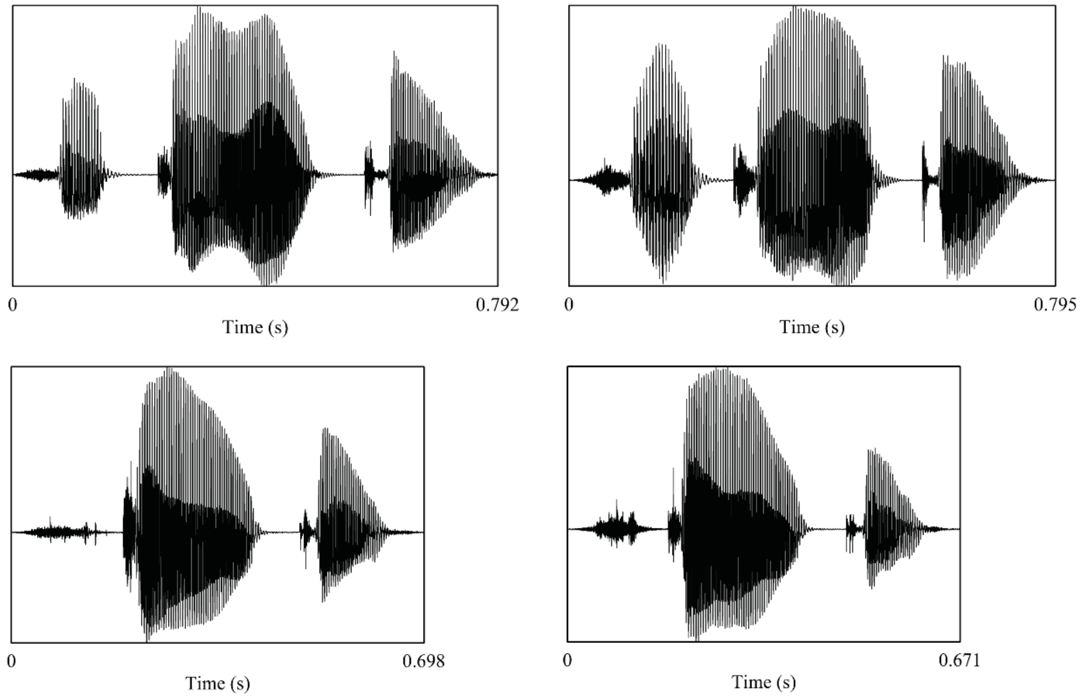

Figure 2. Boxplot of word duration split to speaker and pronunciation variant.