Northumbria Research Link

Both Semantic Diversity and Frequency Influence Children's Sentence Reading

Ascensión Pagán , Megan Bird , Yaling Hsiao & Kate Nation

Routledge
Taylor & Francis Group

OPEN ACCESS   Check for updates

Both Semantic Diversity and Frequency Influence Children's Sentence Reading

Ascensión Pagán[a,b], Megan Bird[a], Yaling Hsiao[a], and Kate Nation[a]

[a]Department of Experimental Psychology, University of Oxford, Oxford, UK; [b]Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK

ABSTRACT

Semantic diversity – a metric that captures variations in previous contextual experience with a word – influences children's lexical decision and reading aloud. We investigated the effects of semantic diversity and frequency on children's reading of words embedded in sentences, while eye movements were recorded. If semantic diversity and frequency reflect different aspects of experience that influence reading in different ways, they should show independent effects and perhaps even different processing signatures during reading. Forty-nine 9-year-olds read sentences containing high/low frequency and high/low diversity words, manipulated orthogonally. We observed main effects of both variables, with high frequency and high semantic diversity words being read more easily. These results show that variations in the amount and nature of contextual experience influence how easily words are processed during reading.

Many experiments have investigated how item-level factors influence reading and its development. A core item-level variable is frequency: the observation that high frequency words are generally easier to process than words lower in frequency. In this experiment, we draw a distinction between a classic perspective on the frequency effect – reflecting the *amount* of experience with a word – and an alternative view which sees the *nature* of experience as also critical. To do this, we monitored children's eye movements as they read sentences containing target words that varied in frequency and semantic diversity.

We begin by outlining what is meant by frequency and semantic diversity. Frequency is defined as the number of times a word appears in a corpus. Frequency effects are abundant in reading tasks, with higher frequency words being processed more easily by adults and children (e.g., Balota et al., 2007; Blythe, Liversedge, Joseph, White, & Rayner, 2009; Joseph, Nation, & Liversedge, 2013). While various theoretical accounts explain the frequency effect differently, they focus on the notion of repetition and accumulated experience with the word (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Reichle, Rayner, & Pollatsek, 1998).

Another theoretical perspective sees the linguistic contexts in which words have been experienced as critical, informed by the finding that variability in experience influences how easily words are identified, beyond frequency (for review, Jones, Dye, & Johns, 2017). Contextual variability has been captured in different ways. Adelman, Brown, and Quesada (2006) found that contextual diversity indexed by document count (the number of unique documents a word appears in across a corpus) predicted unique variance in lexical decision and naming. McDonald and Shillcock (2001) captured

CONTACT Ascensión Pagán ✉ appc1@leicester.ac.uk   Department of Neuroscience, Psychology and Behaviour, College of Life Sciences, University of Leicester, George Davies Centre, Lancaster Road, Leicester LE1 7HA, UK; Kate Nation ✉ kate.nation@psy.ox.ac.uk   Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

Supplemental data for this article can be accessed here

the local lexical environment of a word via its co-occurrence with other words in an adjacent window across a large corpus. Termed contextual distinctiveness, this was a better predictor of lexical decision latencies than frequency. To capture the content of contexts more fully, Hoffman, Ralph, and Rogers (2013) used latent semantic analysis to derive *semantic diversity* (for a related approach, Jones, Johns, & Recchia, 2012). A word's semantic diversity value corresponds to the mean distance in multi-dimensional space between all the contexts it appears in across a corpus. An example of a low diversity word is *spinach* – this tends to occur in a limited range of contexts, all relating to food (Hoffman & Woollams, 2015). In contrast, a high diversity word like *chance* appears in a range of different contexts. Put simply, *spinach* provides a reasonable clue as to the content of a context whereas *chance* does not. Contextual knowledge builds over time as each encounter a person has with a word adds to their database of information about that word. Given greater contextual variation, words high in semantic diversity are likely to map to a range of multiple or nuanced meanings, based on the notion that variation in the meaning of a word is an emergent property of variation in the context in which it is used. Adults find high semantic diversity words easier to identify (e.g., Hoffman et al., 2013; Hoffman & Woollams, 2015; Jones et al., 2017, 2012), as do children (Hsiao & Nation, 2018). This facilitative effect cannot be explained by other important predictors such as frequency, document count and age-of-acquisition. Instead, it shows that variation in the content of the contexts a word has been experienced in previously influences the ease with which it is subsequently identified. As a result, models of lexical processing need to consider learning and the environment in which a word is learned. Accordingly, the Semantic Distinctiveness Model (Jones et al., 2017) assumes that each exposure to a word provides an opportunity for its lexical representation to be updated. If a word appears in a different context, new information is encoded as it is updated. Over time, words occurring in varying contexts with varying contents become more context-independent than words appearing in more similar contexts, allowing them to be identified more easily.

We used an eye movement paradigm to chart the time course of semantic diversity effects during silent sentence reading. By monitoring the pattern of fixations and saccades as people read, the eye movement record can show whether a particular variable influences "early" (word identification) or "later" (meaning integration) stages of processing (Liversedge, Paterson, & Pickering, 1998). Building on a wealth of evidence documenting skilled reading (Rayner, 2009), it is clear that different aspects of eye movement behavior reflect different cognitive processes in children's reading too (e.g., Blythe & Joseph, 2011). Thus, monitoring eye movements as children read words that vary in semantic diversity should inform when and where in processing semantic diversity exerts its influence.

Frequency effects are seen early in processing in children's reading with shorter fixation durations on high frequency words than low frequency words (Blythe et al., 2009; Joseph et al., 2013). Semantic diversity has not been investigated directly using eye movements. However, Plummer, Perea, and Rayner (2014) monitored eye movements as adults read sentences containing words varying in contextual diversity (instantiated as document count) and frequency. High document count words were fixated for less time, even when frequency was controlled, and document count subsumed any effect of frequency. They also reported a post-hoc analysis to explore the effect of semantic diversity on a subset of items. In contrast to document count, semantic diversity was not associated with early processing measures. However, high semantic diversity was associated with *slower* processing on later measures (go-past and total time), consistent with variations in semantic diversity influencing word-to-text integration, rather than word identification. These preliminary findings support the view that semantic diversity is distinct from both frequency and document count in terms of its processing signature.

In this experiment, we examined the time course of frequency and semantic diversity in children. We embedded words that varied orthogonally on the two constructs into the same neutral sentences. This allowed us to investigate their effects on word processing in natural reading, while controlling for any effects of sentential context. Consistent with previous work, we predicted that frequency effects would be seen early in processing. Our prediction for semantic diversity was less clear: If it

influences word identification, it should affect early processing; alternatively, if it reflects discourse-level processing, its influence should emerge later in processing.

Method

Participants

Fifty-three Year 5 children, recruited from two primary schools in Oxford took part in this experiment. First, children's legal guardians gave their informed consent and children gave their assent prior to their inclusion in the study. The Test of Word Reading Efficiency (Torgesen, Rashotte, & Wagner, 1999) was used to exclude any child with below average reading (standard score <85). Three children were excluded on this basis, and one for problems with calibration. The final sample comprised 49 children (M = 9.5 years, SD = 0.5) with normal or corrected-to-normal vision and no history of reading difficulties, and average-to-good reading (M = 115, SD = 13; range = 90–141). This experiment was approved by Oxford University's Research Ethics Committee. Therefore, this study conforms to the recognized standards of Declaration of Helsinki.

Materials

We selected 160 words (4–9 letters, M = 6.7, SD = 1.1) based on their frequency (zipf; log10 (frequency per billion words) and semantic diversity values, taken from the Oxford Children's Corpus (see Hsiao & Nation, 2018). Diversity and frequency were manipulated orthogonally, with 40 words in each condition. In addition, forty sentence-frames were constructed, each with a slot for a target word. As shown in Table 1, the same sentence frame was used to present four different words, one from each condition (see Supplementary Materials). Target word length was matched across each quartet, allowing for a maximum of one letter difference.

To confirm that children of this age know the target words, 61 9–10 year-olds (different from the main experiment) participated in a screening check. We created four counterbalanced lists with a target word in one of the four sentence frames. Approximately 15 children per list rated each sentence on a scale of 1 (easy-to-understand) to 3 (difficult-to-understand). The sentences were considered easy (M = 1.1, range = 1–2) with no difference across conditions (p > .1). To check the predictability of the sentences, we presented the 40 sentence frames up to the point of the target

Table 1. Frequency and semantic diversity values for target words in each condition. Also shown is an example sentence and performance (% correct) on the comprehension question by condition.

			Frequency			Semantic Diversity		
Condition	Example	Comprehension Questions Accuracy Mean (SD)	Mean	SD	Range	Mean	SD	Range
HF/HSD	The family looked for the missing *camera* before their friends arrived.	94 (15)	5.0	0.4	4.4–6.2	2.1	0.1	2.0–2.3
HF/LSD	The family looked for the missing *candle* before their friends arrived.	95 (14)	5.0	0.4	4.3–5.8	1.9	0.1	1.7–2.0
LF/HSD	The family looked for the missing *heater* before their friends arrived.	94 (15)	3.9	0.2	3.6–4.2	2.1	0.1	2.0–2.3
LF/LSD	The family looked for the missing *stereo* before their friends arrived.	96 (12)	3.9	0.2	3.6–4.2	1.9	0.1	1.3–2.0

Note. Embedded target words are in shown in italics (not in the experiment). See supplementary materials for a full list of sentences and target words. T-test comparisons revealed statistically significant differences in frequency between high and low frequency target words (t = 20.15, p < .001) and in semantic diversity for the high and low semantic diversity target words (t = 13.49, p < .001). There was no difference in frequency for the two sets of high frequency words (t = 0.64, p = .53) or the two low frequency sets (t = 0.11, p = .91). Similarly, the two high semantic diversity sets were equivalent in diversity (t = 1.50, p = .14), as were the two low semantic diversity sets (t = 0.88, p = .38). Note that frequency is calculated using Zipf values (the log10(frequency per billion words)).

word and asked 14 additional 9–10 year-olds to write down the word that should come next, emphasizing that the word did not need to complete the sentence. The sentence frames did not predict the target word and there was no difference across conditions ($ts<0.16$).

The four sets of 40 sentences were counterbalanced across four lists using a Latin Square design. Each list included 40 experimental sentences (10 per condition) and two practice sentences. Each sentence occupied one line on the screen (averaging 12 words; range: 8–15) and the target word appeared in the middle of the sentence. To promote reading for meaning, yes/no comprehension questions (not referring to the target) appeared after 50% of sentences, balanced across conditions.

Given the demands of finding words that fitted the semantic diversity x frequency manipulation while controlling for length, and permitted each word in a quartet to fit into the same sentential context without changing predictability, we did not attempt to match on other lexical properties. Semantic diversity is only moderately correlated with other key predictors of word-level reading (Hsiao & Nation, 2018) and both sets of words were equivalent on a range properties, including imageability, number of senses and concreteness (Table 2), but not age of acquisition. We discuss this point in the General Discussion.

Apparatus and procedure

Eye movements were recorded using an SR Research Eyelink 1000 tracker (SR Research Ldt., Ontario, Canada; spatial resolution 0.05°, sampling rate 1,000 Hz). The sentences were presented on a 15" Dell monitor set at a refresh rate of 60 Hz with a 1,024 × 768 resolution and viewing distance of 60 cm. We used a lowercase black font (Courier New, size 12) on a gray background; three characters subtended 1° of visual angle. Reading was binocular but only the right eye was recorded. Forehead and chin rests minimized head movements. To calibrate, children looked at three horizontal fixation points. Calibration accuracy did not exceed 0.25. Children were instructed to read sentences silently. After each sentence, they pressed a button on a games controller to terminate the trial and to respond to the comprehension questions. Following two practice trials, the experimental sentences were presented randomly. The experiment lasted less than 30 minutes.

Results

Accuracy on the comprehension questions was high, with all children scoring at least 75% (M = 89%, SD = 5%). This indicates that children were reading for comprehension and there was no difference across conditions (Table 1).

Pre-analysis, the "clean" function in DataViewer (SR Research) trimmed the eye movement data. Fixations shorter than 80 ms and longer than 1200 ms were deleted. Fixations shorter than

Table 2. Additional lexical properties associated with the high vs. low semantic diversity target words and t-statistics testing group differences.

	High semantic diversity			Low semantic diversity			
	N	M	SD	N	M	SD	t
N letters	80	6.70	1.09	80	6.23	1.14	0.42
Document count[1]	80	2.42	0.62	80	2.32	0.53	1.07
Age of Acquisition[2]	79	6.19	1.68	71	7.21	2.22	− 3.18*
Imageability[3]	30	6.29	0.82	26	6.06	0.81	1.03
Concreteness[4]	80	4.13	0.99	77	4.17	0.96	− 0.22
Number of senses[5]	72	3.85	3.92	69	3.23	2.71	1.09

Notes. N varies depending on whether word appears in norms, max N per condition = 80. Norms taken from: [1]Oxford Children's Corpus, [2]Kuperman et al. (2012), [3]Schock, Cortese, and Khanna (2012), [4]Brysbaert, Warriner, and Kuperman (2014), [5]Wordsmyth Dictionary. *p < .01. Note that document count scores are log-transformed.

80 ms which were located within one character space of the next or previous fixation were merged into that nearby fixation. The data were log-transformed to reduce skew (Baayen, Davidson, & Bates, 2008).

We analyzed the data using linear mixed effects (lme) modeling with the lmer function from the lme4 package (Bates, Maechler, & Bolker, 2012) within R (R Core Team, 2013). For each dependent variable, we built an lme model with frequency and semantic diversity as categorical variables and specified as fixed factors, using the "contr.sdif" (MASS package) function. Contrasts were specified as 0.5/-0.5 for low/high frequency and low/high semantic diversity, such that the intercept corresponded to the grand mean and the fixed effects corresponded to the main effect of the fixed factors. Both participants and items were specified as random factors. A full random structure for subjects and items was specified to avoid being anti-conservative (Barr, Levy, Scheepers, & Tily, 2013). Significance values and standard errors reflect both participant and item variability (Baayen et al., 2008) and following convention, $t > 2$ was considered significant.

Our dependent variables included "early" measures: first fixation duration (initial first-pass fixation on a word) and gaze duration (sum of all consecutive first-pass fixations on a word before leaving it) and "late" measures on the target word: go past time (time from the initial fixation on a word until the eyes move onward) and total time (sum of the duration of all the fixations on a target word, including regressions). Raw data and analysis code are available via the Open Science Framework (https://osf.io/j9f72/) and as Supplementary Materials.

Figure 1 shows descriptive data. As summarized in Table 3, there was a main effect of frequency for all the dependent measures except first fixation duration, with low frequency words receiving longer fixations and slower reading times than high frequency words. There was also a main effect of semantic diversity with longer fixations and slower reading times for low diversity words compared to high diversity words. As for frequency, there was no effect on first fixation duration, but robust effects in gaze duration and go past time; the effect was marginal for total time[1]. There was no interaction between frequency and semantic diversity on any measure ($ts < .3$).
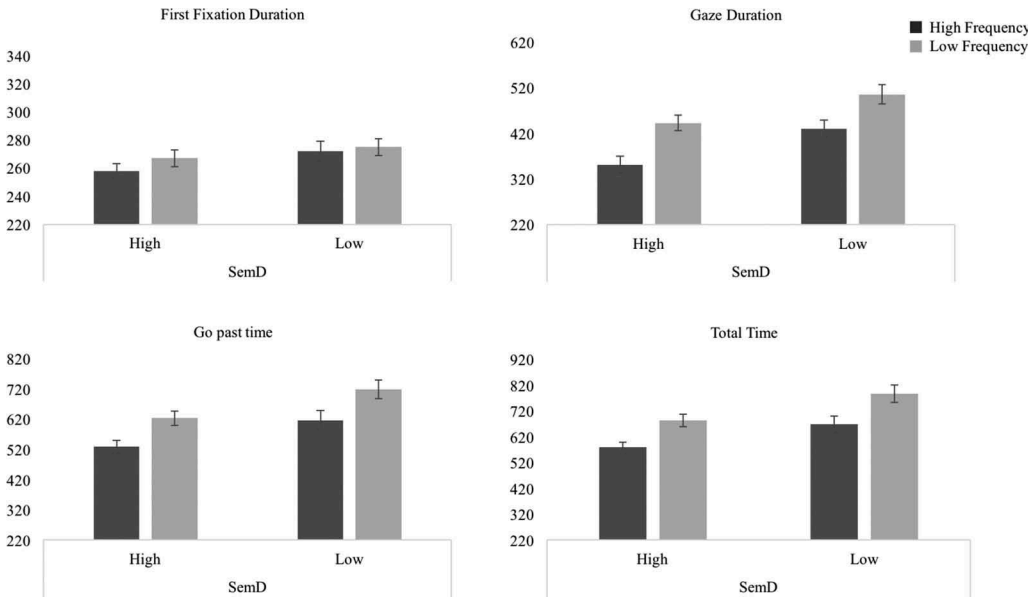

Figure 1. Means (ms) and standard errors (Error bars) for target word processing in each condition.

Table 3. Summary of Lme analyses on the target word for first fixation duration, gaze duration, go past time and total times.

	First Fixation Duration			Gaze Duration			Go past time			Total Time		
	b	SE	t	b	SE	t	b	SE	t	b	SE	t
Intercept (global mean)	5.49	0.02	220.5	5.83	0.05	115.3	6.17	0.07	91.8	6.26	0.06	104.2
FreqLow-High	0.04	0.02	1.7	0.17	0.03	**6.3**	0.20	0.03	**5.81**	0.18	0.04	**4.9**
SDLow-High	0.01	0.02	0.8	0.09	0.03	**3.1**	0.01	0.03	**2.54**	0.07	0.04	1.8
FreqLow-High:SDLow-High	−0.01	0.04	−0.2	0.00	0.06	0.04	0.02	0.07	0.27	0.01	0.09	0.2

Note: Lme model: lmer (ldepvar ~ Frequency*SemanticDiversity + (1 + Freq*SemanticDiversity|pp) + (1 + Freq*SemanticDiversity |items), data = datafile). Freq: lexical frequency; SD: semantic diversity. Following standard conventions, effects were considered significant when $t > 2$ (in bold).

General discussion

This experiment examined the influence of frequency and semantic diversity on children's reading of target words embedded in sentences. We used an orthogonal design to manipulate the two variables and kept the sentence context constant across conditions. To our knowledge, this is the first eye-movement study to investigate the effects of frequency and semantic diversity during silent sentence reading.

As predicted, children processed higher frequency words more easily than lower frequency words, replicating previous findings (Blythe et al., 2009; Joseph et al., 2013). We also found that frequency remains a reliable predictor of children's reading behavior, even when semantic diversity is controlled. Turning to semantic diversity, high diversity words were easier to process than low diversity words, even when frequency was controlled. These findings extend those of Hsiao and Nation (2018) to naturalistic reading and reveal that semantic diversity's processing signature is similar to (but separate from) the effect of frequency. This suggests that the contextual nature of previous experience with a word influences item-level differences in the ease of word identification.

Before reflecting on our findings further, it is important to note that any experiment that manipulates lexical variables in an orthogonal design is always at risk of attributing an effect to a particular variable when in fact the effect should be attributed to some other factor, correlated with the variable of interest but not controlled in the experiment. In our experiment, alongside the key manipulation that required items to vary in semantic diversity but not frequency (and vice versa), each quartet of targets needed to fit equally comfortably within the same sentential context and be matched for length – a key variable to control when using eye movement measures. Given these constraints, we did not attempt to match across conditions for the myriad of other variables that influence reading. Nevertheless, we need to consider whether the semantic diversity effect observed in our data can be "explained" by something else. Comparing the high vs. low semantic diversity words, there was no difference in document count, or in "classic" semantic variables including imageability, concreteness and number of senses (all $ps > .27$). There was however a significant difference in age of acquisition, with high diversity words being earlier acquired than low diversity words (6.17 vs. 7.21 years; note there was a comparable difference in age of acquisition for the high vs. low frequency words: 6.07 vs. 7.29 years). This is not too surprising. There is a modest correlation between the two variables (r = − .40, Hsiao & Nation, 2018) and contextual variability has a close relationship with the order in which words are acquired in infancy, with early acquired words having more semantic connections that those acquired later (Hills, Maouene, Riordan, & Smith, 2010). Hsiao and Nation (2018) used a continuous design to investigate the effect of semantic diversity and age of acquisition on lexical decision and reading aloud. They found that both variables influenced children's reading. This offers reassurance that the two variables are not one of the same, but clearly more research is needed to unpack their relationship. Ultimately, to fully understand semantic diversity requires a different type of experimental design – one that systematically brings about variation in semantic diversity as words are learned and examines the direct consequence of this for subsequent reading behavior (see Johns, Dye, & Jones, 2016; Joseph & Nation, 2018; Pagán & Nation, 2019; Rosa, Tapia, & Perea, 2017). In the meantime, however, our experiment is clear in

showing that semantic diversity influences children's eye movements while reading in a way that cannot be readily explained by frequency, document count, word length, imageability, concreteness and number of senses.

These facilitative effects of high semantic diversity contrast with Plummer et al.'s (2014) findings. In a post-hoc supplemental analysis on a subset of words, they observed an effect of semantic diversity on later processing measures only (go past time and total time), suggesting that it might influence word-to-text integration rather than word identification. Moreover, reading times slowed as diversity increased – the opposite pattern of results to those observed here. As Plummer *et al.* acknowledged, their finding must be interpreted with caution as semantic diversity was not the focus of their experiment: their items were not selected or controlled with semantic diversity in mind. Another important difference concerns the sentential contexts in which words appeared in different conditions. While Plummer *et al.* built a different sentence frame for each word in each condition, we used the same sentence frame across the four conditions. This provided us with excellent stimulus control. Plummer *et al.* tested undergraduate students whereas our participants were children, but further investigation is needed before speculating on potential developmental differences.

Our findings indicate that the nature of previous experience with words, not just the amount of experience, shapes the development of lexical representations. Words high in semantic diversity map to a range of multiple or nuanced meanings, based on the notion that variation in the meaning of a word is an emergent property of variation in the context in which it is used. Accordingly, words that are ambiguous or polysemous tend to be higher in semantic diversity (see Hoffman et al., 2013; Hsiao & Nation, 2018). Importantly however, semantic diversity captures shades of meaning based on contextual usage in a way that is continuous and graded: even words deemed to be unambiguous show variation in semantic diversity (for further discussion, see Hoffman et al., 2013). Relevant to our findings is the notion that one consequence of this greater contextual variation is that a word's representation becomes more context independent over time, meaning that when a high diversity word is experienced in a new context (or out of context), it is easier to identify, even when frequency is matched (Hsiao & Nation, 2018). This fits with the Semantic Distinctiveness Model (Jones et al., 2017) in that contextual variation during experience provides more opportunity for updating, leading to easier identification on subsequent encounters, and with theoretical accounts of the polysemy processing advantage (e.g., Rodd, Gaskell, & Marslen-Wilson, 2004). In line with this, we found shorter fixation durations on high semantic diversity words.

It is also important to consider how semantic diversity influences processes involved in text comprehension and meaning integration. Arguably, what brings a processing advantage for word identification (i.e., context independence) might lead to processing costs when context dependence is desired (i.e., when building a coherent text representation). In our experiment however, semantic diversity behaved similarly in earlier and later processing, with an advantage for words high in diversity throughout. Notably, our sentences were neutral and unpredictable: it is likely that different levels of contextual constraint serve to moderate whether a word's semantic diversity eases or intensifies processing effort, as revealed by later measures tapping word-to-text integration. Future work is needed to investigate how the semantic diversity of individual words impacts on processing when context and predictability are allowed to vary more freely[2].

In conclusion, this experiment extends previous studies by teasing apart the time-course of the effects of frequency and semantic diversity on lexical processing during silent sentence reading. We found that high semantic diversity, like high frequency, was associated with easier processing when reading words in neutral sentences. Our findings call for theoretical accounts of reading that accommodate both learning and contextual variation during experience.

Notes

1. Lme model on the probability of making regressions to the target word showed no differences across conditions (ps > 0.4), indicating that total time was not influenced by the percentage of regressions.
2. Raw data files for the entire sentence are posted to the Open Science Framework (https://osf.io/j9f72/).

Acknowledgments

Conflict of Interest

No potential conflict of interest was reported by the authors.

Funding

References

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. Retrieved from http://CRAN. R-project. org/package= lme4

Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford library of psychology. The Oxford handbook of eye movements* (pp. 643–662). New York, NY: Oxford University Press.

Blythe, H. I., Liversedge, S. P., Joseph, H. S., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*(12), 1583–1591. doi:10.1016/j.visres.2009.03.015

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. doi:10.3758/s13428-013-0403-5

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204. doi:10.1037/0033-295X.108.1.204

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language. *Contextual Diversity in Early Word Learning*, *63*, 259–273. doi:10.1016/j.jml.2010.06.002

Hoffman, P., Ralph, M. A. L., & 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. doi:10.3758/s13428-012-0278-x

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 385.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal Memory and Language*, *103*, 114–126. doi:10.1016/j.jml.2018.08.005

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*(4), 1214–1220. doi:10.3758/s13423-015-0980-7

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 67, pp. 239–283). Amsterdam: Elsevier.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal Experimental Psychology*, *66*(2), 115. doi:10.1037/a0026727

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211.

Joseph, H. S., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychological Review*, *42*(2), 207–223.

364 A. PAGÁN ET AL.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990. doi:10.3758/s13428-012-0210-4

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford, England: Elsevier Science Ltd.

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(3), 295–322. doi:10.1177/00238309010440030101

Pagán, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science*, 43(1), e12705. doi:10.1111/cogs.12705

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(1), 275.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157.

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. doi:10.1016/j.cogsci.2003.08.002

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

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, 44, 374–379. doi:10.3758/s13428-011-0162-0

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. Austin, TX: Pro-ed.