Cut(N) and *cut*(V) are not homophones: Lemma frequency affects the duration of noun–verb conversion pairs[1]

ARNE LOHMANN

Heinrich-Heine-Universität Düsseldorf

This paper tests whether lemma frequency impacts the duration of homographic noun–verb homophones in spontaneous speech, e.g. *cut*(N)/*cut*(V). In earlier research on effects of lemma frequency (e.g. Gahl 2008), these pairs of words were not investigated due to a focus on heterographic homophones. Theories of the mental lexicon in both linguistics and psycholinguistics differ as to whether these word pairs are assumed to have shared or separate lexical representations. An empirical analysis based on spontaneous speech from the Buckeye corpus (Pitt et al. 2007) yields the result that differences in lemma frequency affect the duration of the N/V pairs under investigation. First, this finding provides evidence for N/V pairs having separate representations and thus supports models of the mental lexicon in which lexical entries are specified for word class. Second, the result is at odds with an account of 'full inheritance' of frequency across homophones and consequently with speech production models implementing inheritance effects via a shared form representation for homophonous words. The findings are best accounted for in a model that assumes completely separate lexical representations for homophonous words.

KEYWORDS: conversion, frequency effects, homophones, speech production, word class

1. INTRODUCTION

It is generally acknowledged that frequency has a reductive effect on the pronunciation of words (see Jurafsky 2003 for an overview). How frequency affects the pronunciation of homophones, is, however, less clear. The question is whether homophone pairs are subject to one cumulative frequency effect or whether individual word frequencies lead to differences in their pronunciation. To empirically decide between these alternatives is of great theoretical importance,

---

as it sheds light on the representational status of homophones, i.e. whether homophones have separate lexical representations, or, at least partly, share the same representation.

In a much-noticed article, Gahl (2008) reports that homophonous (but heterographic) pairs of words, such as *thyme* and *time*, are not pronounced the same, but differ in duration contingent on their individual lemma frequencies, with the high-frequency words being pronounced with shorter duration than their low-frequency twins. Other studies failed to find such an effect on homophone pronunciation (e.g. Jurafsky, Bell & Girand 2002, see Section 1.3).

The present study tests whether individual lemma frequency impacts the duration of homographic noun–verb pairs, such as *cut*(N) and *cut*(V) or *face*(N)/*face*(V). These N/V pairs represent an especially interesting group of homophones for such an investigation because of their potential to further our understanding of the mental lexicon. One reason for this lies in the frequency of homophonous N/V pairs. A calculation based on CELEX data (Baayen, Piepenbrock & van Rijn 2001) reveals that approximately 20% of all English nouns and verbs have homophonous counterparts in the other word class. Testing the lemma frequency hypothesis on this group thus means exploring its validity for a substantial share of the lexicon. A second reason is that the representational status of N/V homophones is especially controversial in both linguistic and psycholinguistic theories of the mental lexicon.

1.1 *The status of word class in the mental lexicon*

In linguistic theory, homophonous noun–verb pairs are at the center of the discussion on the status of word class in the mental lexicon. In most standard accounts, pairs like *cut*(N)/*cut*(V) are viewed as resulting from the word-formation process of conversion or zero-derivation, in which a new word instantiating a different word class is derived from a source word (e.g. Quirk et al. 1990: 1558–1567; Bauer, Lieber & Plag 2013: 545–567). These accounts assume homophonous N/V pairs to have two separate lexical representations, with the lexicon consisting of entries specified for word class. However, this view is not universally endorsed. In alternative accounts, it has been argued that word class is an epiphenomenon that comes about through the use of a word in context, but is not represented in the lexicon. Homophonous N/V pairs are assumed to share just one lexical entry. This view is also known as 'lexical underspecification' (Barner & Bale 2002: 775). Broadly speaking, two approaches to underspecification can be distinguished, which differ as to whether syntactic or semantic aspects of word class are emphasized. A view that focuses on the former is put forth within the framework of *Distributed Morphology* (Halle & Marantz 1993, Marantz 1997). In this framework, it is argued that entries in the lexicon are roots that 'are category neutral' (Barner & Bale 2002: 772), with their syntactic category becoming specified only through their use in a sentence. In the lexicon, however, there are 'no nouns, no verbs' (Barner & Bale 2002: 771).

2

Similar proposals have been made in cognitive-functional theories, which focus on semantic aspects of the notion of word class. For example, building on Langacker's *Cognitive Grammar*, Farrell (2001) argues that pairs such as *cut*(N) and *cut*(V) are 'neither nouns nor verbs' in the lexicon, but share a semantically underspecified entry. Word class is only provided through morphosyntactic context, which triggers a 'contextually imposed profiling scenario' (Farrell 2001: 128) and leads to either a 'thing' or a 'process' interpretation for the noun and the verb use, respectively (see Velasco 2009 for essentially the same argument implemented in the framework of *Functional Discourse Grammar*; see Don 2004 for a critique of underspecification).

Both *Distributed Morphology* and the cognitive-functional approaches share the assumption that lexical entries are not specified with regard to word class. Given the assumption of just one shared representation for both the noun and the verb of an N/V homophone pair, information about class-specific usage should not be stored in the lexicon. Consequently, such approaches would not predict N/V homophones to exhibit effects that are due to class-specific frequencies.

1.2 *Homophone representation in language production models and the lemma frequency effect*

In psycholinguistic theory, and models of language production in particular, lexical representations are generally assumed to be specified for word class. However, the representational status of N/V homophones is still debated in such models due to their status as homophones, whose lexical representation is especially controversial. Frequency effects on homophones have a special place in language production research because of their potential to decide between rivaling model architectures. In the following, I will briefly contextualize the lemma frequency hypothesis tested here within the context of different assumptions about homophone representation (for further discussions of these questions, see e.g. Gahl 2008, Middleton, Chen & Verkuilen 2015).

A possible lemma frequency effect is of great theoretical significance because it stands in contrast to so-called 'frequency inheritance', which states that a low-frequency word with a high-frequency homophone 'inherits' the frequency of the high-frequency word. This may play out in an equally high resistance to error, increased speed in naming or the same word duration, as tested here.

Proponents of the inheritance effect explain it via a partly shared representation of homophonous word pairs. In many dominant production models, lexical retrieval is assumed to involve two stages: (a) the retrieval of a semantic and grammatical representation of the word, followed by (b) the retrieval of the phonological form. In the production model by Levelt (1989) (also Levelt, Roelofs & Meyer 1999), probably the most influential model in language production research, these representational stages are termed the lemma and the wordform stage. In this model, effects of frequency inheritance are explained

3

by homophonous words sharing the same wordform due to their identity in phonological form. This shared wordform is claimed to be the only locus of the word frequency effect (Jescheniak & Levelt 1994, Levelt et al. 1999), with the result of homophones being affected by the cumulative frequency of both words. This view is also termed 'full frequency inheritance'.

Since the evidence for inheritance effects is contested, other researchers propagate a view of 'no inheritance' of frequency and therefore do not implement such effects in their models. In these rivaling models, completely separate representations for homophonous words are assumed (e.g. Caramazza et al. 2001, Miozzo & Caramazza 2005). Separate representations mean that both words are subject only to their individual frequencies.

An effect of lemma frequency as found by Gahl (2008, 2009) demonstrates that individual word frequencies impact word duration. It can be explained in an account of 'no inheritance' and thus in models that assume separate representations for homophonous words. This effect is at odds with full frequency inheritance and consequently with models that assume a shared wordform that is the only locus of frequency effects, as in the influential proposal by Jescheniak & Levelt (1994).

A third possibility is what has been termed 'partial inheritance' (see Middleton et al. 2015). This would assume that while there is an influence of lemma frequency, there may still be a certain degree of frequency inheritance between homophones, so that a low-frequency member of a homophone pair would still be influenced by its high-frequency twin, however, without completely mitigating the lemma frequency effect. Such a partial effect is possible in production models that assume a shared representation of homophones at the wordform level, but with frequency affecting not just the wordform but also the lemma level (see Kittredge et al. 2008, Middleton et al. 2015).

What are the predictions with regard to word durations of these three alternatives? Full frequency inheritance would predict the same word durations of homophones, when contextual variables are controlled for. No inheritance would mean that the low-frequency member of a homophone pair should be pronounced with greater duration, contingent on the difference in frequency, with no influence of the frequency of the more frequent twin. Partial inheritance would predict an effect of lemma frequency which leads to a greater duration of low-frequency homophones relative to their high-frequency counterparts, but with the latter still yielding an effect on their duration.

The question of inheritance and representation is especially acute in the case of the N/V pairs under investigation, as these are not only homophonous but also homographic. In previous research, it has been discussed whether homographic homophones exhibit a more pronounced susceptibility to frequency inheritance effects. This could be due to an effect of feedback from a shared orthographic representation activating the phonological forms of both words (see Bonin & Fayol 2002, Gahl 2008). Another possibility is a more general representational difference between homographs and heterographs: heterographic homophones

4

may have separate phonological representations, but homographs may not, as they also share the same orthographic representation (see Biedermann & Nickels 2008). Both possibilities predict that the N/V pairs investigated are characterized by a stronger effect of frequency inheritance, and conversely a weaker or no effect of lemma frequency.

The discussion of the alternative inheritance effects depicted and the different model architectures that may implement them will be taken up again when discussing the empirical results obtained (see Section 4).

1.3 *The present study in the context of previous research about frequency effects on homophone duration*

Frequency effects on the duration of homophones have been investigated in reading experiments, as well as in spontaneous speech collected in corpora. The empirical results are mixed for both. Based on an analysis of data from a list reading paradigm, Whalen (1991) reports a shorter duration of high-frequency words relative to homophones of lower frequency. Guion (1995) reports a similar positive finding for homophones embedded in constructed sentences, but a negative finding when the words were pronounced in citation form in generic carrier phrases. Cohn et al. (2005) fail to find an effect of lemma frequency on duration, testing the pronunciation of homophones both in constructed sentences and also read off lists. All of the experimental studies tested heterographic homophones. These were content words in the case of Guion (1995) and Cohn et al. (2005). Whalen (1991) tested a mix of content and function words.

In previous corpus-based research, effects of lemma frequency on duration were tested on content word homophones, e.g. *thyme* versus *time* (Gahl 2008, 2009), and function word homophones (Jurafsky et al. 2002, Jurafsky 2003). While in Gahl's studies a positive finding for lemma frequency is reported, Jurafsky et al. fail to find empirical support for such an effect. There is one corpus-based study testing frequency effects on N/V homophones in child-directed speech (Conwell 2016), which reports a marginally significant effect of lemma frequency on word duration. An investigation from the regular speech of adults is yet missing.

It is not clear why some studies found duration differences contingent on lemma frequency while others did not, as there is no clear pattern emerging from the differences in results. One possible reason, mentioned in Gahl (2008), may be that function and content words exhibit a differential sensitivity to the lemma frequency effect, since all positive findings are from studies that tested content words or a combination of content and function words, while the comparison of function words yielded null-results (Jurafsky et al. 2002, Jurafsky 2003). Other reasons may be methodological in nature (see the discussion in Gahl 2008: 477–479).

5

Testing the lemma frequency hypothesis on N/V homophones extends previous research along two important dimensions. N/V homophones are content words, the class of words concentrated on mostly. However, in contrast to the homophone pairs tested in previous studies, the N/V pairs are homographs. As noted in the previous section, whether homographic and heterographic homophones are equally susceptible to the lemma frequency effect is a point of discussion. A second dimension potentially interacting with frequency effects is the semantic relation between the homophones. The words tested in previous research were homonyms, e.g. *thyme* versus *time* in Gahl ([2008](#)), i.e. were unrelated semantically. Relatedness in meaning may, however, facilitate frequency inheritance (see the discussion in Jescheniak, Meyer & Levelt [2003](#)). Noun–verb conversion pairs are clearly related in meaning, having come about through a derivational word-formation process (compare the meanings of *cut*(N)/*cut*(V), *attack*(N)/*attack*(V), etc.). In sum, compared with previously tested homophones, noun–verb conversion pairs provide a more stringent testing ground for the lemma frequency hypothesis due to the similarity of the words investigated.

This article tests the lemma frequency hypothesis on N/V homophones from the spontaneous speech of speakers of American English. To preview the results, the main finding is that differences in lemma frequency affect the pronunciation of these homophones, contra frequency inheritance. In an additional analysis, only the low-frequency subsample is tested as to whether the frequency of the high-frequency counterparts influenced the duration of the words in this subsample, as under an assumption of 'partial inheritance'. No evidence for such an effect is found.

2. DATA AND METHOD

2.1 *Data*

The present study is based on data from the Buckeye corpus (Pitt et al. [2007](#)), which contains the spontaneous speech of 40 adults from Columbus, Ohio. In order to extract the durations of suitable N/V pairs, I first compiled a list of search strings consisting of (i) all phonologically homophonous[2] and homographic noun–verb lemma pairs from WebCELEX (Baayen et al. [2001](#)) and (ii) the collection of noun–verb/verb–noun conversion pairs by Bram ([2011](#)). I extracted frequency information from the corpus for all items featuring in one or both of these data sources, both uninflected and including the inflectional ending *-s*, as this can feature on both nouns and verbs. In order to have a reasonable amount of data for both words of each pair, a threshold of at least five occurrences per word-class-specific word was implemented. To arrive at a list of words fulfilling

---

[2] This means also excluding pairs exhibiting suprasegmental differences, so that N/V pairs differing with regard to stress position, e.g. *inCREASE*(V) versus *INcrease*(N) were not considered.

6

that criterion, I first identified those word pairs that occur at least ten times in the corpus and for which at least one noun and one verb occurrence is attested, based on the word class information provided in the corpus annotations. Then, I manually coded word class for all tokens retrieved and kept for further analysis only those pairs for which both words surpassed the threshold of five occurrences. This selection procedure resulted in a list of 63 N/V pairs (see Appendix), which are instantiated by 3,462 tokens.

2.2 *Method and operationalization of variables*

In order to test the effect of lemma frequency on duration, a mixed-effects model was built, predicting the duration of the word tokens. In the following, I explain the operationalization of the variables that entered the models.

2.2.1 *The dependent variable: word duration*

The Buckeye corpus material contains time-aligned word and phoneme-level segmentation. The target words' audio files were extracted from the corpus along with segmentation tiers using scripts in Praat (Boersma & Weenink 2016). The duration of the words was extracted via another Praat script.

2.2.2 *Frequency-related predictors*

Frequency counts were obtained from two large corpora of contemporary American English.[3] Care was taken to accurately capture the lemma frequencies of the words under investigation, by taking all inflected forms of the noun and the verb into account.[4]

The first corpus resource employed is the *Corpus of Contemporary American English* (COCA, Davies 2014), which was chosen as it represents a very large corpus balanced across different genres. As a second resource, the POS-tagged version of the SUBTLEX-US corpus was used, a corpus of US film and television subtitles (Brysbaert & New 2009, Brysbaert, New & Keuleers 2012). Subtitle frequencies have been argued to better represent everyday language experience

---

[3] Frequency counts were not taken from the Buckeye corpus, since it is fairly small (approximately 400,000 words) and contains language collected in only one specific setting. Therefore, it was considered to be not representative of the general language exposure of the speakers recorded in that corpus.

[4] This operationalization avoids a shortcoming in previous research on the lemma frequency effect by Gahl (2008, 2009). Even though the main claim in these studies is about lemma frequency, an actual lemma frequency was not calculated. Focusing on heterographic word pairs, Gahl measured the string frequency of the two members of the pair respectively, e.g. one frequency count for the string *time* and one for the string *thyme*. However, these frequencies are not congruent with lemma frequencies, as inflected forms belonging to the same lemma are ignored. See Baayen, Milin & Ramscar (2016) for a general discussion of the issue of relying on orthographic conventions in performing frequency counts.

7

than corpora that are largely based on written sources (Brysbaert & New 2009: 979). The cognitive validity of both corpus resources has been successfully tested in language processing tasks. Frequency counts obtained from SUBTLEX-US have been shown to accurately predict reaction times in lexical decision tasks (Brysbaert & New 2009, Brysbaert et al. 2012). Similarly, COCA frequencies have been shown to excel in predicting reaction times in an auditory lexical decision task, achieving higher accuracy than other frequency counts, including those from only the spoken section of COCA (Tucker & Brenner 2016). The correlation between the frequency counts across the two corpora is very high ($r_{Pearson} = 0.93$). Because of this high correlation, differences resulting from using one or the other count in the ensuing calculations were found to be only marginal. For that reason only results based on the SUBTLEX-US counts will be reported in the following, unless otherwise noted.

When testing for effects of lemma frequency, a straightforward strategy would be to simply enter the word-class-specific lemma frequencies as a predictor into the model. However, this approach tests for global frequency effects among all words in the sample, but does not test the specific hypothesis of a frequency effect differentiating the homophones of the individual pairs. In fact, the lemma frequency predictor could return a statistically significant result even if all homophones in the sample were pronounced the same, but if there was simply an effect of more frequent strings (e.g. both members of the N/V pair *work*) being pronounced with shorter duration than less frequent strings (e.g. both members of the N/V pair *vote*), as predicted by frequency inheritance accounts. The reason for this is a high correlation between string and lemma frequency in the dataset. I extracted the string frequencies of the word pairs from both corpora by summing up the frequencies for the noun and the verb strings, e.g. summing up the frequencies for *cut*(N) and *cut*(V), excluding other inflected forms. The correlation coefficient of string frequency and lemma frequency is very high ($r_{Pearson} = 0.89$), which is a result of the cumulative string frequency being close to the lemma frequency value of the high-frequency word and the frequency imbalance in the sample, in which the high-frequency members of the pairs contribute more tokens than the low-frequency members.

In order to directly test the hypothesis of lemma frequency differentiating the homophones, it is necessary to employ a predictor that captures the frequency difference between the two members of each pair. To that end, I calculated a logged ratio of the two lemma frequencies, by dividing the lemma frequency of the word-class-specific word by the lemma frequency of its homophone twin:

$$\log_{10} \left(\frac{\text{lemma frequency}}{\text{lemma frequency of homophone twin}} \right)$$

This calculation yields a positive value for the high-frequency member and a negative value for the low-frequency member of each homophone pair, with the size of the value reflecting the size of the difference in frequency. This lemma

8

frequency ratio is not strongly correlated with the string frequency count. Both the difference in lemma frequency and the string frequency of the pair will be entered as predictors into the model in order to test duration differences brought about by a difference in lemma frequency, while still controlling for a general frequency effect. For the two frequency-related variables, the terms LEMMA FREQUENCY RATIO and STRING FREQUENCY will be used in the following.

2.2.3 *Control variables*

The duration of words in spontaneous speech is influenced by a variety of different factors. Any study interested in testing the effect of just one particular variable therefore faces the task of controlling for these other factors. One way to do so is to enter control variables that capture these influences into the regression model as both random and fixed effects. This is also the strategy chosen in the present paper. The choice of covariates is very similar to previous corpus-based research on frequency effects on homophone duration, most notably Gahl (2008, 2009).

First, in order to control for influences specific to either the particular item or speaker, random intercepts for speaker (speaker ID of the 40 different speakers in the Buckeye corpus) and N/V pair (63 homophone pairs tested; see Appendix) were entered into the model. Moreover, random slopes for LEMMA FREQUENCY RATIO by N/V pair and speaker were added. Furthermore, the following variables were employed as fixed-effect control variables.

(i) SPEECH RATE: An obvious determinant of word duration is the rate of speech in the context of the target word. Speech rate was measured as segments per second in the context surrounding the target word (± 10 s), but not including the target word itself. Periods of silence, as marked by the corpus annotators, were ignored in this operationalization, i.e. did not slow down the speaking rate as measured here.

(ii) LENGTH: Another determinant of duration is the phonological length of the word. Two operationalizations of length were tested, namely the length in number of segments and the length in number of slots on the CV-tier. While both are highly correlated, the CV-tier operationalization better fits the data and was therefore chosen as the final length measurement. It was calculated by obtaining the CV-structure of all words from CELEX (Baayen et al. 2001).

(iii) BIGRAM PROBABILITY BASED ON PRECEDING/FOLLOWING WORD: A further factor impacting the duration of a word in spontaneous speech is its predictability from the neighboring lexical context (see, e.g. Bell et al. 2003). Following the procedure of similar studies, contextual predictability was calculated based on the previous and the following word (see Bell et al. 2003, Gahl 2008, 2009). This resulted in two separate predictors, one based on the bigram that includes the preceding word and one based on the bigram that includes the following word. Contextual predictability was calculated by dividing the respective bigram frequency by the frequency of the word

9

preceding or following. The result of this calculation is a ratio that indicates the probability of the target word given either the previous or the following word. Bigram and word frequencies for this calculation were obtained from COCA (Davies 2014).[5] If a bigram did not occur in COCA, it was entered with a frequency of zero.

(iv) PAUSE FOLLOWING: Since a following pause may result in a lengthened pronunciation of the target word, a binary variable was created that captures whether the target word precedes a pause longer than 500 ms (cf. Gahl 2009).

(v) SYNTACTIC POSITION: The noun–verb homophones under investigation occur in different syntactic positions of the sentence, which has consequences for the prosodic processes influencing their pronunciation. On the dimension of duration, the most important prosodic difference between nouns and verbs is that nouns undergo pre-boundary lengthening more frequently, because they occur more often in final position of phrases and clauses. This results in a greater duration of nouns compared with verbs (e.g. Sorensen, Cooper & Paccia 1978). In order to control for these lengthening effects, it was coded whether the target word occurred at the right boundary of a phrase, a clause, or neither. This coding procedure resulted in the three values phrase-final, clause-final or phrase-medial for this variable.

(vi) PITCH RANGE: Another prosodic feature that is likely to impinge on the duration of the target words is whether these are the locus of sentence stress or accent. Accented words are pronounced with greater duration, an effect that is termed 'accentual lengthening' (e.g. Turk & White 1999). Differences in accentuation may be especially important in analyzing N/V pairs, as in intonational phonology it is discussed whether arguments are more likely to be accented than predicates (see, e.g., Ladd 2008: 244–251). Since accent is typically marked by a pitch excursion, the pitch range, as the difference between maximum and minimum pitch, was calculated for each word using a Praat script. This calculation led to the exclusion of 27 tokens, as pitch could not be tracked for these data points in Praat.[6]

(vii) WORD CLASS: Since it remains possible that there are further prosodic differences between nouns and verbs that are not captured by the aforementioned variables, word class was employed as a further control variable. It was coded whether the target word instantiated a noun or a verb. Tokens instantiating other word classes or ambiguous cases were excluded.

---

[5] Bigram frequencies specific to the word class of the target item were retrieved, e.g. for the bigram *to work*, two different frequencies were employed, depending on whether the target word was a noun or a verb.

[6] A manual check reveals that these items were spoken with creaky voice, resulting in irregular F0, which therefore was not picked up by the Praat algorithm.

2.2.4 *Treatment of variables for model building*

Following standard practice, the scalar variables BIGRAM PROBABILITY – PRECEDING WORD, BIGRAM PROBABILITY – FOLLOWING WORD, LENGTH, PITCH RANGE and the frequency counts were log-transformed (to the base of 10) and centered before they were entered into the models.[7] For all of these variables it was found that their distributions were more normal in log space. For SPEAKING RATE, the log-transformation led to a less normal distribution, which is why this variable was not transformed but only centered. In order to further address possible problems of a non-linear relationship between the response variable and the predictors, the dependent variable WORD DURATION was transformed employing Box–Cox power transformation (Box & Cox 1964), following Plag, Homann & Kunter (2017). I used the *boxcox* function of the MASS package (Venables & Ripley 2002) in R to calculate the optimal parameter for this transformation, which was $\lambda = 0.1010101$.

3. RESULTS

3.1 *Results of regression model predicting word duration*

A mixed-effects regression model, as implemented in the packages lme4 (Bates et al. 2014) and lmerTest (Kuznetsova, Brockhoff & Bojesen Christensen 2014) in R (R Development coreteam 2011), was fitted to the Box–Cox-transformed word duration. First, a maximal model including all predictors was fitted, before removing fixed-effect predictors not significantly improving model fit. *P*-values for the predictor variables were calculated via likelihood ratio tests, comparing the fit of the model with and without the variable in question. The random-effects structure with random intercepts for N/V pair and speaker, as well as random slopes for LEMMA FREQUENCY RATIO by both N/V pair and speaker, was kept throughout the model fitting procedure. This means that a design-driven rather than data-driven approach was followed, with no attempt at simplifying the random-effects structure in the case of a possible non-significant contribution of a certain random effect (see Barr et al. 2013 on this point).

Regarding the fixed effects, all predictors yielded *p*-values smaller than alpha = 0.05, except for the variable WORD CLASS, which was found to be blatantly non-significant ($p > 0.4$) and was therefore removed from the model. This result

---

[7] A problem when calculating the logarithm for the contextual predictability variables was that some of the original values were zero, due to non-occurrence in the reference corpus. Therefore, I added 1 before applying the logarithm function. In order to rule out that the results are unduly influenced by this operation, I calculated the reported regression models also on a subset of the data excluding these cases. For this subset, I log-transformed the variables BIGRAM PROBABILITY – PRECEDING WORD and BIGRAM PROBABILITY – FOLLOWING WORD without adding 1. In these alternative models, differences in lemma frequency are still statistically significant. The effect also persists in models featuring the predictability predictors in their original, non-transformed format.

is not unexpected as the prosodic processes leading to differences in duration between nouns and verbs were independently captured through the predictors SYNTACTIC POSITION and PITCH RANGE. All remaining predictors yield effects in the expected directions, as greater length on the CV-tier, greater pitch range, slower speaking rate, lesser predictability and pre-pausal position are all correlated with greater duration of the target words. Moreover, SYNTACTIC POSITION yields a significant effect on word duration, reflecting the expected effect of pre-boundary lengthening, which is larger in clause-final compared with phrase-final position (compare the coefficients for phrase-final versus clause-final in Table 2 below).

Regarding the frequency-related variables, I entered the logged lemma frequency ratio as well as the logged string frequency as predictors into the model. As expected, both STRING FREQUENCY and LEMMA FREQUENCY RATIO are negatively correlated with word duration. Both are significant predictors at alpha = 0.05. Collinearity of variables was checked by calculating Variance Inflation Factors (VIFs), which are <2 for both models, indicating low collinearity.

Summary statistics of the model are provided below. The random-effects summary statistics are given in Table 1, the fixed-effects summary appears in Table 2.

	VARIANCE	SD
N/V PAIR (INTERCEPT)	0.00044	0.020968
N/V PAIR (LEMMA FREQUENCY RATIO)	0.00009	0.009401
SPEAKER (INTERCEPT)	0.00042	0.020431
SPEAKER (LEMMA FREQUENCY RATIO)	0.00002	0.004137
RESIDUAL	0.00204	0.045181

Table 1
Random-effects summary statistics of the mixed-effects model of word duration
($n = 3,435$).

Crucially, the difference in lemma frequency as captured via the LEMMA FREQUENCY RATIO emerges as a statistically significant predictor of duration in the models with all control variables, indicating a shorter duration of the high-frequency homophones relative to their low-frequency twins. The variable STRING FREQUENCY also yields a statistically significant effect on duration. This finding means that in addition to differences in duration between the members of each pair, there is still an effect of overall frequent pairs being pronounced with shorter duration than pairs of lesser frequency, consistent with an expected general effect of frequency on duration.

In order to gauge the respective explanatory power of the random and fixed effects, the conditional and marginal R-squared values were calculated using the R package `piecewiseSEM` (Lefcheck & Freckleton 2016). The conditional

	β	SE	t	$p(\chi^2)$
INTERCEPT	1.7590	0.005	354.51	—
SPEECH RATE	−0.0009	0.001	−2.25	<0.05
LENGTH CV-TIER	0.2282	0.028	8.08	<0.001
BIGRAM-PRECEDING	−0.2789	0.135	−2.06	<0.05
BIGRAM-FOLLOWING	−1.3870	0.300	−4.62	<0.001
SYNTACTIC POSITION=PHRASE-FINAL	0.0128	0.003	4.27	<0.001
SYNTACTIC POSITION=CLAUSE-FINAL	0.0403	0.002	18.15	<0.001
PITCH RANGE	0.0303	0.002	15.63	<0.001
PAUSE FOLLOWING	0.0390	0.004	10.33	<0.001
LEMMA FREQUENCY RATIO	−0.0090	0.003	−3.58	<0.01
STRING FREQUENCY	−0.0199	0.006	−3.11	<0.01

Table 2
Fixed-effects summary statistics of the mixed-effects model of word duration ($n = 3,435$).

(=overall) *R*-squared value of the model is 0.59. Marginal *R*-squared is 0.40, which indicates the share of variance explained by solely the fixed effects.

A separate model was built employing frequency counts from the COCA corpus. This model yields practically the same results, with the crucial predictor LEMMA FREQUENCY RATIO also improving the fit of this model ($p < 0.05$). A comparison of goodness of fit across the two models yields the result that the model employing frequency counts from the SUBTLEX-US corpus is characterized by a slightly lower AIC and explains a larger share of variance. A likelihood ratio test comparing the two models yields a statistically significant result ($\chi^2 = 17.907$, $df = 18$, $p < 0.001$).

3.2 *Additional analyses*

3.2.1 *The size of the lemma frequency effect*

The model results indicate an effect of lemma frequency on the pronunciation of the noun–verb homophones in the sample. However, the model output does not provide a straightforwardly interpretable measure of the size of this effect. In order to get an idea of the difference in duration between the homophones attributed to differences in frequency, I calculated a model in which I replaced the crucial predictor LEMMA FREQUENCY RATIO with a binary variable indicating whether the data point instantiates the low-frequency or the high-frequency member of the N/V pair in question. This model was fitted to the untransformed word durations. The coefficient estimate of this variable indicates a duration difference of 22 ms between the low-frequency and high-frequency homophones.[8] This value can be

---

[8] When calculating the same model based on frequency counts retrieved from the COCA corpus, this coefficient estimate is 16 ms.

13

taken to indicate the average effect of the difference in lemma frequency on the duration of the homophones. Since the size of the frequency difference varies among the pairs in the sample, the difference in duration should also vary. This aspect of the data is captured via the calculation of the scalar lemma frequency ratio in the main model reported (see Table 2), which yields a slightly better model fit than the model employing only a binary predictor of the lemma frequency difference (as indicated by a likelihood ratio test comparing the two models, with the following result: $\chi^2 = 0.11$, $df = 18$, $p < 0.001$).

3.2.2 *A type-based analysis*

While the results testify to a general effect of lemma frequency differentiating the homophones in the dataset, it has to be kept in mind that the distribution across N/V pairs in the sample is not balanced, i.e. some pairs contribute more tokens to the sample than others. This raises the question of whether the effect is truly a general one, or may be due to only certain very frequent pairs.[9] To address this question, an additional model was built on a type rather than token basis. This model was fitted to the Box–Cox-transformed average durations of the word-class-specific words.[10] In this way, each N/V pair contributes the same amount of variance to the dataset. Moreover, the independent variables were averaged. For the scalar variables, average values for each word were calculated and employed as predictors. The categorical predictor SYNTACTIC POSITION was first transformed into a scalar predictor with the values phrase-medial = 0, phrase-final = 1, clause-final = 2, before the average boundary strength per word was calculated. The categorical predictor PAUSE FOLLOWING was averaged by calculating the pause ratio per word type. All predictors were logged to the base of 10 and centered before they entered the model. The model features the same fixed effects as the token model reported above (see Table 2) and a random intercept for N/V pair. See the fixed-effects model output in Table 3.

Crucially, in this model the predictor LEMMA FREQUENCY RATIO is again statistically significant ($p < 0.001$). In conclusion, the lemma frequency effect does not seem to be unduly influenced by the type-token distribution in the dataset.

3.2.3 *Analyses of the effects of lemma frequency and word class*

When introducing the control variables, it was noted that nouns and verbs are subject to different prosodic effects. This means that the durations of the

---

[9] I thank an anonymous reviewer for bringing this potential concern to my attention.

[10] Since the λ value used for transformation of the word durations was negative (λ = −0.06060606), I fitted the model to the Box–Cox-transformed durations multiplied by −1 in order to avoid a switching of signs of the coefficients. It should be noted that this operation causes the intercept to change its sign, which is therefore negative.

VARIABLE	β	SE	t	$p(\chi^2)$
INTERCEPT	−0.7032	0.0006	−1107.63	—
LENGTH CV-TIER	0.0451	0.007	6.92	<0.001
BIGRAM-PRECEDING	−0.4648	0.135	−3.43	<0.001
BIGRAM-FOLLOWING	−1.1790	0.542	−2.18	<0.05
SPEECH RATE	−0.0333	0.015	−2.21	<0.05
SYNTACTIC POSITION	0.0093	0.004	2.27	<0.05
PITCH RANGE	0.0114	0.025	4.60	<0.001
PAUSE	0.0430	0.018	2.37	<0.05
LEMMA FREQUENCY RATIO	−0.0029	0.001	−3.70	<0.001
STRING FREQUENCY	−0.0033	0.002	−2.10	<0.05

Table 3
Fixed-effects summary statistics of a mixed-effects regression model of average word durations ($n = 126$).

homophones in the dataset are affected by both word-class-specific prosody and a difference in frequency. Therefore, it is important to ensure that the statistically significant finding for lemma frequency is truly a separate effect that cannot be reduced to the difference in word class and concomitant prosodic effects. A potential cause for concern is that word class and differences in frequency are not strictly orthogonal predictors. More specifically, in the dataset it is more often the case that the verb lemma is the more frequent member of the pair than the noun lemma: in 46 pairs the verb is more frequent, while the noun is the more frequent member in only 17 word pairs. Verb status predicts a shorter duration independent of frequency differences, because verbs occur more frequently before prosodic boundaries of lesser strength than nouns and may be less frequently accented (see Section 2.2.3). This raises the question of whether the significant result for lemma frequency is an artifact of category-specific prosody.

This seems unlikely, as in the model reported above, relevant prosodic effects have been controlled for by the variables SYNTACTIC POSITION and PITCH RANGE, still the difference in lemma frequency remains a statistically significant determinant of duration. As a further control, since these variables may not capture prosodic differences between nouns and verbs in their entirety, I entered WORD CLASS as a separate variable, which does not yield a statistically significant effect and was therefore removed. However, even when keeping WORD CLASS in the model, the lemma frequency effect persists, indicating an independent effect of lemma frequency beyond prosodic differences between nouns and verbs.

To further scrutinize the independence of the effects of lemma frequency and word class, I analyzed different subsamples of the data, in which either the noun member or the verb member of the homophone pairs was more frequent.

15

If there are independent effects of word-class-specific prosody and frequency, then the effect of word class, i.e. verbs being pronounced with shorter duration, should differ across the three samples: it should be strongest in the verb-frequent subsample, because both variables pull in the same direction, medium in the overall sample and weakest in the noun-frequent sample, in which the two effects are expected to (partially) cancel each other out. In contrast, if there is no independent effect of frequency, one would expect the effect of word class to be the same across the different frequency configurations. The results are clearly in line with the former: I tested the effect of word class in mixed-effects models with WORD CLASS as the only fixed effect and with random intercepts for speaker and N/V pair and random slopes of WORD CLASS by N/V pair and speaker. When fitted to the untransformed word durations, verbs are predicted to be 50 ms shorter than nouns in the overall sample. This difference is reduced to 20 ms in the noun-dominant sample, in line with the assumption that the effect of word class is mitigated by frequency, while in the verb-dominant sample the difference is 61 ms, consistent with the assumption of an additive effect of frequency.

In sum, the results indicate that it is unlikely that the frequency effect is an artifact of its correlation with word class and corresponding prosodic effects.

3.2.4 *Analysis of the low-frequency subsample*

The main result of the regression model reported in Section 3.1 is that the N/V homophones are pronounced differently contingent on differences in their frequency. While this finding indicates an effect of lemma frequency, it does not rule out any inheritance of frequency across the two members of the N/V pairs. The result is also compatible with a scenario of partial frequency inheritance, as laid out in Section 1.2. This would mean that the low-frequency words inherit frequency from their high-frequency twins to a certain degree, which, however, does not neutralize the duration differences brought about by the lemma frequency effect.

In order to more directly test for such inheritance effects, I created a subsample containing only the low-frequency words ($n = 871$), to test whether the duration of these words is influenced only by their own frequency or also by the frequency of their high-frequency twins. Since accounts of frequency inheritance postulate these effects to take place at the wordform level (see Section 1.2), wordform frequencies from SUBTLEX-US were retrieved, which unlike lemma frequencies do not take into account the frequencies of other inflected forms of the words.

Three different frequency counts were tested as predictors of word duration: the frequency of the less frequent word (= actual wordform frequency), henceforth termed LF, the wordform frequency of the high-frequency counterpart (HF) and also a combined frequency count (all logged to the base of 10). In a first step, I built a model predicting word duration (Box–Cox transformed) with the LF count as the only frequency-related predictor. The same control variables as

in the previously reported models were entered as random and fixed effects.[11] This model returns a statistically significant result of LF wordform frequency ($p < 0.05$). To this model I then added the HF count, in order to test whether it significantly affected duration in addition to the LF predictor. This is not the case, as the HF predictor does not significantly improve the model ($p > 0.8$), while LF frequency remains a significant predictor ($p < 0.05$). See the fixed-effects model output in Table 4.

VARIABLE	β	SE	t	$p(\chi^2)$
INTERCEPT	2.2278	0.089	248.96	—
SPEECH RATE	−0.0036	0.001	−2.44	<0.05
LENGTH CV-TIER	0.4238	0.054	7.82	<0.001
BIGRAM-FOLLOWING	−6.0028	2.277	−2.64	<0.01
SYNTACTIC POSITION=PHRASE-FINAL	0.0251	0.008	3.06	<0.01
SYNTACTIC POSITION=CLAUSE-FINAL	0.0642	0.073	8.83	<0.001
PITCH RANGE	0.0602	0.007	9.00	<0.001
PAUSE FOLLOWING	0.0520	0.007	4.81	<0.001
LF-FREQUENCY	−0.0422	0.016	−2.56	<0.05
HF-FREQUENCY	−0.0028	0.015	−0.18	= 0.86

Table 4
Fixed-effects summary statistics of a mixed-effects regression model of word duration for the subsample of low-frequency words ($n = 871$).

In order to further explore a possible influence of the HF count or combined frequency, I tested the corresponding predictors individually by entering these into models without the LF predictor. These calculations return statistically significant results for both of these variables when entered in separate models. However, this is likely to be an effect of a correlation between the LF and the HF and combined frequency counts ($r_{Pearson} = 0.54$ and $r_{Pearson} = 0.78$, respectively). I compared the goodness of fit across the individual models containing only one of the frequency-related predictors each. The model featuring the LF predictor has a better model fit than the models with either the HF count or the combined frequency count. Likelihood ratio tests comparing the models yield the result that this difference is statistically significant (LF frequency versus HF frequency $\chi^2 = 6.4$, $p < 0.001$; LF frequency versus combined frequency $\chi^2 = 4.2$, $p < 0.001$).

[11] The random-effects structure of the model features random intercepts for N/V pair and speaker, but no random slopes. The addition of random slopes for the frequency predictors resulted in nonconvergence of the model. The predictor BIGRAM PROBABILITY BASED ON PRECEDING WORD was omitted from the model, since it was blatantly non-significant. Leaving this predictor in the model does not change the pattern of results as reported here.

17

In summary, adding the frequency counts of the high-frequency twins does not improve models that predict the word durations of the low-frequency homophones. Employing individual word frequency as the only frequency-related predictor is sufficient for that aim. In conclusion, the results do not provide evidence for frequency inheritance.

4. DISCUSSION

4.1 *The present findings in the context of previous studies*

The results show that lemma frequency affects the duration of homographic noun/verb homophones. This finding corresponds to the trend found in child-directed speech by Conwell (2016). In general, the results obtained corroborate the lemma frequency effect for homophonous content words found in earlier corpus studies (Gahl 2008, 2009). Since the N/V pairs tested are also content words, the results obtained tie in with the possibility that the lemma frequency effect may impact content but not function words, a difference possibly arising from different production mechanisms responsible for these two classes of words (see the discussion in Gahl 2008: 479). It should be noted, however, that in the studies on duration differences between homophonous function words (Jurafsky et al. 2002, Jurafsky 2003), only a handful of word pairs were tested. It is therefore a task for future research to further elucidate the possibly different sensitivity to word frequency effects of content versus function words.

An important difference from previous studies is that the homophone pairs tested are homographic. In discussing her results, Gahl (2008) raises the question of whether the finding of a lemma frequency effect in her data may be due to the fact that the pairs compared were heterographic, while in the studies that provide evidence for frequency inheritance, homographs were tested (Jescheniak & Levelt 1994). However, no evidence for frequency inheritance was found in the present study. Conversely, the results suggest that the lemma frequency effect is not contingent on spelling differences.

Moreover, the words that were compared in the present study are not only orthographically identical, but also semantically more similar compared with the word pairs previously studied. This is an important difference, as Jescheniak et al. (2003) raise the question of whether semantic similarity may influence the likelihood of frequency inheritance, with stronger effects expected between semantically more similar homophones. While most previous studies do not disclose all homophone pairs analyzed (Whalen 1991, Guion 1995, Cohn et al. 2005, Gahl 2008, 2009), judging from the examples that are discussed, the homophones analyzed were homonyms, i.e. unrelated semantically, e.g. *thyme* versus *time* in Gahl (2008). In contrast, the N/V pairs investigated here, which can be assumed to come about via the word-formation process of conversion, are clearly related semantically, e.g. *cut*(V)*/cut*(N), or *answer*(V)/*answer*(N) (see the Appendix for a complete list of words tested). The results show that the lemma

18

frequency effect persists in semantically related homophones. In consequence, the findings do not provide evidence for the assumption of a facilitative effect of semantic similarity on frequency inheritance, at least on the dimension of word duration.

All in all, given the greater similarity of the homophone pairs compared with the words investigated in previous studies, the results strengthen the case for a lemma frequency effect on the duration of homophonous words.

4.2 *Implications for the debate on the representation of word class in the lexicon*

The present analysis is immediately relevant for the discussion on whether entries in the lexicon are specified for word class (see Section 1.1). Accounts of lexical underspecification, which assume just one representation for the two members of an N/V pair, would have a difficult time explaining that the noun and the verb are pronounced differently contingent on their individual lemma frequencies. In underspecification accounts, pronunciation differences would need to be explained via contextual factors, but not through inherent differences like individual lemma frequencies of the two members of an N/V pair. However, in the models calculated, the effect of lemma frequency was found to persist when contextual variables were controlled for. Therefore, the empirical evidence obtained supports a theory of the mental lexicon that assumes syntactic specification of entries. The lemma frequency effects can be accounted for naturally in such theories, as both the noun and the verb represent separate entries with their own frequencies of linguistic experience.

The existence of a lemma-specific frequency effect in noun–verb conversion pairs ties in with findings of word-class-specific impairments reported in aphasia research (Baxter & Warrington 1985, Caramazza & Hillis 1991, Hillis & Caramazza 1995). Caramazza & Hillis (1991) provide evidence from the performance of two aphasic subjects who exhibit a category-specific impairment in that they are able to produce nouns but not verbs. This holds even for noun–verb homophone pairs (e.g. *crack*(N) versus *crack*(V)), for which the aphasic speakers made selective errors with only the verb but not the noun. Caramazza & Hillis (1991) interpret this finding as evidence for separate representations of the verb and the noun.

However, these findings have also been discussed in the underspecification literature, arguing that such results could also be explained under the assumption of underspecification. Barner & Bale (2002) argue that this is a mapping problem that affects the insertion of lexical entries into verb frames in contrast to noun frames, but does not speak to different representations. Since this is a possible interpretation of the results for aphasic speakers, one may ask whether it may also be used to integrate the present findings into an underspecification account. Doing so would mean to relegate the lemma frequency effect not to the representation of the word(s) but to the frequency of the mapping into word-class-specific frames, while upholding the claim of a common representation. The problem with this

19

argument is that it requires the mapping frequency to be stored somehow, which means that word-class-specific usage information would be part of the speaker's knowledge about these words. This, however, would undermine the assumption of just one representation, which is the same for both words. Therefore, the present evidence is more straightforwardly accounted for in a model in which lexical entries contain word class information, corresponding to the view put forth about the mental and neural lexicon by Caramazza & Hillis (1991). Further evidence supporting this view has also been reported from an experiment with healthy speakers: Conwell (2015) shows that speakers react differently to auditory presentations of noun versus verb homophones in an EEG experiment. Crucially, these words had been presented in isolation, so that the different reactions could not be due to differences in mapping the words into class-specific contexts.

4.3 *Implications for homophone representation in models of speech production*

In Section 1.2, it was pointed out that the lemma frequency effect stands in direct contrast to the assumption of frequency inheritance across homophones and therefore to speech production models that implement this effect in their architecture. Accounts of full, partial and no frequency inheritance were discussed.

The existence of an effect of lemma frequency on duration is clearly incongruent with 'full frequency inheritance', as such an account would predict the same durations for homophonous words. In consequence, the results are at odds with the model by Jescheniak & Levelt (1994) and Levelt et al. (1999), in which frequency is hypothesized to only affect the wordform level at which both members of a homophonous pair share the same representation.

However, a lemma frequency effect does not per se rule out the possibility of 'partial inheritance', as there may be a certain degree of frequency inheritance, which is not strong enough to neutralize the lemma frequency effect. In order to test for this possibility, a subsample containing only the low-frequency words was created and further analyzed. It was found that adding the frequency of the high-frequency homophone as a predictor does not improve the fit of a model built to predict the durations of the low-frequency words. Hence, no evidence for partial frequency inheritance was found.

In conclusion, the results support an account of 'no frequency inheritance'. The absence of frequency inheritance is best explained in speech production models that assume completely separate lexical representations of homophones, as in the works by Caramazza (e.g. Caramazza et al. 2001, Miozzo & Caramazza 2005). Since, in the present study lemma frequencies (including all inflected forms) were counted, the results are most clearly compatible with models that assume the lemma level to be the locus of the frequency effect. A clearly compatible class of models is two-layer models, which assume separate lemma and word-form representations for homophonous words (see, e.g., Model C in Miozzo & Caramazza 2005). However, the results are not necessarily incompatible with single-layer architectures, which assume just one level of lexical representation

that contains the specific inflected form of the word (see the discussion in Caramazza et al. 2001). The lemma frequency count and the frequency of the category-specific wordform (the uninflected noun/verb, respectively) are highly correlated ($r_{Pearson} = 0.8$ in the SUBTLEX-US data), so that it is not possible to decide which level or layer is ultimately responsible for the frequency effect. In consequence, the data do not allow one to distinguish between single- and dual-layer architectures. Still, the most important point remains: irrespective of the number of layers, the results support the assumption of separate representations of the N/V homophones tested.

Furthermore, the results of the current study are relevant for the question of representational differences between homographic and heterographic homophones, raised in Section 1.2. The absence of evidence for frequency inheritance means also an absence of evidence for the phonological representations of homophones being modulated by a shared orthography. Together with research demonstrating an effect of lemma frequency on heterographic homophones (Gahl 2008, 2009), the results are consistent with an account that assumes separate phonological representations of both heterographic and homographic homophones.

One caveat in differentiating between the three alternatives as regards to frequency inheritance is that the present analysis is based only on homophones, but did not analyze non-homophonous control words. The result of a lemma frequency effect is clearly at odds with 'full frequency inheritance', and, furthermore, no evidence for 'partial inheritance' was found. However, even more conclusive evidence could be provided by an analysis that took into account low-frequency control words, i.e. words of the same frequency as the low-frequency homophones, but lacking a high-frequency counterpart (cf. Middleton et al. 2015: 78). If there is truly no frequency inheritance, then, other things being equal, both groups should be characterized by the same durations. Such an empirical analysis remains a task for future research.

5. CONCLUSIONS AND OUTLOOK

The results reported in this article demonstrate that lemma frequency impacts the duration of homographic noun–verb conversion homophones. Given that N/V homophones make up a sizable portion of the English lexicon, the present study shows that lemma frequency effects are not confined to a small group of homophonous words, but pervade the lexicon.

The results elucidate the role of word class and also generally the status of homophonous words in the mental lexicon and thereby allow insights into its representational structure. In that regard, further questions await exploration. For example, with regard to the degree of specification of lexical entries, a further debate tangential to the current study is taking place. This debate focuses on the representation of polysemous words. The point of discussion is whether polysemous words have only one shared lexical entry of a core sense, which is then specified in context, or whether the individual senses are represented separately.

21

Experimental research involving a variety of processing tasks arrives at mixed results. For example, Frisson & Pickering (1999) find no evidence for a processing difference between different senses and hence for separate representation, while Klein & Murphy (2001) and Foraker & Murphy (2012) do. Studies investigating frequency effects on duration could provide a different kind of empirical evidence to further this debate. Viewing the noun–verb pairs investigated here as being related via polysemy, the existence of a lemma frequency effect points toward separate representation of the noun and the verb sense. However, this would involve comparison of senses only at a coarse level, as each noun and verb lemma is polysemous in itself. For example, for the pair *cut*(N)/*cut*(V), WordNet (Princeton University 2010) lists a number of different senses for the noun and the verb respectively, e.g. for the noun, 'a wound made by cutting' but also 'a share of profits'. It would be an avenue for future research to explore whether word duration is not just sensitive to differences in frequency between the noun and the verb but is additionally dependent on the frequency of the individual senses of each noun and verb.

APPENDIX

List of N/V pairs tested:

act	drive	need	show
answer	end	needs	shows
attack	excuse	notice	sign
break	face	offer	sleep
call	fall	order	sound
calls	fight	pass	stand
care	fire	pay	struggle
cause	focus	plan	study
change	help	play	support
check	hope	practice	talk
control	limit	raise	touch
cook	look	rent	turn
cost	love	respect	vote
cut	matter	ride	waste
deal	mind	run	work
dress	name	set	

REFERENCES

Baayen, R. Harald, R. Piepenbrock & H. van Rijn. 2001. WebCelex, Max Planck Institute for Psycholinguistics. Online resource. http://celex.mpi.nl.

Baayen, R. Harald, Petar Milin & Michael Ramscar. 2016. Frequency in lexical processing. *Aphasiology* 30.11, 1174–1220.

Barner, David & Alan Bale. 2002. No nouns, no verbs: Psycholinguistic arguments in favor of lexical underspecification. *Lingua* 112, 771–791.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.3, 255–278.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. http://cran.r-project.org/package=lme4.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Baxter, Doreen M. & Elizabeth K. Warrington. 1985. Category specific phonological dysgraphia. *Neuropsychologia* 23.5, 653–666.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113.2, 1001–1024.

Biedermann, Britta & Lyndsey Nickels. 2008. Homographic and heterographic homophones in speech production: Does orthography matter? *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior* 44.6, 683–697.

Boersma, Paul & David J. M. Weenink. 2016. Praat. Doing phonetics by computer, version 6.0.14. http://www.praat.org.

Bonin, Patrick & Michel Fayol. 2002. Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology* 14.3, 289–313.

Box, George & David Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2, 211–252.

Bram, Barli. 2011. *Major total conversion in English: The question of directionality*. Ph.D. thesis, Victoria University of Wellington.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41.4, 977–990.

Brysbaert, Marc, Boris New & Emmanuel Keuleers. 2012. Adding part of speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods* 44.4, 991–997.

Caramazza, Alfonso, Albert Costa, Michele Miozzo & Yanchao Bi. 2001. The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.6, 1430–1450.

Caramazza, Alfonso & Argye E. Hillis. 1991. Lexical organization of nouns and verbs in the brain. *Nature* 349, 788–790.

Cohn, Abby, Johann Brugman, Clifford Crawford & Andrew Joseph. 2005. Lexical frequency effects and phonetic duration of English homophones: An acoustic study. *Journal of the Acoustical Society of America* 118, 2036.

Conwell, Erin. 2015. Neural responses to category ambiguous words. *Neuropsychologia* 69, 85–92.

Conwell, Erin. 2016. Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Child Language* 44.3, 734–751.

Davies, Mark. 2014. The Corpus of Contemporary American English: 450 million words, 1990–2012 [Full-Text Corpus Data, Version of 2014]. http://corpus.byu.edu/coca/.

Don, Jan. 2004. Categories in the lexicon. *Linguistics* 42, 931–956.

Farrell, Patrick. 2001. Functional shift as category underspecification. *English Language and Linguistics* 5.1, 109–130.

Frisson, Steven & Martin J. Pickering. 1999. The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 25.6, 1366–1383.

Foraker, Stephani & Gregory L. Murphy. 2012. Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language* 67.4, 407–425.

Gahl, Susanne. 2008. *Time* and *Thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84.3, 474–496.

Gahl, Susanne. 2009. Homophone duration in spontaneous speech: A mixed-effects model. *UC Berkeley Phonology Lab Annual Report* 1, 279–298.

Guion, Susan G. 1995. Word frequency effects among homonyms. *Texas Linguistic Forum* 35, 103–116.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale & Samuel J. Keyser (eds.), *The View from Building 20, Essays in Linguistics in Honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Hillis, Argye E. & A. Caramazza. 1995. Representation of grammatical categories of words in the brain. *Journal of Cognitive Neuroscience* 7.3, 396–407.

Jescheniak, Jörg & Willem J. Levelt. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.4, 824–843.

Jescheniak, Jörg D., Antje S. Meyer & Willem J. M. Levelt. 2003. Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.3, 432–438.

Jurafsky, Daniel. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In Jennifer Hay, Rens Bod & Stefanie Jannedy (eds.), *Probabilistic Linguistics*, 39–95. Cambridge, MA: MIT Press.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of the lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology VII*, 1–34. Berlin/New York: Mouton/de Gruyter.

Kittredge, Audrey K., Gary S. Dell, Jay Verkuilen & Myrna F. Schwartz. 2008. Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology* 25.4, 463–492.

Klein, Devorah E. & Gregory L. Murphy. 2001. The representation of polysemous words. *Journal of Memory and Language* 45.2, 259–282.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2014. lmerTest. http://cran.r-project.org/web/packages/lmerTest/index.html.

Ladd, D. Robert. 2008. *Intonational phonology* 2nd edn. (Cambridge Studies in Linguistics 119). Cambridge: Cambridge University Press.

Lefcheck, Jonathan S. & Robert Freckleton. 2016. piecewiseSEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution* 7.5, 573–579.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1, 1–75.

Marantz, Alec. 1997. No escape from syntax. Don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitriadis & Laura Siegel (eds.), *University of Pennsylvania Working Papers in Linguistics 4.2*, 201–225. Penn Graduate Linguistics Society.

Middleton, Erica L., Qi Chen & Jay Verkuilen. 2015. Friends and foes in the lexicon: Homophone naming in aphasia. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41.1, 77–94.

Miozzo, Michele & Alfonso Caramazza. 2005. The representation of homophones: Evidence from the distractor-frequency effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 31.6, 1360–1371.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. *Buckeye corpus of conversational speech*, 2nd release. Columbus, OH: Department of Psychology, Ohio State University. www.buckeyecorpus.osu.edu.

Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53.1, 181–216.

Princeton University 2010. *'About WordNet'. WordNet*. Princeton University. http://wordnet.princeton.edu.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1990. *A comprehensive grammar of the English language*, 8th impression, standard edn. London/New York: Longman.

R Development coreteam 2011. R: A language and environment for statistical computing. http://cran.r-project.org/.

Sorensen, John M., William E. Cooper & Jeanne M. Paccia. 1978. Speech timing of grammatical categories. *Cognition* 6.2, 135–153.

Tucker, Benjamin V. & Daniel Brenner. 2016. Massive auditory lexical decision: Going big in the auditory domain, Talk held at *Mental Lexicon 2016*, Ottawa.

24

Turk, Alice E. & Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27.2, 171–206.

Velasco, Daniel García. 2009. Conversion in English and its implications for Functional Discourse Grammar. *Lingua* 119.8, 1164–1185.

Venables, W. N. & B. D. Ripley. 2002. *Modern applied statistics with s*, 4th edn. New York: Springer.

Whalen, D. H. 1991. Infrequent words are longer in duration than frequent words. *The Journal of the Acoustical Society of America* 90, 2311.

Author's address: *Department of English and American Studies,*
Heinrich-Heine-Universität Düsseldorf,
Universitätsstrasse 1, 40225, Germany
arne.lohmann@hhu.de