Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 8, 2359–2368
https://doi.org/10.1037/xge0001409

Similar-Sounding Words Flesh Out Fuzzy Meanings

David A. Haslett[1] and Zhenguang G. Cai[1, 2]
[1] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong
[2] Brain and Mind Institute, The Chinese University of Hong Kong

Most words are low in frequency, yet a prevailing theory of word meaning (the distributional hypothesis: that words with similar meanings occur in similar contexts) and corresponding computational models struggle to represent low-frequency words. We conducted two preregistered experiments to test the hypothesis that similar-sounding words flesh out deficient semantic representations. In Experiment 1, native English speakers made semantic relatedness decisions about a cue (e.g., *dodge*) followed either by a target that overlaps in form and meaning with a higher frequency word (*evade*, which overlaps with *avoid*) or by a control (*elude*), matched on distributional and formal similarity to the cue. (Participants did not see higher frequency words like *avoid*.) As predicted, participants decided faster and more often that overlapping targets, compared to controls, were semantically related to cues. In Experiment 2, participants read sentences containing the same cues and targets (e.g., *The kids dodged something* and *She tried to evade/elude the officer*). We used MouseView.js to blur the sentences and create a fovea-like aperture directed by the participant's cursor, allowing us to approximate fixation duration. While we did not observe the predicted difference at the target region (e.g., *evade/elude*), we found a lag effect, with shorter fixations on words following overlapping targets, suggesting easier integration of those meanings. These experiments provide evidence that words with overlapping forms and meanings bolster representations of low-frequency words, which supports approaches to natural language processing that incorporate both formal and distributional information and which revises assumptions about how an optimal language will evolve.

Public Significance Statement
Computational models struggle to represent the meanings of low-frequency words, and most words are low in frequency. This study provides evidence that people compensate for the impoverished semantic representations of low-frequency words by associating them with similar-sounding words. These findings support new approaches to natural language processing, and they provide insight into the organization of the mental lexicon and the evolution of language.

Keywords: distributional semantics, word meaning, language evolution, word frequency

Supplemental materials: https://doi.org/10.1037/xge0001409.supp

About 95% of the nonobsolete words in the Oxford English Dictionary (OED) occur less than once per million tokens. These low-frequency words are processed more slowly and less accurately than higher frequency words (Howes & Solomon, 1951; Preston, 1935), in part because they have underdetermined meanings (e.g., Cuetos et al., 2017). Indeed, a prevailing theory of meaning (the distributional hypothesis) and corresponding models (distributional semantic models) struggle to represent low-frequency words (e.g., McDonald, 2000; Sahlgren & Lenci, 2016). We hypothesize that, to compensate for these deficiencies, low-frequency words (e.g., *evade*) piggyback on representations of more familiar words that are similar in both form and meaning (e.g., *avoid*). If confirmed, this would have implications for natural language processing, the cultural evolution of language, and the organization of the mental lexicon.

This article was published Online First June 12, 2023.
David A. Haslett https://orcid.org/0000-0001-5522-9943
Thanks to Martin Pickering and Patrick Wong for useful comments and suggestions. This research was supported by a PhD fellowship from the Hong Kong RGC (to David A. Haslett) and a GRF grant (project number: 14600220) from the Hong Kong RGC (to Zhenguang G. Cai). All stimuli, data, supplemental materials, and scripts are available at: https://osf.io/cmxer. Both studies were preregistered. The preregistrations are available at: https://osf.io/wrtgu and https://osf.io/vw8ak. An earlier, substantially revised version of this manuscript was published as a preprint on OSF (https://osf.io/ym9vb). It was made available because it was cited in a different manuscript and has not been publicized.
We have no conflicts of interest to disclose.
David A. Haslett served as lead for conceptualization, formal analysis, investigation, and writing–original draft. Zhenguang G. Cai served as lead for funding acquisition, supervision, and writing–review and editing.
Correspondence concerning this article should be addressed to David A. Haslett, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, G20, Leung Kau Kui Building, Shatin, NT, Hong Kong SAR. Email: david.a.haslett@gmail.com

Frequency and Distribution

Most words are low in frequency (Zipf, 1936, 1949), and those words are processed less efficiently than high-frequency words (cf. Brysbaert et al., 2018). Importantly, frequency impacts memory and meaning. People are less accurate in free recall of low-frequency words (Gregg, 1976; Gregg et al., 1980), slower and less accurate when assigning low-frequency words to categories (Monsell et al., 1989), and slower to read the words following low-frequency words in sentences, which suggests difficulty integrating the meanings of low-frequency words (Duffy et al., 1988). Evidence from patients with Alzheimer's disease supports that frequency effects stem from "impoverished semantic representations" (Cuetos et al., 2017, p. 35).

Part of the problem is that distribution does not reliably capture the meanings of low-frequency words. According to the distributional hypothesis, words with similar meanings tend to occur in similar contexts (Firth, 1957; Harris, 1954), and distributional semantic models (DSMs, such as latent semantic analysis and word2vec; Landauer & Dumais, 1997; Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013) quantify the linguistic contexts where a word occurs. Because they can measure semantic similarity, DSMs "have been successfully applied in virtually all fields of cognitive science" (Günther et al., 2019, p. 1007). But low-frequency words occur in few contexts and so have skewed distributions. Lenci et al. (2022) found that even in state-of-the-art DSMs, low-frequency words are idiosyncratic, with different methods or different corpora producing very different representations, and low-frequency words remain a "major problem" in large language models (Schick & Schütze, 2020). For instance, despite being trained on hundreds of billions of word tokens (Brown et al., 2020), ChatGPT uses a significantly smaller vocabulary than humans do (Guo et al., 2023). These models work well for most word tokens (the high-frequency words that make up most of any corpus) but struggle with most word types (the low-frequency words that make up most of any lexicon).

Form and Meaning

If you struggle to represent the meaning of a low-frequency word, like *evade*, then a word that overlaps in both form and meaning, like *avoid*, could be a helpful cue. Form does supplement distribution for low-frequency words, at least in some languages, but that semantic information is limited to grammatical category (Monaghan et al., 2005, 2007). In fact, it has been argued that in an optimal lexicon, semantic neighbors will sound dissimilar to prevent confusion (Gasser, 2004; Monaghan et al., 2011). For example, the meaning of *parameter* has been influenced by "association or confusion with *perimeter*" (OED), and some people bemoan the "common misuse of *parameter*… as a limit or boundary" because "it leads to confusion and the loss of [a] valuable distinction" (Altman & Bland, 1999, p. 1667). According to this view, words that overlap in form and meaning impede processing, so a lexicon that evolves for the sake of efficient communication (cf. Gibson et al., 2019) will avoid substantial overlaps.

We suspect that the costs and benefits of overlapping forms and meanings differ in low-frequency words. Gasser et al. (2005) and Monaghan et al. (2011) agree that the pressure for semantically related words to sound dissimilar is diminished in low-frequency words, but we go further, arguing that substantially overlapping words can be not only tolerable but advantageous. Conflating *evade* with *avoid* or *parameter* with *perimeter* might be preferable to preserving a distinct but impoverished representation. Consistent with this revised view of the optimal lexicon, Haslett and Cai (2022a) found that, in English, semantic neighbors tend to be more similar in form when they are lower in frequency, so people can rely more on similar-sounding words to flesh out meanings when distribution is more likely to fall short. By emphasizing the unwieldy size of human vocabularies, Bolinger (1953) comes to a similar conclusion. Rather than "exercise special care to keep [similar-sounding words] apart," languages "align meaning with form, purely for economy of effort" (p. 325). People know tens of thousands of words (cf. Brysbaert et al., 2016), and most words are low in frequency, so maintaining a precise and unique meaning for each word is inefficient, if not impossible.

This strategy of linking overlapping forms and meanings is speculative, though. There is only limited evidence that overlapping words guide interpretation or facilitate processing. Groups of overlapping words affect how people interpret novel words (e.g., *glep* will tend to be associated with the shared semantic feature of *glow*, *glint*, and *gleam*; Hutchins, 1998), but Kwon (2017) provides evidence that this influence is collective; it does not derive from any particular similar-sounding word (cf. Lupyan & Casasanto, 2015). In contrast, Haslett and Cai (2022b) showed that, when learning novel words designed to resemble English words (e.g., *embraich* resembles *embrace*), people interpreted those novel words as closer in meaning to the similar-sounding words than sentence contexts implied. Meanings were influenced by particular words, not by statistical regularities or sound symbolism (cf. Dingemanse et al., 2015). Of course, it remains to be seen whether findings about novel words generalize to low-frequency but widely known words like *evade*.

Overlapping forms and meanings also facilitate lexical decisions. People recognize target words (e.g., *float*) faster following primes that overlap in both form and meaning (e.g., *boat*) and faster than the additive effects of formal and semantic priming would explain (Bergen, 2004; Pastizzo & Feldman, 2009; Rastle et al., 2000). Even without any priming, people are faster to recognize words that are semantically more similar to their formal neighbors (Marelli & Amenta, 2018; Marelli et al., 2015). In bilinguals, cognates (etymologically related words) confer an analogous advantage. A word in someone's second language (e.g., *boeuf* in French) is easier to process if it has a cognate in their native language (e.g., *beef* in English; Costa et al., 2000), and this effect is greater in low-frequency words (Strijkers et al., 2010). Brysbaert et al. (2016) suppose that cognates within a language similarly mitigate processing costs. For example, *abbess* should be easier to understand if you are familiar with *abbot*, and this pertains directly to word meaning, unlike lexical decisions (e.g., Balota & Chumbley, 1984; D'Anna et al., 1991).

We expect false cognates, which overlap in form and meaning but are etymologically unrelated, to provide the same sort of advantage. That is, a common word, like *avoid*, should bolster the semantic representation of an overlapping but unrelated low-frequency word, like *evade*. We tested this hypothesis in two preregistered experiments. If overlapping words serve as a scaffolding that helps people process otherwise fuzzy meanings, it will suggest ways to improve how computational models represent low-frequency words (rather than producing idiosyncratic representations based on too few contexts), and it will update predictions about how an efficient lexicon should evolve.

Experiment 1

In this preregistered experiment, participants made semantic relatedness decisions about pairs of English words. A cue word (e.g., *dodge*) was followed by a target, either a "free-rider" that overlaps with a higher frequency word in form and meaning (e.g., *evade*, which overlaps with *avoid*) or a control (e.g., *elude*), which is matched with the free-rider on distributional similarity to the cue. If low-frequency words piggyback on representations of overlapping words, then free-riders should be judged to be related to the cues more often and more quickly than controls are. Note that participants were never presented with the higher frequency overlapping words (e.g., they made decisions about *evade* without seeing *avoid*). We are interested in whether overlapping words affect how people represent and access the meanings of free-riders not because of formal or semantic priming but because of the organization of the mental lexicon.

Transparency and Openness

The experimental design, materials, hypotheses, and planned and exploratory analyses (including stopping rules and exclusion criteria) were all preregistered and can be found at https://osf.io/wrtgu. All stimuli, data, supplemental materials, and R scripts are available at https://osf.io/cmxer.

Participants

In a pilot study with the same design as this experiment, 16 participants decided that free-riders were semantically related to cues in about 70% of the trials and that controls were semantically related to cues in about 50% of the trials. We, therefore, recruited 144 new participants from the online crowdsourcing platform Prolific (https://prolific.co/), which gave us greater than 0.8 power at the 0.05 α level, assuming a similar effect size and accounting for an exclusion rate of up to 40% (cf. Thomas & Clifford, 2017). Following our planned exclusion criteria (see below), we ended up with 101 participants (female = 44, male = 57, prefer not to say = 0, other = 0, selected from a drop-down list; $M_{\text{age}} = 35.1$, range 19–50; we did not collect information about ethnicity). All participants were registered as adult native English speakers who grew up in the United Kingdom, having participated in at least 50 previous studies on Prolific, with an acceptance rate of at least 98%. Participants were required to use desktop or laptop computers, not phones or tablets. They were rewarded with £1 for completing the experiment, which took about 7 min.

Materials

We designed 72 cue–target pairs; in each pair, the target has two versions, a free-rider and a control that are matched on distributional similarity to the cue, according to several DSMs (Günther et al., 2015; Mandera et al., 2017). The selection of free-riders was discretionary. In some cases, we automatically detected phonological or orthographic neighbors with similar distributions (e.g., *evade* occurs in similar contexts to *avoid*); in other cases, we used examples of semantic change (e.g., *sheen* has shifted toward *shine*). The unseen overlapping words (e.g., *avoid* and *shine*) are significantly more distributionally similar to the cues than the targets are. We also matched the targets on formal similarity to the cues, as indicated by paired *t*-tests comparing Levenshtein distances. Paired *t*-tests further indicate that the targets do not differ on a wide variety of lexical variables pertaining to form, meaning, and familiarity (gathered from the English Lexicon Project and the South Carolina Psycholinguistic Metabase; Balota et al., 2007; Gao et al., 2022). For details, see the online supplemental materials, which are available on the OSF page (https://osf.io/cmxer).

Importantly, the targets are relatively low in frequency. Most of the targets fall in OED Frequency Band 4 (i.e., 0.1–1.0 occurrences per million words), which are widely recognizable but rare. This is corroborated by word prevalence norms (Brysbaert et al., 2019): Despite being low in frequency, participants typically recognized the target words in a lexical decision task.

To avoid a ceiling effect, most cue–target pairs are not strongly related; the relatedness decisions are subjective, with room for disagreement. For example, you might feel that the cue *adorn* is not semantically related to the targets *varnish* and *lacquer*, but this ambivalence leaves room for the overlapping word *garnish* to influence interpretation of the free-rider *varnish*. To offset the ambiguous relatedness of the items, we included 10 semantically related filler pairs along with 30 unrelated filler pairs, rather than including only unrelated fillers.

Procedure

We implemented the study on the online experiment builder Gorilla (Gorilla.sc; Anwyl-Irvine et al., 2020). Participants were presented with a cue (e.g., *dodge*) for 1 s, followed by a target, either a free-rider (e.g., *evade*) or a control (e.g., *elude*). They decided as quickly as possible whether the target was semantically related to the cue or not (pressing L or A). The trial was completed upon response or timed out 2 s after the onset of the target word. To incentivize fast and accurate responses, feedback was displayed for 3 s following incorrect responses (a red X) and timed-out trials (a *Too slow* warning).

After giving consent, participants read the instructions and completed eight practice trials. They were then randomly assigned to one of four lists, each comprising 36 items (18 free-riders and 18 controls) and 40 fillers, presented in random order. Notice that each participant responded to only half of the 72 items (and each responded to the same 40 fillers). We divided the list into two subsets (counterbalancing target word frequency across the subsets) to keep the experiment short (and to thereby maintain high-quality responses from online participants) while still including as many items as possible.

Results

We first trimmed the data according to our preregistered exclusion criteria. We excluded 43 participants who correctly responded to less than 75% of the fillers (note that this large proportion of excluded participants is not unexpected for online experiments; e.g., Thomas & Clifford, 2017). For the remaining 101 participants, we excluded trials faster than 200 ms or slower than 2.5 *SD* above the mean RT by item (one fast guess and 72 slow outliers), leaving us with a total of 3,545 trials by 101 participants. (The pattern and significance of effects reported below do not change when including all trials by all participants; see the supplemental materials, which are available on the OSF page (https://osf.io/cmxer)).

Using the lmerTest and tidyverse R packages (Kuznetsova et al., 2017; R Core Team, 2022; Wickham et al., 2019), we conducted logit/linear mixed effects (LME) modeling on trial-level responses (related vs. unrelated decisions) and on trial-level RTs for trials with a related response (i.e., excluding 1,026 trials where the participant decided that the cue and target were unrelated), applying treatment coding, with control targets as the reference level. We used a forward model comparison with an α of 0.2 to determine the maximal random effect structure justified by the data (Matuschek et al., 2017).

As we predicted, participants were significantly more likely to decide that free-riders were semantically related to cue words ($M = 0.80$, $SD = 0.40$), compared to controls ($M = 0.62$, $SD = 0.49$) (see Table 1 and Figure 1). They were also significantly faster to decide that free-riders ($M = 811$, $SD = 213$) were semantically related to cue words, compared to controls ($M = 861$, $SD = 228$).

Discussion

In Experiment 1, we provided evidence that words with overlapping forms and meanings influence how the meanings of low-frequency words are represented and accessed. According to the distributional hypothesis, if two target words are equally similar to a cue word in distribution, then people should be equally likely to decide that they have related meanings. We instead found that people are more likely (and faster) to decide that a target is related to a cue when that target overlaps in form and meaning with another word that is itself semantically related to the cue. This finding advances previous research by demonstrating that similar-sounding words influence the interpretation of known words, not just novel words (e.g., Haslett & Cai, 2022b), that form can cue specific word meanings, not just category membership (e.g., Monaghan et al., 2011), and that overlapping forms and meanings can thereby facilitate word meaning access, not just word recognition (e.g., Bergen, 2004). Furthermore, our findings suggest that form supplements distributional information in word meaning representation: Previous experience does not suffice to define the many low-frequency words that populate the English lexicon, so people sometimes fall back on similar-sounding words.

Experiment 2

Rayner and Duffy (1986, p. 199) argue that low-frequency words are hard to integrate into sentences because their meanings are "less complete or well elaborated." We have provided evidence that overlapping words can help flesh out those meanings. Do overlaps therefore facilitate the integration of low-frequency words into sentences? Farmer et al. (2006) found that people read words with more reliable phonological cues faster in sentence contexts. That is, when people expect a noun, they are faster to read *marble*, a noun that is phonologically similar to other nouns, than to read *insect*, a noun that is phonologically similar to verbs. However, those phonological cues pertain to grammatical categories, not specific word meanings. To investigate whether overlapping words (like *avoid*) facilitate the integration of low-frequency words (like *evade*) into sentences, we inserted the cues and targets from Experiment 1 into sentence contexts.

This experiment employed MouseView.js (Anwyl-Irvine et al., 2021), which allowed us to record the incremental reading of sentences in an online experiment. MouseView.js blurs the participant's screen, mimicking peripheral vision, but leaves a small aperture, mimicking the fovea. The aperture is directed by the participant's cursor, so their gaze can be approximated via mouse-tracking. Anwyl-Irvine et al. validated MouseView.js by presenting participants with either pleasant or disgusting images and comparing dwell times in regions of interest (ROIs) for online participants using MouseView.js versus in-lab participants using an eye-tracker (from a study by Armstrong et al., 2020). MouseView.js and the eye-tracker recorded the same pattern of attraction and aversion to stimuli. Anwyl-Irvine et al. conclude that MouseView.js is a reliable alternative to eye-tracking in online experiments that involve deliberate shifts in attention, including reading.

Table 1
Results of Logit/Linear Mixed Effects Modelling of Semantic Relatedness Decisions in Experiment 1

Comparison	β	*SE*	*z/t*	*p*
Related versus unrelated decisions: Decision ~ TargetType + (TargetType + 1|Item) + (1|Subject)				
Intercept	0.71	0.19	3.74	<.001
Free-rider—Control	1.07	0.19	5.75	<.001
RT of related decisions: RT ~ TargetType + (TargetType + 1|Item) + (TargetType + 1|Subject)				
Intercept	885.0	16.62	53.26	<.001
Free-rider—Control	−62.3	11.03	5.65	<.001

Transparency and Openness

The experimental design, materials, hypotheses, and planned and exploratory analyses (including stopping rules and exclusion criteria) were all preregistered and can be found at https://osf.io/vw8ak. All stimuli, data, supplemental materials, and R scripts are available at https://osf.io/cmxer.

Participants

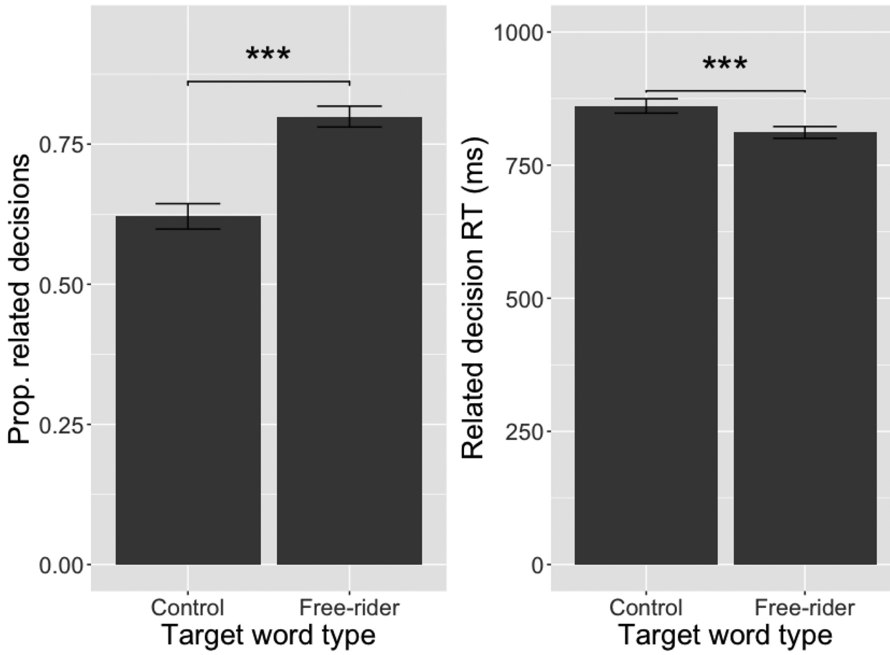
We recruited 144 adult native English speakers from the United Kingdom with the same selection criteria as in Experiment 1, except that participants using Safari browsers were not allowed to participate in order to ensure compatibility with MouseView.js. Participants were rewarded with £1.50 for a task that took about 12 min. Following our planned exclusion criteria (see below), we ended up with 122 participants (females = 85, males = 37, prefer not to say = 0, other = 0, selected from a drop-down list; M_{age} = 36.0, range 18–50; we did not collect information about ethnicity).

Materials

We used the same cue–target pairs as in Experiment 1 but inserted in pairs of sentences. Cue sentences were four words long on average, ranging from two to six words, and target sentences were six words long on average, ranging from five to eight words. For example, the cue *dodge* appeared in the sentence *The kids dodged something*, and the targets *evade* and *elude* appeared in *She tried to evade/elude the officer*. As in Experiment 1, we manipulated the target type within items. That is, the free-riders and controls appeared in identical sentences, except for one item: The free-rider *delve* (which overlaps with *dive*) and the control *immerse* differ in licensing, so

Figure 1

Proportion of Related Decisions (Left) and Related Decision RT (Right) as a Function of Target Type in Experiment 1



Note. Error bars indicate 95% confidence intervals.

delve appears in the sentence *Each student delved into her books*, whereas *immerse* appears in *Each student immersed herself in books*. We used the LSAfun package (Günther et al., 2015) to ensure that the free-riders and controls can be equally well predicted from the cue and target sentences. The conditions do not differ, as demonstrated by paired *t*-tests comparing vectors produced by several semantic spaces. See the supplemental materials, which are available on the OSF page (https://osf.io/cmxer).

As in Experiment 1, we split the items into subsets to keep this online experiment short. We divided the 72 items into four subsets (counterbalanced based on target frequency), for a total of eight lists, each containing nine free-riders, nine controls, and 18 semantically unrelated fillers. Fillers were mismatching sentence pairs from a different subset (e.g., *The kids dodged something* followed by *Helen drew a moat in the blueprint*).

Procedure

We again implemented the study on the online experiment builder Gorilla. After giving consent, participants read the instructions and completed three practice trials. They were then randomly assigned to one of eight lists, each comprising 18 related items and 18 unrelated fillers, presented in random order. In a trial, participants were presented with the cue sentence (e.g., *The kids dodged something*) for 3.5 s. After reading the cue sentence, participants clicked a *Next sentence* button to retrieve the target sentence and situate their mouse on the left side of the screen at a point that coincided with the start of the target sentence. Critically, during the reading of the target sentences, we applied MouseView.js so that participants could view target sentences five or so letters at a time. The sentences were divided into five segments, with one or two words per segment (e.g., *She* | *tried to* | *evade* | *the* | *officer*; see Figure 2). The target word always appeared in isolation, in the middle of the sentence (the third segment), so the ROI is the same in both conditions in all items. They had up to 10 s to read the target sentence and click the *Continue* button in the bottom right corner of the screen; otherwise, the trial timed out. After they finished reading the target sentence, participants decided whether the target sentence was semantically related to the cue sentence or not (by clicking on *Related* or *Not related*) as quickly as possible. If no decision was made within 5 s, the trial timed out. As in Experiment 1, to incentivize fast and accurate responses, feedback was given for 3 s when a response was incorrect (a red X) or when the trial timed out (a *Too slow* warning), and the pair of sentences were then shown on screen for 6 s.

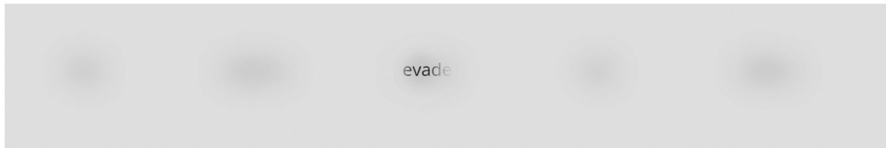
Results

Following data exclusion plans described in the preregistration, we excluded 22 participants who correctly responded to less than 24 of the 36 semantic relatedness decisions. For the data from the 122 remaining participants, we excluded four trials in which the total reading time of the target sentence was less than 2,500 ms. To investigate processing advantages in sentences judged to be semantically related, we included only related decisions in our analysis of first-pass fixation, leaving us with a total of 1,659 trials by 122 participants.

We define first-pass fixation duration as the time stamp when a participant's cursor first enters the ROI subtracted from the time stamp when their cursor next exits the ROI. Contrary to our prediction, there was no significant difference in first-pass fixation on target words between sentences containing free-riders ($M = 863$, $SD =$

Figure 2

A Screen Capture of a Trial in Experiment 2, With the Cursor Fixated on the Free-Rider Evade *in the Sentence* She Tried to Evade the Officer



Note. Non-fixated words to the left and right of the target are blurred.

358) versus controls ($M = 847$, $SD = 332$) (see Table 2 and Figure 3). However, planned exploratory analyses do demonstrate an advantage for free-riders. LME on trial-level data (applying treatment coding, with control targets as the reference level) indicates that participants were significantly faster to read the segments immediately following free-riders ($M = 712$, $SD = 271$) than following controls ($M = 743$, $SD = 286$). For example, first-pass fixations were faster in *She tried to evade the officer* than in *She tried to elude the officer*. (This remains true when analyzing the entire rest of the sentence, not just the following segment, e.g., *the officer* rather than *the*; see the supplemental scripts, which are available on the OSF page (https://osf.io/cmxer)). This lag effect is generally interpreted as easier integration of the preceding word into the sentence (e.g., Duffy et al., 1988; see the discussion below), and this advantage in reading is corroborated by significantly faster total reading time for sentences containing free-riders ($M = 5{,}767$, $SD = 1{,}323$) than controls ($M = 5{,}861$, $SD = 1{,}309$) (i.e., the time between clicking *Next sentence*, situating the cursor on the left side of the screen, and clicking *Continue* in the bottom right corner of the screen). Finally, participants were significantly more likely to judge sentences containing free-riders ($M = 0.81$, $SD = 0.31$) to be semantically related to cues, compared to sentences containing controls ($M = 0.73$, $SD = 0.44$).

Table 2

Results of Logit/Linear Mixed Effects Modelling of Sentence Reading in Experiment 2

Comparison	β	*SE*	*t/z*	*p*
Fixation on target words: Fixation ~ TargetType + (1|Item) + (TargetType + 1|Subject) + (TargetType + 0|Item)				
Intercept	849.01	20.01	42.44	<.001
Free-rider—Control	21.55	15.94	1.35	.180
Fixation on posttarget words: Fixation ~ TargetType + (1|Item) + (1|Subject)				
Intercept	748.14	18.35	40.77	<.001
Free-rider—Control	−31.43	10.63	2.96	.003
Total reading time: RT ~ TargetType + (1|Item) + (1|Subject)				
Intercept	5,903.77	101.51	58.16	<.001
Free-rider—Control	−86.16	39.71	2.17	.030
Related versus unrelated decisions: Decision ~ TargetType + (TargetType + 1|Item) + (1|Subject)				
Intercept	1.30	0.17	7.85	<.001
Free-rider—Control	0.52	0.16	3.35	<.001

Discussion

In Experiment 2, contrary to our hypothesis, we did not find a significant difference in fixation duration on free-riders versus controls. However, participants were faster to read the words immediately following free-riders than following controls. Participants were, for example, slower to read *the officer* following *She tried to elude* (a control) than following *She tried to evade* (a free-rider). This indicates an advantage in integrating free-riders into sentences (e.g., Duffy et al., 1988; Rayner et al., 2007). Typically, lower frequency words are harder to process (e.g., Monsell et al., 1989), and the influence of frequency reliably produces a lag effect, slowing the reading of words following low-frequency words (i.e., catch-up processing or spillover; Kliegl et al., 2006). Higher frequency words with overlapping forms and meanings mitigated this lag effect in free-riders.

MouseView.js is a new technology, and it is important to bear several caveats in mind: MouseView.js has a lower sampling rate than eye-tracking, moves only as quickly and precisely as a participant's hand, and cannot detect parafoveal processing or unconscious fixations. However, the advantage for words following free-riders is robust; it persists when including all trials by all participants. See the supplemental materials, which are available on the OSF page (https://osf.io/cmxer).

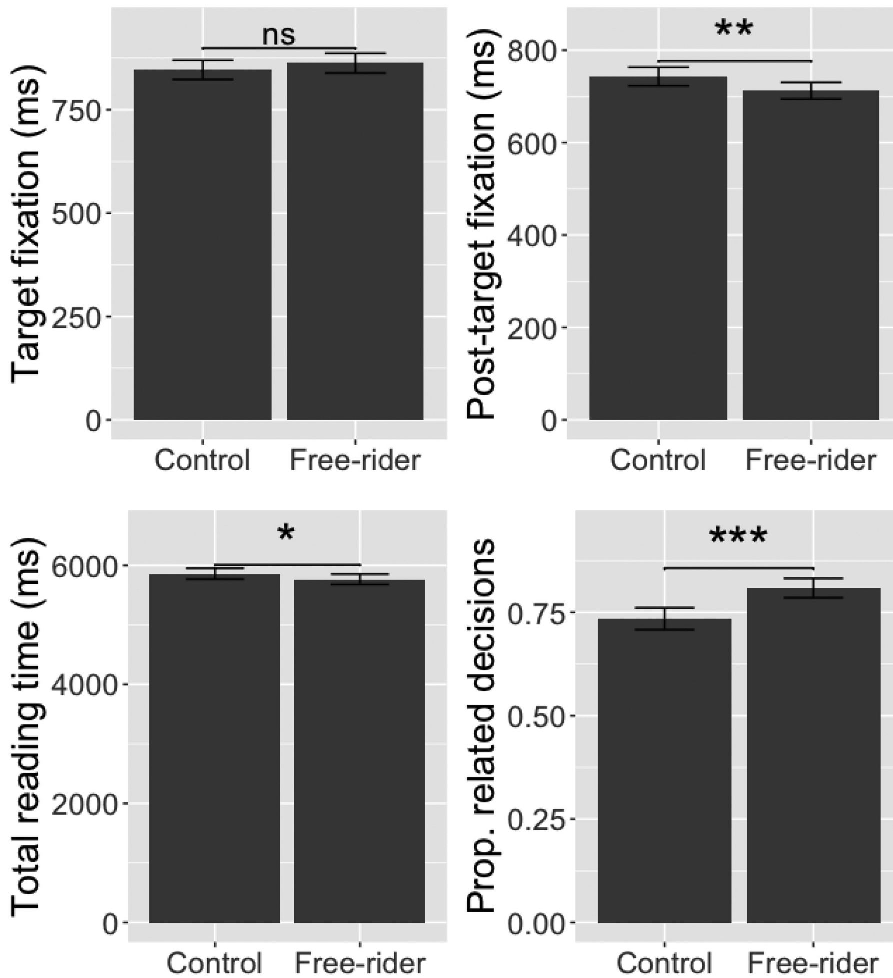
General Discussion

In two experiments, we have provided evidence that low-frequency words co-opt the semantic representations of words that overlap in form and meaning, which influences interpretation and facilitates processing. In Experiment 1, participants decided more often and faster that a cue (e.g., *dodge*) was semantically related to a free-rider (*evade*, which overlaps with *avoid*) than to a control (*elude*) that is equally similar to the cue in distribution. In Experiment 2, participants read sentences containing the same cues (*The kids dodged something*) and targets (*She tried to evade/elude the officer*), and they were again more likely to decide that the sentences containing a free-rider were semantically related to the cue sentences. Although they did not read the free-riders faster than the controls, they did read posttarget words faster following free-riders (e.g., *the officer* following *She tried to avoid/elude*), indicating faster integration of the free-riders into sentences. A low-frequency word like *evade* piggybacks on the semantic representation of an overlapping word like *avoid*, making it easier to process.

Given that participants never encountered the higher frequency overlapping words (e.g., they made decisions about *evade* but never saw *avoid*), priming cannot straightforwardly explain these

Figure 3

Fixation Duration on Target Words (Top Left), Posttarget Words (Top Right), Total Reading Time (Bottom Left), and Proportion of Related Decisions (Bottom Right) as a Function of Target Type in Experiment 2



Note. Error bars indicate 95% confidence intervals.

results. Perhaps free-riders activated overlapping words (e.g., *evade* activated *avoid* rather than vice versa), which subsequently affected processing (a sort of priming feedback loop). However, the speeded decisions in Experiment 1 might make that implausible. Participants made semantic relatedness decisions about free-riders in just over 800 ms on average, which is, for example, a full 200 ms faster than the mean RT in the Calgary semantic decision project (Pexman et al., 2017). These results are more consistent with an explanation in terms of the organization of the mental lexicon, such that overlapping words have already influenced the representations of free-riders and so play only an indirect role in word meaning access. Compare this impact of overlapping words to the polysemy advantage. People recognize polysemous words, with multiple related senses (e.g., *paper* can refer to a sheet, an essay, or a newspaper), faster than unambiguous words (Rodd et al., 2002; Rubenstein et al., 1970). Rodd et al. (2004) argue that the related senses of *paper* share a broad attractor basin in semantic space, enabling faster settling into a stable state.

The same logic of grouping words in semantic space could apply to words that are similar, rather than identical, in form. Consider how in a word association task, unfamiliar prompts, like *diurnal*, often elicit similar-sounding associates, like *urinal* (Postman 1970; Stolz & Tiffany, 1972). These responses are considered nonsensical (e.g., Wolter, 2001), but they might reflect the expedient organization of the mental lexicon; linking low-frequency words to similar-sounding words can activate helpful cues, like *nocturnal*. Consistent with this, Kumar et al. (2022) found the that best computational models of word retrieval supplement distribution with form and, moreover, that people who relied more on phonology in a semantic fluency task (i.e., naming category members, such as animals) successfully retrieved more words (e.g., named more animals). When distribution falls short, people recruit similar-sounding words.

This advantage in word retrieval is consistent with the commonsense claim that words from large families are easier to process than words from small families. For example, once you know what *govern* means, its derivations pose little trouble, whether they are high in frequency, like *government*, or low, like *misgovern* (Goulden et al., 1990; cf. Schreuder & Baayen, 1997). Brysbaert et al. (2016) estimate that while adult native English speakers know well over 40,000 lemmas on average (e.g., *govern* and *governing* are the same lemma), they tend to know "only" 11,000 morphological families (e.g., *govern* and *government* are the same family). Linking related words increases receptive vocabulary without increasing the burden on memory or learning.

The definition of the word family can be broadened to incorporate etymology. As discussed above, words like *abbey*, *abbot*, and *abbess* share a root, which should provide the same sort of advantage as a shared morpheme (Brysbaert et al., 2016). We go further by providing evidence that false cognates, like *evade* and *avoid* (which do not share a root), facilitate comprehension, too. Practically speaking, that line is blurred. Distinguishing cognates from false cognates requires explicit study well outside the normal course of language acquisition. For example, native German speakers often misidentify loan words as "core" German vocabulary (and vice versa) because they lack knowledge about the history of those words, instead relying on surface-level information (Eisenberg, 2011, as cited in Haberland, 2019). It should be unsurprising, then, that we found the same advantage for historically unrelated words as might be expected for conspicuous cognates.

Our results dovetail with the above-mentioned finding that people are faster to recognize words that are more semantically similar to formal neighbors (e.g., *farm* is similar to *farmer*, whereas *numb* is dissimilar to *number*; Marelli et al., 2015). That metric is a significant predictor of recognition times even after controlling for morphological family size, consistent with the impact of false cognates on word meaning access. Recent similar studies (Gatti et al., 2022; Hendrix & Sun, 2021) have employed a DSM that computes the distributional similarity of sublexical strings, allowing it to better represent less frequent words. FastText (Bojanowski et al., 2017) tracks the distributions of *n*-grams, such as the 4-g *vert*, so it incorporates etymological information into word meaning representations, as in *vertical* and *vertigo*. FastText can even concoct representations for unknown words (i.e., that do not occur in its training corpus) based on their formal similarity to other words. By attending to similar-sounding words, a DSM-like fastText can make human-like inferences about the meanings of low-frequency words.

The successes of fastText and the results of our experiments demonstrate how overlapping forms and meanings can facilitate word processing. Some research emphasizes that substantial overlaps cause confusion and so are suboptimal (e.g., Monaghan et al., 2011), but distributional information is unreliable for low-frequency words (e.g., Lenci et al., 2022), and most words are low in frequency, so in many cases, a conflated meaning that is rich and ready to hand will be preferable to an impoverished meaning. Accordingly, we found overlapping forms and meanings to be advantageous when accessing the meanings of low-frequency words. In an optimal lexicon that offsets constraints on memory and learning, words with fuzzy meanings should evolve to sound more, not less, like their semantic neighbors.

Conclusion

The bulk of the English lexicon is low in frequency, yet state-of-the-art computational models struggle to process low-frequency words. We have provided evidence that English speakers compensate for a lack of experience with words by associating them with similar-sounding words that have related meanings. Models of word meaning representation that incorporate these unconventional word families, like fastText, can better explain how people process the many low-frequency words they encounter.

Constraints on Generality

Word form is not the only way to supplement a lack of distributional information. Sensorimotor experience can also bolster representations of low-frequency words (cf. Andrews et al., 2014; Lupyan & Lewis, 2019), so we expect conceptual replications of these experiments to be most successful when using low-frequency words that have abstract meanings, which are deficient in both distributional and sensorimotor information. We also expect effect sizes to vary as a function of participant education level, assuming that more formal education corresponds, on average, to greater familiarity with low-frequency words. Finally, like too much cognitive science research, our experiments pertain specifically to English, so it will be important to see whether the findings generalize to unrelated languages (cf. Blasi et al., 2022).

References

Altman, D. G., & Bland, J. M. (1999). Statistics notes: Variables and parameters. *BMJ*, *318*(7199), Article 1667. https://doi.org/10.1136/bmj.318.7199.1667

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*(3), 359–370. https://doi.org/10.1111/tops.12096

Anwyl-Irvine, A. L., Armstrong, T., & Dalmaijer, E. S. (2021). Mouseview.js: Reliable and valid attention tracking in web-based experiments using a cursor-directed aperture. *Behavior Research Methods*, *54*(4), 1663–1687. https://doi.org/10.3758/s13428-021-01703-5

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Armstrong, T., Stewart, J. G., Dalmaijer, E. S., Rowe, M., Danielson, S., Engel, M., Bailey, B., & Morris, M. (2020). I've seen enough! Prolonged and repeated exposure to disgusting stimuli increases oculomotor avoidance. *Emotion*, *22*(6), 1368–1381. https://doi.org/10.1037/emo0000919

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, *80*(2), 290–311. https://doi.org/10.1353/lan.2004.0056

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Bolinger, D. L. (1953). The life and death of words. *The American Scholar*, *22*(3), 323–335. https://www.jstor.org/stable/41207650

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Agarwal, S.

(2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's Age. *Frontiers in Psychology*, *7*, Article 1116. https://doi.org/10.3389/fpsyg.2016.01116

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning. Memory, and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Cuetos, F., Arce, N., Martínez, C., & Ellis, A. W. (2017). Word recognition in Alzheimer's disease: Effects of semantic degeneration. *Journal of Neuropsychology*, *11*(1), 26–39. https://doi.org/10.1111/jnp.12077

D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, *23*(1), 109–122. https://doi.org/10.1080/10862969109547729

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. https://doi.org/10.1016/j.tics.2015.07.013

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203–12208. https://doi.org/10.1073/pnas.0602173103

Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. *Studies in Linguistic Analysis*, 10–32.

Gao, C., Shinkareva, S. V., & Desai, R. H. (2022). SCOPE: The South Carolina psycholinguistic metabase. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01934-0

Gasser, M. (2004). The origins of arbitrariness in language. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. *26*, No. 26, pp. 434–439). https://escholarship.org/uc/item/34g8355v

Gasser, M., Sethuraman, N., & Hockema, S. (2005). *Iconicity in expressives: An empirical investigation. Experimental and empirical methods*. CSLI Publications.

Gatti, D., Marelli, M., & Rinaldi, L. (2022). Out-of-vocabulary but not meaningless: Evidence for semantic-priming effects in pseudoword processing. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0001304

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363. https://doi.org/10.1093/applin/11.4.341

Gregg, V. H. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). John Wiley & Sons.

Gregg, V. H., Montgomery, D. C., & Castaño, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior*, *19*(2), 240–245. https://doi.org/10.1016/S0022-5371(80)90202-9

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun—An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). *How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection*. arXiv preprint arXiv:2301.07597.

Haberland, H. (2019). On the limits of etymology. *Acta Linguistica Hafniensia*, *51*(1), 90–103. https://doi.org/10.1080/03740463.2019.1625556

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Haslett, D. A., & Cai, Z. G. (2022a). When words impinge: Overlapping sounds and meanings in low-frequency words. In *Architectures and mechanisms for language processing 28*.

Haslett, D. A., & Cai, Z. G. (2022b). New neighbours make bad fences: Form-based semantic shifts in word learning. *Psychonomic Bulletin and Review*, *29*(3), 1017–1025. https://doi.org/10.3758/s13423-021-02037-1

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 157–183. https://doi.org/10.1037/xlm0000819

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *41*(6), 401–410. https://doi.org/10.1037/h0056020

Hutchins, S. (1998). *The psychological reality, variability, and compositionality of English phonesthemes* [PhD Dissertation]. Emory University.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Kumar, A. A., Lundin, N. B., & Jones, M. N. (2022). Mouse-mole-vole: The inconspicuous benefit of phonology during retrieval from semantic memory. In *Proceedings of the annual meeting of the Cognitive Science Society* (pp. 651–657).

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Kwon, N. (2017). Empirically observed iconicity levels of English phonaesthemes. *Public Journal of Semiotics*, *7*(2), 73–93. https://doi.org/10.37693/pjos.2016.7.16470

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's Problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lenci, A., Sahlgren, M., Jeuniaux, P., Cuba Gyllensten, A., & Miliani, M. (2022). A comparative evaluation and analysis of three generations of distributional semantic models. *Language Resources and Evaluation*, *56*(4), 1269–1313. https://doi.org/10.1007/s10579-021-09575-z

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193. https://doi.org/10.1017/langcog.2014.21

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1319–1337. https://doi.org/10.1080/23273798.2017.1404114

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behavior Research Methods*, *50*(4), 1482–1495. https://doi.org/10.3758/s13428-018-1017-8

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of orthography-semantics consistency on word recognition. *Quarterly Journal of Experimental Psychology*, *68*(8), 1571–1583. https://doi.org/10.1080/17470218.2014.959709

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McDonald, S. (2000). *Environmental determinants of lexical processing effort* [PhD thesis]. University of Edinburgh.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*, 1–9.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143–182. https://doi.org/10.1016/j.cognition.2004.09.001

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, *140*(3), 325–347. https://doi.org/10.1037/a0022924

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*(1), 43–71. https://doi.org/10.1037/0096-3445.118.1.43

Pastizzo, M. J., & Feldman, L. B. (2009). Multiple dimensions of relatedness among words: Conjoint effects of form and meaning in word recognition. *The Mental Lexicon*, *4*(1), 1–25. https://doi.org/10.1075/ml.4.1.01pas

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: Concrete/abstract decision data for 10,000 English words. *Behavior Research Methods*, *49*(2), 407–417. https://doi.org/10.3758/s13428-016-0720-6

Postman, L. (1970). The California norms: Association as a function of word frequency. In L. Postman, & G. Keppel (Eds.), *Norms of word association* (pp. 241–320). Academic Press. https://doi.org/10.1016/B978-0-12-563050-4.50009-1

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *The Journal of General Psychology*, *13*(1), 199–203. https://doi.org/10.1080/00221309.1935.9917878

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4–5), 507–537. https://doi.org/10.1080/01690960050119689

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J., & Reichle, E. D. (2007). Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert (2006). *Journal of Experimental Psychology: General*, *136*(3), 520–529. https://doi.org/10.1037/0096-3445.136.3.520

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*(1), 89–104. https://doi.org/10.1207/s15516709cog2801_4

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 487–494. https://doi.org/10.1016/S0022-5371(70)80091-3

Sahlgren, M., & Lenci, A. (2016). *The effects of data size and frequency range on distributional semantic models*. arXiv preprint arXiv:1609.08293. https://doi.org/10.18653/v1/D16-1099

Schick, T., & Schütze, H. (2020). Rare words: A major problem for contextualized embeddings and how to fix it by attentive mimicking. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(5), 8766–8774. https://doi.org/10.1609/aaai.v34i05.6403

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*(1), 118–139. https://doi.org/10.1006/jmla.1997.2510

Stolz, W. S., & Tiffany, J. (1972). The production of "child-like" word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Verbal Behavior*, *11*(1), 38–46. https://doi.org/10.1016/S0022-5371(72)80057-4

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197. https://doi.org/10.1016/j.chb.2017.08.038

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*(1), 41–69. https://doi.org/10.1017/S0272263101001024

Zipf, G. (1936). *The psychobiology of language*. Routledge.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

Received June 20, 2022
Revision received February 3, 2023
Accepted February 15, 2023 ■