Received: 11 August 2020 | Revised: 11 September 2021 | Accepted: 4 November 2021

DOI: 10.1111/lnc3.12444

REVIEW ARTICLE

WILEY

Frequency effects in reading are powerful – But is contextual diversity the more important variable?

Catherine L. Caldwell-Harris

Department of Psychology, Boston University, Boston, Massachusetts, USA

Correspondence
Catherine L. Caldwell-Harris, Department of Psychology, Boston University, 64 Cummington St., Boston, MA 02215, USA.
Email: charris@bu.edu

Abstract

For decades word frequency has been one of the most important variables in psycholinguistics. Frequent words are more easily recognized and processed more efficiently than rare words. In the fields of word recognition and psycholinguistics, all researchers are reminded to statistically control for word frequency. But is that advice still correct? Are other variables which are correlated with word frequency more important for human language processing? These questions have arisen because of the recent construction of text corpora of billions of words. Also important is the growing practice of archiving word recognition data in databases accessible for anyone to mine. A key result is that words that typically appear in restricted contexts are processed less efficiently than words appearing in diverse contexts. But the new variable of contextual diversity hasn't simply replaced word frequency. This paper traces the history of contextual diversity findings, including the twists and turns towards a more sophisticated understanding of what makes words easy to learn and process. Myriad findings of the last 20 years are discussed: the rational theory of memory, spacing effects in learning, phrase frequency effects, the neural basis of repetition suppression, and why reverse frequency effects are observed in semantic aphasia. Methods reviewed include artificial language learning, event-related potentials, and eye movement studies. The result is a new appreciation that

word processing skills emerge from complex brain networks which include information about words' typical contexts of occurrence.

1 | INTRODUCTION

For many decades, from the 1970s through the early 21st century, the paradigm of visual word recognition was the petri dish of cognitive psychology. Using psychology students as subjects, psychologists were able to develop and test the myriad theories that are now foundational to scientific psychology, including theories of parallel versus serial processing, subliminal perception, visual pattern recognition, and how words prime each other. One of the most powerful findings was the robustness of word frequency effects. But why are more frequently encountered words easier to recognize? In the last 2 decades, researchers have wondered if frequency effects aren't what they appeared to be. Higher word frequency is correlated with faster reaction times, but is this an illusory correlation?

2 | FREQUENCY OF OCCURRENCE

Practice makes perfect. The more we play a sport or instrument, or sing a song, utter a phrase, or drive to a location, the easier and more efficient it is to do the same action the next time. The frequency with which we perform mental activities works the same way. The more frequently we process a word, the more rapidly and accurately we can identify that word and its meanings when it is next encountered (Preston, 1935; see reviews in Behrens & Pfänder, 2016; Brysbaert et al., 2018; Diesel, 2007; Divjak & Caldwell-Harris, 2015; Divjak, 2019).

Higher frequency words are recognized more quickly than lower frequency words. This has come to be known as the "word frequency" (WF) effect (Howes & Solomon, 1951; Monsell, 1991). Frequency effects have been found in production and comprehension in every domain studied by psycholinguists, from small units of letters and phonemes to morphemes, and word to multiword expressions (MacLeod & Kampe, 1996). In the last decade, four book-length treatments have been published on how the frequency of occurrence of words and phrases influence language processing (Behrens & Pfänder, 2016; Divjak, 2019; Divjak, & Gries, 2012), and second language acquisition (Madlener, 2015).

A common experimental paradigm to test frequency effects in words is the lexical decision task. Words and nonwords (e.g., *deck*, *dcek*) are displayed on a computer screen. Research participants must press a button if the displayed item is a legitimate word or not. The length of time to press the button (also called response latency) is the measure of processing time. Another common task is naming; research participants read out loud the word as quickly as possible, with a voice-activated microphone triggering the calculation of response latency. For these and similar tasks, the frequency of occurrence of the words usually has a strong correlation with response latency and other measures (see review in Divjak, 2019).

What is the mechanism underlying word frequency effects? It is intuitive that the more frequently we encounter words, the more chance we have to learn them and build robust mental representations and efficient neural pathways (Monsell, 1991). These efficient neural pathways begin with early visual processing of the word, such as recognition of the shapes of letters. Letters activate mental representations for the words they are part of; ultimately the words' meanings are activated, and the word is recognized.

Figure 1 illustrates the historic assumptions of scholars: each time we hear a word (or phrase), its memory representation becomes stronger (more entrenched); facilitating retrieval, learning, and enhancing learning when it is encountered again. This can be labeled frequency$_{rep}$, referring to the frequency in the sense of repetition. Figure 1 illustrates the correlation between how quickly a human can make yes/no decisions in a task like lexical decision, with the frequency of occurrence of words in text corpora. These are the two endpoints of the path in the diagram. The underlying processes that cause this correlation are learning and neural entrenchment, shown as the intermediary steps in Figure 1.

Frequency is so important for learning that it has long been included in theoretical models of human learning and representation. Early theorists conceived of the mental lexicon as a mental dictionary, with lexical access being the result of a frequency-ordered search (e.g., Murray & Forster, 2004). Theorists viewed a word's frequency to be part of a word's representation in the brain's mental dictionary (Morton, 1979). Dictionary-like storage for the lexicon began to seem too static with the development of artificial network models, particularly connectionist networks (Rumelhart & McClelland, 1982). These networks do not have data structures where word meanings are stored. Instead, information associated with words is distributed across a large set of processing units. The weighted links between units encode word information during the learning of words; the links continue to be updated as words are used (Seidenberg, 2005).

Connectionist networks explain frequency effects in any human endeavor X (where X, of course, includes word reading) in terms of more redundant connections, shorter connections,

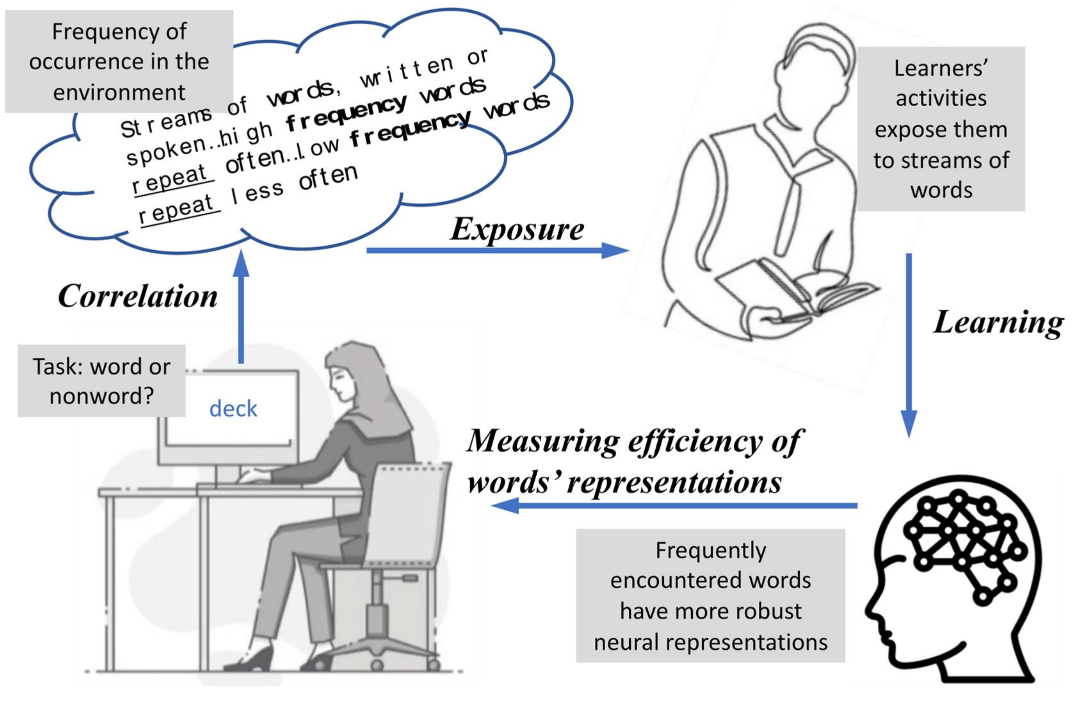

FIGURE 1 Leftmost components depict the data available to researchers, which are frequency of occurrence in texts, and research participants' word recognition speeds. Rightmost components are the proposed causal mechanisms

and special intermediary processing units (called hidden units) dedicated to processing X, and larger/stronger attractors for X (Seidenberg, 2005). Monsell (1991) noted that connectionist models brought "... the problem of learning to center stage. The result is to give the frequency of occurrence its rightful place as a fundamental shaper of a lexical system; always dynamically responsive to experience..." (p. 150).

The idea that $\text{frequency}_{\text{rep}}$ makes mental representations more efficient is intuitively appealing. But over the last decade, researchers have conducted mega studies with large numbers of participants, correlating word recognition processing times to an array of different variables, such as age-of-acquisition, imageability, and familiarity (Balota et al., 2012; Brysbaert et al., 2016). One finding is that subjective frequency (i.e., people's estimates of their exposure to words) is more predictive of processing time than objective text counts, especially for low-frequency words. The relationship between frequency and processing also differs for people with large versus small vocabularies (Kuperman & Van Dyke, 2013). In particular, poor readers may not benefit as greatly from exposure to words as do advanced readers (Kuperman & Van Dyke, 2013). Explorations with text corpora and large databases will be the focus of this article: is the relationship between $\text{frequency}_{\text{rep}}$ and processing efficiency an illusory correlation?

3 | THE STUNNING FINDINGS FROM THE LAST DECADE: DISCOVERING CONTEXTUAL DIVERSITY

Contemporary doubts about the causal role of frequency emerged as an accidental discovery. Researchers have long used the Brown Corpus (Francis & Kucera, 1982), one million words compiled from electronic files of books and newspapers. With one million words of text, even well-known words might only occur once or not at all. For example, Lovelace (1988) noted that the words *cucumber, fraternal, lettuce, nag,* and *quilt* had an occurrence of 0 (see also Brysbaert & New, 2009; Brysbaert, et al., 2018). In the early years of the 21st century, corpora of millions of words were constructed, such as the COCA database of 300 million words (Davies, 2009) and the WaCky corpus, which has databases of one billion or more words in several languages (Baroni et al., 2009). Large corpora were data-mined to extract frequency counts but also to develop new measures of words' attributes. Statistical techniques were used to determine which of many lexical variables would best predict the results of the lexical decision and naming response times available in stored databases (e.g., Balota, Yap, Cortese, et al., 2007; Balota et al., 2012). Large amounts of easy-to-analyze data allowed considerable inventiveness as researchers used the large corpora to construct different types of predictor variables.

3.1 | Measuring contextual distinctiveness

In the earliest explorations of this type, Lund and Burgess (1996) and McDonald and Shillcock (2001) measured lexical variation with a construct called *Contextual Distinctiveness*. This is a corpus-derived summary measure of the frequency distribution of the contexts in which a word occurs. McDonald and Shillcock predefined a set of target words and a set of k context words. Target words were those for which naming and lexical decision times were already available in a database. Computer programs then searched large text corpora to find the target words in their contexts of occurrence. The researchers then identified all words in the context of the target word, using a window of words on either side. The size of the window was a parameter, k, that could vary from four to 10 words preceding or following the target word. The next step was to determine how often each of these *k* context words occurred in the window centered around the target word. This resulted in *k* vectors, where each component of the vector was the frequency of a context word. This co-occurrence vector provided a summary of the word's contextual behavior. Principal components analysis was then used to extract a single summary variable of words' contextual behavior, called contextual distinctiveness.

Words with low contextual distinctiveness occurred in a wide variety of contexts, such as function words and multi-purpose terms like *thing*, *day,* and *very*. These are also usually the words with very high word frequency (high frequency$_{rep}$). Words occurring in somewhat restricted contexts are more specialized terms such as *ostrich* or *mountain*; these tend to be medium frequency. Examples of words that are almost entirely stuck in one specific context include *amok* (*run amok*) and *paragon* (*paragon of virtue*); these have very low word frequency.

McDonald and Shillcock (2001) noted that contextual distinctiveness is highly correlated with the log of word frequency, $r = -0.82$. The noteworthy finding was that contextual distinctiveness was a better predictor of lexical decision and naming latencies than word frequency.

3.2 | Why contextual distinctiveness helps

Why do contextually distinctive words have a processing advantage when humans speedily access the meaning of words? Repeated exposure to an item plausibly strengthens its representation in memory, as discussed above. But human memory shows spacing effects (Glenberg, 1979). An example familiar to any student is the rapid forgetting of course concepts after a test is over. This can be lamented as a waste of study effort. But forgetting is adaptive from an evolutionary perspective (Pirolli & Card, 1999). If our hunter-gatherer ancestors encountered new information in just one context on a single occasion, that information wasn't general enough to be worth retaining. Consistent with this, people have the poorest recall for items that have been encountered during massed practice. The best recall is for items that have been encountered repeatedly and separated in time and space (Carpenter & DeLosh, 2005; Dempster, 1989). This is why cramming for a test yields less enduring learning than spaced study periods.

These observations and related findings prompted the rational analysis of memory (Anderson & Milson, 1989; Anderson & Schooler, 1991). In a study of people's recall of daily newspaper events, memory was best for items that recurred in the newspapers across weeks or months, and poorest for events with the same frequency packed into a short period (Anderson & Schooler, 1991). Memory experts have now concluded that memory is best conceptualized not as a storage device, but as a dynamic system organized according to the *principle of likely need* (Anderson & Milson, 1989; see also Pagán & Nation, 2019). This pithy phrase targets the conditions that facilitate memory retention: will this information be needed in the future?

But would this principle apply to words encountered spread out over different contexts versus massed into a single context? Adelman et al. (2006) used a simple measure called contextual diversity (CD): the number of passages or documents in which a word occurs. Like McDonald and Shillcock (2001), they found that CD was a better predictor of lexical decision and naming times than word frequency, using data from English Lexicon Project (Balota et al., 2007).

These findings were confirmed using eye-tracking during free reading. Researchers embedded target words that varied in word frequency and contextual diversity into narrative texts. All first pass and later reading times were longer for lower CD words when word frequency was statistically controlled (Plummer et al., 2014). The same pattern occurred in Chinese both when manipulating the contextual diversity of words, and when manipulating the contextual diversity of the characters (Chen, Huang, et al., 2017; Chen, Zhao, et al., 2017). Contextual diversity effects have also been found in applied classroom settings (Rosa et al., 2017) and studies using grade-school children (Perea et al., 2013).

3.3 | Creative computing

The research took an inventive turn when Brysbaert and New (2009) drew on the numerous electronic databases of subtitled TV and movies. Which source of text would be a more reliable measure of how your brain processes language: texts from books and newspapers, or texts from TV/movie dialog? It was the TV/movie dialog database, especially for young adults. Note, however, that written texts were sometimes better corpora from which to extract frequency and diversity counts for older adults, suggesting that researchers should incorporate a combination of sources in their investigations (Brysbaert & Ellis, 2016).

Using subtitles had the additional advantage of creating textual databases in any language which had a sizable film or TV industry. This allowed researchers to explore how variables other than frequency may influence response latencies in word recognition experiments. For example, Brysbaert and New (2009) operationalized contextual diversity (CD) as the number of films in which a given word occurred. CD was a better predictor of single-word processing times than was word frequency. Baayen (2010) had a similar finding using subtitle databases.

3.4 | The paradox of repetition suppression

The phenomenon of repetition suppression provides another way to understand the limitations of $\text{frequency}_{\text{rep}}$. Repeating a prior stimulus leads to weaker and less sustained neural activation, as observed via functional magnetic resonance imaging (fMRI; see Grill-Spector et al., 2006). To understand repetition suppression, consider what occurs at a neural level when a stimulus is repeated. As has been discussed already, repetition facilitates learning and leads to more efficient processing (Monsell, 1991). This more efficient processing leads to repetition priming: when a stimulus is immediately repeated, or repeated with a brief delay, the second occurrence is easier to recognize (Butler & Berry, 2004). But this ease-of-recognition brings with it reduced neural activation, called repetition suppression (Grill-Spector et al., 2006).

High levels of neural activation typically strengthen the connections between neurons. But reduced neural activation following repetition means less strengthening of connections. The result is that the stimulus is encoded with a less robust and/or less efficient neural pattern. This illuminates the prior discussion of spacing effects. Consider how much neural strengthening occurs for a word that you read once each day for three days, compared to reading the word three times in a row in a few seconds. The latter word "burst" is treated by your brain as tantamount to a single occurrence (Hollis, 2020). The result is less learning for items that are encountered repeatedly in a short time.

3.5 | But is semantic diversity even more important than contextual diversity?

Alternative explanations for frequency effects turned another corner with Jones et al. (2012), who investigated whether *semantic* diversity facilitates lexical processing more than contextual diversity. Consider a word being repeated in different but highly similar documents, such as the word *mortgage* appearing in two loan documents. A repetition more strongly signals a new context if mortgage appears in a document unrelated to finance; as in the phrase *mortgage the country's future*. It is plausible that words may gain more robust encoding if they appear in semantically diverse contexts.

Jones et al. (2012) reasoned that similarity between any two documents could be defined as the proportion of words shared. Those authors ignored order and filtered out function words, using methods pioneered by Landauer and Dumais (1997). They defined a word's semantic distinctiveness as the average dissimilarity over all of the documents in which the word occurred. In a related effort, Johns et al. (2012) used semantic distinctiveness, word frequency, and contextual distinctiveness to predict humans' recognition accuracy of spoken language words with varying noise levels. Semantic distinctiveness predicted more variance in response times than the other two variables.

At a conceptual level, how is semantic diversity different from contextual diversity? My take on the literature is that semantic diversity is a type of contextual diversity; one in which semantic content changes across the two contexts. There could in principle be other types of contextual diversity. For example, semantic content could be similar across two documents, but the language register could change from formal to colloquial, or the grammar could be that of a native vs. foreign language speaker (see Jones et al., 2017).

3.6 | Semantic diversity explains why frequency effects are not always observed

Word frequency effects are sometimes absent or even reversed in patients with semantic aphasia. A key deficit in semantic aphasia is word retrieval problems (Hoffman et al., 2011). Semantic aphasia is common in Broca's aphasia and involves lesions to the inferior frontal gyrus (IFG). The IFG performs computations that match the information in working memory with the stimulus information that is required to perform a task. Consider the task of naming objects in a picture. As people examine the pictured object, several possible ways to identify the object typically become mentally activated. For example, a picture of a shirt may activate concepts of clothing, fabric, and words such as *shirt*, or *blouse*. In word naming tasks, processes mediated by the IFG select a single word, from this set of activated terms, that is most appropriate for the picture. Lacking an intact IFG is one reason people with semantic aphasia have difficulty in picture naming and other word retrieval tasks.

Why does difficulty with item selection result in an advantage in naming low-frequency words? Hoffman et al. (2011) provided the example of *dog*, a semantically diverse, high frequency word that can appear in a variety of contexts, as it has both concrete and metaphorical usages. When semantically diverse words are processed, their varied senses are activated, as are their lexical associates. A poorly functioning IFG will be swamped and impaired at selecting the correct word. In contrast, low frequency words typically have highly specific meanings and activate fewer associates, easing the selection problem.

3.7 | Putting measures to the test with artificial language learning

One challenge for researchers is that frequency and diversity measures are often correlated. These measures can be independently varied using artificial languages, one of the standard tools for testing hypotheses about learning (see Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, 2016). Jones et al. (2012) showed research participants three-word "sentences" composed of invented words in the artificial language called Xaelon. Participants also saw an image describing sentence meaning. In these tasks, people generally figure out word meaning from making inferences over a large set of trials. During testing, participants had to judge whether a word was a legitimate word in Xaelon. Response latencies were faster for the Xaelon words which occurred in varied semantic contexts, compared to words that were repeated in semantically similar contexts. Semantic diversity thus does seem to be important for creating efficient lexical representations.

A variant of this technique is to use naturalistic passages containing novel words, which allows free reading, an ecologically valid task. Johns et al. (2016) embedded novel words in ordinary passages. When tested later, participants were faster and more accurate at recognizing words that were encountered in distinct discourse contexts, compared to redundant contexts. This result has also been found for research participants reading such manipulated texts in their non-native language (Frances et al., 2020).

In an interesting twist, Pagán and Nation (2019) used this technique along with eye tracking to study word learning. Adults learned words faster when $\text{frequency}_{\text{rep}}$ was high even if words were clumped together in the same context. An explanation for why similar contexts may benefit initial word learning is anchoring (Mak et al., 2020). Anchoring helps stabilize novel word representations by reducing contextual noise. However, Mak et al. (2020, Experiment 2) further showed that as learning progresses, high diversity boosts lexical processing. Therefore, the ideal learning technique from this research is to learn words first via less diverse (i.e., anchoring) contexts, then practice them in diverse environments which help enrich lexical representations.

4 | DELVING DEEPER, IDENTIFYING MECHANISMS

4.1 | Diversity effects arise naturally in a learning model that includes context

Baayen (2010) developed a computational model to explain morphological processing. Called the naive discrimination reader (NDR), the NDR shares features with connectionist models (as in Seidenberg, 2005), using an error-driving learning algorithm to map from inputs to outputs. The outputs are representations of meanings.

In the NDR model, lexical meanings are learned from contextually rich input. These are letter bigrams and trigrams drawn from a window of four words rather than from words in isolation. The activation of a meaning on a given trial is obtained by summing the weights on connections linking active letters and letter pairs to that meaning. The author assumed that lexical learning is contextual because words are usually encountered in multi-word utterances rather than in the word lists on which most models of reading tend to be trained.

The NDR model correctly predicted frequency effects for derived and inflected words, even though morphemes are not explicitly represented in the model. Baayen reasoned that the model would show contextual diversity effects since words were learned in context. He computed contextual diversity in ways similar to the vector models used by prior researchers. The contextual diversity accounted for substantially more variance in word recognition efficiency than did word frequency. Another success of the model was that it also predicted phrase frequency effects. Contextual diversity is an emergent property of the NDR model's learning features.

4.2 | Contextual diversity and word frequency effects have distinct markers of neural processing

Both contextual diversity and word frequency facilitate word recognition. But are the facilitative effects of CD and word frequency due to the same or different underlying mechanisms? This question was pursued by Vergara-Martínez et al. (2017) using event-related potentials (ERPs), a technique in which electrical signals from the scalp are measured. When these signals are averaged across persons and stimuli, components can be identified and linked to mental processes. A key component in language processing is the N400, which is a negative-going signal that reaches its maximum amplitude at approximately 400 ms. after a word is displayed. This component of the ERP signal is typically largest when people are presented with an anomalous word in a sentence (Kutas & Van Petten, 1988). It seems that the large negative amplitude indexes the difficulty of fitting words' meanings into their context. When words are presented in isolation, the standard word frequency effect is that high-frequency words have small N400s (easy to process) and low-frequency words have high N400s.

Vergara-Martínez et al. (2017) observed different ERP patterns for high-CD versus high-frequency words. High-CD words had a larger ERP negative component from 225 to 325 ms post-stimulus onset than did low-CD words. This result goes in the opposite direction of the standard N400 word frequency effects, which have low negative amplitudes. Scalp distribution of the contextual diversity effect resembled the ERP effects associated with semantic richness. This is consistent with the tendency for higher CD words to be semantically rich, due to having diverse meanings that can appear in different contexts.

This difference in scalp distribution and the timing difference associated with high CD versus high-frequency words suggested to the authors that CD is related to the semantic properties of a word, while word frequency is related to lexical properties. This led Vergara-Martínez et al. (2017) to argue that CD and word frequency effects originate from different sources during word processing. This matches widespread agreement that CD is sensitive to semantic components of a word, more so than word frequency (Johns, 2021). But at the same time that Vergara-Martínez et al. (2017) were arguing for fundamental differences between CD and word frequency, other researchers noted the heuristic value of a different position: to understand how CD and word frequency are different, it can help to examine how they are similar. This is explored in the next section.

4.3 | Are word frequency and contextual diversity the same thing, measured at different scales?

Jones et al. (2017, p. 251) asked, “Are two paragraphs from the same article really from the same context? What about speech, time, and the physical environment? How humans dice up their experience into discrete contextual units is of fundamental importance to understanding a mechanism based on CD.”

Hollis (2020) took up this challenge. He noted that word frequency and contextual diversity are identical measures when the size of the context around a target word is 0. He then computed multiple versions of contextual diversity, defining context as windows of text surrounding a word. Window sizes of 4–32 were ideal for maximizing the correlation between CD and word processing measures. With a window of size 1, the target word is counted regardless of context. Hollis explained: "CD_1 and CD_{doc} represent two extremes along a continuum of defining the width of a context window... When treated this way, it seems a bit unusual to ask whether repetition, measured by WF or contextual variation (indexed by CD_{doc}) has primacy as a psychological construct. Both variables are measuring the same concept, just at different scales... There is no CD/WF dichotomy because we affect and are affected by multiple scales of context simultaneously... each may have more or less relevance to explaining behavioral performance in a particular task" (p. 16).

As part of exploring context window size, Hollis identified one arena where CD and frequency have different values, even though they are typically correlated. These are words that are "bursty," meaning words have high frequency_{rep} within a specific context, but have low contextual diversity. Examples are first names, scientific terms, and other specialized vocabulary items. The implication is that bursty words, which may also be low-frequency, will be more difficult to process in isolation. Hollis (2020) argues that a great deal of the benefits of contextual diversity can be attributable to "burstiness," because words that are typically restricted to specialized contexts, regardless of their high or low frequency_{rep}, will be processed differently than words occurring in many contexts. Building on this idea, researchers developed the idea of textual diversity and word prevalence, explored in the next section.

4.4 | Textual diversity and word prevalence: How widely is a word known?

Are words that occur across semantically distinct contexts more varied in meaning, as in homonyms and polysemes? This is what Hoffman et al. (2011) reasoned given their finding of reversed word frequency effects in aphasiac patients, described above, and also explored in a subsequent study (Hoffman et al., 2013). Another research team analyzed the Hoffman et al. (2013) data and showed that words with low semantic diversity are words that occur in very specialized contexts (Cevoli et al., 2021). Semantic diversity helps explain single-word naming data for low-frequency words because some low-frequency words are well known by people despite their low frequency (e.g., *hinge, diverge, sideswiped*). Other low-frequency words are so specialized in meaning that some readers may never have encountered these words in their history of reading. Cevoli et al. (2021) advocated the term *textual diversity* to refer to whether words appear in specialized versus general-topic documents.

Similar to this is the concept of *word prevalence* defined as the number (or proportion) of people who know a word (Brysbaert et al., 2019). Word prevalence helps distinguish low-frequency words which are widely known (*uppercase, hoodie*) from those known by very few people (e.g., *scourge, whicker*). Brysbaert et al. (2019) obtained word prevalence ratings by asking 220,000 diverse people (via a crowdsourcing study) to provide known/not-known decisions for over 61,000 English words. Word prevalence predicted additional variance in response times, relative to other metrics, particularly at the low end of the word frequency range.

An interesting aspect of word prevalence is the insight it affords into the words known by different groups of people. Brysbaert et al. (2019) provide tables of the words which are known by most British respondents but less known by Americans (and vice-versa). The American-known words include food items brought to the US from other cultures (*ziti, tamale, kabob, kielbasa*), holidays (*kwanza*), and animals native to North American (*crawdad, chigger*). The British-known words were inscrutable to this American writer but included *tippex, biro,* and *tombola*. Words known by most males but not females (and vice-versa) were also illuminating. Words known by men included military and martial arts terms (*bushido, katana, strafe*), and terms related to science (*azimuth, parsec, thermistor*) and computer technology (*servo, degauss, checksum*). Words mostly known by women included fabrics (*tulle, sateen, damask, chenille*) and clothing (*bandeau, espadrille*).

5 | FUTURE RESEARCH AND OPEN QUESTIONS

5.1 | Is it time to curtail studies on single-word processing?

Experiments on single words presented via a computer screen (and before computers were common in laboratories, via tachistoscope) has led to decades of theoretical advances due to the relative ease of this research and its reliability (i.e., no "replication crisis" for lexical decision as a paradigm). Researchers built on each other's work and were able to rapidly test new theoretical innovations, including those reviewed here: connectionist networks, different types of computational models, ERPs, and insights that connected single word reading to spacing effects in learning.

But humans usually do not usually read single words. Humans read for meaning, not to make word/nonword decisions. Here I review insights from eye movement research comparing word frequency to contextual diversity measures.

Fixation times and eye movement regressions display standard word frequency effects, but eye tracking also reveals a word frequency X predictability interaction (Ashby et al., 2005). The interaction is that word frequency effects are weaker in predictable contexts. Consider that in typical reading, predictable context is the norm. Word frequency will be less important because readers use context to make predictions about upcoming words. Recall the finding that words high in contextual diversity are processed more efficiently than are words occurring in specific contexts. But "processed more efficiently" means processed efficiently as isolated stimuli. This suggests that words encountered in diverse contexts evolve to have relatively context-free representations. Words encountered in specific contexts are processed most easily when they are encountered in their typical contexts.

Consistent with this is research comparing the correlations between eye movement latencies, lexical decision times, and naming times (Kuperman et al., 2013). Kuperman et al. (2013) noted that the predictability of words in context plausibly plays a powerful role in reading natural text, which they investigated as follows. After parsing out the effects of word frequency and word length, lexical decision response time and naming times shared between 7% and 44% of the unique variance in reaction time, depending on the frequency range of the words used. This suggests considerable similarity in these lexical decision and naming tasks. Eye movement latencies share between 5% and 17% of the variance with lexical decision times, but only when target words are presented in neutral sentences. When target words were drawn from corpus studies in which eye movements to all words were analyzed, the amount of shared variance was only 0.2%. The implication is that very different processes occur for reading narrative text compared to single words.

Predictability in eye movements was linked to contextual diversity in a study by Yan et al. (2018). Like Kuperman et al. (2013), these authors also analyzed eye-tracking corpora. Crucially, they demonstrated that CD could predict eye movement patterns, but the correlation was strongest when the sentence had little constraining context, making it resemble single word reading. CD was not a good predictor of eye-movement data when constraining context was present. The authors concluded that words' lexical representations include information about words' typical contexts of occurrence. They wrote: "Our account unifies WF, CD, and predictability effects, framing CD and WF as proxies to the probability a word will be needed and highlighting the role of context as an important component of the generative process of word distributions" (p. 1209).

5.2 | Phrase frequency effects

Researchers have touted the importance of words' history of use for understanding their representation, but the interesting case of multiword expressions has not yet been addressed (although see Schmidtke et al., 2021). Research on the importance of context cannot ignore words that rarely occur outside of a familiar expression, such as *amok* (*run amok*), *paragon* (*paragon of virtue*), or *wedlock* (*born out of wedlock*).

When low CD words are presented in isolation for visual word recognition studies, readers aren't given the ideal retrieval cues to retrieve word identity. Retrieval times are thus slowed. High CD words are easier to retrieve without a specific context because their representations don't include being a member of a multiword utterance.

Including phrasal context in words' representations explains phrase frequency effects (Baayen, 2010; Christiansen & Arnon, 2017). When words that are part of familiar phrases are tested in lexical decision or naming studies with their phrase displayed, then the frequency of the whole phrase matters more than the frequency of the target word (Caldwell-Harris et al., 2012; Caldwell-Harris & Morris, 2008). This suggests that language processing may involve considerable retrieval of large chunks from memory, rather than primarily grammatically guided integration of single words (see discussion in Daelemans & Van den Bosch, 2005).

5.3 | What if researchers just want to control for word difficulty confounds…?

Psycholinguists need to match their language stimuli for word frequency because of the confound of greater recognition difficulty of low-frequency words. Should this advice be updated to include contextual diversity, semantic diversity, word prevalence, all of these, or some other measure? Unless researchers are specifically interested in comparing different measures against each other, it is probably sufficient to match on a single measure. This can be word prevalence or contextual diversity if those norms are available in the language of interest, or word frequency, as norms for word frequency now exist in many languages. But – research continues, so this advice could evolve.

6 | SUMMARY AND IMPLICATIONS OF A HISTORY OF EVOLVING IDEAS

6.1 | Summary

Across 20 years, the journey of how words repeat across texts has twisted and turned in useful and sometimes unexpected ways. Psycholinguists began with simple counts of occurrence (i.e., word frequency). With the advent of large electronic text corpora in the early years of the century, researchers were able to operationalize constructs like Contextual Diversity (or CD_{doc}, the number of different documents in which words occur). The benefit of contextually diverse words was connected to the psychological principle of likely need. Computer simulations deepened our understanding of contextual/semantic diversity in human word learning.

New ways to identify contexts were developed, from text windows of 4–32 words to Reddit conversations. These metrics were then tested and refined using large online databases of visual word recognition response times. Researchers defined new constructs like word burstiness and word prevalence. Defining what constitutes "context" is an ongoing question (Jones et al., 2017).

6.2 | Lessons for 21st-century psycholinguists

One lesson is for researchers to focus less on structure and more on process. Structure refers to attributes of words, such as words' frequency counts in corpora. Process is how words are used in context (Jones et al., 2017), and how words help readers make predictions about upcoming material in a text (e.g., the concept of expectancies and the principle of likely need, Anderson & Schooler, 1991). The backdrop to this shift in emphasis is the growing understanding that there are no frequency effects nor diversity effects in the brain. Variables such as word frequency, word prevalence, and contextual/textual/semantic diversity are variables measured in text corpora; they point to underlying processes that give rise to language learning and processing. Exploring and understanding these underlying processes is an on-going goal of psycholinguistic research.

For our lives beyond the psychology of language: What about the more general case of facts and information that seem to be fixed and "stored"? What about our "self"? Language resembles other psychological processes: context-dependent, dynamic, and constantly being modified about expectations of immediate use and future relevance.

ACKNOWLEDGEMENTS

I thank Marc Brysbaert, Greg Hollis, Matthew Mak, and Michael Jones for helpful comments on a previous version of the manuscript. I thank Alison Morris (deceased 2014) for years of fun word recognition research, always using word frequency, in the decade before the discovery of context diversity.

ORCID

Catherine L. Caldwell-Harris https://orcid.org/0000-0002-9830-1156

REFERENCES

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703–719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408.

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–461.

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

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: Large scale analysis of lexical processes. In J. S. Adelman (Ed.), *Visual word recognition*. (vol 1, pp. 90–115). Psychology Press.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–226.

Behrens, H., & Pfänder, S. (Eds.). (2016). *Experience counts: Frequency effects in language*. Walter de Gruyter.

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology*, *30*, 1240–1263.

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.

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.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458.

Butler, L. T., & Berry, D. C. (2004). Understanding the relationship between repetition priming and mere exposure. *British Journal of Psychology*, *95*(4), 467–487.

Caldwell-Harris, C. L., Berant, J. B., & Edelman, S. (2012). Measuring mental entrenchment of phrases with perceptual identification, familiarity ratings, and corpus frequency statistics. In S. T. Gries, & D. Divjak (Eds.), *Frequency effects in cognitive linguistics (vol. 1): Statistical effects in learnability, processing and change* (pp. 165–194). Walter De Gruyter.

Caldwell-Harris, C. L., & Morris, A. L. (2008). Fast pairs: A visual word recognition paradigm for measuring entrenchment, top-down effects, and subjective phenomenology. *Journal of Consciousness and Cognition*, *17*, 1063–1081.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*(5), 619–636.

Cevoli, B., Watkins, C., & Rastle, K. (2021). What is semantic diversity and why does it facilitate visual word recognition? *Behavior Research Methods*, *53*(1), 247–263.

Chen, Q., Huang, X., Bai, L., Xu, X., Yang, Y., & Tanenhaus, M. K. (2017). The effect of contextual diversity on eye movements in Chinese sentence reading. *Psychonomic Bulletin & Review*, *24*(2), 510–518.

Chen, Q., Zhao, G., Huang, X., Yang, Y., & Tanenhaus, M. K. (2017). The effect of character contextual diversity on eye movements in Chinese sentence reading. *Psychonomic Bulletin & Review*, *24*(6), 1971–1979.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in cognitive science*, *9*(3), 542–551.

Daelemans, W., & Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge University Press.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+). *Design, architecture, and linguistic insights. International journal of corpus linguistics*, *14*(2), 159–190.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*(4), 309–330.

Diesel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*, 108–127.

Divjak, D. (2019). *Frequency in language: Memory, attention and learning*. Cambridge University Press.

Divjak, D., & Caldwell-Harris, C. L. (2015). Frequency and entrenchment. In E. Dabrowska, & D. Divjak (Eds.), *Cognitive foundations of language* (pp. 53–74). Mouton De Gruyter.

Divjak, D., & Gries, S. T. (Eds.). (2012). *Frequency effects in language representation (vol. 244)*. Walter de Gruyter.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. (2016). The relationship between artificial and second language learning. *Cognitive Science*, *40*(4), 822–847.

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports*, *10*(1), 1–11.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Houghton-Mifflin.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2011). Semantic diversity accounts for the missing word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience*, *23*, 243–2446.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hollis, G. (2020). Delineating linguistic contexts, and the validity of context diversity as a measure of a word's contextual variability. *Journal of Memory and Language*, *114*, 104146.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *416*, 401–410.

Johns, B. T. (2021). Disentangling contextual diversity: Communicative need as a lexical organizer. *Psychological Review*, *128*(3), 525–557.

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*(4), 1214–1220.

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *Journal of the Acoustical Society of America*, *132*, EL74–EL80.

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. *Psychology of Learning and Motivation*, *67*, 239–283. Academic Press.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, *66*, 115–124.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, *66*(3), 563–580.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802–823.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. *Advances in Psychophysiology*, *3*, 139–187.

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.

Lovelace, E. A. (1988). On using norms for low-frequency words. *Bulletin of the Psychonomic Society*, *26*(5), 410–412.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208.

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 132–142.

Madlener, K. (2015). *Frequency effects in instructed second language acquisition*. Walter de Gruyter.

Mak, M. H., Hsiao, Y., & Nation, K. (2021). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language*, 118, 104203.

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*, 295–322.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Lawrence Erlbaum.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. Kolers (Ed.), *Processing of visible language* (pp. 259–268). Springer.

Murray, W., & Forster, K. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721–756.

Pagán, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science*, *43*(1), e12705.

Perea, M., Soares, A. P., & Comesaa, M. (2013). Contextual diversity is a main determinant of word-identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675.

Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 275–283.

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *The Journal of General Psychology*, *13*, 199–203.

Rosa, E., Tapia, J. L., & Perea, M. (2017). Contextual diversity facilitates learning new words in the classroom. *PLoS One*, *12*(6), e0179004. https://doi.org/10.1371/journal.pone.0179004

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94.

Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2021). CompLex: An eye-movement database of compound word reading in English. *Behavior Research Methods*, *53*(1), 59–77.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*(5), 238–242.

Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(3), 461–474.

Yan, S., Mollica, F., & Tanenhaus, M. K. (2018). A context constructivist account of contextual diversity. In C. Kalish, M. Rau, J. Zhu & T. Rogers (Eds.), *40th annual meeting of the cognitive science society* (pp. 1205–1210). Cognitive Science Society.

AUTHOR BIOGRAPHY

Catherine Caldwell-Harris, Associate Professor at Boston University, received her PhD from the University of California, San Diego. At BU she directs the Psycholinguistics Laboratory. She is also Director of Research at the Center of Mind and Culture, a Boston-based research institute. Dr. Caldwell-Harris' expertise includes Asperger's Syndrome, cross-cultural psychology, bilingualism, foreign language learning and immigration. Her interest in word and phrase frequency arose when doing repetition blindness experiments. She documented how illusory words can be created by inserting word fragments into brief RSVP displays. Another visual illusion involved reversals: If the words *house white* are shown in rapid succession, viewers reported the displayed phrase was *white house*. In the domain of bilingualism,

Dr. Caldwell-Harris is best known for showing that autonomic arousal is lower when listening to emotional phrases in a foreign language, compared to phrases in a native language. She has studied lying, joking, trolly dilemmas, and why American and Chinese cultures differ in expressing *I love you*. Her study of Russian immigrants to the U.S. documented how young children "Englishify" the household, while older immigrants' friendships maintain their native language. She recently constructed an agent based model of factors influencing language-learning outcomes for immigrants to the U.S.

How to cite this article: Caldwell-Harris, C. L. (2021). Frequency effects in reading are powerful – But is contextual diversity the more important variable? *Language & Linguistics Compass*, e12444. https://doi.org/10.1111/lnc3.12444