

UC San Diego

UC San Diego Previously Published Works

Title

Perceptual features predict word frequency asymmetry across modalities

Permalink

<https://escholarship.org/uc/item/63h684d1>

Journal

Attention, Perception, & Psychophysics, 81(4)

ISSN

1943-3921

Authors

Lau, Sin Hang
Huang, Yaqian
Ferreira, Victor S
[et al.](#)

Publication Date

2019-05-01

DOI

10.3758/s13414-019-01682-y

Peer reviewed

Perceptual Features Predict Word Frequency Asymmetry Across Modalities

Sin Hang Lau*, Yaqian Huang*, Victor S. Ferreira, and Edward Vul

University of California, San Diego

Author Note

Sin Hang Lau, Department of Psychology, University of California, San Diego;
Yaqian Huang, Department of Linguistics, University of California, San Diego; Victor S.
Ferreira, Department of Psychology, University of California, San Diego; Edward Vul,
Department of Psychology, University of California, San Diego.

The first two authors contributed equally.

Correspondence concerning this article should be addressed to Sin Hang Lau, Department
of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093.

E-mail: aubreylau@ucsd.edu

Abstract

The relationships between word frequency and various perceptual features have been used to study the cognitive processes involved in word production and recognition, as well as patterns of language use over time. However, little work has been done comparing spoken and written frequency against each other, which leaves open the question of whether there are modality-specific relationships between perceptual features and frequency. Words have different frequencies in speech and written texts, with some words occurring disproportionately more often in one modality than the other. The current study investigates whether perceptual features predict this frequency asymmetry across modalities. Our results suggest that perceptual features such as length, neighborhood density, and positional probability differentially affect speech and writing, which reveals different online processing constraints and considerations for communicative efficiency across the two modalities. These modality-specific effects exist above and beyond genre differences. This work provides arguments against theories that assume words differing in frequency are perceptually equivalent, as well as models that predict little to no influence of perceptual features on top-down processes of word selection.

Keywords: perceptual features, word frequency, language production, rational model

Perceptual Features Predict Word Frequency Asymmetry Across Modalities

Word frequency is an estimate of how often a word occurs in an average person's life. It is often calculated from collections of written texts or transcribed dialogues, or both. The observation that words differ in frequency is not a trivial one, as word frequency has been found to have profound impacts on our understanding of real-time language processing and long-term language change through behavioral and corpus studies. Hence, investigating why some words are used more frequently than others is important to our understanding of the fundamental properties of language and patterns of language use. In particular, it is critical to understand the causes and characteristics of the close relationship between word frequency and words' perceptual features. Even though data on word frequency in different modalities are available, not much work has systematically contrasted word frequency in speech and writing. Yet, one of the fundamental properties of human language is that we use it in multiple modalities. Due to this gap in the literature, there are at least two unanswered modality-specific questions on word frequency: Do words differ in spoken and written frequency? If so, what features predict whether or not a word tends to be used more often in one modality or the other?

Although the general relationships between word frequency and perceptual features have been thoroughly examined, the possibility of modality-specific effects have not. The process of producing language is roughly described as transforming non-verbal messages to sequences of sounds or letter strings. It is generally agreed that production involves stages including formulating non-linguistic conceptual messages, mapping messages to words and their grammatical features, and mapping words to their phonological or orthographic features (Levelt, 1989). Crucially, language production models make different predictions about the influence of

lower-level perceptual features, as well as the extent to which we should find modality-specific effects.

Some models treat production stages as discrete, top-down, and sequential (e.g., Fromkin, 1971; Garrett, 1975, 1980, 1982; Levelt, 1989), whereas others view the stages as interactive and parallel (e.g., Dell & Reich, 1981; Harley, 1984). Consider that in a model in which all stages of word production are independent and take place in a purely top-down and serial manner, lower-level perceptual features (such as phonological and orthographic features) should have little to no interaction with, or even access to, message formation and word selection. Specifically, Landauer and Streeter (1973) point out that earlier studies on word frequency effects assume the *perceptual equivalence hypothesis*, which proposes that high and low frequency words do not differ in perceptual dimensions and behavioral effects of word frequency are due to response bias. In contrast, if production stages interact with one another, we might see lower-level perceptual features affecting word selection. Further, it may be possible to observe modality-specific effects, such that phonological and orthographic features differentially affect speaking versus writing.

The goal of the present paper is to extend previously established relationships between perceptual features and word frequency by exploring whether words that are spoken more often than they are written and words that are written more often than they are spoken differ on perceptual dimensions. To this end, we investigate whether general measures of length, neighborhood density, and positional probability predict whether a word is spoken or written relatively more. Based on our interest in modality-specific differences, we further examine whether the difference in phonological and orthographic measures of the same perceptual feature

predicts this frequency asymmetry across modalities (e.g., whether words with more letters than phonemes are spoken or written more). Our study does not distinguish whether modality-specific effects are due to online processing factors or long-term language change, but it is likely that both processes are at play. Nevertheless, our emphasis on contrasting word frequencies in different modalities is still informative to production models. In particular, most studies comparing production models have focused on investigating speech errors (see Bock, 1987 for an overview). Hence, the modality-specific approach employed here provides another avenue to examine language production using a broader range of language use.

As far as we know, the current work is the first attempt to directly contrast speech and writing by investigating the difference in word frequency in the two modalities, rather than looking at the modalities separately (see Frauenfelder, Baayen, & Hellwig, 1993 for separate analyses on the relationship of neighborhood density and frequency in written text and speech, and Hayes, 1988 on the relationship between word rank and frequency). Modality-specific effects, especially if observed above and beyond genre differences, illustrate that the multimodal nature of language use and the influence of perceptual features must be considered in production models to enrich our understanding of processing and language as a communication system.

Our investigation is focused on three perceptual features closely tied with word frequency: length, neighborhood density, and positional probability. We refer to these features as low-level perceptual features, as their calculations are based on the surface features of words. However, studies have shown that their effects may propagate to other levels of processing in perception and production (see Jescheniak & Levelt, 1994 for word frequency effects in production, and Forster & Chambers, 1973 in comprehension).

Word length is the most direct feature to be detected by listeners or readers in real-time speech and writing. The number of syllables has an inhibitory effect, such that words with more syllables are recognized slower. In contrast, the number of letters can either be facilitatory or inhibitory; words with more letters are recognized faster for words containing 3-5 letters but slower for words containing 8-13 letters, and there is no significant length effect for words with 5-8 letters (New, Ferrand, Pallier, & Brysbaert, 2006). Neighborhood density of a word is a canonical measure of lexical similarity, corresponding to the number of words that differ from a target word by one perceptual unit, either phoneme or letter, and predicts how words are pronounced by English speakers. There is evidence that words with more phonological neighbors are more likely to be phonetically reduced in duration in connected speech, which makes them easier to produce (Gahl, Yao, & Johnson, 2012). In contrast, others report that those words are instead more prone to coarticulation and hyper-articulation, which reflect efforts to make words sound clearer and easier to identify (Scarborough, 2012; Scarborough, 2013; Wright, 2004). These results suggest that a listener-directed account in word perception and a production-based account of phonetic reduction of predictable forms co-exist. In the spoken modality, at least, words with more neighbors tend to require more clarity in communication (Scarborough & Zellou, 2013) and lexical similarity must be assessed online with active reference to phonological neighborhood (Scarborough, 2012). Positional probability is an estimate of the probability of the sequences of sounds or letters in each of the positions of a word. Phonotactic probability has been shown to influence word perception (Pitt & McQueen, 1998), and the sensitivity to positional probability has been found as early in development as nine months into infancy (e.g., the highly probable CVC sequence in English; Jusczyk, Luce, & Charles-Luce,

1994). Moreover, Vitevitch, Luce, Pisoni, and Auer (1999) showed that nonsense words with high probability segments and sequences are responded to more quickly than nonsense words with low probability segments and sequences, indicating that probabilistic phonotactic information is not only encoded in the mental representation of listeners and speakers but is actively involved in the parsing and processing sequences in speech streams. Taken together, word length and neighborhood affect the lexical level of word recognition, whereas neighborhood and positional probability affect the sublexical level of sequence parsing and articulation. Thus, in this paper, we broadly refer to these features as low-level perceptual properties that influence various stages of word production and perception, though the loci of their effects are not limited to the sublexical level.

Existing research suggests that there is a close relationship between word frequency and perceptual features. For example, frequent words tend to be shorter (Zipf, 1936) and have more neighbors (Landauer & Streeter, 1973). Recall that the calculation of positional probability takes frequency of occurrence into account, such that frequent words tend to contain more probable sequences than infrequent words (see the method section for a detailed explanation of calculations). These relationships can arise from two possible mechanisms: cognitive constraints on online language processing, and long-term language change. Cognitive constraints on online language processing refers to the cognitive processes involved in producing words. Shorter lengths and more probable sequences positively contribute to the accessibility of a word, making its pre-articulatory planning and retrieval faster and easier. Findings supporting this view include shorter naming times and higher naming accuracy rates for objects with shorter names (e.g., Klapp, Anderson, & Berrian, 1973; Roelofs, 2002; Santiago, MacKay, Palma, & Rho, 2000), as

well as for words with higher-probability phonological sequences (e.g., Goldrick & Larson, 2008, Vitevitch, 2002; Vitevitch, Armbruster, & Chu, 2004). On the other hand, the results on the effect of neighborhood density are mixed, which we review separately below.

The other possibility is that relationships between word frequency and perceptual features come from long-term language change, referring to the factors affecting whether or not a word is changed to have different features for processing and articulation over time. This idea is not implausible given recent findings showing that form-meaning correspondences are not as arbitrary as previously assumed. For instance, the distinctiveness of word forms reflects the distinctiveness of perceptual experiences of the referent concepts (Lynott & Connell, 2013), so word forms should change over time in accordance with how our perceptual experiences in the world change. Moreover, language users generally follow a few principles to maintain communication efficiency. Zipf (1949) explains the inverse relationship of word length and frequency by the *least effort principle*, which suggests that language users as a community efficiently assign shorter words to more frequently used meanings because they are easier and more economical to produce. The same principle can account for the tendency to reuse more probable sequences as well. Piantadosi, Tily, and Gibson (2011) revised Zipf's explanation to an efficiency principle that shortens the most predictable but not necessarily the most frequent words, which captures the finding that shorter words tend to contain more probable sequences. Indeed, it has been shown in both a corpus study and a behavioral experiment that the abbreviated form of a word is more likely to be used in predictable contexts (Mahowald, Fedorenko, Piantadosi, & Gibson, 2013). In line with these results, phonetic reduction also occurs as a function of the target word's frequency and its conditional probability given the

previous word (i.e., frequent and probable words are more likely to be reduced; Jurafsky, Gregory, & Raymond, 2001). In addition, social factors independent of processing may affect the frequency of occurrence of a concept or its conditional probability over time, which may lead to changes in the perceptual features of words.

Note, however, that behavioral data do not clearly indicate whether neighborhood density facilitates or inhibits word production and recognition. In speech production, some report that words with more phonological neighbors are named faster and more accurately (e.g., Baus, Costa, & Carreiras, 2008; Vitevitch, 2002; Vitevitch & Sommers, 2003), but these results are not always replicated and others have even found the opposite pattern (see Sadat, Martin, Costa, & Alario, 2014 for a review on phonological neighborhood effects). Although results from naming latencies are inconclusive, a larger neighborhood facilitates the retrieval accuracy of the target word in both speaking and writing (Goldrick, Folk, & Rapp, 2010). In visual word recognition, Pollatsek, Perea, and Binder (1999) showed that a larger neighborhood facilitates lexical decision but inhibits reading by increasing gaze duration for the target word (see Andrews, 1997 for a review of orthographic neighborhood effects). Nonetheless, there are clear inhibitory effects in spoken word recognition. That is, words with more phonological neighbors are consistently recognized more slowly and less accurately than words with fewer neighbors (Luce & Large, 2001; Luce & Pisoni, 1998; Vitevitch & Luce, 1999). A brief summary is that tasks requiring different modalities, as well as different measures of the same process could produce different neighborhood effects. These contrasting results may be due to differences in activation status of the neighbors, determined by multiple factors including frequency, semantic distance, and phonological overlap. Under an interactive activation framework where representations of words

with similar forms are co-activated and in competition with one another for selection, strongly active neighbors would generally yield a net inhibitory effect, whereas weakly active neighbors would yield a net facilitatory effect (Chen & Mirman, 2012). Additionally, there are cross-linguistic differences in the extent to which orthography is consistent with phonology, and stronger lexical competition is found in languages with relatively consistent mappings and rich morphology (e.g., Andrews, 1997; Vitevitch & Stamer, 2006).

Regardless of how neighborhood precisely affects language processing, the existence of both facilitatory and inhibitory effects reported in the studies above possibly reflects opposing forces in communication, rather than just considerations of production ease and economy for the producer. Words with more neighbors may be easier to produce, both because of the greater practice that comes from being in a denser neighborhood, and as a consequence of diachronic forces favoring reusing words that are easier to produce or reshaping words over time to have easier-to-produce features (e.g., shorter lengths and more probable sequences), and thus expanding certain neighborhoods that conform to those features. At the same time, words with more neighbors also create potential confusion and processing difficulties in comprehension. Rational models suggest that the existence and persistence of some amount of ambiguity serves a useful communicative function and reflects that an efficient communication system takes into account both the producer and the comprehender's interests. The amount of ambiguity that remains (e.g., uncertainty about what word is produced in the perceptual aspect, or uncertainty about the meaning of a polysemous or homophonous word in the semantic aspect) is a balance between the opposing forces of clarity and ease, so that utterances are easy enough to produce but also not too difficult to comprehend (Piantadosi, Tily, & Gibson, 2012). The structure of our

lexicon also reflects the opposing forces of arbitrariness, which contributes to the superior flexibility and complexity of human language compared to animal communication but hinders learning, versus systematicity and iconicity, which promote learning but constrain abstraction and only serve a small subset of communicative goals (Dingemans, Blasi, Lupyan, Christiansen, & Monaghan, 2015). Therefore, the existing amount of arbitrariness, systematicity, and iconicity in our lexicon reflected by word features is the result of the linguistic community's ongoing attempt to strike a balance between learnability and expressiveness.

To investigate whether there are processing and communicative concerns that differentially affect the two modalities, our main analysis examines if the three perceptual features of interest predict whether a word is used disproportionately more often in one modality rather than the other. One potential confound in our investigation is the possibility that modality-specific effects simply reflect genre differences instead of processing features or communicative concerns specific to speech versus writing. Different genres have different task goals and require varying degrees of formality, which inevitably affects word choice even within the same modality. Additionally, some genres may be more common in one modality than the other, leading to skewed data if genre is not considered in cross-modality comparisons. In light of these limitations, we include a secondary analysis to statistically separate formality differences across genres from any modality-specific effect we observe.

Method

Materials

Frequency information in the study was obtained from the spoken and written frequency-per-million lists compiled by Leech, Rayson and Wilson (2001) based on the British

National Corpus (BNC). The BNC is based on 100 million tokens, containing different genres of written and spoken sources, for example, imaginative and informative texts and conversational and task-oriented dialogues. For simplicity's sake, we refer to these four genres as informal writing, formal writing, informal speech, and formal speech, respectively. After excluding phrases that were tagged as words, words with only one single phoneme or character (for the purpose of bigram/biphone positional probability measures), abbreviations, special characters, and words that have no frequency information in either modality, we conducted our primary analyses contrasting spoken and written frequencies on 4515 types. Our secondary analyses on genre comparisons were conducted using 1408 types with frequency information available in all four genres and the same exclusion criteria as the primary analyses.

In the primary analyses, the asymmetry of spoken and written log frequencies was the variable of interest, so the dependent variable was calculated as the difference between these two log frequency measures for each word, that is, the spoken-written frequency difference¹. In the secondary analyses, the asymmetries in frequencies between different genres for both within- and cross-modality were examined. The four different genres (i.e., formal writing, informal writing, formal speech, and informal speech) constitute six possible pairwise comparisons. The dependent variable for these six comparisons was therefore calculated as the difference between the log frequencies of a word in the two genres in a particular comparison (e.g., log frequency in informal speech minus log frequency in formal writing). The current work evaluates if these asymmetries in frequencies are predicted by three main families of perceptual features: length, neighborhood density, and positional probability.

¹ We added 1 to each raw frequency value as a smoothing method for log transformation: $\log(\text{spoken frequency}+1) - \log(\text{written frequency}+1)$, because there were entries with 0 raw per-million frequency in either modality. Method adopted from Brysbaert & Diependaele (2013).

We first compiled data for the phonological and orthographic measures for each perceptual feature. The raw values were then standardized into z-score units (by subtracting the mean and dividing by the standard deviation of each measure) to facilitate later modality-specific comparisons, as some measures may be more variable in one modality than the other. Length was calculated by counting the number of letters or phonemes in a word. Neighborhood density and positional probabilities were obtained from CLEARPOND (Marian, Bartolotti, Chabal & Shook, 2012), a cross-linguistic online corpus tool. This tool is based on the SUBTLEX-US corpus, which contains 51 million tokens from subtitles (Brysbaert & New, 2009). Neighborhood density was calculated as the number of words that differ from the target word by an edit (addition, deletion, or substitution) distance of one in either phonemes or letters. For instance, *face* [feɪs] and *fact* [fækt] are orthographic neighbors, whereas *fish* [fɪʃ] and *fig* [fɪg] are phonological neighbors. Bigram and biphone positional probabilities were calculated by dividing the sum of the log frequencies of all words containing the target bigram/biphone unit at the N and N+1 positions by the sum of the log frequencies of all words containing any unit at the N and N+1 positions. For example, the bigram positional probability of the string *ca* in *cat* was calculated by dividing the sum of the log frequencies of all words containing *ca* at the first and second position (e.g., *cat*, *captain*, *cap*, *cape*, *captivate*, etc.) by the sum of the log frequencies of all words containing two or more letters (e.g., *ant*, *bee*, *citrus*, *dog*, *egg*, etc.) Then, we proceeded to use the same method to calculate the positional probability for the next bigram (e.g., *at* in *cat*). We then log transformed the probabilities for all bigram units in a single word (smoothed by adding $10e-6$) and took the average to get the positional probability for each word, controlling for word length. The same was done for biphone positional probabilities.

After compiling data for each modality, we calculated an average measure and a phonological-orthographic difference measure for each family of the perceptual features. The average measures were calculated by taking the average of the phonological and orthographic measures of each feature (e.g., “ownership” has a phonological length of 0.22 and an orthographic length of 1.04 in z-score unit, so its average length is 0.63), which allowed us to investigate the general effect of each feature averaging across modalities. In addition to the average measures, we calculated a phonological-orthographic difference measure for each feature (i.e., phonological length - orthographic length, phonological neighborhood - orthographic neighborhood, and bigram probability - biphone probability). When evaluating the effects of these difference measures, we included the corresponding average measures as covariates to partial out the general effect of each feature before examining modality-specific effects. This was justified by the high correlations between the phonological and orthographic measures for all features (see Figure 1 and 2).

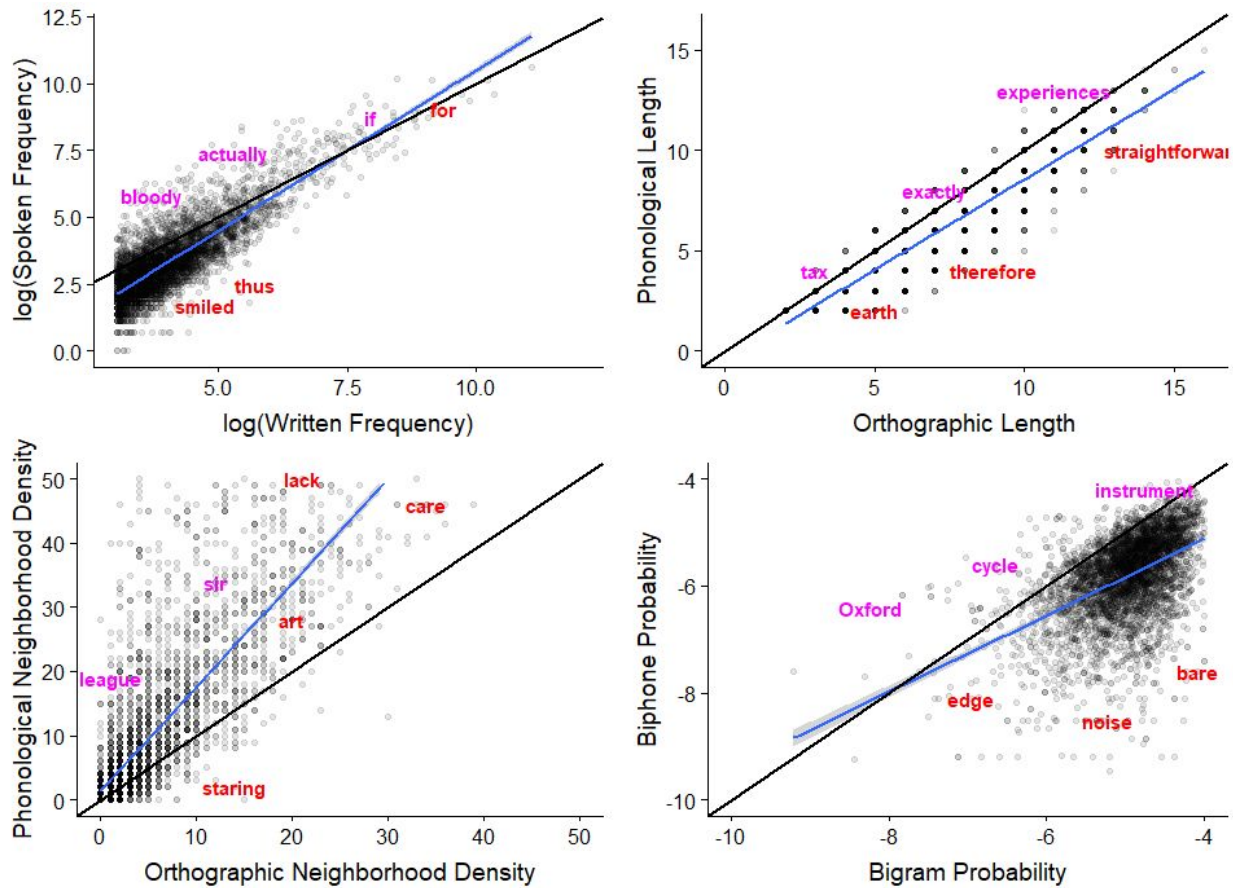


Figure 1. Correlations between phonological and orthographic measures of frequency, length, neighborhood density, and positional probability. The blue lines are the linear model regression lines, and the black lines are the identity lines.

Furthermore, Figure 2 shows the full correlation matrix between all raw, average, and difference measures of frequencies and perceptual features. As can be seen by comparing the correlations inside the black square against the correlations elsewhere, orthogonalizing the raw perceptual measures to average and difference measures successfully reduced the correlations between our variables of interest, and thus is conducive to producing more interpretable results.

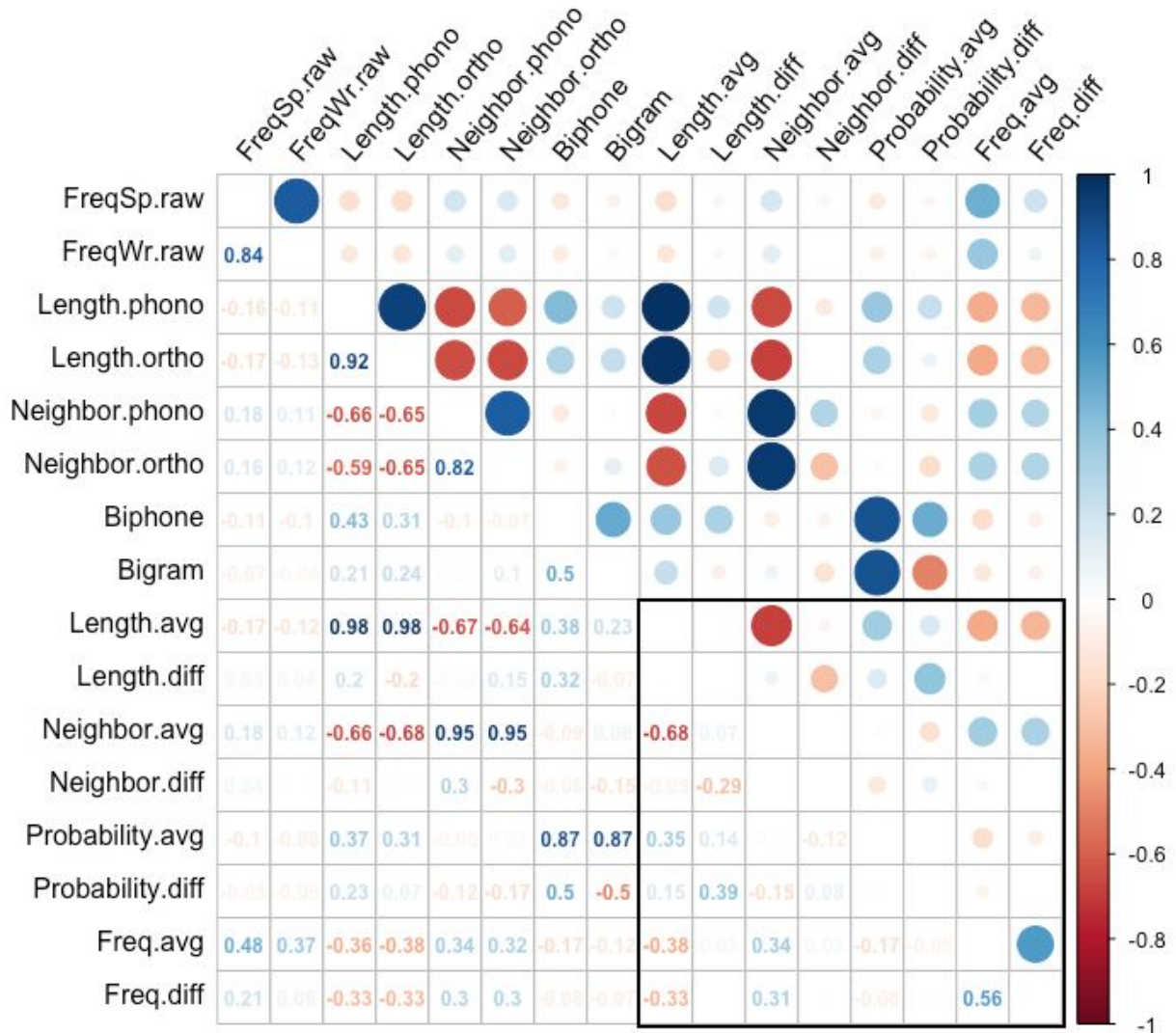


Figure 2. Visualization of the full correlation matrix of all variables. Color dots above the diagonal and correlation values below the diagonal are used to represent the strength and direction of the correlations. Red shows a negative correlation; blue shows a positive correlation. Larger dots and darker colors indicate stronger correlations than smaller dots and lighter colors. The black square in the bottom right corner outlines the orthogonalized measures used in our analyses.

Analyses

In our primary analyses, we used multiple regression to examine the effects of the three families of perceptual features in predicting the spoken-written frequency asymmetry. Therefore, the full model uses the difference between spoken and written log frequencies as the dependent variable, all perceptual measures (i.e., length, neighborhood, and positional probability average and difference measures) as the independent variables, and the average frequency of a word across modalities and part-of-speech (PoS) as covariates (i.e., spoken-written log frequency \sim average log frequency + PoS + average length + average neighborhood density + average positional probability + length difference + neighborhood density difference + positional probability difference). To evaluate the effect of each measure, we performed model comparisons between the full model against models dropping one term at a time while keeping all the other terms. That is, we are reporting the partial effects of average frequency and each perceptual measure in predicting the spoken-written frequency asymmetry. Readers who are interested in the marginal effects of perceptual measures may refer to Figure 2. Results from F tests and coefficients are reported in the results section (see the Appendix for a table of the coefficients, standard errors, t-, and p-values for all predictors).

In our secondary analyses, we aimed to first establish the relationships between perceptual features and frequency differences across genres. Recall that there are four genres, yielding six possible pairwise comparisons. For each pairwise comparison, we used multiple regression to extract the slope estimates for the three families of perceptual features in predicting the frequency difference between two genres, including the average frequency between the two genres and part-of-speech (PoS) as covariates (i.e., the difference between two genres' log frequencies \sim average log frequency + PoS + average length + average neighborhood density +

average positional probability + length difference + neighborhood density difference + positional probability difference). Our main interest is not in the perceptual features of words frequently used in each genre per se but in estimating whether there are additive effects of formality and modality. Nevertheless, the coefficients and their significance levels from all comparisons are reported graphically in Figure 3 alongside an example interpretation for interested readers.

Up until this point, we have been referring to genre as a construct that is the combination of the formality of the task and the modality used. After establishing the relationships between perceptual features and genres, we used multiple regression to separate these relationships into formality and modality effects and compared their magnitudes. This addresses whether there is any modality effect when formality is controlled for. To this end, we ran a regression on all four genres together. Specifically, for each word we calculated how its frequency in each genre deviates from its average frequency across all genres. Then we predicted these difference scores using the interactions of average frequency and each of our perceptual features with indicator variables coding for formality (+1 for informal, -1 for formal) and modality (+1 for spoken, -1 for written). These interaction terms capture the additive coefficient components of formality and modality differences across the set of all four genres. We report the coefficients and standard errors for these interaction terms.

The data and the analysis code can be found at <http://osf.io/qz4hp>.

Results

To interpret the coefficients reported in the results, a positive coefficient indicates that an increase in the measure of interest disproportionately increases the frequency with which a word

is spoken rather than written, whereas a negative coefficient indicates that an increase in the measure of interest is in favor of a word being written rather than spoken.

Average Frequency

Words with a higher average frequency tend to occur disproportionately more often in speech than in writing (the average frequency coefficient was positive, $\beta = 0.37$), $F(1, 4490) = 1370.32, p < .001$. This shows that producers are more likely to use more common words in speech than in writing. Below we focus on reporting the relationships between perceptual features and the spoken-written frequency asymmetry.

Length

Shorter length leads to faster and easier production (Bates et al., 2003; Szekely et al., 2004). If producers are more concerned with time constraint and the ease of production in one modality than the other, length should be a significant predictor of the spoken-written frequency difference, and it was. In our model comparison, we found a significant effect of average length, $F(1, 4490) = 44.09, p < .001$; longer words are generally written more than they are spoken (the average length coefficient was negative, $\beta = -0.10$). We also found a significant effect of phonological-orthographic length difference, $F(1, 4490) = 24.30, p < .001$; words with more phonemes than letters (i.e., words that are relatively phonologically complex) occur disproportionately more frequently in writing than in speech (the length difference coefficient was negative, $\beta = -0.14$). These results suggest that speakers are relatively more concerned with production ease than writers.

Neighborhood Density

Higher neighborhood density may contribute to production ease but may also lead to potential confusion. Thus, it is a good indicator of the relative importance of clarity and ease. If the general concern for these two dimensions differ across modalities, neighborhood density should be a significant predictor of the spoken-written frequency difference, and it was. We found a significant effect of average neighborhood density, $F(1,4490) = 17.23, p < .001$; words with more phonological and orthographic neighbors are used more often in speech than in writing (the average neighborhood density coefficient was positive, $\beta = 0.06$).

Though the average neighborhood density effect showed that speakers are generally more concerned with ease than writers, there should still be hints of concern for clarity if producers of both modalities are truly rational. If that is the case, producers should prefer words with relatively fewer neighbors in the modality in use, and the phonological-orthographic neighborhood difference should be a significant predictor of the spoken-written frequency difference. This modality-specific effect is indeed what we found. The effect of phonological-orthographic neighborhood difference was significant, $F(1,4490) = 15.45, p < .001$; words with more phonological than orthographic neighbors occur more frequently in writing than in speech (the coefficient of the phonological-orthographic neighborhood density difference was negative, $\beta = -0.07$). However, this neighborhood difference should be interpreted with caution, as we show in our later analyses that this is likely to be a formality effect more so than a modality effect. Nevertheless, the average neighborhood effect is robust.

Positional Probability

High positional probability has to do with both ease and high predictability. If these two characteristics differentially affect the two modalities, positional probability should be a

significant predictor of the spoken-written frequency difference, and it was. We found a significant effect of average positional probability, $F(1,4490) = 10.14, p = .001$; words containing more probable sequences of phonemes and letters appear more frequently in speech than in writing (the average positional probability coefficient was positive, $\beta = 0.04$). This effect should be interpreted with caution, as we show later that this is more likely to be a formality effect.

We showed that ease and predictability are relatively more crucial to speech than writing, but it could still be the case that producers prefer words that are particularly probable in the modality in use compared to the other. If that is true, the biphone-bigram difference should be a significant predictor of the spoken-written frequency difference. Indeed, we found this modality-specific effect, $F(1,4490) = 31.30, p < .001$. Words with more probable phonological sequences relative to their orthographic sequences tend to have a higher spoken frequency than written, and vice versa (the coefficient of the biphone-bigram difference was positive, $\beta = 0.06$).

Genre Comparisons

To examine whether the modality effects reported above simply reflect formality differences, we compared how the three families of perceptual measures influence frequency differences across all possible combinations of genres (i.e., formal versus informal speech, formal speech versus formal writing, informal speech versus formal writing, etc., yielding six pairwise comparisons in total). The results are shown in Figure 3.

Figure 3 shows how each perceptual feature relates to the difference in frequency between the genre for the column minus the frequency of the genre for the row. A positive coefficient indicates that larger scores on the perceptual feature predict disproportionately higher

frequencies for the genre of the column (rather than the row). For instance, considering the length.average panel (top left), -0.66* in the bottom left square indicates that the coefficient for average length in predicting the difference between informal speech and formal writing is negative; this means that longer words tend to be disproportionately more frequent in formal writing than in informal speech.

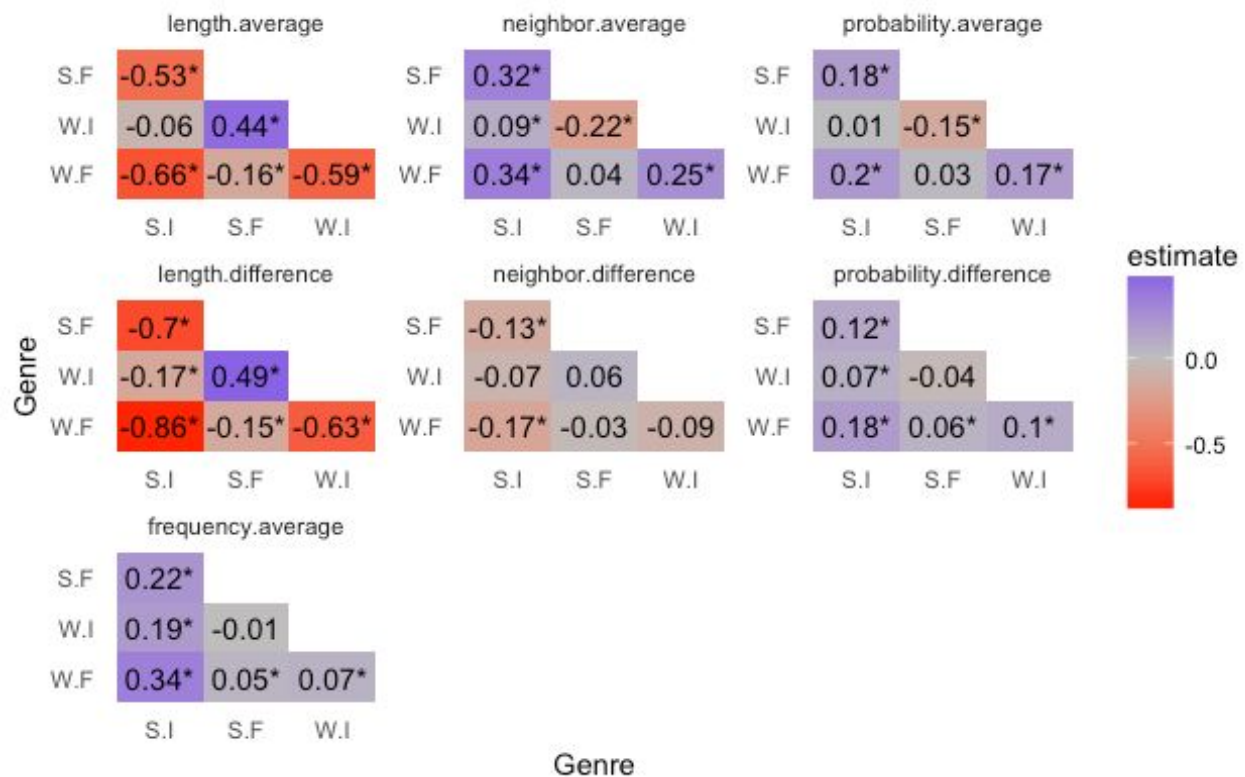


Figure 3. The coefficients for each perceptual measure and average frequency (panels) in predicting the frequency difference between the genre of the column and the genre of the row (S: Speech, W: Writing; I: Informal, and F: Formal; S.F: formal speech, W.I: informal writing, and so on). Asterisk indicates $p < .05$. Purple indicates a positive slope, red indicates a negative slope, with darker colors indicating larger effects.

As is most clearly seen in the panels of average length and phonological-orthographic length difference, the specific slopes may be fruitfully decomposed on the assumption that there is some additive effect of a formality difference (Informal - Formal), and some additive effect of a modality difference (Spoken - Written). This allows us to separate out any modality effects, and compare their magnitudes to the effects of formality. Figure 4 plots these effects, showing how formality and modality differences interact with perceptual features in predicting disproportionately high and low frequencies in each genre.

As Figure 4 shows, across all perceptual features, the effect of modality difference (Spoken - Written) are in the same direction, and smaller than, the effects of formality difference (Informal - Formal). Specifically, words with higher average frequency (across all genres) are more likely to be used in informal, rather than formal, language ($\beta = 0.07$, $SE = 0.008$), and more likely to be spoken, rather than written ($\beta = 0.05$, $SE = 0.008$), suggesting that producers tend to use more common words in speech and in informal contexts.

Longer words are less likely to be used in informal ($\beta = -0.30$, $SE = 0.02$) and spoken ($\beta = 0.07$, $SE = 0.02$) language. Moreover, words with more phonemes than letters are also less likely to be used in informal ($\beta = -0.35$, $SE = 0.03$) and spoken ($\beta = -0.09$, $SE = 0.03$) language. These results are in line with our primary analysis, suggesting that producers are more concerned with the ease of production when speaking (rather than writing), and in informal (rather than formal) contexts.

Words that have more phonological and orthographic neighbors are more likely to be used in informal ($\beta = 0.16$, $SE = 0.02$) and spoken ($\beta = 0.05$, $SE = 0.02$) language, suggesting that producers are relatively less concerned with avoiding ambiguity in speech and in informal

contexts. Additionally, words with disproportionately more phonological (relative to orthographic) neighbors are more likely to be used in formal language ($\beta = -0.04$, $SE = 0.02$). Although the coefficient estimate for modality suggests that these words tend to occur more frequently in writing, the effect is indistinguishable from zero ($\beta = -0.01$, $SE = 0.02$). Altogether, these neighborhood results suggest that the relative phonological/orthographic distinctiveness of a word only affects the frequency asymmetry across formalities but not modalities, whereas the general concern for clarity and ease affects both.

Words with more probable phonological and orthographic sequences are more likely to be used in informal language ($\beta = 0.06$, $SE = 0.01$) but are not more likely to occur in either modality ($\beta = 0.002$, $SE = 0.01$). However, words with disproportionately more probable phonological sequences (relative to their orthographic sequences) are more likely to be used in informal ($\beta = 0.06$, $SE = 0.01$) and spoken ($\beta = 0.04$, $SE = 0.01$) language. These results suggest that producers are generally concerned with predictability regardless of whether they are speaking or writing, but they are more likely to pick words with more probable sequences in the modality in use (relative to the modality not in use).

In sum, we found large formality effects, but modality effects remained after controlling for formality. Additionally, formality and modality differences trended in the same direction, such that words that are used more often in informal contexts and words that are spoken more share similar features, and words that are used more often in formal contexts and words that are written more share similar features. Note that the modality effects were calculated after factoring out formality, which suggests that speech may inherently be more informal than writing, even when the formality of the task itself is controlled for.

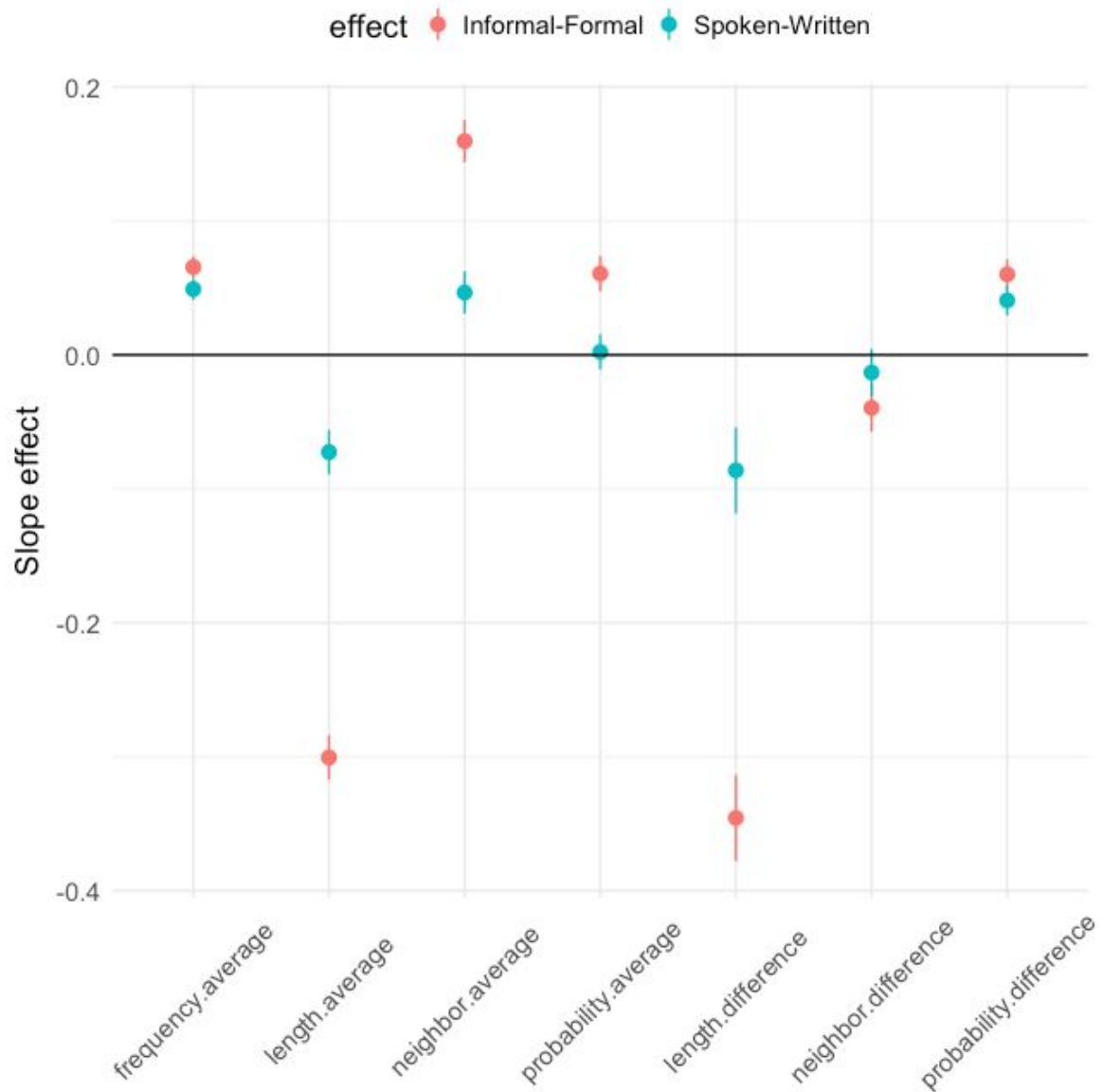


Figure 4. The slope estimates of the interaction terms for formality (Informal - Formal) and modality (Spoken - Written) differences in each perceptual feature.

Discussion

In this study, we investigated the relationships between perceptual features and the difference in spoken and written word frequency. Specifically, we examined the general and

modality-specific effects of length, neighborhood density, and positional probability in predicting frequency asymmetries across modalities. Second, we conducted genre comparisons to investigate the relationship between modality effects and formality effects across genres.

Our results demonstrated three main points: First, the previously found effects of perceptual features differentially affect word frequency in different modalities; words that are spoken disproportionately more than they are written are generally shorter and have more neighbors. Second, perceptual features have additional modality-specific effects even after controlling for their general effects; words that are spoken disproportionately more than they are written tend to have fewer phonemes than letters and more probable phonological sequences (relative to orthographic sequences). Third, formality effects are large, but the aforementioned modality effects exist above and beyond genre differences. The only two effects that were found to be formality rather than modality effects were the average positional probability and the phonological-orthographic neighborhood difference effects; that is, producers tend to use words with more probable sequences and more orthographic neighbors (relative to phonological neighbors) in informal language, regardless of modality. Potential implications for production models and communication efficiency are discussed below.

Our observations that words indeed differ by spoken and written frequencies and that the differences can be predicted by various perceptual features suggest that examining the multi-modal nature of language production may lend insights into existing theories. These findings are in line with Landauer and Streeter's (1973) rejection of the assumption that words of different frequencies are perceptually equivalent. It is important to note that results in support for perceptual equivalence may have come from examining word frequency effects in one modality

only. More broadly, the lack of perceptual equivalence also has implications for production models. Our results suggest that lower-level perceptual features influence word choice in different modalities, though we make no commitment to whether the driving force is online processing or long-term language change. This is in line with models that allow interactions between production stages (e.g., Dell & Reich, 1981; Harley, 1984), but not with entirely top-down models where perceptual features belong to the last stage of production and have no interaction with previous stages such as message formation and word selection (e.g., Fromkin, 1971; Garrett, 1975, 1980, 1982).

Even though the production stages for speech and writing may not be drastically different, producers may have different concerns when using different modalities. Here we propose that the time pressure in producing speech leads speakers to be relatively more concerned with the ease of production than writers, resulting in the preference for perceptual features that have facilitatory effects on production from both online processing and long-term language change perspectives. Shorter length and higher positional probability are favorable factors for online processing, facilitating word production (Goldrick & Larson, 2008; Klapp et al., 1973; Roelofs, 2002; Santiago et al., 2000; Vitevitch, 2002; Vitevitch et al., 2004). Likewise, Zipf's least effort principle (1949) and Piantadosi et al.'s (2011) revision of it also predict that words with these features are likely to be reused over time due to their contribution to production ease and high predictability. On the other hand, not only is writing not subject to time constraints as stringent as speaking is, corpus data also reflect written texts that have likely undergone revision. Hence, content and clarity may be more important to writing, resulting in a lesser degree of concern for the ease of production.

We have demonstrated so far that producers have different degrees of concern for clarity and ease in speech versus writing, but it is not the case that producers are only concerned about one dimension in each modality and not the other. In fact, our results on neighborhood density provide suggestive evidence that producers try to strike a balance between the opposing forces of clarity and ease in both modalities, as suggested by rational models (Piantadosi et al., 2012). Though our review of neighborhood density shows that its effects on real-time processing are inconclusive and depend on many different factors, speakers may prefer words with more neighbors in general when they have facilitatory effects, possibly because disambiguation through the means of selecting a relatively distinct word is inefficient and unnecessary. The function of disambiguation in speech can be served by other contextual cues, such as gesture and prosody. But conditioned on the tendency to use words with more neighbors, speakers also tend to pick words that sound more distinct (i.e., words with relatively fewer phonological neighbors than orthographic). Likewise, even though writers tend to use words with fewer neighbors, they also pick words that are orthographically more distinct (i.e., words with relatively fewer orthographic than phonological neighbors). In other words, the seemingly contradictory findings on the general and modality-specific (phonological-orthographic difference) effects of neighborhood illustrate that both speakers and writers attempt to find the middle ground between easy yet completely confusing and effortful yet completely clear communication, given the availability of other contextual cues. However, these results should be interpreted with caution, as the effect of the phonological-orthographic difference in neighborhood density shrinks to zero when formality is controlled for, whereas the general effect of neighborhood density is robust.

In the introduction, we noted that genre differences within and across modalities may confound our results. Thus, we included some secondary analyses to dissociate formality effects from modality effects to the extent that was possible with our data. Indeed, we found a huge formality effect, suggesting that formal and informal words differ by perceptual features. Even though the effects of average positional probability and neighborhood difference measures shrank after controlling for formality, the average length, average neighborhood density, length difference, and positional probability difference effects reported above still survived this analysis. This demonstrates that words that are spoken disproportionately more and words that are written disproportionately more do have different perceptual features, which do not reduce to genre differences. Despite our attempt to control for genre differences, it may be the case that formality is a factor that is impossible to fully control for. Even given a matched genre, such as a research report, the words used in a conference talk versus those used in a manuscript would be vastly different in formality, with speech being more conversational and informal than writing on average. This is also supported by our data showing that words that occur disproportionately more in informal contexts and in speech share similar features. To a certain extent, this formality difference simply is the difference between how producers choose words in speech and writing. As a result, some genres may occur more in one modality than the other, because the distinctive features of speech and writing may make one modality better-suited for some communicative goals than the other. Nevertheless, our findings suggest that modality effects are not simply genre effects. Thus, it is worth characterizing how spoken and written language differ.

All in all, our first attempt at directly contrasting spoken and written word frequencies has yielded the discovery of several modality-specific effects above and beyond genre

differences. We encourage future research to take into account the importance of perceptual features in different modalities, in order to inform existing theories of word production and recognition.

Conclusion

The current study elaborated on previously established word frequency effects by contrasting spoken and written word frequency. We provided evidence that perceptual features such as length, neighborhood density, and positional probability predict whether a word occurs disproportionately more often in speaking or writing. Further, we found modality-specific effects such that phonological and orthographic measures of the same feature differentially affect production in different modalities. These results still stand after controlling for formality differences across genres. Our results are in line with production models that allow for interactions between lower-level perceptual features and other top-down processes, as well as rational models of communication which suggest producers negotiate between the opposing forces of clarity and ease.

References

- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4), 439-461.
- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2), 344-380.
- Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, 23(6), 866-888.
- Bock, J. K. (1987). Exploring levels of processing in sentence production. In G. Kempen (Ed.), *Natural language generation* (pp. 351-363). Dordrecht: Martinus Nijhoff.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera 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, M., & Diependaele, K. (2013). Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behavior Research Methods*, 45(2), 422-430.
- Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, 119(2), 417.
- Dell, G.S., & Reich, P.A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.

- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603-615.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12, 627–635.
- Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood Density and Frequency Across Languages and Modalities. *Journal of Memory and Language*, 32(6), 781–804.
- Fromkin, V. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, 47(1), 27-52.
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language*, 66(4), 789-806.
- Garrett, M.F. (1975). The analysis of sentence production. In G.H. Bower (Ed.) *The psychology of learning and motivation* (Vol. 9). New York: Academic Press.
- Garrett, M.F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1). London: Academic Press.
- Garrett, M.F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.) *Normality and pathology in cognitive functions*. London: Academic Press.
- Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, 62(2), 113-134.

- Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, 107(3), 1155-1164.
- Harley, T. A. (1984). A Critique of Top-down Independent Levels Models of Speech Production: Evidence from Non-plan-internal Speech Errors. *Cognitive Science*, 8(3), 191-219.
- Hayes, D. P. (1988). Speaking and Writing: Distinct Patterns of Word Choice. *Journal of Memory and Language*, 27(5), 572-585.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp.229-254). Amsterdam: John Benjamins.
- Jescheniak, J. D., & Levelt, W. J. (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.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33(5), 630.
- Klapp, S. T., Anderson, W. G., & Berrian, R. W. (1973). Implicit speech in reading: Reconsidered. *Journal of Experimental Psychology*, 100(2), 368-374.
- Landauer, T. K., & Streeter, L. A. (1973). Structural Differences Between Common and Rare Words: Failure of Equivalence Assumptions for Theories of Word Recognition. *Journal of Verbal Learning and Verbal Behavior*, 12(2), 119-131.

- Leech, G., Rayson, P., Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press, Cambridge, MA.
- Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16(5-6), 565-581.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 136.
- Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, 45(2), 516-526.
- Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313-318.
- Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8): e43230.
- New, B., Ferrand, L., Pallier, C., Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic bulletin & review*, 13(1), 45-52.
- Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 1142.

- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280-291.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon?. *Journal of Memory and Language*, 39(3), 347-370.
- Roelofs, A. (2002). Syllable structure effects turn out to be word length effects: Comment on Santiago et al.(2000). *Language and Cognitive Processes*, 17(1), 1-13.
- Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, 68, 33-58.
- Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language and Cognitive Processes*, 15(1), 1-44.
- Scarborough, R. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua*, 122(2), 164-176.
- Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*, 41(6), 491-508.
- Scarborough, R., & Zellou, G. (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, 134(5), 3793-3807.

- Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, 51(2), 247–250.
- Vitevitch, M. S. (2002). The Influence of Phonological Similarity Neighborhoods on Speech Production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 28(4), 735–747.
- Vitevitch, M. S., Armbruster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 514.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*, 40(3), 374-408.
- Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and language*, 68(1-2), 306.
- Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, 31(4), 491-504.
- Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, 21(6), 760-770.
- Wright, R. (2004). Factors of lexical competition in vowel articulation. In: J. J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI*. Cambridge: Cambridge University Press.
- Zipf, G. (1936). *The Psychobiology of Language*. London: Routledge.

Zipf, G. (1949). *Human Behavior and the Principle of Least Effort*. New York: Addison-Wesley.

Appendix

Output of the Full Model in the Primary Analyses

Term	Estimate	SE	t-statistic	p-value
(Intercept)	-2.1031312	0.0434639	-48.3880311	< 2e-16***
frequency.avg	0.3661279	0.0098906	37.0177929	< 2e-16***
PoSAdv	0.1517775	0.0458870	3.3076348	0.0009483***
PoSConj	-0.4019741	0.1204072	-3.3384573	0.0008493***
PoSDet	-1.1936820	0.1910114	-6.2492709	4.51e-10***
PoSDetP	-0.4247657	0.1138133	-3.7321287	0.0001922***
PoSEx	-0.2444088	0.6185548	-0.3951288	0.6927666
PoSInf	-1.6877377	0.6190035	-2.7265397	0.0064251**
PoSInt	2.3203934	0.3571059	6.4977741	9.04e-11***
PoSNeg	-1.2272320	0.6177841	-1.9865064	0.0470378*
PoSNoC	0.0145848	0.0276397	0.5276773	0.5977494
PoSNoP	0.0887995	0.0493292	1.8001407	0.0719055
PoSNoP-	-0.5160687	0.1798139	-2.8700155	0.0041237**
PoSNum	1.2354932	0.1379545	8.9558004	< 2e-16***
PoSOrd	0.3328159	0.2527365	1.3168493	0.1879564
PoSPrep	-0.5436470	0.0842287	-6.4544131	1.20e-10***
PoSProN	0.3831377	0.1041216	3.6797138	0.0002362***
PoSVerb	0.0932906	0.0311025	2.9994519	0.0027195**
PoSVMOD	-0.2321921	0.1893041	-1.2265565	0.2200537
length.avg	-0.0988304	0.0148849	-6.6396253	3.52e-11***
length.diff	-0.1360004	0.0275908	-4.9291881	0.0000009***
neighbor.avg	0.0590042	0.0142139	4.1511626	0.0000337***
neighbor.diff	-0.0655992	0.0166915	-3.9300959	0.0000862***
probability.avg	0.0396789	0.0124581	3.1849774	0.0014576***
probability.diff	0.0592684	0.0105943	5.5943795	2.35e-08***