Word Frequency Can Affect What You Choose to Say

Mark Koranda*, Martin Zettersten* & Maryellen C. MacDonald
(mjkoranda, zettersten, mcmacdonald @wisc.edu)
Department of Psychology, University of Wisconsin-Madison,
1202 W. Johnson Street, Madison, WI 53706 USA
*equal authorship contribution

Abstract

Though communicative goals clearly drive word choice in language production, online demands suggest that accessibility might play a role, too. If the benefits of accessibility are important enough to communication, more accessible words (high-frequency words) might be chosen over more accurate, less accessible ones. We used a novel artificial language learning paradigm to test whether high-frequency words are preferred over low-frequency words at a cost of meaning accuracy. Participants learned eight words which corresponded to precise angles on a compass. On test trials, participants viewed angles lying in-between two trained angles and were asked to produce a word for the angle. Across two experiments, we showed that participants extended their use of high-frequency words to more distal angles compared to low-frequency words. In cases of competition between high- and low-frequency words, the former tended to win out even when less accurate, suggesting that accessibility can compromise some accuracy.

Keywords: lexical access; accessibility; word frequency; artificial language; lexical bias; language production

Introduction

Language affords many ways to communicate an idea. For example, if we tell someone about a new pet, we could say *cat* or *kitten,* that *she* or *Mittens* always seems to be *hungry or starving,* that her behavior is *meowing* or *crying.* Producers typically make these and other choices rapidly and unconsciously on the basis of many competing forces, but we know relatively little about the direct trade-offs at work. For example, *cat* is a much more common word that *kitten,* but *kitten* may more accurately reflect the intended message. A word's accessibility—the ease of retrieval from memory owing to frequency, salience, repetition or other factors—and fidelity to the message are two broad forces in language production which, while often studied in isolation or in parallel, have not been directly compared to date.

To the extent that talking is for communicating messages, producers are necessarily sensitive to meaning accuracy. However, several language production accounts also suggest that producers are motivated by the need to manage the effortful task of production, making implicit choices about utterance forms that improve efficiency (Jaeger & Levy, 2007; MacDonald, 2013). For example, some aspects of lexical production appear to be guided by the ease with which lexical constituents can be retrieved from memory and planned for an utterance (Bock, 1987). High-frequency words are produced more quickly in single-word production, are ordered earlier in multi-word production compared to low-frequency words (Bock, 1982; Fenk-Oczlon, 1989), and yield fewer hesitations and errors compared to low-frequency words (Jescheniak & Levelt, 1994).

These results point to clear effects of frequency on some lexical processes, but data for effects of frequency on choice of word are more scarce. Using an artificial language learning paradigm (ALL), Harmon and Kapatsinski (2017) showed that participants extend high-frequency morphemes to novel grammatical categories. What remains unclear is the strength of this effect, specifically whether a high-frequency word's accessibility in memory could make it a tempting production choice over a more accurate, low-frequency word. To the best of our knowledge, no experiments have demonstrated a frequency effect in lexical selection at a cost of meaning precision. Indeed, an experimental test of this claim would be difficult in natural language, because word frequency is confounded with a number of linguistic properties (Bock, 1982).

In order to investigate the effects of word frequency on word choice without frequency's common natural language confounds, we designed an ALL task in which we manipulated word frequency, controlling for meaning. We then tested participants with new meanings that required learners to extend the meanings of the words they previously learned, in a task that allowed us to examine the relative weight of word meaning and word frequency on lexical selection. If production choices are driven primarily by meaning and effects of accessibility are limited to choices between synonymous words, then word frequency should not affect production choices. However, if accessibility affects production choices more generally, then participants should sometimes produce high-frequency words even when the low-frequency alternative is more accurate.

Experiment 1

To see whether producers would compromise meaning for accessibility, we trained them on four high-frequency and four low-frequency words referring to equidistant compass directions. We then tested them on untrained parts of the semantic space in the task and investigated how participants extended a word for a principal (trained) compass direction to refer to novel directions. By comparing the degree to which participants extended high-frequency compared to low-frequency words to more distal directions, we tested how producers balance word frequency and meaning in their word choices.

Method

Participants

39 University of Wisconsin-Madison undergraduates (25 female; mean age: 18.8 years, *SD* = 0.84; 38 native speakers of English) participated for course credit.

Stimuli

Each participant was taught eight novel words drawn randomly from a set of 18 (*pim, dak, vorg, yeen, grah, skod, gled, veek, blit, peka, sarp, minada, hoon, clate, noobda, gorm, frabda, mog*) developed by Amato and MacDonald (2010). Words were randomly assigned to eight equidistant angle orientations across the 360-degree face of a compass image: 15°, 60°, 105°, 150°, 195°, 240°, 285°, and 330° (see Figure 1). Each participant received a different assignment of words.

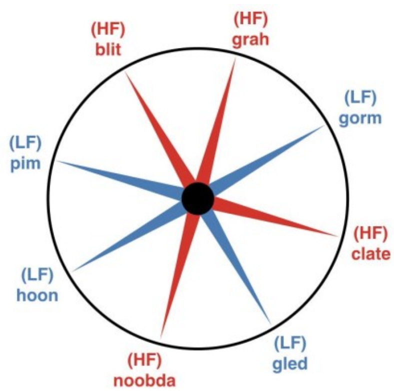

Figure 1: The eight principal compass directions and example words learned during training in Experiment 1. HF = high-frequency words; LF = low-frequency words.

Each direction was assigned to a *high-frequency* or a *low-frequency* category in one of two counterbalanced compass arrangements (see Figure 1 for one arrangement). The arrangement of low-frequency/ high-frequency words was created to maximize the number of compass regions in which a high-frequency word was adjacent to a low-frequency word while still providing a baseline in which two low-frequency or two high-frequency words were adjacent to one another.

Design & Procedure

Participants were told that they were playing a game in which their job was to help elves hunt for gold, by indicating a search direction for the prize. The experiment consisted of a Training Phase, in which participants were taught novel words for the 8 principal compass directions (see Figure 1), and a Test Phase, in which participants were tested on new directions that varied in distance from the principal compass directions.

Training Phase. Participants were first presented with each compass direction and its word, and they typed each of the novel words into a text box. Next participants completed a *Word Learning* training in which they were presented with one of the eight compass directions and chose which of two words matched that direction. Participants typed their response into a text box prompt and received immediate feedback on their answer. Critically, half of the words (the *high-frequency* words) occurred four times more frequently (as both a target and as a foil) than the other half of words (the *low-frequency* words).

Once participants achieved 80% accuracy on a 20-trial block, word knowledge was tested in the *Word Recall* phase. Participants were prompted to recall each word via typed responses. If participants made an error, they returned to the Word Learning phase. The Training Phase continued until participants achieved 100% accuracy on all 8 words during the Word Recall trials. Thus, all participants entered the Test Phase having learned the word for each principal compass direction, but having experienced high-frequency words four times more frequently than low-frequency words.

Test Phase. The Test Phase contained four blocks. In the *Low Competition Block*, participants described randomly generated compass directions that were clearly nearer to one of the 8 principal compass directions than to others (see Figures 2A & 3A). Each test stimulus direction was 1° - 11° away from the nearest principal compass direction. During this block, the 4:1 ratio of high-frequency to low-frequency words was maintained. Participants saw a compass direction near (within 1° - 11° of) each high-frequency word 12 times and a compass direction near of each low-frequency word 3 times, for a total of 60 test trials. For each trial, participants were asked to type a direction word into the text box based on the compass to direct a group of elves towards a hidden treasure. Trials timed out after 5s if participants did not begin typing.

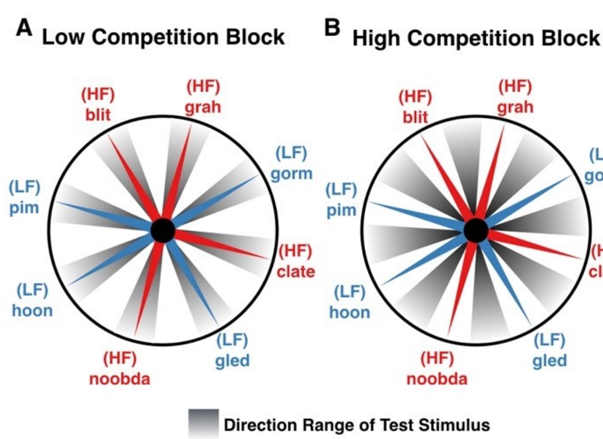

Figure 2: Gray shading indicates directions tested in the (A) Low Competition block and (B) High Competition block.

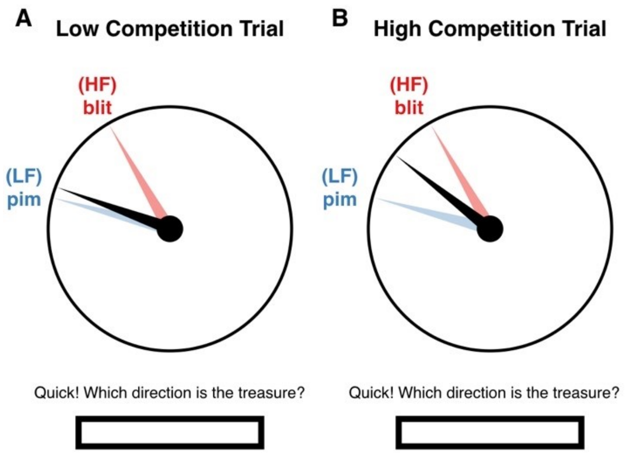

Figure 3: Example of a test trial in the (A) Low Competition block and (B) High Competition block. Participants saw only the direction in black. The two nearest compass directions and words in light blue (low-frequency) and light red (high-frequency) are added for illustration purposes and were not visible to participants.

To incentivize fast and accurate performance, participants received feedback in the form of a score after each trial, with points proportional to participant's accuracy (how close the word was to the typed compass direction) and speed (how quickly participants completed typing the word). Participants' base score varied from 0 to 45 points based on the distance of the tested angle from the word entered, with closer labels yielding higher points (45 points = no difference between tested angle and the entered word's compass direction; 0 points = tested angle is 45° or more away from the entered word's compass direction). This base score was then scaled based on the speed of participants' responses. For example, a difference in reaction time of 300ms corresponded to a change in base score by 0-2 points. Thus, while both speed and accuracy were emphasized, the scoring system weighed accuracy more heavily than speed in assigning points. If participants did not complete typing before the trial timed out or if their response was a word that named a direction more than 45° from the indicated compass direction, they received a score of 0.

In the *High Competition Block*, participants were tested with randomly generated compass directions that were close to the midline between two principal compass directions, creating competition between two words that could guide the elves (see Figures 2B & 3B). For each of the 8 sections of the compass lying between two principal compass directions, participants saw 8 compass directions sampled in between two compass directions, within an interval of 23°. Thus, each test stimulus direction was 12° - 22° away from the nearest principal compass direction. On 6*8 = 48 of the 64 test trials (*low-frequency/high-frequency* trials), the compass direction was more ambiguously between a low-frequency and a high-frequency word (though the compass direction always lay objectively closer to one principal compass direction than another). The trial design and feedback were otherwise identical to trials in the Low Competition Block.

Participants next completed two test blocks constructed to re-test participants' knowledge of the words for the 8 principal compass directions. In these blocks, participants were given the 8 principal compass directions in random order (30°, 75°, 120°, 165°, 210°, 255°, 300°, and 345°). The *Unambiguous Block* preserved task demands of the previous trials, in that from the participants' perspective, these trials were not differentiated from the preceding test trials. In the *Final Naming Block*, the trial structure was identical to the previous test trials, except that trials did not time out and participants did not receive feedback (i.e., these trials were described as being separate from the treasure hunting game).

Results

Word Training Performance

Participants' accuracy across all word learning blocks was high (M = 95.2%, SD=3.1%). All participants reached the 80% word learning criterion by the second block, thus passing on to recall test. On average, participants completed approximately 5 learning blocks (M = 4.59, SD = 1.93) before reaching the required perfect performance on the recall test, progressing to the testing portion of the game.

Final Word Retention

Unambiguous Test Block. Recall accuracy for the individual words in the timed test was marginally greater for high-frequency words (M = 91.0%, 95% CI = [86.7%, 95.4%]) relative to low-frequency words (M = 84.0%, 95% CI = [78.0%, 90.0%]), $t(38) = 1.99$, $p = .054$. Reaction (word typing) times were significantly shorter for high-frequency words (M = 2347ms, 95% CI = [2175ms, 2518ms]) compared to low-frequency words (M = 2628ms, 95% CI = [2352 ms, 2904ms]), $t(38) = -2.46$, $p = .018$.

Final Naming Block. In the final word recall task emphasizing only accuracy, without time pressure, accuracy was identical for high- and low-frequency words (M = 97.4%, 95% CI = [94.3%, 100%]). There was a marginal difference in reaction times between high-frequency (M = 2755ms, 95% CI = [2452ms, 3058ms]) and low-frequency words (M = 3080ms, 95% CI = [2695 ms, 3466ms]), $t(38) = -1.81$, $p = .08$. These results suggest that participants still maintained high accuracy on both high- and low-frequency words at the end of test, with a slight advantage for high-frequency words emerging under time constraints.

Test Performance

Our main question was whether word frequency experience during training would increase the likelihood of participants overextending high-frequency words during test, including in situations when a more accurate trained word (closer on the compass) was available. To investigate participants' tendency to overextend words, we focused specifically on *low-frequency/high-frequency* trials, in which a compass direction was tested in between a low-frequency and a high-frequency trained direction, so a high- and low-frequency word were in competition. We considered participants' likelihood of choosing the word for the nearest compass direction, dependent on whether that compass direction was a high- or a low-frequency word, while controlling for the distance from the nearest learned compass direction. As a conservative test, we focused exclusively on trials in which participants chose one of the two principal direction words within 45° of the stimulus direction (94.2% of responses). All of the patterns of findings remain identical if all *low-frequency/high-frequency* trials are considered.

We fit a logistic mixed-effects model predicting the likelihood of choosing the nearest word from Word Frequency (centered; High = -0.5 vs. Low = -0.5) and the distance of the stimulus from the nearest compass direction. We included by-subject and by-item random intercepts as well as by-subject random slopes for word frequency and distance. As expected, the likelihood of choosing the nearest word decreased with increasing distance from the nearest principal compass direction, b = -0.23, Wald 95% CI = [-.25, -.20], $z = -17.07$, $p < .0001$. Crucially, controlling for distance from the nearest principal direction, participants were more likely to use the nearest word when it was a high-frequency word compared to a low-frequency word, b = .71, Wald 95% CI = [.30, 1.12], $z = 3.39$, $p < .001$. This effect corresponded to an estimated 3.12° shift (95% CI = [1.31°, 4.93°]) in participants' decision boundary for high-frequency words as compared to low-frequency words.

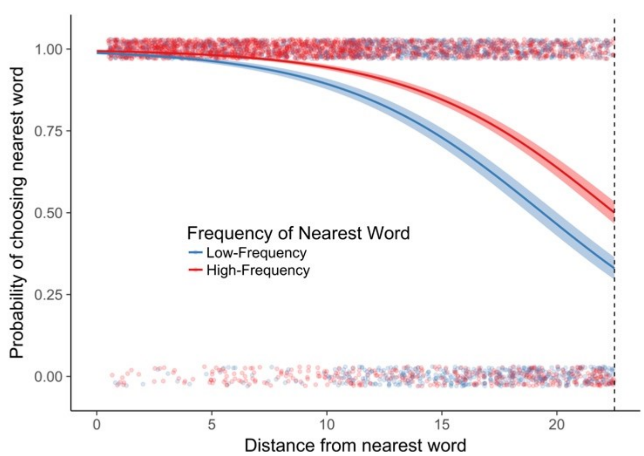

Figure 4. Probability of choosing the nearest word for high- and low-frequency words on low-frequency/high-frequency trials.

Next, we investigated participants' speed in responding on trials in which they chose the nearest word (thereby maximizing accuracy). We fit a linear mixed-effect model predicting participants' reaction times from Word Frequency (centered; High = -0.5 vs. Low = -0.5) and Distance from Nearest Principal Direction with the same random effects structure as above. Participants responded faster when the nearest principal direction had a high-frequency word (*M* = 2213ms, 95% CI = [2130ms, 2296ms]) compared to a low-frequency word (*M* = 2677ms, 95% CI = [2594ms, 2832ms]), *b* = -402.5, Wald 95% CI = [-522.5, -282.7], $F(1, 40.7) = 43.04$, $p < .001$.[1] Participants responded slower as the test stimulus's distance from the nearest principal direction—and therefore the competition between two principal direction words—increased, *b* = 10.2, Wald 95% CI = [4.9, 15.4], $F(1, 37.0) = 14.1$, $p < .001$. The effects were similar for both the Low Competition block and the High Competition block (no block by frequency or distance interaction).

Experiment 2

Experiment 1 showed that participants produce high-frequency words at a cost of semantic accuracy. In Experiment 2, we sought to replicate these findings with a different arrangement of high-frequency and low-frequency words in training, and a different sequence of test trials, intermixing high and low competition test trials. We predicted that the Experiment 1 results would generalize to these new parameters, such that producers' choices would again be influenced by word frequency.

Method

Participants

44 University of Wisconsin-Madison psychology undergraduate students (26 female; mean age: 18.5 years, *SD* = .87; 43 native speakers of English) participated for course credit.

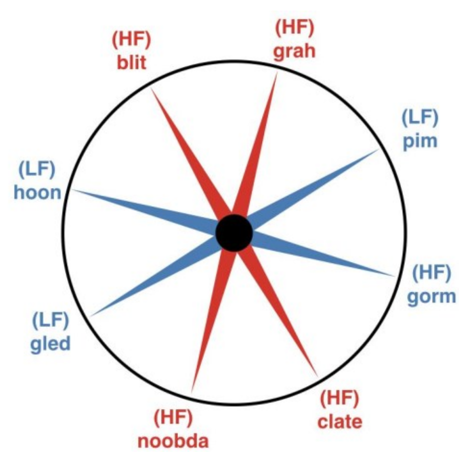

Figure 5: Arrangement of the eight principal compass directions learned in Experiment 2. HF = high-frequency words; LF = low-frequency words.

[1] Error degrees of freedom obtained through Kenward-Rogers approximation (Judd, Westfall, & Kenny, 2012), which can lead to non-integer estimated degrees of freedom.

Design

Two changes were made to the design of Experiment 1. First, we varied the arrangement of high-frequency and low-frequency words among the principal compass directions (LF, HF, HF, LF, LF, HF, HF, LF) to reduce the likelihood that idiosyncratic properties of the Experiment 1 arrangement drove frequency effects, and to ensure that participants encountered a higher proportion of trials that required using a low-frequency word (see Figure 5). Second, Low Competition and High Competition test trials were randomly intermixed during the Test Phase.

Results

Word Training Performance

Participants' accuracy across all pair learning blocks was high (M = 95.8%, SD=3.3%). On average, participants completed around 5 pair learning blocks (M = 4.36, SD = 2.62) before progressing to the Test Phase.

Final Word Retention

Unambiguous Test Block. Timed recall for the individual words after test was greater for high-frequency words (M = 92.6%, 95% CI = [88.4%, 96.8%]) relative to low-frequency words (M = 84.1%, 95% CI = [77.7%, 90.5%]), $t(43) = 2.35$, $p = .02$. Reaction times were marginally shorter for high-frequency words (M = 2366ms, 95% CI = [2161ms, 2572ms]) compared to low-frequency words (M = 2541ms, 95% CI = [2334ms, 2748ms]), $t(43) = -1.93$, $p = .06$.

Final Naming Block. In the untimed word recall task, accuracy was slightly higher for high-frequency (M = 98.9%, 95% CI = [97.3%, 100%]) compared low-frequency words (M = 94.9%, 95% CI = [91.4%, 98.3%]), $t(43) = 2.21$, $p = .03$. There was a marginal difference in reaction times between high-frequency (M = 2684ms, 95% CI = [2444ms, 2923ms]) and low-frequency words (M = 2982ms, 95% CI = [2666 ms, 3299ms]), $t(43) = -1.74$, $p = .09$.

There was no main effect of experiment version (Experiment 1 vs. Experiment 2) on accuracy and reaction times and no interaction between experiment version and frequency for either block, suggesting the general learning patterns were similar across experiments.

Test Performance

To test the impact of frequency on participants' overextension tendencies, we fit the same model as in Experiment 1. Controlling for angle distance from the nearest compass direction, participants were more likely to use the nearest word when it was a high-frequency word as compared to a low-frequency word, b = 1.34, Wald 95% CI = [.74, 1.95], $z = 4.36$, $p < .0001$. This effect corresponded to an estimated 7.19° degree shift (95% CI = [3.96°, 10.43°]) in participants' decision boundary for high-frequency words as compared to low-frequency words. There was no interaction between Low vs. High Competition trials.

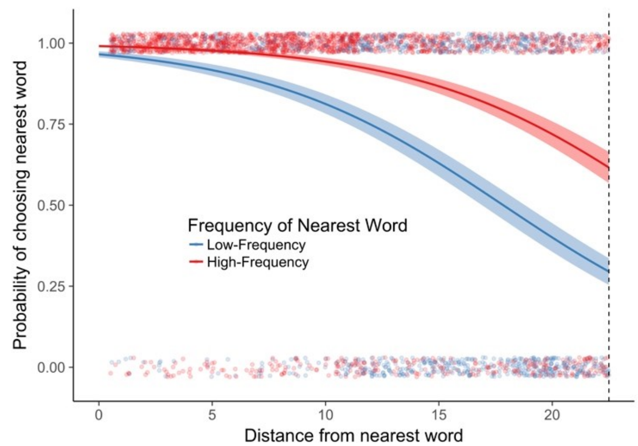

Figure 6. Probability of choosing the nearest high- or low-frequency word on low-frequency/high-frequency trials.

Next, we tested participant's response speed on test trials in which they chose the nearest word, using the same model as in Experiment 1. Similar to Experiment 1, participants responded faster when the nearest principal direction had a high-frequency word (M = 2287ms, 95% CI = [2182ms, 2393ms]) compared to a low-frequency word (M = 2568ms, 95% CI = [2462ms, 2715ms]), $b = -222.9$, $F(1, 40.7) = 11.42$, $p = .002$. Participants responded more slowly as the test stimulus' distance from the nearest principal direction increased, $b = 9.0$, Wald 95% CI = [2.8, 15.3], $F(1, 40.8) = 8.0$, $p = .007$.

General Discussion

In two experiments, we showed for the first time that language producers are influenced by word frequency in their productions to the point that they may favor high-frequency words over low-frequency alternatives that better capture an intended meaning. Response durations for high-frequency words were significantly shorter than for low-frequency words, replicating the frequency effect on production speed. On critical trials where the compass pointed near a category boundary between a high-frequency and a low-frequency word, participants extended high-frequency words 3-7 degrees further than low-frequency words in 22.5 degree spans, even though the points they earned strongly depended on the accuracy of their production choice. These results suggest that producers will compromise accuracy in meaning for accessibility. Moreover, these studies showed that the benefit for high-frequency words does not stem from poor knowledge of the low-frequency words, as participants were initially trained to 100% correct on all words and also performed close to ceiling for all words on final recall tests.

These results provide a substantive contribution to the literature on frequency effects in language production. To date, most studies have shown relative retrieval advantages for high-frequency over low-frequency words, such as naming latency, and lower error rates (Jescheniak & Levelt, 1994). Our study greatly extends the reach of frequency effects beyond speed and beyond extensions of word meanings that do not compromise accuracy (Harmon & Kapatsinski, 2017) to showing that frequency can affect lexical choice even at some cost in meaning accuracy. These results are therefore consistent with production accounts in which efficiency of production drives certain production choices (Jaeger & Levy, 2007; MacDonald, 2013).

One limitation of our method is that participants typed instead of spoke their responses, and a natural future direction is replication with spoken language. A second limitation is that the frequency effects found here arose during production of recently-learned novel words rather than in natural language, where frequency differences are established over a lifetime of experience. Because word frequency tends to be confounded with length and other factors in natural language (Piantadosi, Tily, & Gibson, 2011), a natural language extension of this paradigm could examine effects of accessibility driven by some combination of frequency, length and other factors, for example if English producers described a compass direction as "west" (a high frequency, short word) instead of a rarer, longer, but more accurate expression such as "west-northwest".

Another important direction for future research is an exploration of the role of feedback and time constraint on producers' willingness to extend learned words to new points on the compass. Instructions to respond quickly and accurately are common in language production studies, but in the current experiments participants received a graded reward (points) contingent on accuracy and speed. While directly incentivized to weigh both constraints to some degree, in practice, the speeded advantage of a high- over low-frequency response would lead to 0-2 additional points relative to an average score of 46. Future work will investigate whether frequency shifts participants' responses even when the emphasis of the task is placed solely on accuracy with no incentives for speed.

The current experiments do not rule out the possibility that producers have remapped the semantic boundaries for high-frequency vs. low-frequency words, as opposed to shifting their usage of the words in accordance with their accessibility. While this may be part of our effect, we find it unlikely that participants have dramatically shifted their perception of the semantic boundaries for two reasons. First, participants received immediate feedback after each response. Second, response times increased as competition increased, suggesting a form of competition between alternative words. If participants had instead extended the semantic range for the high-frequency words, then both near and more distant test points would be within a semantic space for this word and would not be expected to differ in response time.

In sum, these two studies showed that producers balance both the accessibility of a word and its accuracy for conveying a meaning when constructing an utterance. The fact that online production may sacrifice some inaccuracy for the benefit of accessibility may contribute to additional insight into word substitution speech errors, where there is some controversy concerning the degree to which speech errors tend to yield a higher frequency word substituting for a lower frequency word (Harley & MacAndrew, 2001). These frequency-driven effects observed here may also prove useful in understanding the use of vague words such as "thing" when another word is more accurate, but likely less accessible.

Acknowledgements

This research was supported by NIH T32DC005359 awarded to MK, by NSF-GRFP DGE-1256259 awarded to MZ and the Wisconsin Alumni Research Fund awarded to MM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NSF or the Wisconsin Alumni Research Fund.

References

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1-47.

Bock, K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, *26*(2), 119–137.

Fenk-Oczlon, G. (1989). *Word frequency and word order in freezes* (Vol. 27). https://doi.org/10.1515/ling.1989.27.3.517

Harley, T. A., & MacAndrew, S. B. G. (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research*, *30*(4), 395–418. https://doi.org/10.1023/A:1010421724343

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22–44. https://doi.org/10.1016/j.cogpsych.2017.08.002

Jaeger, T. F., & Levy, R. P. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems* (pp. 849–856).

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.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3636467/

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108