2022. Lee, C.; Lew-Williams, C.; Goldberg, AEG. *Proceedings of the Cognitive Science Society Proceedings*.

Accessibility factors that lead to good-enough language production

Crystal Lee, Casey Lew-Williams, & Adele E. Goldberg
{cl40, caseylw, adele}@princeton.edu
Department of Psychology, Princeton University
Princeton, NJ 08544

Abstract

Accessibility plays a major role in speech production. Here we investigate and measure four factors that influence speakers to produce one word over another more optimal word form. Three experiments asked participants to label images of insects and instruments. Participants were incentivized to produce an accurate specific label (e.g., *bee*), over a more general label (e.g., *insect*), so that specific labels were more optimal. Each of three experiments manipulated a different factor that could influence accessibility – word frequency, priming, and interference – and all experiments additionally varied whether labels had to be produced under time pressure or not. Results showed that each variable significantly influenced the accessibility of labels: participants produced more specific labels when those labels were higher frequency, when they were primed, when a visually-similar label had *not* been primed, and when participants were unconstrained by time pressure. These findings demonstrate that multiple factors influence the accessibility of familiar words during production, regularly leading participants to rely on "good-enough" rather than optimal options to convey their message.

Keywords: lexical production; accessibility; language production; word frequency; priming; interference; time pressure

Introduction

Speakers (and signers) are incredibly adept at communicating an idea or message. However, the process of producing language is nontrivial. As we plan a message, we must select words that accurately convey our intended message, and we do not always choose the most optimal labels. For example, if someone quotes a line spoken by Romeo by saying, "A *flower* by any other name…", the comprehender may well understand the utterance, even though Romeo actually spoke of a *rose*. What causes speakers to access and produce non-optimal words? Here, we seek to compare the role of different factors influencing what has recently been referred to as "good-enough" language production (Ferreira & Griffin, 2003; Koranda, Zettersten, & MacDonald, 2022; for review see Goldberg & Ferreira, 2022).

Good-enough language production refers to the production of non-optimal utterances that nonetheless approximate the intended message. While work on language as involving a "noisy channel" investigates how listeners recover from non-optimal language (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008), our focus is different. The current work investigates factors that give rise to good-enough production in the first place. Specifically, we presume that words that are non-optimal, but intended to be "good-enough," are produced when the speaker (or signer) is unable to access the more optimal form. Nearly every conversation showcases the pervasiveness of good-enough production, as speakers' utterances are rarely uniformly optimal. Good-enough language production thus highlights the role of accessibility from memory during language production.

Accessibility, or the speed and accuracy of retrieval, is not a new concept. Factors that are known to influence lexical accessibility include: frequency, priming, salience, and animacy (e.g., Bonin et al., 2019; Coltheart et al., 1979; Gordon, 1983; Lorch, Balota, and Stamm, 1986). A majority of this research has relied on lexical decision tasks, which involve deciding whether a string is a real word or a nonsense word (Meyer & Schvaneveldt, 1971), in order to control for the myriad factors that potentially influence lexical access. Yet lexical decision tasks are particularly unnatural, except perhaps in the context of playing Scrabble or Wordle, since words are generally accessed in order to implicate an associated concept, and nonsense "words" serve no such purpose.

Far fewer studies have used naturalistic production tasks (but see e.g., Bock, 1986, 1987; Arnold, 2010; Harmon & Kapatinski, 2017; Tachihara & Goldberg, 2020). A notable recent such study taught participants a set of novel words, each referring to a different degree on a compass (Koranda, Zettersten, & MacDonald 2022). The words varied in whether they were presented with higher or lower frequency. Participants were awarded points based on how accurately they used the novel words and how quickly they produced them. Koranda et al. demonstrated that when participants needed to describe an angle that lay between two points that had been labeled by a high- and a low-frequency word, participants were more likely to extend the higher-frequency word to apply to the unlabeled angle. The result confirms that frequency influences the choice of words, presumably because higher frequency words are easier to access. The current work extends Koranda et al.'s study in the following ways.

Rather than teaching participants novel words, we capitalize on the natural variability in frequency of familiar words. We also test other factors that may impact accessibility: priming and competition. Finally, we combine data across the three studies in order to detect a possibly subtle influence of time pressure.

Previous work has found that priming a lexical item increases the speed with which the item is recognized in lexical decision tasks (e.g., Scarborough, Cortese, & Scarborough, 1977), and we aim to determine whether the same effect is evident in the current more naturalistic picture-naming task. Work on the role of competition in accessibility is complex. In lexical decision tasks, people tend to be slower to recognize words with many phonological neighbors which presumably compete for activation (e.g., Cluff & Luce, 1990). This effect, however, may be influenced by the lexical decision task itself, insofar as determining that a string is a non-word requires that real-word neighbors must be rejected. In the semantic domain, the role of competition is often discussed in terms of "codability." For instance, since *sofa* has a nearly synonymous competitor, *couch*, it is considered less codable than a word such as *apple* which has no close competitors (e.g., Zettersten & Lupyan, 2020). In Experiment 3, we compare productions in which a competitor has been made salient (via priming) or not.

Finally, to simulate the variable pace of natural conversations, in each experiment, we manipulated whether or not participants faced a time limit to respond. We predicted that participants would be more likely to produce non-optimal, "good-enough" labels under time pressure.

Experiments

In each of three experiments, the dependent measure is the type of response provided during the production phase of the experiment: how likely participants are to produce a specific picture label (e.g., *bee*) compared to a general description (e.g., *insect*), an accurate but less optimal choice in the experimental context. Images that required labeling came from two general categories – insects and instruments – and participants saw each category in a different block of the experiment, order counterbalanced.

All three experiments manipulated whether participants were required to produce labels under time pressure. This was manipulated across two blocks: the labels for the first 8 images in one block could be produced without time pressure; then 8 speeded trials were presented in another block. Participants were warned they needed to produce labels for the images within 3 seconds or the trial would end.

All three experiments included target words that varied in their frequencies of occurrence in the 500-billion-word Corpus of Contemporary American English (Davies, 2008). Log-frequency of target words was included in all analyses as a fixed factor. Experiment 2 additionally manipulated whether specific labels were primed before they were elicited as target words. Finally, Experiment 3 manipulated whether a visually-similar competitor labels was primed before target words were elicited.

Participants

A total of 240 adult participants were recruited online from Cloud Research for this study, 80 for each of the 3 experiments. 31 participants were excluded for failing to achieve 75% accuracy in either the first or third phase of the experiment. 21 participants were also excluded for extensive knowledge of musical instruments. Of the remaining participants, the first 40 participants who successfully completed each experiment and filled out our counterbalancing lists were included (mean age = 38.4; SD = 10.4; female = 61). All participants were paid between $1.50 - $3.60 USD.

Figure 1: Example stimuli used across the three experiments. From left to right, top to bottom: *bassoon, accordion, tuba, harp.*

Stimuli

Eight images of insects (e.g., *moth, millipede)* and 8 images of instruments (e.g., *trombone, clarinet*) were included in all experiments. These images and labels were chosen from a norming study examining the codability of each image. Images and labels were chosen to meet the following criteria: 1) variation in frequency, 2) variation in codability. The choice of labels was varied across tasks in order to institute the priming or competition under investigation in Experiments 2 and 3. Other items remained constant. Log-frequencies of each correct label were calculated based on the number of occurrences in the COCA corpus (Davies, 2008) (see Table 1)

Table 1: Example stimuli from Experiment 1.

Word	Log-frequency (Davies, 2008)
Ukelele	1.75
Trombone	2.81
Clarinet	3.07
Harp	3.38

Procedure

Each participant was exposed to 2 blocks of 3 phases each. In each block, the primary production phase was presented after a lexical decision phase (included solely for the purpose of priming), and before a 2AFC phase. One block of trials involved labels and images of insects, the other, instruments. The order of the general categories (insects or instruments) were counterbalanced across participants.

Lexical decision phase. An initial lexical decision task was included in all three experiments, as an attention check and to induce priming in Experiments 2 and 3. Participants were presented with 6-8 nonce words and 3-4 real words (one at a time) and were asked to choose whether each was a real word of English or not. Participants were excluded if they did not perform correctly on at least 6/8 trials (75% correct).

Production phase. The production task provides the key dependent measure: word choice. Participants were asked to type the labels of entities shown in a series of images. Images of insects (or instruments) were presented one at a time; the order of category (insects or instruments) was counterbalanced across participants. Participants were told in advance that accurate specific labels (such as *bee*) would earn a bonus of $0.10. They were also informed that they were allowed to respond with the name of the general category (i.e., *insect* or *instrument*), and would earn $0.05 each time they did this. Therefore, specific responses were incentivized and more optimal compared to general labels. Incorrect responses received no reward or penalty. Participants only needed to type at least the first 4 letters of their answer in order to be correct, to alleviate spelling/typing concerns. For each of the two general categories, participants responded to 8 images without any time pressure, followed by a second set of 8 images which required responses within 3 seconds. No images were repeated.

2AFC phase. Following the production phase, participants were asked to perform a 2-alternative forced choice task when presented with two images and a label. This was used to determine whether participants were able to recognize the label and image pairings of the target items when retrieval demands were eliminated. Participants clicked on one of two images in response to "click on the *label.*" There were a total of 8 trials in each block. Side of presentation for each test image was randomized, as were foil images. Participants who scored fewer than 6/8 in this phase were excluded from analysis.

Experiment 1

Experiment 1 manipulated only speed of response and the lexical frequencies of the target words. Based on results of Koranda, Zettersten, & MacDonald (2022) on the production of novel words, we predicted that participants would produce more specific labels when those labels were more frequent. We also predicted participants would produce more specific labels when no time pressure was imposed, indicating that accessibility issues to word forms may arise due to timing constraints. In Experiment 1, the initial lexical decision task was only used as an attention check. It contained no words that were related to the main Production phase.

Method

Stimuli

8 images of insects (*moth, millipede, dragonfly, beetle, termite, flea, cockroach, fly)* and 8 images of instruments (*ukulele, trombone, clarinet, harp, bongos, tuba, accordion, flute*) were used. These items were chosen because they varied in frequency.

Results

Data exclusion

Performance on the 2AFC task was used as an exclusion criterion (scores less than 6/8) and only 1 participant was excluded. In addition, production data was excluded that were clearly typos, omissions (lack of response) or incomprehensible responses (6.7% of data).

Production phase

After excluding obvious uninterpretable responses and omissions, the most frequent responses were specific labels (49.2%). Incorrect specific labels were the next most frequent (28.1%, e.g., typing *wasp* for *bee*), followed by general labels (22.6%).

As planned, to analyze participants' responses we implemented a logistic mixed-effects model to predict participants' use of specific responses (e.g., *insect*) on the basis of two fixed factors -- time pressure (speeded or un-speeded), the log-frequency of the specific label--and their interaction. We included random intercepts of subject, image category, as well as random slopes for time pressure for subjects. We attempted to additionally include the following random effects, but removed them due to failures to converge: a random slope of image label by image category, a random intercept of image label, and a random slope of frequency by subject.

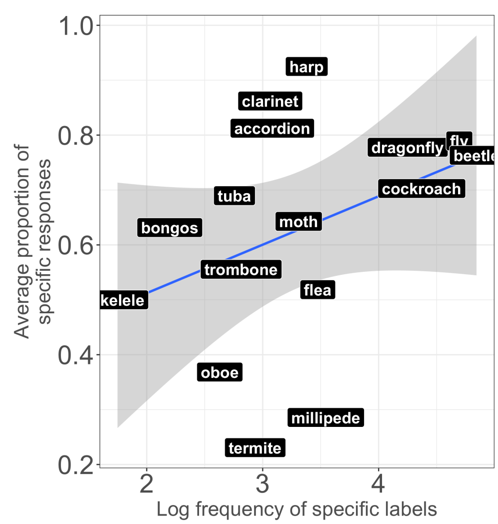

Figure 3: Average proportion of specific responses by log-frequency of items.

As shown in Figure 2, results reveal a significant main effect of log-frequency (Odds Ratio = 17.24, 95% CI = [9.32, 31.89], *p*-value = 0.001), meaning that participants were

2022. Lee, C.; Lew-Williams, C.; Goldberg, AEG. *Proceedings of the Cognitive Science Society Proceedings*.

more likely to produce a specific label such as high-frequency *fly* compared to a lower-frequency label like *flea*. We did not see any effect of time pressure in Experiment 1 nor an interaction with frequency.

2AFC task

The average accuracy in identifying the correct image when provided with a label was near ceiling (97.8%). We take the high accuracy on this task to suggest that the labels used were generally recognizable when retrieval demands were eliminated.

Experiment 2

In Experiment 2, we tested whether priming of a subset of specific labels during the lexical decision task will increase the production of those labels during the production task.

Method

Stimuli

The stimuli in this experiment were: *moth, millipede, dragonfly, beetle, termite, flea, cockroach,* caterpillar as insects; and *french horn, trombone, clarinet, harp, bongos, tuba, accordion, flute* as instruments. Changes in stimuli from Experiment 1 were made to avoid phonological primes (e.g., removing *fly* from the stimuli so we did not mistakenly prime *dragonfly*) and to maintain a similar frequency distribution.

Design and procedure

Experiment 2 used a very similar procedure as Experiment 1, except that the lexical decision task included a subset of specific labels in order to prime those words in the production task. The lexical decision task also presented participants with 8 words total (rather than 6), one at a time, including 4 nonce words and 4 real words. The real words primed 4 out of the 8 specific labels of images shown in the subsequent production phase. Two lists were employed to counterbalance which specific labels were primed.

Results

Data exclusion

13 participants were excluded for accuracy below 75% on the lexical decision task, and three additional participants were excluded for below 75% accuracy on the 2AFC task. Finally, production data that were clearly typos, omissions (lack of response) or incomprehensible responses were excluded (9.69% of data).

Production phase

Participants showed a similar overall pattern of responses as in Experiment 1: they produced specific labels most often (56.2%), followed by incorrect specific labels (26.5%). General labels were produced 17.3% of the time.

We again implemented a logistic mixed-effects model to predict participants' use of specific responses (e.g., *insect*). We included predictors of time pressure (speeded vs. un-speeded), whether the label was primed or not, their interaction, and an interaction between priming and log-frequency of the label. We also included random intercepts of subject and image category, as well as random slopes of time pressure per subject. We attempted to include the following random effects, but the model failed to converge: a random slope of image label by image category, a random intercept of image label, and a random slope of frequency by subject.

We found a significant main effect of priming (Odds Ratio = 0.05, 95% CI = [0.01, 0.22], *p*-value = 0.001), meaning that participants were more likely to produce a specific label when that label was primed. We also found a main effect of speed (Odds Ratio = 0.52, 95% CI = [0.28, 0.97], *p*-value = 0.038), indicating that specific labels were more likely to be produced when participants were not under time pressure. Finally, we found an interaction between priming and frequency (Odds Ratio = 2.89, 95% CI = [1.88, 4.44], *p*-value = 0.001). This means that priming had more influence on infrequent labels compared to frequent labels. We did not find a significant interaction between priming and speed.

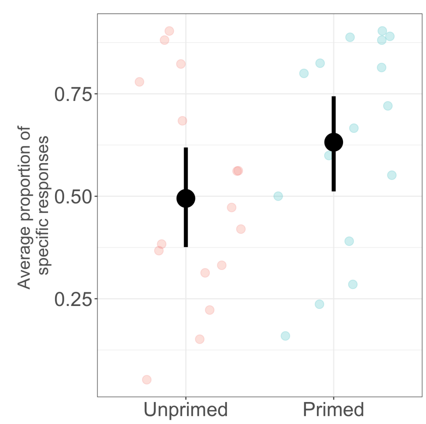

Figure 4: A scatterplot of average proportion of specific responses by item (whether it was primed vs. un-primed) from Experiment 2. The point represents the mean, and the black line represents the 95% confidence interval.

2AFC task

The average accuracy in this phase was close to ceiling (98.6%), again suggesting that the labels used in this experiment were generally recognizable when retrieval demands were eliminated.

Experiment 3

In Experiment 2, we demonstrated that priming a label influenced participants' abilities to access that word during a subsequent lexical production task. In a final experiment, we investigated whether priming a label interferes with the access of words that name other visually similar images, creates competition across lexical items. In particular, we hypothesize that priming a label of an image (e.g., *wasp*) prior to the introduction of a visually confusable image (e.g., *bee*) will interfere with the production of *bee*, due to competition from the primed term (*wasp)*.

Method

The design used in Experiments 1 and 2 was repeated with the following adjustments.

Stimuli

The stimuli used in Experiment 2 were adjusted in order to include two pairs of visually confusable items (insects: *millipede, caterpillar; wasp, bee; other insects were moth, tick, dragonfly, mosquito*; instruments: *oboe, clarinet; trombone, trumpet; other instruments were french horn, bassoon, violin, harp*).

Design and procedure

Experiment 3 used the same procedures as Experiment 2, where the lexical decision task included a subset of specific labels used in the production task in order to prime those words. This phase was designed to induce competition between visually similar items through priming of a competitor during the initial lexical decision task. Four of the 8 trials primed real words. These words primed 4 of the 8 specific labels of images seen in the subsequent production phase. These real word pairs were chosen to manipulate competition. In one set, pairs of words were chosen to create pairs of visually similar labels (i.e., were visually confusable) that consisted of one high-frequency and on low-frequency word (e.g., *bee* and *wasp*). The other set of words consisted of four words of similar frequencies, but which were not visually-similar (these labels were also seen in the production phase). The remaining four trials presented nonce words used in Experiment 2. Four lists were used in order to counterbalance the order of presentation of image categories (whether insects or instruments was presented first/second) and whether the set of images was visually confusable or not.

Results

Data exclusion

As before, performance of lower than 75% accuracy on either the lexical decision or the 2AFC task was used as an exclusion criterion; this excluded 10 participants in total (5 for each task). In addition, production data that were typos, omissions (lack of response) or incomprehensible responses were again excluded (7.19% of data).

Production phase

Overall, once again, the responses followed the same pattern: specific labels were the most common (55.0%), followed by erroneous specific labels (26.1%), and then general labels (17.2%).

We again implemented a logistic mixed-effects model to predict their production of specific responses. We included predictors of time pressure (speeded vs. un-speeded), whether the label was presented with a competitor or not during the lexical decision phase (i.e., confusability across labels), their interaction, and an interaction between confusability and log-frequency of labels. Random intercepts of subject, image category, as well as random slopes of time pressure per subject were included. We attempted to include the following random effects, but the model failed to converge: a random slope of image label by image category, a random intercept of image category, and a random slope of frequency by subject.

We found a significant main effect of confusability (Odds Ratio = 0.09, 95% CI = [0.03, 0.29], p-value < 0.001). This suggests that when participants are primed with a visually similar label, they were less likely to produce the specific target word. In addition, we found a significant interaction between confusability and log-frequency (OR = 2.21, 95% CI = [1.63, 3.00], p-value < 0.001). No other significant effects were found.

2AFC task

Average accuracy on the 2AFC was again high (95.8%), again indicating that the labels used were generally recognizable when accessibility demands were eliminated.

Experiments 1-3: Investigating time pressure

In each of three experiments, we experimentally manipulated the speed of production by participants. That is, in one block per experiment, participants had an unlimited amount of time to produce a label, and in a second block, participants had only three seconds. We found an effect of time pressure only within Experiment 2, but not within Experiment 1 or 3. In order to determine if the effect of time pressure was present in a larger data set, we collapsed our data across all three experiments. This was possible because the experimental procedure and design were kept constant.

We conducted a logistic mixed effects-model to again predict the production of specific responses. We included a fixed effect of time pressure (speeded vs. un-speeded) and log-frequency of the specific labels. We also included a random intercept of image category, a random intercept of experiment (Experiment 1-3), and a random intercept of subject with random slopes of time pressure. We attempted to include a random slope of log-frequency of specific labels by subject but the model failed to converge.

We found a significant effect of time pressure, such that participants were more likely to produce a specific label when they were not under time pressure (OR = 0.68, 95% CI = [0.53, 0.87], p-value = 0.002). This suggests that time pressure during language production affects the accessibility of word forms.

General discussion

The overall aim of the current experiments was to examine the role of lexical accessibility in a picture naming task, when specific terms were incentivized, but not required for accuracy. In particular, we investigated frequency, priming, competition, and time pressure as potential factors influencing word choice during naturalistic production. Across three experiments, we found four key factors that influence the accessibility of word forms during lexical selection.

Experiment 1 replicates previous research on this topic: frequency influences lexical selection (Koranda, Zettersten, & MacDonald, 2022): Participants were more likely to produce specific labels that were higher frequency. Experiment 2 demonstrates a significant effect of repetition priming: participants were more likely to produce a specific label if that label had been primed. Experiment 3 provides evidence that priming labels to induce competition across visually similar labels negatively impacts the accessibility of competitor labels; this effect was evident when a high-frequency word was primed and a low-frequency word was subsequently needed in the production task (e.g., priming *bee* inhibited *wasp* responses). The effect was unidirectional: High-frequency words (e.g., *bee*) remained accessible despite priming of a low-frequency competitor (*wasp*). Finally, we find evidence for an effect of time pressure in Experiment 2, and in the final analysis including data from all three experiments: Specific responses were more likely when participants could take their time to respond. The results overall suggest that word frequency and priming increase accessibility while competition between words and time pressure negatively influence accessibility during production.

Our work converges with previous work on how frequency of word forms can affect accessibility during word choice processes (Koranda, Zettersten, & MacDonald, 2022; Harmon & Kapatsinski, 2017) and with other work on retrieval advantages for highly frequent words (Jescheniak & Levelt, 1994). We also find complementary evidence for time pressure and priming's influence on lexical accessibility (Forbach et al., 1974; Ferreira & Griffin, 2003). Furthermore, our results provide novel evidence that competition negatively affects accessibility. Competition within lexical selection is not a new topic (see Spalek, Damian, & Bölte, 2013), but we extend prior work to demonstrate that competition during lexical selection influences "good-enough" production. Participants were less likely to produce a specific label when a distinct label for a similar image had been primed. The effect was stronger when a higher frequent competitor was primed, leading to fewer specific low-frequency responses.

Interestingly, the effect size of frequency was the largest in the three experiments, demonstrating that frequency is an important factor in determining the accessibility of word forms during naturalistic production. Priming, competition, and speed also influenced production choices. Speakers appear to rely on good-enough language productions rather than optimal choices when the optimal choices are: infrequent, not observed previously (unprimed), in competition with a recently primed label, or produced under time pressure.

One limitation to our study is that we used typing as a proxy for speech production. While many studies have used online procedures for production (e.g., Koranda, Zettersten, & MacDonald, 2022), it remains an open question as to whether these results will be replicated using in-person and/or spoken methods. Another limitation is that some participants may not have been familiar with the images or labels used in the production phase. While performance on the 2AFC task was reassuringly high, the task sets a low bar for familiarity, since participants only needed to assign each label to one of two very different images. Future research should gauge item-specific knowledge using more subtle tasks.

The current work has taken for granted that general category labels (*insect* or *instrument*) are "good-enough" descriptions of the images, since they are accurate but not optimal. Many of the participants' incorrect specific label responses that we considered errors may also actually be "good-enough" descriptions of the images provided. That is, speakers are likely to find "There's a bee!" a good-enough way to warn someone away from a wasp, despite the inaccurate use of *bee*. Further research should examine the situational context of language production to further understand when and how good-enough language arises.

Conclusion

Accessibility issues influence speakers' word choices. Our results demonstrate that frequency, priming, competition, and time pressure drive accessibility, influencing lexical selection when people label what they see. Increased frequency of labels and repetition priming improves lexical access to an intended label, while reduced accessibility from time pressure or competition across lexical items leads to good-enough language production.

References

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187-203.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575.

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

Bonin, P., Gelin, M., Dioux, V., & Méot, A. (2019). "It is alive!" Evidence for animacy effects in semantic categorization and lexical decision. *Applied Psycholinguistics*, *40*(4), 965-985.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*, 804-809.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 551.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *The Quarterly Journal of Experimental Psychology*, *31*(3), 489-507

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA).*

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis) selection. *Psychological Science, 14*(1), 86-90.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*(2), 337-339.

Goldberg, A. E., & Ferreira, F. (2022). Good enough language production. *Trends in Cognitive Sciences.*

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 24-44.

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

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*, 824.

Koranda, M., Zetterzten, M., & MacDonald, M. C. (2022). Good-enough production: Selecting easier words instead of more accurate ones. *Psychological Science.*

Lorch, R. F., Balota, D. A., & Stamm, E. G. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre-or postlexical access?. *Memory & Cognition*, *14*(2), 95-103.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, *90*(2), 227.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 1.

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes, 28*, 597-614.

Tachihara, K. & Goldberg, A.E. (2020) Cognitive accessibility predicts word order of couples' names in English and Japanese. *Cognitive Linguistics 31(2):* 231-249.

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: More nameable features improve category learning. *Cognition, 196,* 104135.