No Acoustic Correlates of Grammatical Class: A Critical Re-Examination of Sereno and Jongman (1995)

Arne Lohmann

Department of English and American Studies, Heinrich Heine University Düsseldorf,
Düsseldorf, Germany

Keywords

Grammatical categories · Acoustic correlates of stress · Stress assignment · Homophones

Abstract

The present article describes an extended replication of a reading study by Sereno and Jongman (1995) that reported acoustic differences between noun and verb pronunciations of English disyllabic, non-stress-shifting homophone pairs, for example, *answer* (v) versus *answer* (n). The original findings point to a gradual influence of the typical stress patterns of both grammatical categories, with noun readings exhibiting a tendency toward trochaic and verb readings toward iambic pronunciation. However, the effects found by Sereno and Jongman (1995) did not consistently reach statistical significance and were based on a very small sample size. Employing a considerably larger group of speakers, the current replication fails to find the reported effects of the grammatical category. The null result obtained can be accounted for within speech production models that assume identical metrical templates for the homophone pairs tested and no direct influence of the grammatical category on the phonetics of stress assignment.

© 2020 The Author(s)
Published by S. Karger AG, Basel

Introduction

In an article on phonetic effects of grammatical category, Sereno and Jongman (1995) report acoustic differences between noun and verb readings of phonologically homophonous, non-stress-shifting disyllabic word pairs, for example, *answer* (v) versus *answer* (n) or *design* (v) versus *design* (n). Crucially, the homophones were read off word lists, this way controlling for the difference in prosodic position between both grammatical categories in discourse (Sorensen,

Cooper, & Paccia, 1978). The main result reported is that the noun pronunciations – while not altering the stress pattern of the words in a categorical fashion – exhibit a tendency toward trochaic stress, while the verb tokens exhibit a tendency toward an iambic pronunciation, reflecting the general difference in position of stress between nouns and verbs (e.g., Davis & Kelly, 1997). These trends manifest themselves in subtle shifts between the first and the second syllable on three acoustic correlates of stress, namely, duration, intensity, and F0.

Sereno and Jongman (1995, p. 61) interpret these findings as showing that “an underlying pattern of stress assignment based on grammatical class usage is operative in speech production.” This presents a challenge for current speech production models, as these do not contain a mechanism for the stated interaction between grammatical category and stress assignment. Given that the target word pairs were non-stress-shifting, stress assignment is assumed to result from the activation of metrical templates that are identical between the noun and the verb and should consequently lead to identical pronunciations (see e.g., Meyer & Belke, 2007, pp. 477–479).

With 57 citations according to a Google Scholar search on May 17, 2019, the article by Sereno and Jongman (1995) is fairly widely cited. A number of theoretical implications of the findings have been discussed: First, since the acoustic differences reflect the typical stress patterns of English nouns and verbs, a number of studies view the finding as providing evidence for a direct link between grammatical category and stress assignment. For example, Phillips (2006) states that “speakers of English recognize the two stress patterns as inherent to the respective grammatical word classes” (p. 38, see also Ishikawa, 2006; Oh, 2011). Second, the existence of acoustic differences directly contingent on grammatical class has been interpreted as providing cues that may help in word recognition and disambiguation between nouns and verbs (Behrens, 2014; Gaskell & Marslen-Wilson, 2001; Herron, 1998). Third, the finding reported is frequently mentioned as demonstrating the existence of acoustic cues for the learning of grammatical categories in first language acquisition (Christiansen & Monaghan, 2006; Monaghan, Christiansen, & Chater, 2007; Onnis & Christiansen, 2008).

Though having given rise to far-reaching conclusions, the article by Sereno and Jongman (1995) provides only limited empirical evidence for the effects reported. Only five speakers were recorded in the original study. Their pronunciations do not consistently show the effect of grammatical category (see the following section for details). The present paper therefore investigates whether firm evidence for the category effect can be acquired when broadening the empirical basis via an extended replication of the original experiment. The re-examination of a theoretically important finding ties in with calls for replication studies in the discussion about a “reproducibility crisis” in psychology and the social sciences (see e.g., Camerer et al., 2018).

A Closer Look at Sereno and Jongman’s (1995) Original Results

In Sereno and Jongman’s (1995) study, the set of target words consisted of 16 word pairs, which were read by five speakers. Grammatical category was treated as a within-subject variable, i.e., all speakers read both the noun and the

verb of each pair. To capture differences in the acoustic realization of stress, Sereno and Jongman (1995) calculated the ratio of the first to the second syllable of the target words for the acoustic parameters duration, F0, and intensity.

Sereno and Jongman (1995, p. 65) report the results of three repeated-measures ANOVAs, one for each acoustic parameter, calculated on the entire data set of productions of the five participants. Interestingly, in none of the three ANOVAs is the difference in grammatical category a significant main effect. One significant interaction effect involving category was found: the ANOVA calculated for intensity ratio reveals a significant interaction between category and position of stress, indicating that trochaic nouns have a higher intensity ratio than trochaic verbs, while the iambic word pairs did not exhibit a significant difference (Sereno & Jongman, 1995, p. 65). This result is in line with the prediction of a more trochaic pronunciation of nouns. Furthermore, Sereno and Jongman (1995, p. 65) report a main effect of category-specific frequency, showing that noun-dominant pairs were spoken with a greater intensity ratio than verb-dominant pairs.

Sereno and Jongman (1995) calculated separate ANOVAs for the individual speakers. The most important result is that only one speaker exhibits a main effect of category at the standardly accepted probability threshold of $\alpha = 0.05$, with nouns showing a more trochaic and verbs a more iambic pronunciation. This speaker shows this contrast on the acoustic parameter of duration only. Furthermore, statistically significant interaction effects either between category and position of stress or between category and frequency bias were found. These reveal selective effects of grammatical category for four of the five speakers, which are restricted to the acoustic parameter of intensity (see also Sereno & Jongman, 1995, p. 68 for a complete overview).

In sum, neither the ANOVA calculated on the entire data set nor the ANOVAs for the individual speakers provide very robust evidence for noun readings having a more trochaic and verb readings a more iambic pronunciation. Nevertheless, with only one exception, the effects reported represent trends in the predicted direction. A possible explanation for the rather elusive results may be too little statistical power due to a fairly small number of speakers and target word productions ($n = 160$). The goal of the current replication study is therefore to investigate whether the trends found in the original study can be corroborated when increasing power through recording a larger number of speakers, each of whom produces twice the number of target words (see Participants and Stimuli).

Materials and Methods

Participants

In order to estimate the number of participants necessary for a sufficiently powered analysis, a power analysis based on the original data provided in Appendix B of Sereno and Jongman (1995) was carried out. More specifically, the power to find a statistically significant main effect of the grammatical category of equal effect size as in the original experiment was calculated by employing the *simr* package (Green & MacLeod, 2016) in *R* (R Core Team, 2014). Based on that analysis, it was aimed at recruiting participants until 40 usable recordings were obtained, as with this number, power for all analyses is more than sufficient (>95% for all acoustic parameters of interest).

Table 1. Word pairs used as stimuli (numbers in parentheses indicate the noun ratio)

Trochaic pairs		Iambic pairs	
Sereno & Jongman (1995)	newly added pairs	Sereno & Jongman (1995)	newly added pairs
Welcome (0.097)	Wonder (0.190)	Embrace (0.157)	Divide (0.138)
Notice (0.167)	Cover (0.244)	Reply (0.208)	Return (0.274)
Handle (0.190)	Answer (0.514)	Escape (0.271)	Excuse (0.463)
Favor (0.401)	Envy (0.516)	Neglect (0.381)	Support (0.474)
Struggle (0.460)	Comment (0.685)	Report (0.486)	Reward (0.604)
Rescue (0.499)	Murder (0.769)	Control (0.679)	Supply (0.654)
Poison (0.620)	Access (0.820)	Dispute (0.839)	Account (0.657)
Practice (0.877)	Battle (0.858)	Debate (0.846)	Command (0.679)

In recruiting participants to meet the goal of 40 recordings, 11 additional speakers were excluded from the analysis because they were aware that they repeated noun/verb homophones between the two lists. Following Sereno and Jongman (1995, p. 63), awareness was tested by questioning participants after the experiment. The 11 speakers who were excluded noticed that words were repeated, as, when asked, they were able to name at least one pair of words occurring in the two lists they read.

The participants were undergraduate students at the University of Alberta, Edmonton, who participated in the experiment in return for course credit. All participants were native speakers of English of the West Canadian variety and gave informed consent in written form. Of the 40 recordings whose productions were considered for acoustic analysis, 33 were female and 7 male. The participants of the original study by Sereno and Jongman (1995) were students at Brown University in Rhode Island, so that certain dialectal differences cannot be ruled out. However, there is no principled reason to assume that such differences would impinge on the effect investigated.

Stimuli

Target Words

In their original experiment, Sereno and Jongman (1995) tested the pronunciation of 16 disyllabic noun-verb pairs. These pairs were phonologically homophonous across word class. Crucially, this entails that none of them were stress shifting noun-verb pairs. The word pairs were chosen so that half of them had trochaic stress (e.g., *ANswer*) and half of them iambic stress (e.g., *deSIGN*). The original data set was furthermore balanced with regard to category-specific frequency. This was done in a binary fashion, that is, for one half of the stimuli in the group of iambic and trochaic words, respectively, usage frequency was greater for the noun, while for the other half usage frequency was greater for the verb (based on frequencies retrieved from Francis & Kučera, 1982). Sereno and Jongman (1995, p. 69) report an effect of category-specific usage frequency as a secondary finding. Noun-dominant word pairs were characterized by a trend toward trochaic pronunciation and verb-dominant pairs showed a tendency toward iambic pronunciation.

In the current replication, the same 16 noun-verb pairs were employed, and 16 further pairs were added. Of the additional pairs, eight had trochaic and eight iambic stress. With regard to the balancing of category-specific frequency, it was aimed at using pairs that cover the entire range from very noun-dominant to very verb-dominant. In order to achieve that, the frequency bias of the word pairs was operationalized in a scalar fashion, first calculating the noun ratio for the 16 words in the original sample. This was done by retrieving the noun and verb lemma frequencies of the word pairs

from the Corpus of Contemporary American English (Davies, 2014) and dividing the noun frequency by the cumulative noun and verb frequency. The 16 additional word pairs were chosen so that the final group of target words covers the range from a noun ratio of close to 0 to close to 1 in both the iambic and the trochaic group. Please see Table 1, which provides the word pairs of the individual groups along with their noun ratio (in parentheses).

List Creation

The lists the target words were read off were designed as described by Sereno and Jongman (1995, p. 63). In order to trigger a category-specific reading of the target words, two lists of filler words were created, one containing only unambiguous nouns, and the other containing only unambiguous verbs. The target words were interspersed into these lists. In order to keep the target-filler ratio the same as in the original study, the number of filler items per list was doubled, resulting in lists of 150 unambiguous nouns or verbs, respectively, plus the 32 target words. Length in number of syllables as well as the stress pattern of the filler words were matched across the noun and verb lists, replicating the distribution provided by Sereno and Jongman (1995, p. 63). This means that the final lists contained 10 monosyllabic, 120 trisyllabic, 18 quadrisyllabic, and 2 pentasyllabic filler items. Of the polysyllabic words, 68 were stressed on the first syllable, 68 stressed on the second syllable, and 4 stressed on the third syllable. Filler items were selected with the help of the CELEX electronic database (Baayen, Piepenbrock, & van Rijn, 2001).

The target items were ordered relative to the filler items as described by Sereno and Jongman (1995, p. 63) with certain adjustments that led to a more careful matching of context between both categories: First, the filler items for the noun list were put in a random order. For the verb list, a matching order of filler items with regard to length in number of syllables and position of stress was created. The target words were then interspersed into the lists as in the original design, that is, with at least 4 filler items separating any 2 target items. Crucially, the target items were interspersed into the same positions in the noun and the verb list, which means that – regarding the mentioned phonological properties of the surrounding words – each homophonous noun-verb pair occurred in the same context.

Sereno and Jongman (1995) used only one pair of lists that was read by all speakers. In order to vary the context, the target words occurred in four list pairs that differed with regard to the position of the target words relative to the fillers. For each list pair, the filler word context was kept constant between the noun and the verb list, as described above.

Procedure

Each participant read both a noun and a verb list belonging to the same list pair. The 40 participants were balanced across the four different list pairs, so that each pair of lists was read by 10 participants. The order in which the noun and the verb list was presented was varied between participants, so that for each list pair, five participants read the verb list first and five read the noun list first.

The word lists were displayed on a computer screen distributed over 14 slides. Each slide displayed 13 words of which either two or three words were target items. Neither the first nor the last word on any given slide was a target item. As in the original experiment, the reading of both lists was preceded by instructions that explicitly mentioned the grammatical category of the words on the list: “You will see words displayed on the screen. Please read the words out aloud in the order they are displayed, starting with the topmost word. All of the words in this part are English {verbs/nouns}.” Participants pressed the button of a button box in front of them to proceed from one slide to the next. After reading the first list, the participants participated in unrelated production and perception tasks, after which they read the second list.

Participants were recorded in a sound-attenuated booth at the Alberta Phonetics Laboratory at the University of Alberta in Edmonton, AB, Canada.

Treatment and Analysis of Data

Exclusion of Data

The 40 participants produced 2,560 pronunciations of the target words, of which 28 data points were excluded, because a different word or word form was produced instead of the target word, for example, *poisonous* instead of *poison*.

Segmentation

The target items were spliced out of the recordings using *Praat* (Boersma & Weenink, 2016). Word and segment boundaries were determined in a two-step process. First, the words were automatically segmented at the word and segment level using the forced alignment software MAUS (Kisler, Reichel, & Schiel, 2017). In a second step, research assistants, trained in phonetic analysis but unaware of the goals of the study, manually checked and corrected word and phoneme level boundaries using *Praat*, which included the marking of the crucial boundary between the two syllables for each word. To ensure that segmentation was consistent between the individual research assistants, segmentation criteria based on cues in the spectrogram and the waveform were developed in several training sessions with the author. The research assistants worked on the words spliced out of context, so that the grammatical category of the word was unknown to them when segmenting the data. The syllable boundary was determined as described by Sereno and Jongman (1995, p. 64). For the majority of the target words, the syllable boundary was set at the left boundary of the first postvocalic consonant (*account, battle, command, comment, cover, debate, divide, favor, neglect, notice, poison, reply, report, return, reward, struggle, supply, support*). For the remaining target words, the syllable boundary was determined as the left boundary of the second consonant following the nucleus of the first syllable (*access, answer, envy, control, dispute, embrace, escape, excuse, handle, murder, practice, rescue, welcome, wonder*).

Operationalization and Acoustic Analysis

As in the study by Sereno and Jongman (1995), the acoustic parameters duration, F0, and intensity were measured for each syllable of the target words. The acoustic software *Praat* (Boersma & Weenink, 2016) was employed for these measurements. For F0 and intensity, the maximal values of the syllable were taken.¹ Duration was calculated in milliseconds, F0 in Hertz², and intensity per syllable in decibels.³

To capture differences in stress between the two syllables of the target words, ratios of the first to the second syllable were calculated for the duration measurements (duration of first syllable/duration of second syllable). For F0 and intensity, the difference (Δ) between the two syllables was calculated, e.g., intensity of the first syllable minus intensity of the second syllable. The difference values were calculated in decibels for intensity and in semitones for F0.⁴

¹ As in Sereno and Jongman (1995), the mean values of F0 and intensity were also measured for each syllable. These operationalizations did not reflect the difference in position of stress, however, which is why they are not discussed further in the following. This is most likely a result of considerable differences in the phonological structure between the stressed and the unstressed syllables of the individual target words. Many of the stressed syllables in the data are heavy syllables that contain more consonant segments (including obstruents) than the unstressed syllables. This introduces stretches of low intensity and unreliable pitch tracking, which consequently affects the mean F0 and intensity measures for these syllables.

² For the female speakers, the F0 settings of *Praat* were adjusted to 75–300 Hz, for the female speakers the setting 100–500 Hz was used.

³ As a further acoustic parameter, the difference in spectral balance between both syllables was calculated as a further parameter (Sluijter & van Heuven, 1996). It yields the same nonsignificant result for differences in the grammatical category as the other acoustic parameters. See the online supplementary material provided for details of this analysis (for all online supplementary material see www.karger.com/doi/10.1159/000506138).

⁴ Sereno and Jongman (1995) based their analyses on ratios for all three acoustic parameters. Employing ratios instead of differences for F0 and intensity yields the same pattern of nonsignificant results.

Fig. 1. Acoustic correlates of stress in trochaic and iambic target word pronunciations grouped by grammatical category (observed values, the horizontal line indicates the median, the dots indicate the mean, the notches indicate the 95% confidence interval around the median, the hinges indicate the first and third quartiles, the whiskers range from the hinge to the smallest/largest value that is at most 1.5 interquartile ranges of the hinge).

Results

Overview of Acoustic Measurements

Please see the following overview of the results provided in Figure 1 in the form of boxplots, which show the distributions for the observed noun and verb pronunciations in the trochaic and the iambic subsample of the data. The distribution of the observed values shows that the difference in stress position between trochaic and iambic words manifests itself on all acoustic dimensions considered. This can be gleaned from the uniformly lower ratios or difference values for the iambic words compared to the trochaic words.

The boxplots do not show any conspicuous differences between the two grammatical categories, as the mean values and distributional characteristics between nouns and verbs are very similar across the acoustic parameters. *T*-tests indicate that there is no statistically significant contrast between noun and verb pronunciations in any subgroup in the data (see Table A1 in the Appendix). In the following section, a possible effect of grammatical category is investigated in detail via the calculation of multiple regression models.

Models

Linear mixed-effect models were built, one for each acoustic parameter analyzed, using the *lmer* function of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014) in *R* (R Core Team, 2014). The dependent variables of these models are the ratio between the two syllables for the parameter duration and the difference values for F0 and intensity. The random-effect structure of the models includes a random intercept for homophone pair and a random intercept for speaker nested in one of four conditions as defined via the four list pairs (see List Creation). Furthermore, random slopes for grammatical category by these grouping variables were added.

Table 2. Fixed-effect model output for the models fitted to Δ intensity and Δ F0

Fixed-effect predictors	Δ Intensity ($n = 2,532$)				Δ F0 ($n = 2,446$)			
	β	<i>SE</i>	<i>t</i>	<i>p</i>	β	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	3.146	0.607	5.187		-0.997	0.624	-1.599	
Category = verb	-0.325	0.215	-1.513	0.13	-0.044	0.229	-0.192	0.85
Stress = iambic	-5.371	0.571	-9.399	<0.01	-2.616	0.566	-4.626	<0.01
Noun ratio	0.449	0.466	0.962	0.34	0.027	0.461	0.058	0.95

Table 3. Fixed-effect model output for the models fitted to the duration ratio ($n = 2,532$)

Fixed-effect predictors	Duration ratio			
	β	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	0.802	0.073	11.035	
Category = verb	-0.003	0.006	-0.422	0.67
Stress = iambic	-0.521	0.070	-7.397	<0.01
Noun ratio	-0.010	0.057	-0.166	0.87

The main variables of interest, grammatical category and the noun ratio of the target word (log-transformed) were entered as fixed effects. The stress pattern (iambic vs. trochaic) of the word pair was also entered as a fixed effect. List order (noun list first vs. verb list first) was entered as a further fixed-effect control variable. Interaction effects between category and noun ratio, and between category and stress pattern were tested.

Model fitting was performed as follows: First maximal models with all random and fixed effects were built, as specified above. This maximal model was then input to the *step* function of the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2014) in *R*. The *step* function performs an automatized, stepwise, backward elimination procedure and includes variables only in the model in case they lead to a decrease in the Akaike information criterion of the model and a significant improvement in model fit (tested via a log-likelihood test). As a further, alternative model fitting strategy, model fitting was performed manually. In the manual fitting process, the random-effect structure was not reduced during model fitting, following a design-based approach (Barr, Levy, Scheepers, & Tily, 2013). Nonsignificant main or interaction effects were omitted starting with those terms with the highest *p* values until only significant effects remained. Both model fitting strategies arrived at the same result with regard to fixed effects: Across all acoustic parameters, the only predictor variable kept after model fitting is position of stress. None of the variables of interest, namely, grammatical category and noun ratio, are statistically significant, neither are interactions of these variables with position of stress. Also, the control variable list order does not significantly affect the dependent variables in any of the models.

Tables 2 and 3 provide more detailed results of models that contain the variables grammatical category and noun ratio, as well as position of stress as fixed effects. The results displayed are the fixed-effect coefficients of models with maximal random-effect structure, as specified above.⁵

As an additional analysis, in order to rule out that the newly added word pairs masked an effect in the pairs used in the original study, all models were also calculated on the subsample of the 16 original word pairs employed by Sereno and Jongman (1995). This analysis yields the same pattern of results as the analysis based on the entire sample: None of the variables of interest produce a statistically significant effect, neither as main effects, nor in interaction with other variables.⁶

Discussion

The general pattern of results obtained is that of little evidence for an effect of grammatical category on the acoustic correlates of stress in disyllabic noun-verb homophones. A closer look at the results reported in the original paper by Sereno and Jongman (1995) had revealed a rather elusive effect, which did not consistently reach statistical significance (see A Closer Look at Sereno and Jongman's (1995) Original Results). Every null result presents a challenge for interpretation as an absence of evidence does not necessarily allow for the conclusion of an absence of an effect. With quantitative analyses, the question of whether to conclude that there is no effect is intertwined with the notion of statistical power. The present replication extended both the number of target words produced by each speaker, as well as the number of participants overall, so that its statistical power was considerably greater than the one of the original study. In fact, the power analysis conducted indicates that the replication study was overpowered, that is, it would have found an even a smaller effect than the one reported in the original study, in case it existed. However, the statistical analysis does not yield a statistically significant effect on any of the acoustic parameters tested. This suggests that the effect is either nonexistent or extremely small in size. Consequently, no evidence for a general, or "systematic effect (of grammatical category) on the modulation of stress" (Sereno & Jongman, 1995, p. 71, the text in parentheses was added by the author) is obtained. The current failure to reproduce a frequently-cited effect underscores the importance of replication studies, as it adds to meta-analyses in the social sciences and psychology that indicate that a large percentage of effects reported can in fact not be reproduced (Open Science Collaboration, 2015; Camerer et al., 2018).

A positive finding, as reported in the original study, would have been of great theoretical importance, as it would have required to make significant adjustments to how stress assignment and implementation is modeled in current speech production theories. Most speech production models assume that lexical

⁵ The sample size for the analysis of the F0 ratio is smaller than for the other acoustic parameters, as fundamental frequency could not be tracked for 86 data points.

⁶ Further additional analyses, viz., a model comparison via Bayes factors, and an analysis for each individual speaker, are provided in the online supplementary material of this article.

stress is assigned via the retrieval of a metrical frame that is part of the lexical entry and which consists of syllables marked for stress (see e.g., Roelofs & Meyer, 1998). With non-stress-shifting noun-verb pairs, identical metrical frames are activated, which would explain the null results obtained by the present replication. One way to implement a positive finding is to assume the existence of a further metrical frame stored at the level of the grammatical category that interacts with the metrical information stored with the lexical entry. Another possibility would be to assume that co-activated words of the same grammatical category – most of which have category-typical metrical frames – may affect phonetic implementation via a cascading mechanism. However, since grammatical category did not affect pronunciation in the current data, the results are consistent with the way stress assignment is implemented in established speech production models (see Meyer & Belke, 2007, pp 477–479, for an overview).

It is important to note that a null result does not contradict findings which show that speakers possess knowledge of the typical stress patterns of nouns and verbs. A number of studies have demonstrated that speakers are able to apply such knowledge in a variety of tasks, for example, in the assignment of stress to nonce words (Davis & Kelly 1997; Guion, Clark, Harada, & Wayland, 2003). These results can be interpreted as showing that speakers know about the distributional biases of different metrical frames across grammatical categories in the lexicon. However, this knowledge does not seem to affect the level of phonetic implementation once a metrical frame has been retrieved.

Appendix

Table A1. Mean values and t test results of acoustic measures by stress position and category

	Trochaic word pairs				Iambic word pairs			
	nouns	verbs	<i>t</i>	<i>p</i>	nouns	verbs	<i>t</i>	<i>p</i>
Duration ratio	0.823	0.814	0.487	0.626	0.274	0.280	−0.755	0.451
Δ Intensity	2.911	2.234	1.929	0.054	−2.762	−2.737	−0.102	0.919
Δ F0	−0.985	−1.113	0.479	0.632	−3.647	−3.552	−0.281	0.779

Acknowledgement

I thank Benjamin Tucker for letting me conduct the recordings for the present study at the Alberta Phonetics Laboratory. Thanks also goes to Matthew Kelley, who helped me in selecting suitable filler items. Furthermore, I wish to express my gratitude to the research group “Spoken Morphology” at the Heinrich Heine University of Düsseldorf as well as the audience at the 12th Mediterranean Morphology Meeting for interesting discussions of the present material. I am indebted to Jessica Nieder and Simon Stein for helpful comments on previous versions of this paper. Viktoria Schneider and Claudiu Raum deserve to be thanked for doing most of the manual segmentation of the recordings. All remaining errors are mine.

Statement of Ethics

The author declares that all subjects have given their written informed consent. The author furthermore declares that the study protocol has been approved by the Research Ethics Board of the University of Alberta, where the experiment was carried out.

Disclosure Statement

The author has no conflicts of interest to declare.

Funding Sources

This study was funded by the German Research Foundation (*Deutsche Forschungsgemeinschaft*, DFG grant LO 2135/1-1).

References

- Baayen, H. R., Piepenbrock, R., & van Rijn, H. (2001). WebCelex. Online resource. <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. <http://CRAN.R-project.org/package=lme4>
- Behrens, S. J. (2014). Understanding language use in the classroom: A linguistic guide for college educators. Bristol, UK: Multilingual Matters. <https://doi.org/10.21832/9781783091768>
- Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. <http://www.praat.org>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Christiansen, M. H., & Monaghan, P. (2006). Discovering verbs through multiple-cue integration. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 88–107). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195170009.003.0004>
- Davies, M. (2014). The Corpus of Contemporary American English (COCA): 450 million words, 1990–2012. Brigham Young University. <https://corpus.byu.edu/>
- Davis, S. M., & Kelly, M. H. (1997). Knowledge of the English noun-verb stress difference by native and nonnative Speakers. *Journal of Memory and Language*, 36(3), 445–460. <https://doi.org/10.1006/jmla.1996.2503>
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, 44(3), 325–349. <https://doi.org/10.1006/jmla.2000.2741>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, 46(Pt 4), 403–427. <https://doi.org/10.1177/00238309030460040301>
- Herron, D. T. (1998). Context, Word-class, and Prosody in the Recognition of Open- and Closed-class Words (PhD thesis). University of California at San Diego, San Diego, CA.
- Ishikawa, K. (2006). English Stress Placement by Japanese Students: Effects of Syllable Structure and Noun-Verb Stress Differences. *English Literature Review*, 50, 38–54.

- Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347. <https://doi.org/10.1016/j.csl.2017.01.005>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest. <http://cran.r-project.org/web/packages/lmerTest/index.html>
- Meyer, A. S., & Belke, E. (2007). Word form retrieval in language production. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 471–487). Oxford, New York: Oxford University Press.
- Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55(4), 259–305. <https://doi.org/10.1016/j.cogpsych.2006.12.001>
- Oh, E. (2011). English-Speaking Children’s Knowledge of Stress Patterns Paper presented at ICPhS 2011, 1–36. <https://doi.org/https://doi.org/10.1002/9781118266892.ch1>
- Onnis, L., & Christiansen, M. H. (2008). Lexical categories at the edge of the word. *Cognitive Science*, 32(1), 184–221. <https://doi.org/10.1080/03640210701703691>
- Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Phillips, B. S. (2006). Word frequency and lexical diffusion. Palgrave studies in language history and language change. Basingstoke England, New York, NY: Palgrave Macmillan.
- R Core Team. (2014). R: A Language and Environment for Statistical Computing}. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 24(4), 922–939. <https://doi.org/10.1037/0278-7393.24.4.922>
- Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, 38(1), 57–76. <https://doi.org/10.1177/002383099503800103>
- Sluijter, A.M., & van Heuven, V.J. (1996) Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4 Pt 1), 2471–2485. <https://doi.org/10.1121/1.417955>
- 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](https://doi.org/10.1016/0010-0277(78)90019-7)