Testing the storage of prosody-induced phonetic detail via auditory lexical decision – a case study of noun/verb homophones

Arne Lohmann[1] and Benjamin V. Tucker[2]

[1]Universität Leipzig

[2]University of Alberta

Address for correspondence:

Arne Lohmann, Department of British Studies, Universität Leipzig, Beethovenstraße 15, 04107 Leipzig, Germany, Email: arne.lohmann@uni-leipzig.de

Abstract

This article reports the results of an auditory lexical decision task, testing the processing of phonetic detail of English noun/verb conversion pairs. The article builds on recent findings showing that the frequent occurrence in certain prosodic environments may lead to the storage of prosody-induced phonetic detail as part of the lexical representation. To investigate this question with noun/verb conversion pairs, ambicategorical stimuli were used that exhibit systematic occurrence differences with regard to prosodic environment, as indicated by either a strong verb-bias, e.g., *talk* (N/V) or a strong noun-bias, e.g., *voice* (N/V). The auditory lexical decision task tests whether acoustic properties reflecting either the typical or the atypical prosodic environment impact the processing of recordings of the stimuli. In doing so assumptions about the storage of prosody-induced phonetic detail are tested that distinguish competing model architectures. The results are most straightforwardly accounted for within an abstractionist architecture, in which the acoustic signal is mapped onto a representation that is based on the canonical pronunciation of the word.

Keywords: Homophones, grammatical category, phonetic detail, prosody

1. The Storage of Phonetic Detail

An important question in research on the mental lexicon, particularly debated in recent years, is whether lexical entries contain information about phonetic detail. This question is raised by the observation that the pronunciation of words exhibits a high degree of variability, especially in spontaneous speech. Ample reports demonstrate the production of multiple pronunciation variants for individual words, as a number of processes - often lumped together under the label ‘reduction’ - lead to phonetically very different realizations of the same word. Corresponding to that observation it has been found that putatively homophonous words exhibit systematic differences in pronunciation, raising the question of whether the same representation can really be assumed for them. One example of such an observation is Gahl (2008), who shows that high-frequency and low-frequency homophones, e.g., *time* vs. *thyme*, differ in acoustic realization. Another example is differences in pronunciation between homophones that differ in morphological complexity, e.g. *frees* vs. *freeze* (Seyfarth, Garellek, Gillingham, Ackerman, & Malouf, 2017).

These observations have raised questions regarding the nature of representations that are used during speech perception and word recognition in particular. One important question in this context is with respect to the level of detail assumed at the level of lexical form representation. Existing models differ as to whether detailed representations close to the surface phonetic signal are used, or whether representations are abstract (see, McQueen, 2005 and Mitterer & McQueen, 2009 for overviews). At the one end of the continuum, abstract form representations are assumed onto which the speech signal is mapped during word recognition. In such models, the abstract units of representation are often phonemes (e.g., Nearey, 2001), or phonologically underspecified feature representations (see Lahiri &

Marslen-Wilson, 1991), but representations do not contain information about the phonetic variability with which a word may be produced.

At the other end of the continuum are episodic, exemplar-based models in which the representation of a word is assumed to be made up of episodic memories of pronunciations of a word, i.e., individual tokens of experience with the word that are stored in a phonetically detailed fashion (Goldinger, 1997; Pierrehumbert, 2001). The distinction between abstract and detailed representations has implications for the question of whether pronunciation variants of a word are assumed to be stored. While the architectural properties of exemplar-based accounts can naturally account for the storage of phonetically detailed pronunciation variants, in abstractionist models usually only one abstract form is assumed to be stored that corresponds to the word’s canonical pronunciation (but see Ranbom & Connine 2007 for an abstract, phoneme-based account that allows for more than one phonologically-specified representation). So far, the evidence accumulated by prior research is mixed and does not unambiguously support a particular model architecture (see Ernestus, 2014 for an overview).

The present paper contributes to the discussion about the representation of phonetic detail by reporting a speech perception experiment of English noun/verb homophones that employs phonetically different pronunciation variants of these homophones. The English lexicon contains a large number of noun/verb pairs that are considered phonologically homophonous, e.g., *kiss* (N)/*kiss* (V), which are the product of the word-formation process of conversion between these two grammatical categories. A number of studies have documented systematic acoustic differences between the noun and the verb pronunciation of such pairs (e.g., (Conwell, 2017; Lohmann, accepted for publication; Sorensen, Cooper, & Paccia, 1978). The pronunciation differences observed are mainly the result of the difference in prosodic position nouns and verbs occur in. While prosody is usually considered a post-lexical process, recently, evidence has been obtained showing that prosody-induced phonetic detail may become part of the lexical representation (further discussed in section 1.1).

These findings raise the question of whether two phonetically different pronunciation variants are stored for noun/verb homophone pairs, a noun pronunciation and a verb pronunciation, which differ with regard to prosody-induced phonetic detail. Such a separate storage of different, phonetically rich representations is consonant with the assumption of exemplar-based models. In contrast, abstractionist models would assume the storage of a string of abstract symbols, most likely phonemes, which would be the same for the noun and the verb. The present article investigates these assumptions via a lexical decision task in which adult speakers of English process an ambicategorical noun/verb signal.

1.1 Prosody and Lexical Representation

As mentioned above, the main cause for phonetic differences found in noun/verb pairs such as *talk* (V) and *talk* (N) is the fairly systematic occurrence of nouns and verbs in different prosodic positions, this being the result of the different sentence positions the two grammatical categories occupy. Nouns occur more frequently in final position of larger syntactic constituents, such as phrases or clauses, which are usually co-extensive with prosodic constituents or domains (see Sorensen et al., 1978). Domain-final position can be considered a prosodically prominent position, characterized by prosodic effects of domain-final lengthening and accentuation (see (Lohmann, accepted for publication; Sorensen et al., 1978). Due to their frequent occurrence in that position in discourse, nouns are characterized by corresponding acoustic characteristics, for example, a systematically greater acoustic duration of nouns as compared to verbs.

While acoustic differences brought about via sentence-level prosody are usually regarded to be a post-lexical process, initial evidence has been obtained suggesting that prosody-induced phonetic detail affects lexical representation. Schweitzer et al. (2015) show that the phonetic implementation of a prosodic phenomenon, in their case the occurrence of a particular tone on a word, is influenced by the frequency with which the particular lexical item occurs with this prosodic event. This finding suggests an at least partly “lexicalised storage of intonation” (Schweitzer et al., 2015: 67), in which the frequent occurrence of a word in a certain prosodic environment affects its representation. Corresponding with that assumption, Sóskuthy and Hay (2017) show that words which frequently occur in utterance-final position - a position which induces strong domain-final lengthening effects - exhibit a lengthened pronunciation even when occurring in other positions within the utterance. These findings are in line with the idea that phonetic detail initially triggered by sentence prosody becomes stored in the lexicon. This storage is driven by frequency, as the frequent co-occurrence of a form in a certain prosodic environment seems to be essential. Such a ‘rub-off’-effect of sentence prosody onto representation can be accounted for within the architecture of exemplar-based approaches to the mental lexicon, which explicitly implement the frequency-sensitive storage of phonetic detail (e.g., Goldinger, 1997; Pierrehumbert, 2001). In exemplar-based models each word is considered to be represented by a ‘cloud’ of exemplars consisting of many individual tokens of experience. This exemplar cloud may comprise different, phonetically rich pronunciation variants whose strength of representation is contingent on how often a particular pronunciation variant is processed. Evidence supporting this idea has been provided by Bürki, Ernestus, and Frauenfelder (2010) and Ranbom and Connine (2007). Their findings suggests that each pronunciation variant is separately represented and there is a "frequency count for each variant" (Bürki et al., 2010: 421). However, the phenomena investigated in these studies involved pronunciation variants that differed on the segmental level. The present paper tests whether similar effects can be found when the pronunciation variants differ only with regard to subphonemic detail.

If these variants were part of the speaker's representation what would be the assumptions for the representation of noun and verb pronunciations of a pair such as *talk* (N/V)? Depending on the frequency of noun versus verb pronunciations for a pair, either the prosodically prominent noun pronunciation or the less prominent verb pronunciation should be more strongly represented in the lexicon. In that regard it is important to note that noun/verb pairs differ considerably with regard to the frequency of each grammatical category. For example, while *voice* is considerably more frequent as a noun, *talk* exhibits a strong verb-bias. Within an exemplar-based framework, *talk* and *voice* should therefore differ with regard to the representation of the two pronunciation variants. In contrast, within abstractionist frameworks as described above, these pronunciation variants would not be expected to be stored, as these models do not assume the storage of information below the phoneme and would typically assume only the representation of the canonical pronunciation. Frequency differences between the noun and the verb would therefore be irrelevant for representation in such frameworks.

1.2 The Present Study

The present study contributes to the exploration of the frequency-sensitive storage of pronunciation variants of noun/verb conversion pairs via an auditory lexical decision task in which participants react to the presentation of an ambicategorical signal e.g., *talk*. The stimuli used are recordings of pairs that show a strong frequency bias toward one of the two categories, i.e., the noun or the verb of the pair has a considerably greater discourse frequency. Within the architecture of exemplar-based models, one of the two pronunciation variants can therefore be assumed to have a stronger, i.e., more entrenched representation than the other. The competing model architectures discussed above give rise to different assumptions with regard to how fast and accurately these stimuli are processed. Within the architecture of exemplar-based models, which assume a frequency-sensitive storage of pronunciation variants, one would expect that the pronunciation variant of a form that matches the category-bias of the pair is processed faster and more accurately. Based on this assumption we can formulate Hypothesis 1.

Hypothesis 1: The pronunciation variant of a form that matches the category-bias of the corresponding N/V pair leads to more accurate responses and faster reaction times. For example, the recording of a verb-biased pair, e.g., *talk*, should be reacted to faster when presented in verb pronunciation than a noun-biased stimulus, e.g. *voice*.

Within an abstractionist architecture, there should be no such effect, as phonetically detailed pronunciation variants are not assumed to be stored.

In abstractionist accounts, one would assume that the speech signal is mapped onto units at an abstract representational level, e.g., phonemes, in the process of lexical access. (see McQueen 2005, Ernestus, 2014). Phonetic detail may influence this process but not in the same way as expressed in H1. A number of studies have shown that the better the match between the speech signal and the stored representation, the faster and more accurate the recognition process (see McQueen, 2005: 257-263). Since in abstractionist accounts the representation is assumed to be based on a canonical pronunciation, and the noun pronunciations, due to their being pronounced in a prosodically more prominent position are phonetically less reduced, it should be easier to map this form onto an abstract representation. What may further contribute to the advantage of noun pronunciations is that these match the prosodic context of presentation better. The noun pronunciations were produced in utterance-final position, that is, before a marked prosodic boundary, while the verb pronunciations were produced in utterance-medial pronunciation with no prosodic boundary following. In the lexical decision task, the stimuli are presented in isolation, so that a prosodic boundary after the word is expected, which is an expectation fulfilled by the noun pronunciations. In sum, within abstractionist accounts it can be assumed that the mapping process of the signal onto an abstract level of representation is more effortless for the noun as compared to the verb pronunciations, which leads to the expectation expressed in hypothesis 2.

Hypothesis 2:

The noun pronunciation variants are processed faster and more accurately.

Note that the two hypotheses are not mutually exclusive. It is possible that supporting evidence for both effects is found, i.e., a general effect of noun pronunciations being easier to process and the interaction between category-bias and pronunciation variant expressed in Hypothesis 1. Such an outcome would not clearly favor one or the other model architecture, but may suggest a hybrid architecture that combines abstract generalizations and exemplar storage (see also Ernestus, 2014: Section 4). We will turn to a discussion of possible model architectures when discussing the results.

2. Method

2.1 Participants

In total 76 participants were recruited, all of whom were undergraduate students at the University of Alberta in Edmonton, Canada, who participated in the experiment in return for course credit. All participants were native speakers of Western Canadian English and gave informed consent in written form. Of the 76 participants, 59 were female and 17 male and the average age was 20.6 years.

2.2 Creation of stimuli

Stimuli are monosyllabic noun/verb conversion pairs with a filled onset that show a strong frequency bias toward either the noun or the verb and were selected with the help of the CELEX database (Baayen, Piepenbrock, & van Rijn, 2001). Usage frequencies were extracted from the COCA corpus (Davies, 2014). The final selection comprises a total of 52 target word pairs, of which 26 were noun-biased and 26 verb-biased (see appendix for a complete list of stimuli words). The operationalization of the category-bias was that at least

80% of all uses of the corresponding lemma in COCA were of the dominant category. Pseudoword filler items were created using the Wuggy software (Keuleers & Brysbaert, 2010), creating monosyllabic pseudowords that were phonotactically possible words of English and whose pseudoword status was only determined by the coda consonant(s), e.g., *juft* or *telk*. Target and filler items were read by a female native speaker of Western Canadian English, born and raised in Edmonton, Canada. All pairs were pronounced in a domain-final noun and a domain-medial verb environment by embedding them in sentences. For example, for the noun/verb pair *phone* (N) / *phone* (V), the words were pronounced in the following sentences: *Robin has a new phone.* vs. *She tried to phone her father*. All sentences were pronounced three times and the clearest pronunciation, based on auditory inspection, was selected and spliced out of the audio signal in order to be presented in isolation in the experiment.

2.3 Acoustic Analysis of Stimuli

As mentioned above, the most important prosodic difference between nouns and verbs is the much greater likelihood for nouns to occur in final position of prosodic domains. This difference is reflected in the two different target word productions in which an ambicategorical string was pronounced, see above. What are the acoustic consequences of these differences in position? Generally, domain-final position is a position that is prosodically more prominent than domain-medial position: words that occur in domain-final position are typically accented and are affected by domain-final lengthening. Regarding the latter, one would therefore expect a considerable difference in overall duration between the noun and the verb pronunciation. Furthermore, it is important to note that domain-final lengthening affects in particular the vowel of the stressed syllable and the final coda consonants (Turk & Shattuck-Hufnagel, 2007). Hence, in the case of monosyllabic stimuli, duration differences should be primarily visible in the rhyme, but not in the onset consonant(s). Differences in the relative duration between the rhyme and the onset of lengthened versus non-lengthened pronunciations can therefore be expected. These should manifest themselves in a lengthened rhyme of the noun pronunciations as compared to the verb pronunciations, while the onset should be of similar duration between the two grammatical categories. In a related perception study, Kemps, Ernestus, Schreuder, and Baayen (2005) show that shifts in relative duration between onset and rhyme affect the processing of Dutch plural suffixed or unsuffixed word stems. They argue that the relative durations of individual syllable parts are represented in the lexicon and therefore affect processing in case an unexpected relative duration is processed by the listener. The same logic can be applied to the processing of noun/verb homophones in the present experiment, as these shifts in relative durations may be a prosodic effect that affects lexical representation (see section 1.1 above). Since the stimuli were chosen so as to instantiate ambicategorical strings with a strong frequency-bias toward either the noun or the verb, one of the two pronunciations can be considered to be less expected than the other one, as it is less commonly encountered in discourse and should consequently have a weaker representation.

In order to test whether the recordings of the stimuli are characterized by the aforementioned differences, we analyzed their overall duration as well as the relative durations of the individual syllable parts. In particular, the duration of the onset and of the rhyme were measured, as the noun pronunciations should be marked by a lengthened rhyme in particular, which should lead to smaller ratio of the onset relative to the overall duration of the word. In a first step, the stimuli were segmented, marking word and phoneme level boundaries, by employing the automatic forced alignment software MAUS (Kisler, Reichel, & Schiel, 2017). Word and segment durations were then calculated using Praat scripts (Boersma & Weenink, 2016). In order to illustrate the shifts in relative duration between the rhyme and the onset, an onset ratio was calculated by dividing the duration of the onset by the overall word duration for the recording of each word. The two images in Figure 1 illustrate the verb (top) and the noun (bottom) pronunciation of *stay*.

INSERT FIGURE 1 HERE

The two pronunciations in Figure 1 are marked by clear differences in overall duration and different relative durations of the onset and the vowel, with a greater onset ratio for the verb pronunciation (upper panel). This pattern holds also for the other N/V pairs, see the mean word durations and mean onset ratios reported in Table 1. Since domain-final position is also correlated with accentuation, acoustic correlates of accentuation may also be detectable. Since accentuation means a pitch accent on the target word, more F0 movement should be detectable in the noun as opposed to the verb pronunciations. We calculated the degree of F0 movement by subtracting the minimum F0 from the maximum F0 (in Hz). The result of this calculation is a considerably greater F0 range of the noun as compared to the verb stimuli. See below Table 1, which provides the mean values based on all stimuli words for F0 range, word duration and onset ratio, along with results from *t*-test calculations (two-sided, paired *t*-tests).

INSERT TABLE 1 HERE

2.4 Procedure

The experiment was conducted in a sound-attenuated booth at the Alberta Phonetics Laboratory at the University of Alberta in Edmonton, Canada. The experiment was implemented in the experimental presentation software E-Prime 2.0 Professional (Schneider, Eschman, & Zuccolotto, 2015). Responses were indicated by pressing buttons on a Serial Response Box.

Participants were presented with 52 target word pairs and 52 pseudoword pairs, both presented in the noun and the verb pronunciation. This means that each participant processed 104 target items and 104 filler items. To avoid the presentation of the same N/V pair with the two types of pronunciations in close succession, the experiment consisted of two blocks. The two blocks differed with regard to the type of pronunciation that was chosen for a particular noun/verb type. In each block, of the 52 target word types, 26 were presented with the verb pronunciation and 26 with the noun pronunciation. The two types of pronunciation were equally distributed across the verb-dominant and the noun-dominant groups. That is, within each dominance group, half of the auditory stimuli (=13) were noun pronunciations and the other half verb pronunciations. In the second block of the experiment the noun/verb pair by pronunciation assignment was reversed. That means, if the noun pronunciation of *plant* was presented in the first block, in the second block the verb pronunciation of *plant* was presented. Within each experimental block the word and pseudoword recordings were presented in random order. The order of the two blocks was randomly varied between participants (41 participants saw the one order, 35 participants the other order). Stimuli were presented over MB Quartz QP headphones calibrated with a 1kHz tone to a level of 81dB. While the experiment was conducted so that each participant reacted to both pronunciation variants for each N/V pair, due to the strong possibility of priming effects, in the ensuing analysis only the results of the first block for each participant are reported. The analyses of the Block 2 data as well as analyses based on the entire dataset are provided in the form of supplementary material accompanying this article. With regard to the hypotheses tested, the results of the data of the second block do not differ from the results of the data from the first block, which are reported here. See also the brief discussion in the supplementary material.

At the beginning of each experimental session participants were provided with brief instructions displayed on the computer screen in front of them. They were asked to decide whether a given pronunciation is a word of English or not. They were instructed to press a button with their dominant hand to select “word” and a button with their non-dominant hand to indicate “not a word”. The presentation of each stimulus was preceded by a 500ms “+” fixation mark at screen center. Participants were given 3000ms to respond after which the presentation timed out and the next stimulus recording was presented. After completion of the experiment, the participants were asked a number of questions about their demographic and linguistic background.

2.5 Modeling reaction times and accuracy

In order to test for effects of the variables of interest, mixed-effect models were built and fitted to the reaction times and the accuracy of classification. Regarding the models fitted to reaction times, two different models were built: Model 1 was fitted to the reaction times measured from the stimulus onset, Model 2 was fitted to the reaction time measured from the stimulus offset. Both methods are used in auditory lexical decision tasks (see e.g., Ussishkin, Dawson, Wedel, & Schluter, 2015). Considering the reaction time from the offset of the stimulus presentation is particularly relevant in the present context, since it is a way to control for differences in duration between the stimuli (Goldinger, 1996). As shown above, the noun pronunciations are of considerably greater duration compared to the verb pronunciations, which is likely to affect reaction times, if measured from stimulus onset.

In order to analyze accuracy, a mixed-effects logistic regression model was built predicting whether the stimulus was correctly classified as a word. All models calculated tested the following fixed-effect predictors:

Pronunciation variant: Noun vs. verb pronunciation (see sections 2.2 and 2.3 for details)

Category dominance: Noun vs. verb dominance (whether the word pair is more frequently used as a noun or as a verb, see section 2.2)

String frequency: Frequency of the word pair (string frequency counting both noun and verb instances) in the Corpus of Contemporary American English (Davies, 2014). This variable was log-transformed before it entered the model.

Concreteness ratings: As a further control variable, concreteness ratings for all noun/verb pairs were retrieved from Brysbaert et al. (2014). Words denoting more concrete concepts have been shown to be responded to faster across a variety of experimental paradigms including auditory lexical decision tasks (e.g., Tucker et al., 2019). It is well known that nouns and verbs differ with regard to concreteness (e.g., Black & Chiat, 2003), so that the same difference may also hold between noun-dominant and verb-dominant N/V pairs. This turns out to be the case with our set of stimuli, as the noun-dominant pairs are characterized by a concreteness rating of 4.411 on average compared to an average concreteness rating of 3.373 of the verb-dominant pairs.

3. Results

Response accuracy and reaction times (see Procedure in Section 3.4 above) were extracted from the E-Prime software. Only the data points that presented reactions to the target stimuli were considered for analysis, excluding reactions to the filler items. For the analysis reported below only first block responses were considered (see Section 2.4.). The data was inspected with regard to the distributional properties of the response times. Measured from stimulus onset, the response times exhibit an almost normal distribution, except for a small percentage of very early responses, viz. <200ms after the onset of the stimulus. Since these very early responses disturb the pattern of a normal distribution and are very likely button presses that do not indicate a proper processing of the signal, data points with response times below that threshold were excluded from further analysis, which resulted in the omission of 0.3% of the data. The analysis of the accuracy of responses is based on this slightly trimmed dataset. For the analysis of response times, only those data points were considered that were accurately classified as words, which resulted in the exclusion of a further 12% of the data, so that the reaction time analysis was based on 3,434 button presses.

The observed reaction times are provided in the form of box plots that are overlain with beeswarm scatterplots showing individual datapoints (see Politzer-Ahles & Piccinini, 2018). These plots visualize the distribution of the onset and offset reaction times by category dominance and pronunciation variant.

INSERT FIGURE 2 HERE

The plots for the onset and the offset reaction times show essentially the same pattern: There seems to be a general effect of verb pronunciations to be processed more slowly (compare the left and the right boxplot within each panel). Secondly, this effect of verb pronunciations to be pronounced more slowly seems to be slightly more pronounced in the group of verb-dominant pairs (compare left versus right panels). Crucially, the observed reaction times do not indicate an interaction effect of faster reactions to pronunciation variants that matched the category-bias of the stimulus.

As stated above, also the accuracy of responses was analyzed. The following bar plot (Figure 3) illustrates accuracy rates by category dominance and pronunciation variant.

INSERT FIGURE 3 HERE

The results for accuracy correspond to the results for the reaction times. Generally, accuracy is lower for the verb pronunciations than the noun pronunciations. This effect seems to be stronger in the group of verb-dominant words.

To statistically analyze the data, mixed-effect models were built employing the *lmer* function of the *lme4* (Version 1.1.19, Bates, Maechler, Bolker, & Walker, 2014) and *lmerTest* packages (Version 3.0.1, Kuznetsova, Brockhoff, & Christensen, 2014) in *R* (Version 3.5.1, R Core Team, 2014). Models were fitted with all aforementioned variables as fixed-effects, including an interaction term between *pronunciation variant* (noun vs. verb pronunciation) and *category dominance* (noun vs. verb). Note that hypothesis 1 predicts an interaction effect between these variables: reaction times should be shorter (and accuracy greater) for pronunciation variants that match the category-bias of the stimulus. Random intercepts for subject and word pair were added, as well as random slopes for *pronunciation variant* by subject and word pair. The model was checked with regard to potentially problematic collinearity by calculating Variance Inflation Factors using the *vif.mer* function. The calculation shows that collinearity is not an issue with VIFs <2.5.

Model fitting was conducted by backward exclusion of non-significant fixed-effects, tested via the comparison of models with and without the crucial predictor employing likelihood-ratio tests. The random effects structure in both models was not trimmed during model fitting, following a design-based approach (Barr, Levy, Scheepers, & Tily, 2013). The reaction time models were fitted to the log-transformed reaction times. During model fitting the variable *string frequency* was removed from both models. *Concreteness* yielded a significant or marginally significant effect and was therefore kept in the models reported. *Pronunciation variant* and *category dominance* are either significant as main effects and/or involved in a significant interaction with each other so that these variables and their interaction was kept. See the output of both models in the following tables (Table 2 and 3).

INSERT TABLES 2 AND 3 HERE

In both models fitted to reaction times, concreteness is a marginally significant main effect, indicating that words with a higher concreteness rating are responded to faster, as expected. The effect of pronunciation variant indicates a slower response to verb pronunciations in both models. This main effect is statistically significant only in the model fitted to offset reaction times. Category dominance is not significant as a main effect in either model, but its interaction with pronunciation variant is. This interaction term indicates that the effect of slower responses to verb pronunciations is particularly strong in the group of verb-dominant words. See also Figure 4 below which illustrates the effect of pronunciation variant in the individual conditions.

A logistic regression model was fitted to whether the stimulus was accurately responded to or not. The results of this analysis are provided in the following table. Positive coefficients indicate a heightened probability of accurate responses. Due to issues of non-convergence, the random slopes for category had to be omitted from this model.

INSERT TABLE 4 HERE

The regression output (Table 4) shows a significant effect of concreteness, with more concrete words being more accurately classified. Pronunciation variant is significant as a main effect showing that verb pronunciations are less accurately classified. This effect interacts with category dominance (see significant interaction term) indicating that verb pronunciations are especially prone to be wrongly classified in the group of verb-dominant pairs.

The following figure (Figure 4) plots the effects of pronunciation variant by category dominance in the three models reported. The interaction effect of pronunciation variant by category dominance reported in the three models can be gleaned from the uniformly steeper slope for the effect of pronunciation variant in the group of verb-dominant pairs as compared to the slope for the effect in the noun-dominant group.

INSERT FIGURE 4 HERE

4. Discussion

Having presented the results, we will now interpret them in light of the hypotheses tested. Hypothesis 1 predicts an interaction between category dominance and pronunciation variant, in the sense that the pronunciation variant matching the category-specific frequency bias of the stimulus item should result in faster response times and higher accuracy. The lexical decision experiment did not yield evidence for this effect. While the statistical analysis shows an interaction between category dominance and pronunciation variant, it is not in the direction predicted. Contrary to the predictions of hypothesis 1, the interaction effect shows that the general effect of verb pronunciations being reacted to more slowly and less accurately is particularly pronounced in the group of verb-dominant words. We will turn to an explanation of this result below.

The predictions of hypothesis 2 are largely borne out. In general, noun pronunciations are responded to faster and more accurately. Regarding response times, this effect is less pronounced in the reaction time data measured at stimulus onset compared to the offset reaction times. This may be explained by the systematic difference in stimulus durations. The short duration of the verb pronunciations compared to the noun pronunciations means that the presentation was faster for the verb pronunciations and therefore afforded the participants more time to react if reaction times are measured from the onset, which may mitigate the general effect of verb pronunciations to be responded to more slowly.

The effect of a faster and more accurate recognition of the noun pronunciation variant can be explained within models that assume the storage of just one, abstract form representation for both the noun and the verb of the pairs. Within abstractionist models, lexical representations of words are strings of abstract symbols, for example phonemes, which in most versions of such models are representations of the canonical pronunciation (see McQueen 2005). Such model architectures would thus assume identical representations of the noun and the verb of the tested pairs that do not contain information about phonetic detail. During speech processing, pronunciation variants are mapped onto this stored representation (see McQueen 2005, Ernestus, 2014). In the present experiment, this mapping process can be assumed to differ between the noun and the verb pronunciation. Recall that the noun pronunciations were produced in prosodically prominent domain-final position, while the verb pronunciations occurred in domain-medial contexts. The phonetic effects of this difference are acoustically shorter pronunciations with a more abrupt ending for the verb pronunciations as compared to the noun pronunciations, as the former were spliced out of a continuing intonational phrase. The noun pronunciations can therefore be considered to be more similar to the canonical pronunciation than the verb pronunciations. The degree of similarity of a stimulus to the representation based on the canonical pronunciation can be assumed to affect the mapping process, in the sense that the greater the phonetic distance between the stimulus and the representation, the less accurately and time-consuming this mapping will be (Ernestus, 2014). As expressed by McQueen (2005: 260) “lexical representations are activated in a graded fashion, in response to their goodness of fit to the available input." Since the noun pronunciation variant is a better fit to an abstract representation based on the canonical pronunciation, the mapping process will be easier and faster to perform, which corresponds to the results obtained. The same negative correlation between phonetic distance and processing speed has been observed in studies on the processing of reduced speech. Reduced pronunciations of words take longer to process and are harder to identify when presented in isolation (e.g., Kemps, Ernestus, Schreuder, & Baayen, 2004; Tucker, 2011) and can be assumed to exhibit a considerable phonetic dissimilarity to a form representation that is based on the unreduced pronunciation of the word (Ernestus, 2014).

In summary, the difference in processing ease between the verb and the noun pronunciation variant can be explained within an abstractionist architecture. Such an account can also explain the observed absence of the interaction effect expressed in hypothesis 1, as pronunciation variants differing with regard to phonetic detail would not be assumed to be stored.

In discussing the theoretical implications of the present findings, it is important to contextualize the present findings against the backdrop of previous studies which investigate the effect of phonetic detail on lexical processing. First, in our opinion, the present failure to find an interaction between pronunciation variant and frequency of instantiation does not contradict previous findings of frequency effects of pronunciation variants as reported by Ranbom and Connine (2007) and Bürki et al. (2010). In these studies, the pronunciation variants differed on the segmental level, unlike the stimuli in the present study. Both the findings of the present as well as of these previous studies could be accounted for within abstractionist models that allowed for the storage of phonologically specified pronunciation variants, but do not assume the storage of phonetic detail.

Two further relevant studies are Conwell (2015) and Conwell & Morgan (2012). While, unlike the present study, neither of these studies investigates frequency effects of pronunciation variants, both articles report differences in processing between phonetically different noun and verb pronunciations of an ambicategorical stimulus. Based on their findings, both Conwell & Morgan (2012) and Conwell (2015) entertain the possibility of storage of phonetic detail. However, it is important to note that their results do not force one to adopt that conclusion, which is acknowledged by the authors. Conwell & Morgan (2012) show that infants react to the changing acoustics between noun and verb pronunciations of N/V pairs in a habituation task. This effect could be a mere reaction to a change in the acoustic signal, but does not necessarily require the assumption of representational differences (see also Conwell & Morgan, 2012: 109). Conwell (2015) finds differences in the neural response between noun and verb pronunciations in an ERP experiment and fails to find this effect when testing nonce-words as controls. She therefore argues that the effect is unlikely to be only a reaction to acoustic differences but speaks to differences in lexical representation. However, since the processing of nonce words can be considered to be shallower than the processing of real words, it remains possible that the effect seen in real words is a reaction to the change in the acoustic signal, which is not visible for nonce words as these may not be processed with the same attention to detail. In conclusion, the results obtained by the two studies could be explained without necessarily assuming a phonetically detailed storage of pronunciation variants.

A further relevant line of research can be found in studies by Salverda, Dahan, & McQueen (2003), Kemps et al. (2004, 2005) and Blazej & Cohen-Goldberg (2015). In these studies, it was shown that listeners exhibit robust effects of attending to phonetic detail in speech perception. However, it is important to note that these studies investigated effects of a prosodic match or mismatch between an acoustic stimulus and a target. For example, Salverda et al. (2003) show that listeners are sensitive to durational cues in polysyllabic words. They show that when a free-standing pronunciation of the word *ham* is spliced into the word *hamster*, participants in an eye-tracking paradigm exhibited more fixations of the competitor picture of *ham*, compared to a condition in which the pronunciation of *ham* was spliced in from the embedded word *hamster*. The acoustic difference between the two conditions is that *ham* is produced either with word-final lengthening when the word *ham* is produced, or with polysyllabic shortening when *ham* is part of the word *hamster*. The experiment thus compared conditions in which one type of stimulus presented a prosodic mismatch to one that presented a prosodic match, as the lengthened pronunciation of the syllable *ham* does not match the prosodic environment in which it is part of the prosodic word *hamster*. Similar findings have been reported by Kemps et al (2004, 2005) and Blazej & Cohen - Goldberg (2015). These studies show that phonetic detail brought about by the prosodic environment can serve as a cue for lexical disambiguation. However, this is not compelling evidence for storage of phonetic detail, but can be interpreted as the listeners possessing knowledge of prosodic processes, which is utilized in word recognition (see also the discussions in Salverda et al. 2003, Shatzman & McQueen, 2006 and McQueen 2005: 262-263).

The current experiment is different from these studies as it was designed to test a hypothesis about the frequency of exposure of certain pronunciation variants, not merely a prosodic match/mismatch condition. It is important to note, however, that the present experiment's results are also informative with regard to effects of prosodic mismatch on word recognition. The verb pronunciations which were produced without pre-boundary lengthening due to their being produced in mid-sentence position can be considered a mismatch to the presentation context in which words were presented in isolation, as in this context a prosodic boundary after the word would be expected. The result of a clear effect of verb pronunciations being responded to more slowly and less accurately may thus also be driven by a prosodic mismatch effect in which word boundaries are not cued by lengthening, which reduces their activation (see McQueen 2005: 263). From this perspective the results of the current study therefore tie in well with previous work, as they provide evidence for an effect of a prosodic match/mismatch with the environment of presentation, but do not necessitate the assumption that phonetic detail is part of the lexical representation (see also Shatzman & McQueen, 2006).

Nevertheless, the results also allow for an alternative interpretation, namely that phonetic detail is stored but is not accessed in the hypothesized way in the current experiment, as its access could be sensitive to context. Cohen & Kang (2018) argue that the perceptual system in word recognition is flexible in the sense that not all potentially available cues are always used, but their employment depends on the nature of the task. More specifically, Cohen & Kang (2018) put forth a return-on-investment account. This account states that "that listeners allot cognitive resources to process linguistic information with an eye towards which information is most helpful in decoding the speech signal" (Cohen &

Kang 2018: 66). For the role of phonetic detail in word recognition this may mean that it is employed when its use presents a gain in the task at hand. The task reported here, which was a lexical decision task, required a holistic processing of the entire word to match it with possible entries in the lexicon. In such holistic processing, attention to phonetic detail may be less useful than in other tasks that require attention to smaller, sublexical components (see also Cohen & Kang 2018: 69-70). In the experiments by Kemps et al (2004, 2005), Salverda et al. (2003) and others (see above), an acoustic signal ambiguous between two competing words or wordforms was presented, which was disambiguated by the presence or absence of either a second syllable or a suffix. In such tasks it was therefore important to pay close attention to any cues to the presence/absence of these constituents to disambiguate the signal. It is therefore possible that stored phonetic detail is capitalized upon in these studies, but not in the present one. In summary, an alternative explanation for the present results is an account that proposes the flexible use of acoustic detail depending on task demands.

A finding that remains to be discussed is that the verb-dominant stimuli in the experiment were especially strongly affected by the presentation of a pronunciation variant differing from the canonical pronunciation. We note, however, that this finding is restricted to the Block 1 data, which means that it is of a somewhat spurious nature (see supplementary material of this article). An explanation for this effect may have to do with the widely-reported observation of verbs being harder to process than nouns, which has been reported for a wide variety of tasks (see Gentner 1981, Vigliocco, Vinson, Druks, Barber, & Cappa, 2011 for overviews). This effect has been shown to not be a direct effect of grammatical category, but to be caused by semantic differences between nouns and verbs, with nouns denoting more concrete, imageable referents, typically objects, compared to verbs, which typically denote actions. The difference in processing ease may be particularly relevant in a lexical decision task, as typical action-denoting verbs require the expression of further arguments in discourse, so that a presentation in isolation can be considered to be a less natural presentation for verbs as compared to nouns. While the stimuli we used were ambicategorical, it is likely that the difference in category dominance of the N/V pairs employed is correlated with semantic differences that increase the processing ease of the noun-dominant pairs we used in the experiment. Our statistical control of the semantic dimension of concreteness may not have been sufficient to capture these differences in their entirety. If, as seems plausible, the verb-dominant stimuli are harder to process, a presentation in a pronunciation variant fairly dissimilar to the canonical pronunciation may be particular taxing for the processing system and therefore result in low accuracy and slow recognition speed. Conversely, the slowing-down effect of a pronunciation dissimilar to the canonical form may not unfold the same force with noun-dominant words, as their processing is facilitated by semantic properties. We further investigated the idea of verb-dominant ambicategorical stimuli being harder to process than noun-dominant ones by analyzing reaction times to N/V homophone pairs in the MALD database (Tucker et al., 2019). In MALD, in which all words are presented in citation form pronunciation, verb-dominant ambicategorical stimuli are processed about 15msec more slowly than noun-dominant stimuli, which supports the claim that the observed processing difference between nouns and verbs can also be detected in ambicategorical noun/verb pairs that differ with regard to category dominance.

5. Conclusion

The present article reports the results of an auditory lexical decision task in which noun and verb pronunciations of ambicategorical stimuli are presented. Two hypotheses differentiating between the assumptions of an exemplar-based model and the ones of an abstractionist model were tested. The results are most straightforwardly accounted for within an abstractionist architecture, in which the acoustic signal is mapped onto an abstract representation that is based on the canonical pronunciation of the word.

Funding information

This work was supported by the Deutsche Forschungsgemeinschaft under grant LO-2135/1-1.

Acknowledgements

We are grateful to Melina Sinclair for assistance in creating the experiment reported. We thank the members of the DFG Research Unit FOR 2373 “Spoken Morphology” for discussing the ideas presented in this article on multiple occasions. Furthermore, we are indebted to the members of the Alberta Phonetics Laboratory, in particular Filip Nenadic and Matthew Kelley, for providing helpful feedback and support. The research reported was approved by the University of Alberta Research Ethics Board.

References

Baayen, H. R., Piepenbrock, R., & van Rijn, H. (2001). WebCelex. Online resource. Retrieved from http://celex.mpi.nl

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Black, M., & Chiat, S. (2003). Noun-verb dissociations: a multi-faceted phenomenon. *Journal of Neurolinguistics*, *16*, 231–250.

Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 50–68. https://doi.org/10.1037/a0038509

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer software]. Retrieved from http://www.praat.org

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one “fenêtre” in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, *62*(4), 421–437. https://doi.org/10.1016/j.jml.2010.01.002

Cohen, C., & Kang, S. (2018). Flexible perceptual sensitivity to acoustic and distributional cues. *The Mental Lexicon*, *13*(1), 38–73. https://doi.org/10.1075/ml.16029.coh

Conwell, E. (2015). Neural responses to category ambiguous words. *Neuropsychologia*, *69*, 85–92. https://doi.org/10.1016/j.neuropsychologia.2015.01.036

Conwell, E. (2017). Token Frequency Effects in Homophone Production: An Elicitation Study. *Language and Speech*, 23830917737108. https://doi.org/10.1177/0023830917737108

Conwell, E., & Morgan, J. L. (2012). Is It a Noun or Is It a Verb? Resolving the Ambicategoricality Problem. *Language Learning and Development*, *8*(2), 87–112. https://doi.org/10.1080/15475441.2011.580236

Davies, M. (2014). The Corpus of Contemporary American English: 450 million words, 1990-2012.

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

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41.

Goldinger, S. D. (1996). Auditory Lexical Decision. *Language and Cognitive Processes*, *11*(6), 559–568. https://doi.org/10.1080/016909696386944

Goldinger, S. D. (1997). Words and Voices: Perception and Production in an Episodic Lexicon. In K. Johnson (Ed.), *Talker variability in speech processing* (pp. 33–66). San Diego: Academic Press.

Kemps, R. [Rachel], Ernestus, M., Schreuder, R., & Baayen, H. R. (2005). Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition*, *33*(3), 430–446.

Kemps, R. [Rachèl], Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*(1-3), 117–127. https://doi.org/10.1016/S0093-934X(03)00425-5

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech and Language*, *45*, 326–347.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest [Computer software]. Retrieved from http://cran.r-project.org/web/packages/lmerTest/index.html

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological appraoch to the recognition lexicon. *Cognition*, *38*, 245–294.

Lohmann, A. (accepted for publication). Nouns and verbs in the speech signal - are there phonetic correlates of grammatical category? *Linguistics*.

McQueen, J. M. (2005). Speech Perception. In K. Lamberts & R. L. Goldstone (Eds.), *Handbook of cognition* (pp. 255–275). London: SAGE.

Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 244–263. https://doi.org/10.1037/a0012730

Nearey, T. M. (2001). Phoneme-like units and speech perception. *Language and Cognitive Processes*, *16*(5-6), 673–681. https://doi.org/10.1080/01690960143000173

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. J. Hopper (Eds.), *Typological studies in language: v. 45. Frequency and the Emergence of Linguistic Structure* (pp. 137–157). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Politzer-Ahles, S., & Piccinini, P. (2018). On visualizing phonetic data from repeated measures experiments with multiple random effects. *Journal of Phonetics*, *70*, 56–69. https://doi.org/10.1016/j.wocn.2018.05.002

R Core Team (2014). R: A Language and Environment for Statistical Computing} [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ranbom, L., & Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. https://doi.org/10.1016/j.jml.2007.04.001

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. https://doi.org/10.1016/S0010-0277(03)00139-2

Schneider, W., Eschman, A., & Zuccolotto, A. (2015). E-Prime Professional (Version 2.0) [Computer software]. Pittsburgh: Psychology Software Tools Inc.

Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication*, *66*, 65–81. https://doi.org/10.1016/j.specom.2014.09.006

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, *88*(2), 1–18. https://doi.org/10.1080/23273798.2017.1359634

Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, *17*(5), 372–377. https://doi.org/10.1111/j.1467-9280.2006.01714.x

Sorensen, J. M., Cooper, W. E., & Paccia, J. M. (1978). Speech timing of grammatical categories. *Cognition*, *6*(2), 135–153. https://doi.org/10.1016/0010-0277(78)90019-7

Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition*, *166*, 298–313. https://doi.org/10.1016/j.cognition.2017.05.032

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, *39*(3), 312–318. https://doi.org/10.1016/j.wocn.2010.12.001

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The Massive Auditory Lexical Decision (MALD) database. *Behavior Research Methods*, *51*(3), 1187–1204. https://doi.org/10.3758/s13428-018-1056-1

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001

Ussishkin, A., Dawson, C. R., Wedel, A., & Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, *30*(9), 1096–1115. https://doi.org/10.1080/23273798.2015.1005635

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Appendix

Stimuli words (grouped by category dominance)

Noun-dominant	Verb-dominant
rule	keep
plant	might
wheel	kill
store	raise
tape	feel
heat	wipe
rock	grope
cash	shake
fish	hide
trip	stay
sense	push
star	pick
cloud	reach
film	pay
branch	win
voice	broach
house	cheat
term	stop

soap	feed
phone	look
box	rinse
milk	dig
book	talk
side	help
group	spread
school	need

Table 1

Mean values of noun and verb pronunciations of stimuli words for the acoustic parameters word duration, onset ratio, and F0 range

	Noun pronunciation	Verb pronunciation	*t*	*df*	*p*
Mean word duration (ms)	354.2	214.5	17.7	103	<0.001
Mean onset ratio	0.275	0.346	-8.4	103	<0.001
Mean F0 range (Hz)	90.2	45.1	3.0	86	<0.05

Table 2

Model output of mixed-effect regression model fitted to offset response times (N=3,434)

Random effects	Variance	Std. Deviation
Speaker (intercept)	0.047	0.217
Speaker (pron. variant)	0.005	0.068
Word pair (intercept)	0.027	0.166
Word pair (pron. variant)	0.029	0.169
Residual	0.114	0.338

Fixed effects	Coefficient estimate	Std. Error	*t*	*p*
Intercept	6.220	0.119	52.128	
Category dominance = verb	0.007	0.055	0.130	0.90
Pronunciation variant = verb	0.331	0.038	8.727	<0.001
Concreteness rating	-0.492	0.025	-1.936	<0.1
Category dominance = verb x Pronunciation variant = verb	0.111	0.053	2.092	<0.05

Table 3

Model output of mixed-effect regression model fitted to onset response times (N=3,434)

Random effects	Variance	Std. Deviation
Speaker (intercept)	0.013	0.113
Speaker (pron. variant)	0.002	0.045
Word pair (intercept)	0.005	0.068
Word pair (pron. variant)	0.011	0.103
Residual	0.047	0.216

Fixed effects	Coefficient estimate	Std. Error	*t*	*p*
Intercept	6.811	0.058	116.819	
Category dominance = verb	-0.030	0.025	-1.194	0.239
Pronunciation variant = verb	0.007	0.023	0.288	0.775
Concreteness rating	-0.025	0.013	-1.970	<0.1
Category dominance = verb x Pronunciation variant = verb	0.078	0.033	2.390	<0.05

Table 4

Model output of logistic regression model fitted to accuracy of responses (N=3,939)

Random effects	Variance	Std. Deviation
Speaker (intercept)	0.447	0.668
Word pair (intercept)	1.812	1.346

Fixed effects	Coefficient estimate	Std. Error	*z*	*p*
Intercept	1.294	1.174	1.102	
Category dominance = verb	0.994	0.523	1.901	<0.1
Pronunciation variant = verb	-0.946	0.187	-5.063	<0.001
Concreteness rating	0.535	0.260	2.061	<0.05
Category dominance = verb x Pronunciation variant = verb	-1.792	0.271	-6.612	<0.001

Figure 1. Verb pronunciation (upper panel) and noun pronunciation (lower panel) of *stay* (the vertical line indicates the boundary between the onset and the vowel)

Figure 2. Reaction times by category dominance and pronunciation variant (Upper panel: reaction times measured from stimulus offset, Lower panel: reaction times measured from stimulus onset; the horizontal line in the boxplot indicates the median value, the blue dot indicates the mean value; the differently colored transparent dots are scatterplots of individual datapoints)

Figure 3. Accuracy rates by category dominance and pronunciation variant (whiskers indicate +-1 standard error around the mean)

Figure 4. Partial effects plot illustrating the effect of pronunciation variant by category dominance as reported in Models 2-4 (upper panel Model fitted to offset RTs, reported in Table 2, medium panel Model fitted to onset RTs, reported in Table 3, lower panel Model fitted to accuracy of responses, reported in Table 4)

Figure 1

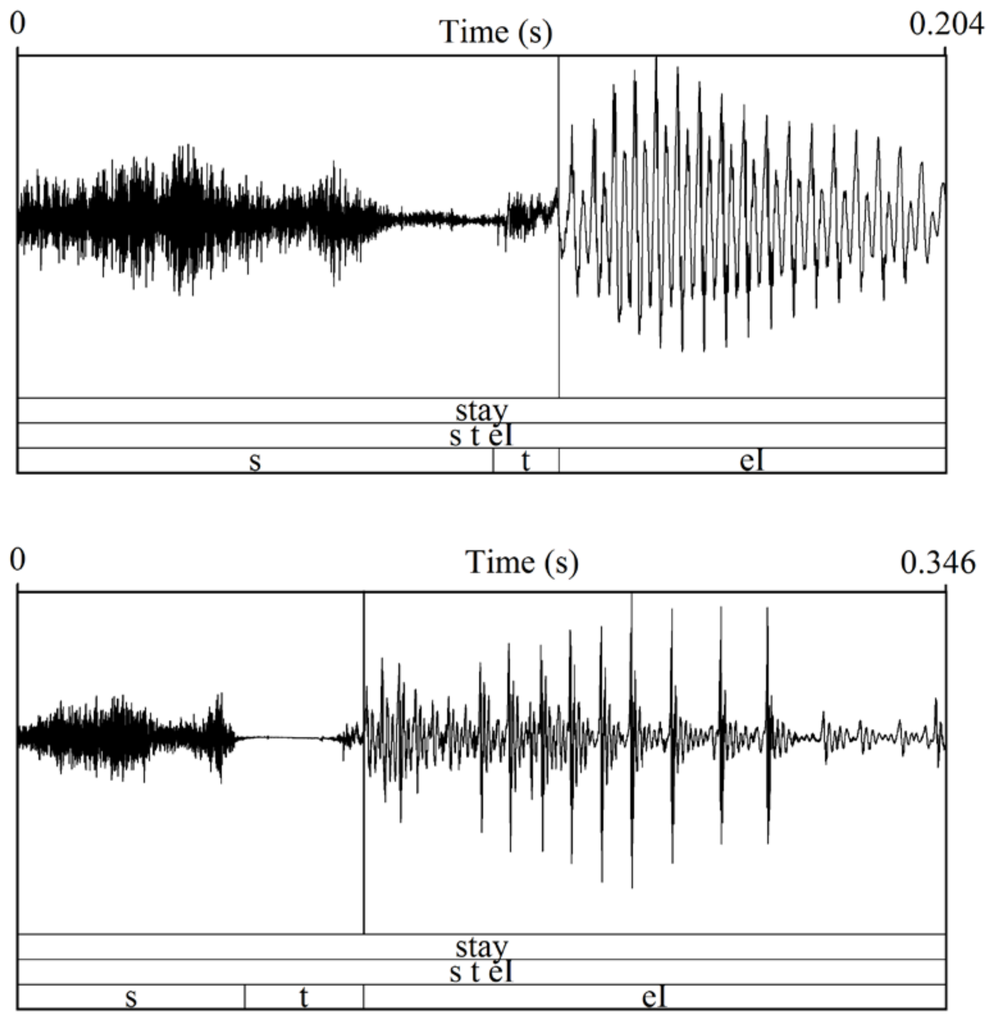


Figure 2

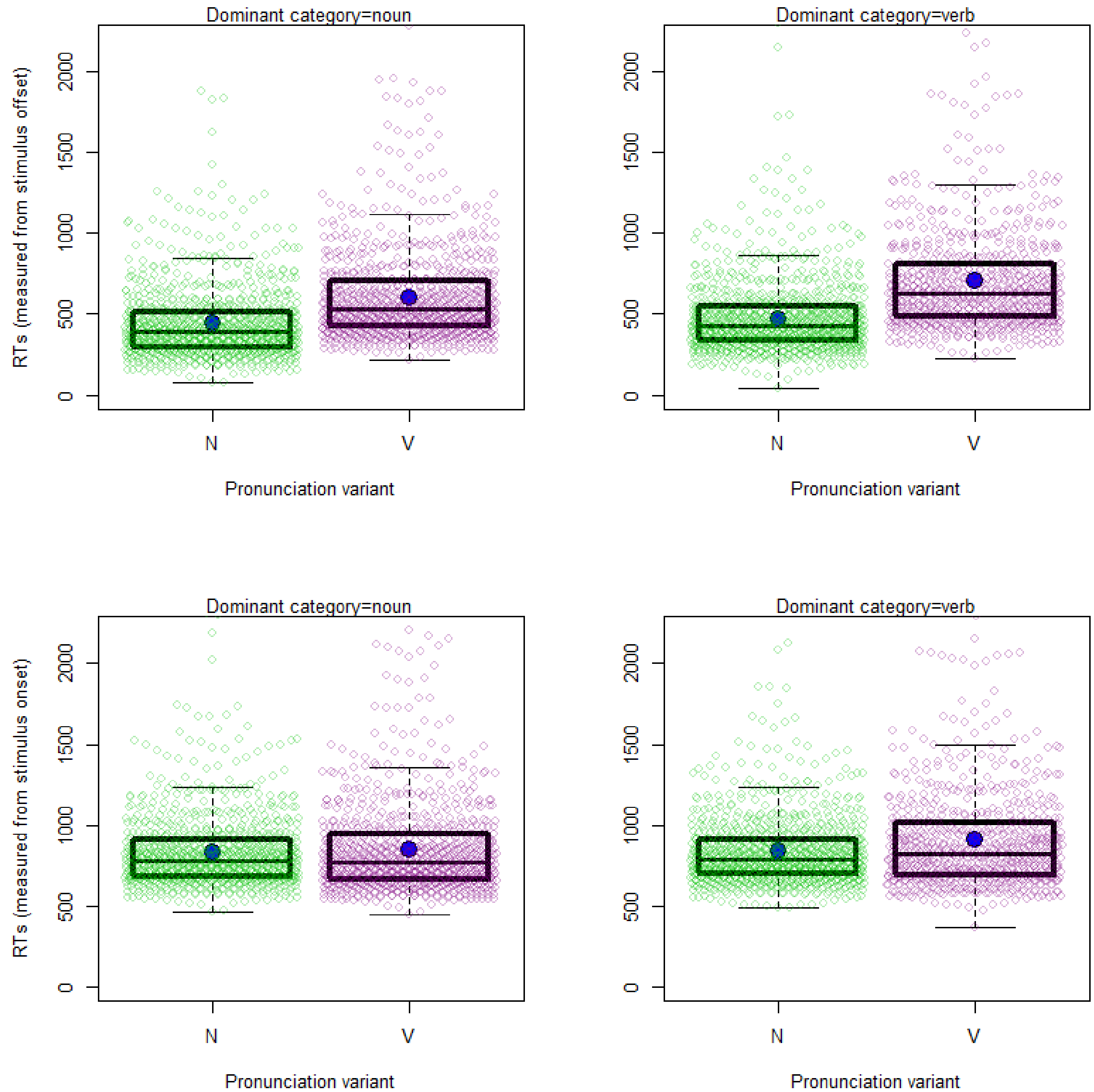

Dominant category=noun
Dominant category=verb
RTs (measured from stimulus offset)
RTs (measured from stimulus onset)
0
500
1000
1500
2000
N
V
Pronunciation variant

Figure 3

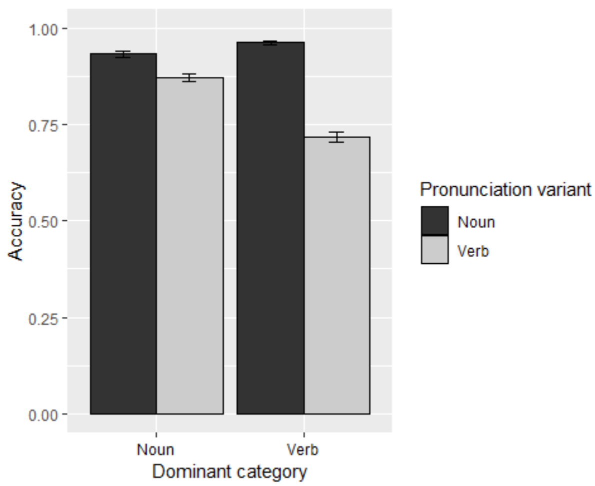
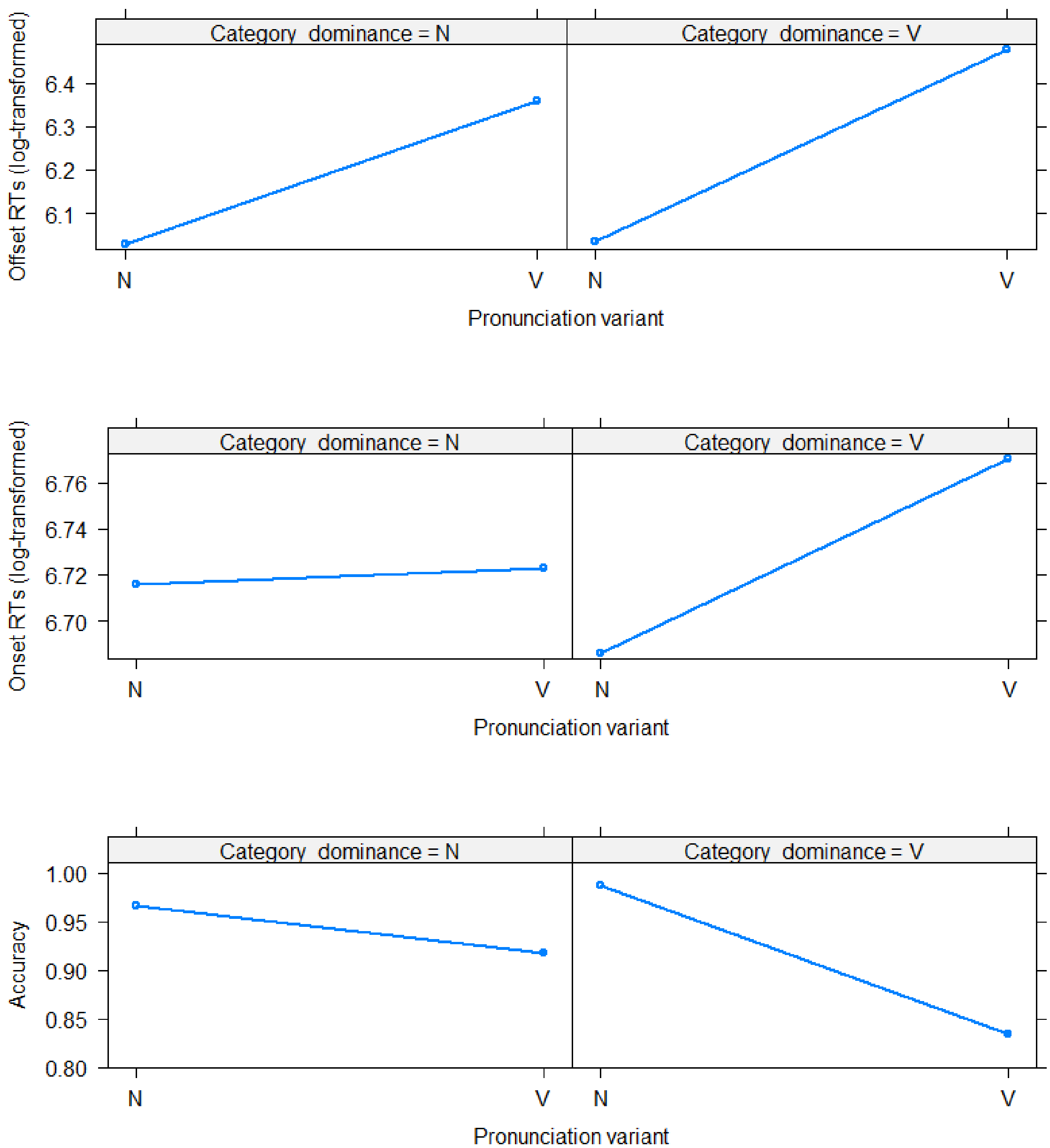

1.00
0.75
0.50
0.25
0.00
Accuracy
Noun
Verb
Dominant category
Pronunciation variant
Noun
Verb

Figure 4


Category dominance = N
Category dominance = V
Offset RTs (log-transformed)
6.4
6.3
6.2
6.1
N
V
N
V
Pronunciation variant
Category dominance = N
Category dominance = V
Onset RTs (log-transformed)
6.76
6.74
6.72
6.70
N
V
N
V
Pronunciation variant
Category dominance = N
Category dominance = V
Accuracy
1.00
0.95
0.90
0.85
0.80
N
V
N
V
Pronunciation variant