

The ERP signature of the contextual diversity effect in visual word recognition

Marta Vergara-Martínez¹ · Montserrat Comesaña² · Manuel Perea^{1,3}

Published online: 3 January 2017
© Psychonomic Society, Inc. 2016

Abstract Behavioral experiments have revealed that words appearing in many different contexts are responded to faster than words that appear in few contexts. Although this contextual diversity (CD) effect has been found to be stronger than the word-frequency (WF) effect, it is a matter of debate whether the facilitative effects of CD and WF reflect the same underlying mechanisms. The analysis of the electrophysiological correlates of CD may shed some light on this issue. This experiment is the first to examine the ERPs to high- and low-CD words when WF is controlled for. Results revealed that while high-CD words produced faster responses than low-CD words, their ERPs showed larger negativities (225–325 ms) than low-CD words. This result goes in the opposite direction of the ERP WF effect (high-frequency words elicit smaller N400 amplitudes than low-frequency words). The direction and scalp distribution of the CD effect resembled the ERP effects associated with “semantic richness.” Thus, while apparently related, CD and WF originate from different sources during the access of lexical-semantic representations.

Keywords Lexical organization · Contextual diversity · Word recognition · ERPs

✉ Marta Vergara-Martínez
Marta.Vergara@uv.es

¹ Estructura de Recerca Interdisciplinària de Lectura (ERI-Lectura), Universitat de València, València, Spain

² Human Cognition Lab, CIPsi, School of Psychology, University of Minho, Braga, Portugal

³ BCBL: Basque Center on Cognition, Brain, and Language, Donostia-San Sebastián, Spain

One of the most replicated findings in the literature on visual-word recognition is that word identification times are faster (and more accurate) for high-frequency words than for low-frequency words (see Forster & Chambers, 1973; Preston, 1935; Rubenstein, Garfield, & Millikan 1970; Solomon & Postman, 1952, for early evidence). Similarly, during normal reading, fixation durations are shorter for high-frequency words than for low-frequency words (e.g., Inhoff & Rayner, 1986; Rayner & Duffy, 1986). For decades, word-frequency (WF; i.e., the number of times a word appears in a lexical database) has been considered the most important lexical factor in visual-word recognition and reading, and it plays a pivotal role in all computational models of visual-word recognition (e.g., the resting level of activation of word units in interactive activation models is a function of word-frequency; see Davis, 2010; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981) as well as in all leading computational models of eye movement control during reading (e.g., EZ-Reader model: Reichle, Pollatsek, Fisher, & Rayner 1998; SWIFT model: Engbert, Nuthmann, Richter, & Kliegl, 2005).

In an influential study, Adelman, Brown, and Quesada (2006) reported that “contextual diversity” (CD), which was defined as the proportion of contexts (documents) in which a word appears in a lexical database, was a better predictor of word identification times than WF in two widely used behavioral tasks (lexical decision and naming). In the past years, the effect of CD has received increasing attention in the field of word recognition. The basic finding is that the higher the number of contexts in which a word appears, the faster the word identification times (see also Cai & Brysbaert, 2010; Dimitropoulou, Duñabeitia, Avilés, Corral, & Carreiras, 2010; Perea, Soares, & Comesaña, 2013; Soares et al., 2015, for converging evidence). This effect is not restricted to single word identification tasks. During sentence reading, fixation durations are shorter for

higher CD words than for lower CD words matched in WF (Plummer, Perea, & Rayner, 2014).

A fundamental and unanswered issue is to clarify the nature of the CD effect. In this experiment we aim to address this question by using a highly sensitive experimental tool: the recording and analysis of the event-related potentials (ERPs). As CD and WF tend to be correlated (i.e., high-frequency words tend to be words that appear in many contexts and vice versa), one might argue that they essentially reflect the same underlying structural processes: Each exposure to a word will influence its accessibility, allowing it to be processed more quickly. (Note that WF stands for the number or raw frequencies, whereas CD filters out repeated encounters of the word in the same documents.) This interpretation would have little implications for models of visual-word recognition and reading. As indicated by Plummer et al. (2014), “models could easily substitute word-frequency with contextual diversity without any serious theoretical implications” (p. 280). Alternatively, one might argue that CD effects have a semantic origin. Adelman et al. (2006) indicated that “whereas WF is subject to effects of structural variables, CD seems more likely to be influenced by semantic variables” (p. 816) and “temporal, as well as semantic aspects of context, contribute to the CD effect” (p. 822). In latent semantic analysis (LSA; Landauer, 2001), a psychological model intended to explain the learning and representation of words and other sources of knowledge, the meaning of a word is conceptualized as “an irreversible mathematical melding of the meanings of all the contexts in which it has been encountered” (Landauer, 2001, p. 1). Within this framework, two uses of the same word are never identical in meaning, as their precise connotation in each case depends on the immediate linguistic and environmental context. Therefore, CD may as well have a crucial impact on the way meaning is built for that particular word. In this line, Hoffman, Lambon Ralph, and Rogers (2013) claimed that words that appear in a wide range of diverse contexts might be more variable in meaning than the words that appear in a restricted set of contexts. In other words, higher CD words could be semantically richer than lower CD words.

How can we tease apart the “lexical/structural” versus “semantic” accounts of the CD effect? Word-recognition experiments that only collect behavioral data cannot be used to disentangle the two explanations proposed for the CD effect since both lexical/structural and semantic manipulations produce facilitative effects. That is, high-frequency words yield shorter response times than low-frequency words in word recognition experiments. Likewise, semantically richer words produce shorter response times than semantically poorer words (number of semantic features, number of semantic associates: Buchanan, Westbury, & Burgess, 2001; Duñabeitia, Avilés, & Carreiras, 2008; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Pexman, Lupker, & Hino, 2002;

Rabovsky, Sommer, & Abdel Rahman, 2012; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012; number of senses/meanings: Borowsky & Masson, 1996; Rodd, Gaskell, & Marslen-Wilson, 2002; Rodd, 2004; Woollams, 2005; Yap, Tan, Pexman, & Hargreaves, 2011; concreteness: Kanske & Kotz, 2007; Kounios & Holcomb, 1994; Schwanenflugel, 1991, but see Barber, Otten, Kousta, & Vigliocco, 2013). Thus, a facilitative effect of CD in the word identification times can be readily accommodated by the two accounts.

The ERPs have the potential to provide a critical measure of neural processing (time course, amplitude, and scalp distribution) related to the underlying cognitive processes of the CD effect. Many studies have investigated the temporal dynamics of lexical and semantic influences during word recognition, mainly focusing on the N400 component. The N400 is a negative deflection starting around 200 ms and reaching its peak amplitude around 400 ms after stimulus onset, which is maximal over centro-parietal electrode sites. For words presented in isolation, the N400 has been associated with lexical-semantic processing and the modulation of its amplitude reflects processing costs during the retrieval of properties associated with a word form stored in memory (Holcomb, Grainger, & O’Rourke, 2002; Kutas & Federmeier, 2000). In this line, the amplitude of the N400 component is modulated by WF: low-frequency words elicit larger N400 amplitudes than high-frequency words (Barber, Vergara, & Carreiras, 2004; Smith & Halgren, 1987; Van Petten & Kutas, 1990; Vergara-Martínez & Swaab, 2012; Vergara-Martínez, Perea, Gómez, & Swaab, 2013). Although the N400 effects are often characterized in the 300–500-ms time window, it is not rare to observe WF effects in earlier time windows (e.g., see Hauk & Pulvermüller, 2004; Hauk, Davis, Ford, Pulvermüller & Marslen-Wilson, 2006).

Crucially, the amplitude of the N400 is also modulated by semantic factors (e.g., concreteness, number of associates, number of semantic features), but in the *opposite* direction to that of the WF effect in word recognition experiments. Larger N400 amplitudes have been found for concrete than for abstract words (Barber et al., 2013; Holcomb, Kounios, Anderson & West, 1999; Kanske & Kotz, 2007; West & Holcomb, 2000). Larger N400 amplitudes have also been reported for words with many semantic features or associates than for those with few semantic features or associates (Amsel, 2011; Laszlo & Federmeier, 2011; Müller, Duñabeitia, & Carreiras, 2010; Rabovsky et al., 2012; but see Amsel & Cree, 2013; Kounios et al., 2009, for an opposite pattern due to explicit semantic task demands). ERP effects related to semantic richness have been found to be distributed over anterior scalp electrodes (concreteness effects: Adorni & Proverbio, 2012; Barber et al., 2013; Holcomb et al., 1999; Kanske & Kotz, 2007; West & Holcomb, 2000; semantic richness: Amsel, 2011; Müller et al., 2010; but see Rabovsky et al., 2012, for centro-

parietal localization of semantic richness effects). As occurs with the WF effect, ERP effects of “semantic richness” are not necessarily confined to the classic N400 interval (300–500 ms), but they have been found to peak at earlier latencies (Amsel, 2011; Rabovsky et al., 2012).

In sum, prior ERP experiments have revealed a dissociation between lexical/structural versus semantic factors in the N400 component: while low-frequency words produce more negativity than high-frequency words, words that are richer in semantic factors (e.g., concreteness, number of associates) produce more negativity than words with less semantic richness—note that both WF and the measures related to “semantic richness” are facilitative in behavioral and eye-tracking experiments. This experiment makes use of the dissociation regarding the N400 component to examine whether the facilitative CD effect is driven by lexical/structural or by semantic processes. We measured the ERPs during a lexical decision experiment (i.e., the most common laboratory word recognition task) with words that varied in the number of contexts they appeared in (high-CD vs. low-CD words) while WF and other psycholinguistic characteristics were controlled for. The predictions are clear-cut. Larger negativities for low CD than for high CD words, along with a centro-parietal distribution (in line with the canonical N400 WF effect) would favor a “lexical/structural” interpretation of the CD effect (i.e., CD would just be another signature of word frequency). Alternatively, larger negativities for high-CD than for low-CD words would favor a “semantic” interpretation of the CD effect. Furthermore, we scrutinized the ERP segments to better characterize the CD effect—note that, although the time course of orthographic and lexical semantic effects in visual word recognition, as measured by the ERP technique, seem to converge on the 300–500 ms time window, the limits of the N400 are far from certain (see Laszlo & Federmeier, 2014, for a metareview of the time course of orthographic, lexical, and semantic factors during visual word recognition). Finally, for comparison purposes, we also measured the ERPs for a set of words that only differed in WF with the experimental low-CD words. This enabled us to compare the effect of CD with the more canonical WF effect with the same participants.

Method

Participants

Twenty-three undergraduate and graduate students of the University of Valencia (14 women) participated in the experiment in exchange for a small gift. All of them were native Spanish speakers with no history of neurological or psychiatric impairment, and with normal (or corrected-to-normal) vision. Ages ranged from 18 to 40 years ($M_{\text{age}} = 26$ years, $SD = 5.9$). All participants were right-handed, as assessed with a

Spanish abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971).

Materials

We selected 70 Spanish words from the EsPal subtitle database (Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013). This database provides not only the token account of each word (i.e., word frequency) but also the proportion of films in which a word appears. As in previous research, the CD variable was operationalized as the proportion of films (documents) in which a word appears (see also Soares et al., 2015). There were 35 high-CD words (i.e., words that occur in a high percentage of films) and 35 low-CD words (words that occur in a low percentage of films). To establish two differentiated word groups regarding CD, the words were selected from a range of high-frequency values. The two conditions only differed significantly in CD ($p < .0001$) and were carefully matched for a number of sublexical, lexical, and semantic variables (see Table 1). To perform a follow-up analysis on word frequency, a second group of 35 low-

Table 1 Mean values for sublexical, lexical, and semantic characteristics of the stimuli (obtained from the Espal database; Duchon et al., 2013)

	CD comparison			
	Range	High CD	← <i>p</i> →	Low CD
Contextual diversity ^a	0–100	41.8 (4.6)	5.3 ⁻¹⁶	27.7 (4.4)
Subtitle-frequency ^b	0–33,170	155.1 (5.2)	.07	139.3 (5.1)
imageability	1.2–7	4.9	.12	5.4
Concreteness	1.9–6.8	4.7	.15	5.1
Age of acquisition ^c	1–11	4.6	.14	5.2
Number of letters	4–8	5.5	.52	5.7
Levenshtein distance	1–12	1.5	.30	1.6
Mean positional bigram frequency (token)	0–58,827	6584	.29	5535
Most frequent lexical category		Noun = 29		Noun = 29
		Verb = 5		Verb = 5
		Adjective = 1		Adjective = 1

Note. The results of the pairwise comparisons between conditions are indicated in *p* values

^a Contextual diversity is defined as the percentage of films containing the word

^b Frequency per million

^c Extracted from the Alonso, Fernandez, and Diez (2015) norms

The Zipf values (in brackets) are the $\log_{10}(\text{freq})+3$. This logarithmic scale of frequency values allow for cross-linguistic studies and straightforward frequency ranking. For contextual diversity the range is 0.4–4.9, for subtitle-frequency the range is 0.3–7.5 (van Heuven, Mandera, Keuleers, & Brysbaert, 2014)

CD words was selected from a range of low-frequency values. Both word groups of low-CD were matched for a number of sublexical, lexical, and semantic variables (see Appendix A), so that the two conditions only differed significantly in WF ($p < .0001$). For the purposes of the lexical decision task, we also created 105 orthographically legal pseudowords (by replacing 2–5 letters from the original words, depending on their length) using Wuggy (Keuleers & Brysbaert, 2010). The list of words/pseudowords is presented in Appendix B.

Procedure

Participants were seated comfortably in a dimly lit and sound-attenuated chamber. All stimuli were presented on a high-resolution monitor positioned at eye level 80 cm in front of the participant. The stimuli were displayed in white lowercase Courier New 24-pt. font against a dark-gray background. Participants performed a lexical decision task: they had to decide as accurately and rapidly as possible whether or not the stimulus was a Spanish word. They pressed one of two response buttons (YES/NO). The hand used for each type of response was counterbalanced across subjects. Reaction times (RTs) were measured from target onset until the participant's response.

The sequence of events in each trial was as follows: A fixation cross (“+”) appeared in the center of the screen for 1,000 ms. This was followed by a 200-ms blank screen which, in turn, was replaced by a stimulus (word or pseudoword) in lowercase letters that remained on the screen for 400 ms. The trial finished when the participant responded or 1,500 ms had elapsed. A blank screen of random duration (range: 700–1,000 ms) was presented after the response. To minimize subject-generated artifacts in the EEG signal during the presentation of the experimental stimuli, participants were asked to refrain from blinking and eye moving from the onset of the fixation cross to the end of the trial. Each participant received the stimuli in a different random order. Sixteen warm-up trials, which were not further analyzed, were presented at the beginning of the session and were repeated if necessary. The whole experimental session lasted approximately 20 minutes.

EEG recording and ERP analyses The electroencephalogram (EEG) was recorded from 29 Ag/AgCl electrodes mounted in an elastic cap (EASYCAP GmbH, Herrsching,

Germany) according to the 10/20 system. These electrodes were referenced to the right mastoid and re-referenced off-line to the averaged signal from two electrodes placed on the left and right mastoids. Eye movements and blinks were monitored with electrodes placed on the right lower and upper orbital ridge and on the left

and right external canthi. The EEG recording was amplified and bandpass filtered between 0.01–100 Hz with a sample rate of 250 Hz by a BrainAmp (Brain Products, GmbH, Gilching, Germany) amplifier. An off-line bandpass filter between 0.01 and 20 Hz was applied to the EEG signal. Impedances were kept below 5 k Ω during the recording session. All single-trial waveforms were segmented and screened offline for amplifier blocking, drift, muscle artifacts, eye movements, and blinks. This was done for a 500-ms epoch with a 100-ms prestimulus baseline. Trials containing artifacts and/or trials with incorrect lexical decision responses were not included in the average ERPs or in the statistical analyses. These processes led to an average rejection rate of 9.2% of all trials (7.9% due to artifact rejection; 1.3% due to incorrect responses). A t test on the number of included trials per condition showed no difference between conditions, $t(22) = 1.4$, $p = .175$. ERPs were averaged separately for each of the experimental conditions, each of the subjects, and each of the electrode sites.

To characterize the CD effect in terms of the time course, polarity, and scalp distribution of its electrophysiological signature, the statistical analyses were performed on the mean voltage values between 225 and 325 ms, and on the full montage of 27 scalp electrodes. The selection of this time epoch was based on the results of running repeated-measures t tests at every 4-ms intervals between one and 500 ms at all 27 scalp sites for CD (high/low). To correct for multiple comparisons, we applied the following criterion: if a sequence of 15 consecutive t -test samples exceeded the .05 significance level, then an onset latency for a given experimental contrast was considered significant and reliable (see Guthrie & Buchwald, 1991; see Fig. 1). As a result, one time window of interest was identified: 225–325 ms. The full set of 27 electrodes was included in the analyses by dividing the electrode montage into seven separate parasagittal columns along the anterior–posterior axis of the head (see Fig. 2; Massol, Midgley, Holcomb, & Grainger, 2011). We performed four separate repeated-measures analyses of variance (ANOVAs), one on each of the three pairs of lateral columns and on the midline column. The lateral column analyses (referred to as Col. 1, Col. 2, Col. 3, extending outward) included the factor anterior–posterior (AP) over dorsal electrode sites (three, four, or five levels) and the factor hemisphere (HEM) over rostral electrode sites. The midline column analysis only included the AP factor with three levels. In sum, the analyses of variance (ANOVAs) included the factors CD, AP, and HEM (on three pairs of columns; CD and AP on the midline column). Effects for the AP and HEM factors are reported when they interact with the experimental manipulation. Interactions between factors were followed up with simple effect tests.

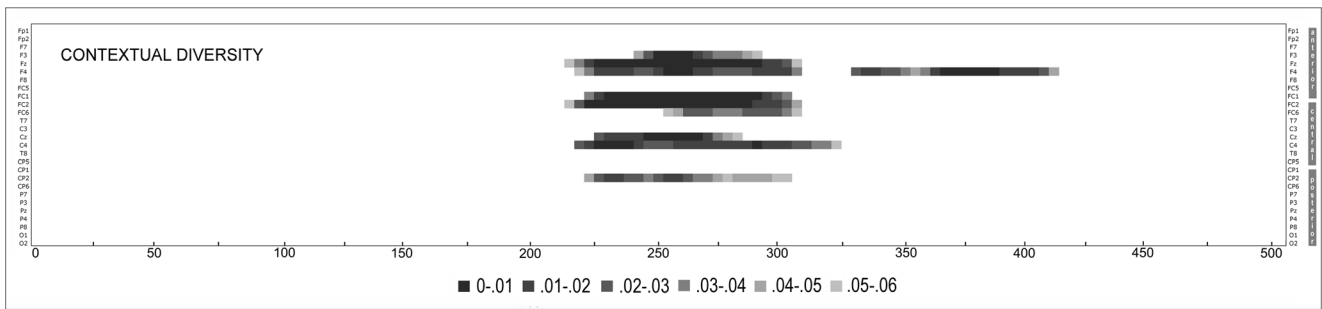


Fig. 1 Results of the univariate statistical analyses of the time course of contextual diversity. The plots convey the results of repeated-measures *t* tests at every 4 ms interval between zero and 500 ms at all 27 scalp sites

(listed in an anterior-posterior progression). *P* values are coded from lighter (*light gray*: .05–.06) to darker (*black*: <.01) and corrected for multiple comparisons (e.g., Guthrie & Buchwald, 1991)

Results

Behavioral results

Incorrect responses (1.3% of the data) and lexical decision times less than 250 ms or larger than 1,500 ms (less than 0.4% of the data) were excluded from the latency analyses. The mean lexical decision times and percent errors were submitted to separate *t* tests (contextual diversity: high CD vs. low CD) over participants (*t*₁) and items (*t*₂).

The statistical analyses on the latency data revealed that, on average, high-CD words (601 ms; *SD* = 123) were responded to faster than the low-CD words (614 ms; *SD* = 121), *t*₁(22) = 4.01, *p* = .001; *t*₂(68) = -2.05, *p* = .044. The statistical analyses

on the error data did not reveal any effects of contextual diversity (both *t*s < 1).

Therefore, the behavioral data replicated the same pattern of data as in the previous experiments where CD has been manipulated: response times were shorter for high- than for low-CD words.

ERP results

Figure 3 shows the ERP waves of contextual diversity (CD) in 11 representative electrodes. The ERPs show a negative potential peaking around 100 ms, which was followed by a slower positivity (P2) ranging between 100 and 250 ms. Following these early potentials, a large and slow negativity peaking around 350 ms can be seen at both anterior and posterior areas (N400). After the N400 component, the waves remain positive until the end of the epoch (500 ms).

Starting around 200-ms poststimuli, high-CD words show larger negative amplitudes compared to low-CD words over anterior scalp areas. This effect lasts approximately until 400-ms poststimuli. The results of the ANOVAs on the averaged voltage values in the 225–325 ms time window, and across the different electrode columns, are reported below.

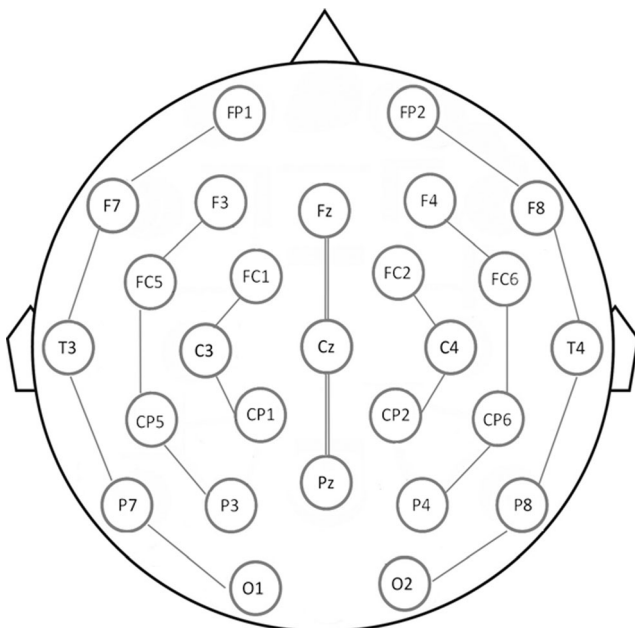


Fig. 2 Schematic representation of the electrode montage. Electrodes are grouped into four columns (midline and extending outward 1, 2 and 3 columns) for statistical analysis

225–325 ms epoch The analysis on midline and Column 1 showed a main effect of CD, midline: $F(1, 22) = 5.77, p = .02, \eta^2_p = .208$; Col. 1: $F(1, 22) = 8.20, p = .009, \eta^2_p = .272$; in Col. 2 the main effect was close to significance: $F(1, 22) = 4.01, p = .058, \eta^2_p = .154$, which was modulated by a significant interaction between CD and AP distribution, midline: $F(1, 22) = 5.54, p = .009, \eta^2_p = .201$; Col. 1: $F(1, 22) = 4.28, p = .026, \eta^2_p = .163$; Col. 2: $F(1, 22) = 5.74, p = .006, \eta^2_p = .207$. This interaction showed that the CD effect was located over anterior scalp areas: words with high-CD values elicited larger negativities than words with low-CD values, midline: Fz: $F(1, 22) = 12.28, p = .002$; Cz: $F(1, 22) = 6.66; p = .017$; Pz: $F < 1$; Col. 1: FC1/FC2: $F(1, 22) = 13.10; p = .002$; C3/C4: $F(1, 22) = 6.97; p = .015$; CP1/CP2: $F(1, 22)$

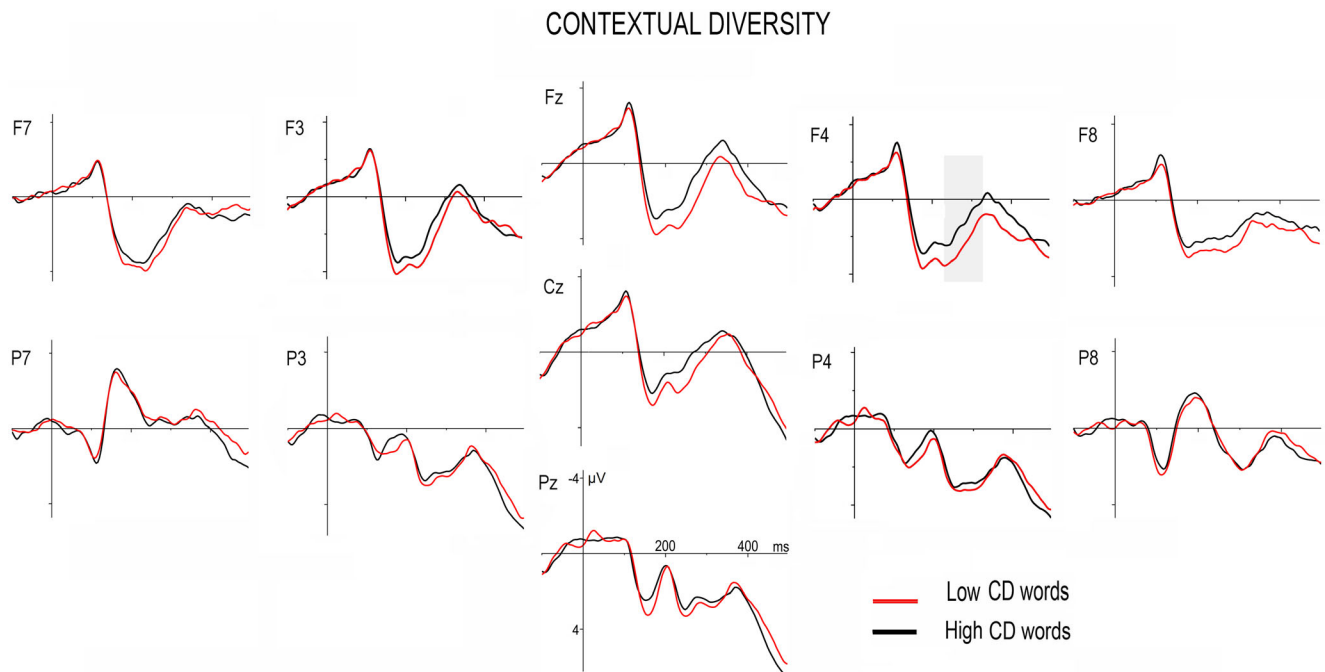


Fig. 3 Grand average ERPs to words in the two CD conditions (low and high) in eleven representative electrodes. The 225–325 ms time epoch is highlighted by the colored bar (Color figure online)

= 4.62; $p = .043$; Col. 2: F3/F4: $F(1, 22) = 6.96$; $p = .015$; FC5/FC6: $F(1, 22) = 5.98$, $p = .023$; CP5/CP6: $F(1, 22) = 2.08$; $p = .16$; P3/P4: $F < 1$.

For the interested reader, the results of the WF manipulations are presented in Appendix A—as in prior research, we found higher N400 amplitudes for lower- than for higher-frequency words.

Discussion

This experiment aimed to shed some light on the nature of the contextual diversity (CD) effect (i.e., lexical/structural vs. semantic) by examining its electrophysiological signature. As expected, the behavioral data were consistent with previous findings: high-CD words were responded to faster than low-CD words. But the central finding was on the ERP data: high-CD words elicited larger negative amplitudes than low-CD words. This constitutes a *reversal* in the direction of the CD effect when contrasted to the WF effect (see Figs. 4 & 5). Note that the findings of numerous studies that have manipulated WF mainly consist of high-frequency words eliciting smaller negative amplitudes than low-frequency words (see Vergara-Martínez & Swaab, 2012, for recent evidence), a pattern that has also been replicated in the present study for the same participants (see the follow-up analysis of WF included in Appendix A).

Our finding of a reversal of the ERP effects of CD compared to WF has important implications regarding the

assimilation of both factors into a common facilitative mechanism in visual-word recognition. If the effects of CD and WF were similar instances of the same underlying lexical/structural processes (facilitating lexical access in the same way), high-CD words would have elicited smaller negative amplitudes than low-CD words. Instead, the direction of the CD effect in the ERP results resembles that obtained in ERP experiments that manipulated factors related to “semantic richness” (i.e., larger negativities for the semantically richer words; e.g., see Rabovsky et al., 2012; West & Holcomb, 2000). Namely, ERPs for high CD words were more negative-going than ERPs for low CD words between 225 and 325 ms after word onset. Importantly, the CD effect obtained in the current experiment cannot be explained in terms of other semantic variables such as concreteness or imageability, as the experimental words were matched in these and other psycholinguistic factors (see Table 1). Although the latency and duration of the CD effect (225–325 ms) is consistent with the time course of different variables affecting lexical-semantic processing, it is outside the common interval of the N400 (300–500 ms). Nevertheless, the limits of the N400 are far from certain (see Laszlo & Federmeier, 2014, for a metareview) because few studies make the effort to really determine the onset of the effect. One might argue that, despite the many reliable effects obtained across large time intervals when the data are analyzed in aggregate, this may also result from large effects peaking very early or very late within the interval. In fact, when we conducted the statistical analyses on a broader time window (225–450 ms), the results also showed

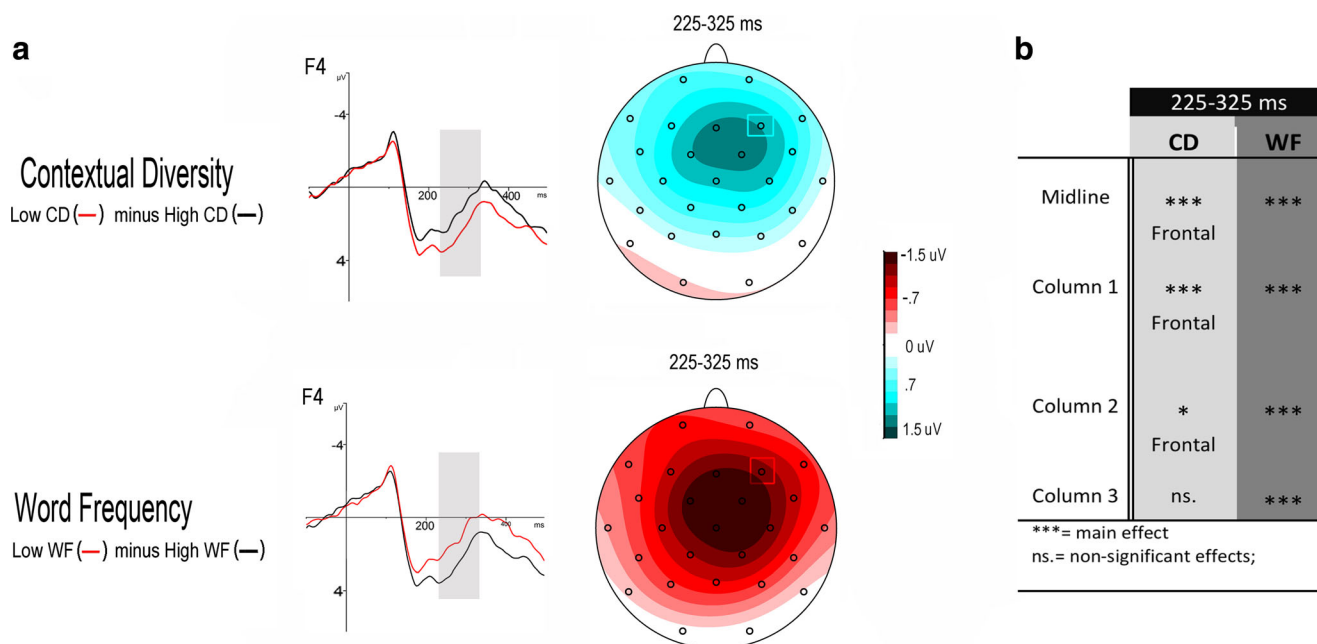


Fig. 4 **a** Topographic distribution of the CD effect (calculated as the difference in voltage amplitude between the ERP responses to low-minus high-CD words) and of the WF effect (calculated as the difference in voltage amplitude between the ERP responses to low-minus high-WF words) in the 225–325-ms time epoch. **b** Summary of

contextual diversity (CD) and word frequency (WF) effects in each electrode column. Significant ($p < .05$) main effects are reported. When there is a significant interaction between CD or WF and AP distribution and/or hemisphere, effects at specific locations are reported (Color figure online)

a significant effect of CD over frontal electrodes,¹ confirming that this method may overestimate the impact of significant effects throughout the course of processing. Hence, the latency and polarity of the CD ERP effects could be interpreted in terms of an (early) N400 modulation. Compared to the CD effect, post hoc analyses of the WF effect revealed a longer duration (150–500 ms; see Fig. 5 and Appendix A). The transient effect of CD could be explained as the result of larger semantic networks that become temporally active for words that appear in many contexts. This is, words that appear in a diverse set of contexts (i.e., high-CD words) could develop a “larger and more varied set of semantic associations, many of which will be irrelevant in any specific situation” (Hoffman, Rogers, & Lambon Ralph, 2011, p. 2442). A similar reasoning has been previously used in the interpretations of the interplay between orthographic neighborhood size (ON) and WF effects in the ERP waves (Vergara-Martínez & Swaab, 2012). The finding of a shorter timing of the ON than the WF effect was explained in terms of the interaction between the transient activation of orthographic neighbors at a lexical-semantic

level and the specific characteristics of the stimulus item during visual word recognition. Note that larger N400 amplitudes for words with many orthographic neighbors relative to words with few orthographic neighbors (Holcomb, Grainger, & O’Rourke, 2002; Laszlo & Federmeier, 2009, 2011; Vergara-Martínez & Swaab, 2012) has also been interpreted in terms of a wider activation at the semantic level of representation from orthographically similar words.

The anterior-scalp distribution of the CD effect further suggests a different underlying neural substrate of CD, when compared to that of WF (central-scalp distribution; see Figs. 4 and 5). This distribution is consistent with previously found N400 effects related to “semantically richer words” observed mainly in frontal electrodes (concreteness: Adorni & Proverbio, 2012; Barber et al., 2013; Holcomb et al., 1999; Kanske & Kotz, 2007; West & Holcomb, 2000; semantic richness: Amsel, 2011; Müller et al., 2010; but see Rabovsky et al., 2012, for centro-parietal distribution of semantic richness effects). This frontal distribution has been linked with top-down control of semantic memory in prefrontal brain areas (Adorni & Proverbio, 2012). One explanation of this pattern is that activity from long-term memory is specifically enhanced for words related to richer concepts (in terms of more semantic features or number of different contexts in which the concept is typically found).

Notably, despite the fact that CD and WF produced electrophysiological effects, their behavioral counterpart was only

¹ The statistical analyses on the 225–450 ms epoch also revealed a significant interaction of CD and AP throughout the analysis of midline column, Column 1, and Column 2: midline: $F(1, 22) = 7.65, p = .002, \eta^2_p = .258$; Fz: $F(1, 22) = 7.37, p = .01$; Cz: $F(1, 22) = 2.36, p = .13$; Pz: $F < 1$; Col. 1: $F(1, 22) = 5.61, p = .009, \eta^2_p = .203$; FC1/FC2: $F(1, 22) = 7.51, p = .01$; C3/C4: $F(1, 22) = 2.60, p = .12$; CP1/CP2: $F(1, 22) = 1.35, p = .25$; Col. 2: $F(1, 22) = 6.43, p = .003, \eta^2_p = .226$; F3/F4: $F(1, 22) = 5.03, p = .03$; FC5/FC6: $F(1, 22) = 2.28, p = .08$; CP5/CP6: $F < 1$; P3/P4: $F < 1$.

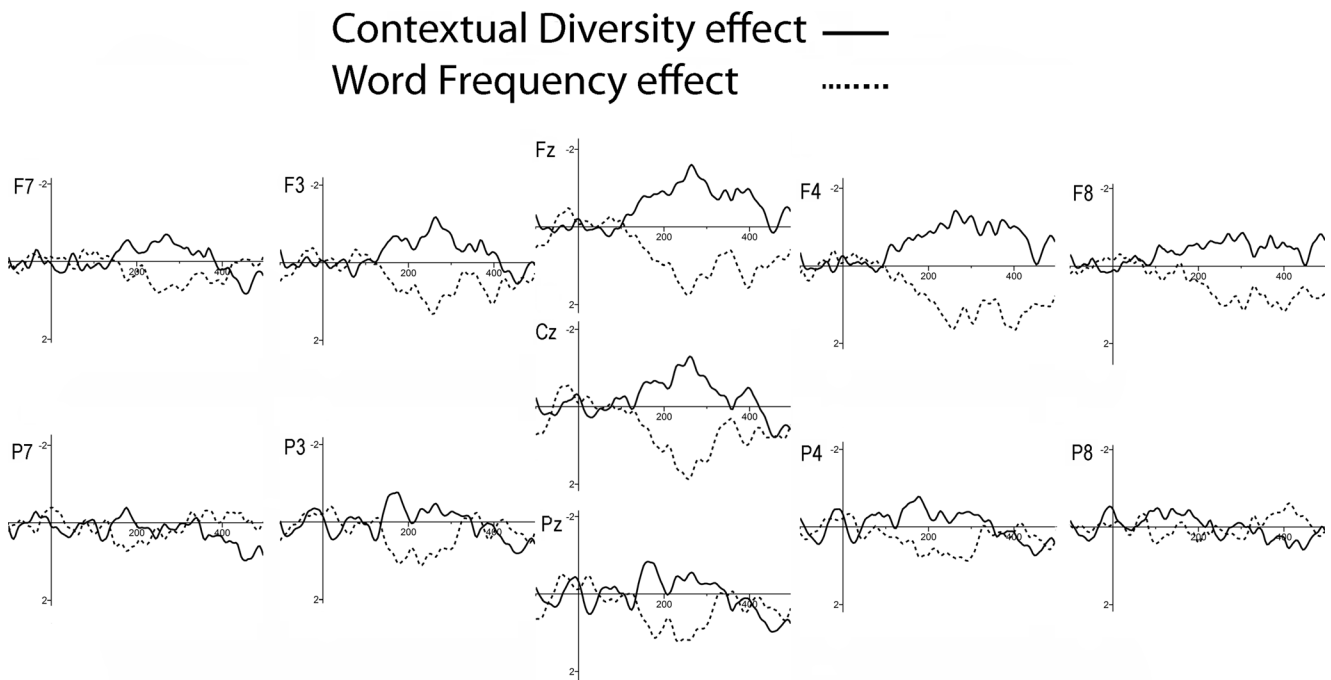


Fig. 5 Difference waveforms of contextual diversity and word frequency for 11 representative electrodes. The CD effect is calculated as the difference in voltage amplitude between the ERP responses to low

versus high CD words. The word frequency effect is calculated as the difference in voltage amplitude between the ERP responses to low- versus high-frequency words

obtained for the CD effect. In the lexical decision process, a “wordness” index may take advantage of the larger activation of the semantic networks for high-CD than for low-CD words (as shown by larger negativities for high-CD vs. low-CD words), thus producing faster response times for high-CD than for low-CD words. However, the effect of WF was not significant in the response time data. One potential reason why the behavioral WF effect was not apparent may have to do with the range of frequencies employed in the present study—note that our main goal was to maximize the differences in CD while controlling for WF. The Zipf values of WF for the high- and low-WF words were above 4.5 points, which is an upper limit for producing floor effects in lexical decision times (see Keuleers, Diependaele, & Brysbaert, 2010; Perea et al., 2013). Although ERP measures may be sensitive enough to capture the impact of subtle differences of WF on different levels of word processing (as shown by the sustained effect of WF), it is possible that this effect was not strong enough to differentially/functionally feed onto the lexical decision counterpart. All in all, the most relevant finding was the opposite pattern of the CD ERP effect when contrasted to the classic WF effect, a result that could be accommodated in a semantic enrichment interpretation of the CD facilitative effects in lexical processing.

Our findings contribute to the interpretation of the N400 as the result of different mechanisms or neural generators (divergent on time course and scalp signature) that may be differently involved in lexical-semantic retrieval, integration

processes, or during the activation of semantic features in word reading (see Kutas & Federmeier, 2011, for review). One of these mechanisms would be sensitive to the strength of the memory traces regarding the specific characteristics of a word (lexical/structural: WF). A different mechanism would be more related to the semantic properties of a word’s subset network composed by interconnected/similar features at different levels of processing (CD; see Laszlo & Federmeier, 2011). The first mechanism may operate as an interface between the brain’s internal model of the environment (built upon the extraction of statistical regularities) and the encountered information. From a connectionist perspective of semantic memory the mismatch between predicted and real observations would be described as the “implicit prediction error” (Elman, 1990; McClelland, 1994), and has been proposed by Rabovsky and McRae (2014) to be reflected by the N400 amplitude. Within this framework, words that are encountered frequently are more prone to be expected (and would elicit lower implicit prediction error, reflected in smaller N400 amplitudes) than words that are rarely encountered (which would elicit larger implicit prediction error, reflected in larger N400 amplitudes). As the strength of activation of (lexical) representations adopts a relative value due to the continuous updating of the brain’s internal model, the N400 effects related to the WF manipulation on out-of-context words can be overridden when the same words are presented in highly constraining contexts (Van Petten & Kutas, 1990). Indeed, effects from measures that represent the properties of single

items (larger N400s related to orthographic neighbor frequency and frequency of the top associate) have been reported to vanish in the second presentation of the words (Laszlo & Federmeier, 2011).

The larger negativities obtained for words with high CD could result from a second mechanism that is not determined by the actualization of an internal model according to experience or context, but rather to properties of the comprehension network at a semantic level of processing. Support for this idea comes from the finding that the N400 amplitude effect of ON size and number of lexical associates survive despite the repetition of the items, or when the stimuli are embedded in highly constraining sentences (Laszlo & Federmeier, 2009). The larger N400 amplitudes for words with many orthographic neighbors (or with many lexical associates) points to an enhanced activation of the semantic properties of the subset network for a particular item. This semantic level of processing seems to take precedence over structural/lexical processing in specific scenarios. In a reading experiment with unbalanced bilinguals on L1 and L2 word processing, Midgley, Holcomb, and Grainger (2009) presented words blocked by language and found a larger N400 for the L1 words compared to the L2 words. This is apparently an unexpected finding as the L1 was the preferred language for the participants (i.e., they were more frequently exposed to words in L1 than in L2). Midgley et al. (2009) concluded that the enhanced N400 amplitudes in L1 words reflected the larger degree of coactivation of similar representations at different levels of processing (orthographic and semantic) taking place in the native language compared to the nonnative language.

What would be the functional difference between the effects of word frequency and contextual diversity? On the one hand, the word-frequency manipulation seems to capture the consequences of mere repetition over word learning and word processing: a word's memory trace is strengthened on each occurrence, boosting the efficiency of access on subsequent presentations. Indeed, repeated words elicit smaller N400 amplitudes compared to the first word presentation, as occurs with the word-frequency effect (see Besson, Kutas, & Van Petten, 1992; Nagy & Rugg, 1989). Conversely, the manipulation of contextual diversity (i.e., the number of contexts in which a word appears) seems to capture the way in which the meaning of words is represented, specifically, the variability in meaning that is enhanced across the multiple contexts in which a word is presented. Using both a corpus-based study and a learning experiment with an artificial language, Jones, Johns, and Recchia (2012) reported that words are encoded better across multiple contexts when “the current episodic context provides novel information about the words not already contained in memory” (p. 120), thus demonstrating the importance of CD in lexical organization (see also Recchia, Johns, & Jones, 2008, for further evidence). It may be important to note here that Råling, Holzgrefe-Lang, Schröder, and Wartenburger (2015)

recently reported a study that exploited the functional meaning of the N400 as a way to disentangle the impact of two different variables (semantic typicality vs. age of acquisition [AoA]) on semantic processing during an auditory category-member-verification task. Of relevance to our study was that the pattern of the AoA effect resembled that of CD: AoA elicited behavioral facilitative responses (faster reaction times for early acquired targets), whereas its ERP counterpart consisted of early acquired targets eliciting larger early N400 amplitudes than the late acquired targets. Although Råling et al. (2015) did not discuss their results in the terms of richer semantic representations eliciting larger negativities, we believe that there are reasons to assume not only the existence of richer semantic representations of early acquired words but also to characterize the underlying relations between AoA, contextual diversity, and semantic enrichment (something that lies beyond the scope of this research study). For example, AoA may reflect not only the strength of network connections but also the quality of those words' representations (see Ellis & Lambon Ralph, 2000, for simulations in a connectionist model). The idea is that words that are learned relatively late would not be completely comparable with those acquired earlier due to the continuing loss of the network's plasticity over life. Likewise, in the area of language learning, Hills, Maouene, Riordann, and Smith (2010) analyzed the impact of contextual diversity in both acquisition and lexical processing. Hills et al. found that a word's contextual diversity, which was defined as the number of unique word types a word co-occurs with in caregiver speech, not only predicted the order of early word learning, but was also highly correlated with the number of unique associative cues for a given target word in adult free association norms.²

In a nutshell, when compared to raw WF, CD may capture the semantic enrichment produced by encountering the words across multiple and different contexts. Converging evidence for this account can be found in the field of human memory (see Hicks, Marsh, and Cook, 2005; Parmentier, Comesaña, & Soares, 2017). For instance, Hicks et al. (2005) found that CD and WF effects contributed independently to recall and posited the locus of CD effect at the level of associative connections between a to-be-remembered word and its episodic context. That is, the higher the number of contexts in which a given word appear, the higher the competition between the contexts as retrieval cues for this word (see Reder et al., 2000).

To sum up, this ERP experiment demonstrated that contextual diversity is not an epiphenomenon (or simply another indicator) of word frequency. Instead, the effects of contextual diversity are better explained as a function of semantically related factors: words that appear in many contexts may be richer in shades of meaning than the words that occur in few different contexts. Therefore, word frequency should not

² The values of AoA for the stimuli in the present study are shown in Tables 1 and 2—note that AoA did not correlate with either CD or WF.

simply be replaced with contextual diversity in models of visual-word recognition and reading. While Although apparently associated, word frequency and contextual diversity originate from different sources during the access of lexical-semantic representations, as evidenced by its dissociating role at eliciting opposite ERP signatures. Additional research should examine in greater depth the interplay between word frequency and contextual diversity during word learning.

Author note The research reported in this article has been partially funded by Grants PSI2011-26924 (Spanish Ministry of Economy and Competitiveness) and GV/2014/067 (Conselleria d'Educació, Investigació, Cultura i Esport de la Generalitat Valenciana).

Appendix A: Analysis of the word frequency effect

It was important for this study to assess the WF ERP effect by comparing words of high versus low frequency within the present experimental context. Both word groups were matched for a number of sublexical, lexical, and semantic variables (see Table 2), so that the two conditions only differed significantly in WF ($p < .0001$). The analyses of the behavioral and ERP measures are described below.

Table 2 Mean values for sublexical, lexical, and semantic characteristics of the stimuli (obtained from the Espal database; Duchon et al., 2013)

	WF comparison			
	Range	High WF	← p →	Low WF
Contextual diversity ^a	0–100	27.7 (4.4)	.13	25.8 (4.4)
Subtitle-frequency ^b	0–33,170	139.3 (5.1)	4.9 ¹⁵	77.2 (4.8)
Imageability	1.2–7	5.4	.42	5.2
Concreteness	1.9–6.8	5.1	.13	4.7
Age of acquisition ^c	1–11	5.2	.83	5.2
Number of letters	4–8	5.7	.85	5.7
Levenshtein distance	1–12	1.6	.71	1.6
Mean positional bigram frequency (token)	0–58,827	5535	.43	6,232
Most frequent lexical category		Noun = 29 Verb = 5 Adjective = 1		Noun = 29 Verb = 4 Adjective = 2

Note. The results of the pairwise comparisons between conditions are indicated in p values

^a Contextual diversity is defined as the percentage of films containing the word

^b Frequency per million

^c Extracted from the Alonso, Fernandez, and Díez (2015) norms

The Zipf values (in brackets) are the $\log_{10}(\text{freq})+3$. This logarithmic scale of frequency values allow for cross-linguistic studies and straightforward frequency ranking. For contextual diversity the range is 0.4–4.9, for subtitle-frequency the range is 0.3–7.5 (van Heuven et al., 2014)

Behavioral results

Incorrect responses (1.5% of the data) and lexical decision times less than 250 ms (less than 0.4% of the data) were excluded from the latency analyses.

The difference between the latencies of high- and low-frequency words was 3 ms (both $t_s < 1$; mean high WF: 614 ms; $SD = 121$; mean low WF: 611 ms; $SD = 123$). The statistical analyses on the error data did not reveal any effects of word frequency (both $t_s < 1$). In sum, word frequency did not affect word identification times when matched in contextual diversity, at least in the range of frequencies (i.e., medium–high frequency) employed in the current study (see Perea et al., 2013, for a similar null finding in lexical decision).

ERP results

The ERP analyses paralleled the analyses presented in the main text (see the EEG Recording and ERP Analyses section). Trials containing artifacts and/or trials with incorrect lexical decision responses were not included in the average ERPs or in the statistical analyses, a process that led to an average rejection of 8.6% of all trials (7.1% due to artifact rejection; 1.5% due to incorrect responses). A t test on the number of included trials per condition showed no difference between conditions: $t < 1$. ERPs were averaged separately for each of the experimental conditions, each of the subjects, and each of the electrode sites. The statistical analysis was performed on the mean ERP values in the 225–325-ms time window. The analyses of variance (ANOVAs) included the factors WF, AP, and HEM (on three pairs of electrode columns; WF and AP on the midline column; see Analysis section in the main text for specification of the electrode montage).

Figure 6 shows the ERP waves for the word-frequency comparison in 11 representative electrodes. Starting around 150 ms until 500-ms poststimuli, low-WF words show larger negative amplitudes compared to the high-WF words with a widespread scalp distribution. The results of the ANOVA on the averaged voltage values in the 225–325-ms time window, and across the different electrode columns, are reported below.

225–325 ms epoch Main effects of WF were obtained on each column, midline: $F(1, 22) = 18.07, p < .001, \eta^2_p = .451$; Col. 1: $F(1, 22) = 21.33, p < .001, \eta^2_p = .492$; Col. 2: $F(1, 22) = 14.21, p = .001, \eta^2_p = .393$; Col. 3: $F(1, 22) = 7.54, p = .012, \eta^2_p = .256$, with larger negative amplitudes for low-frequency words than for high-frequency words.

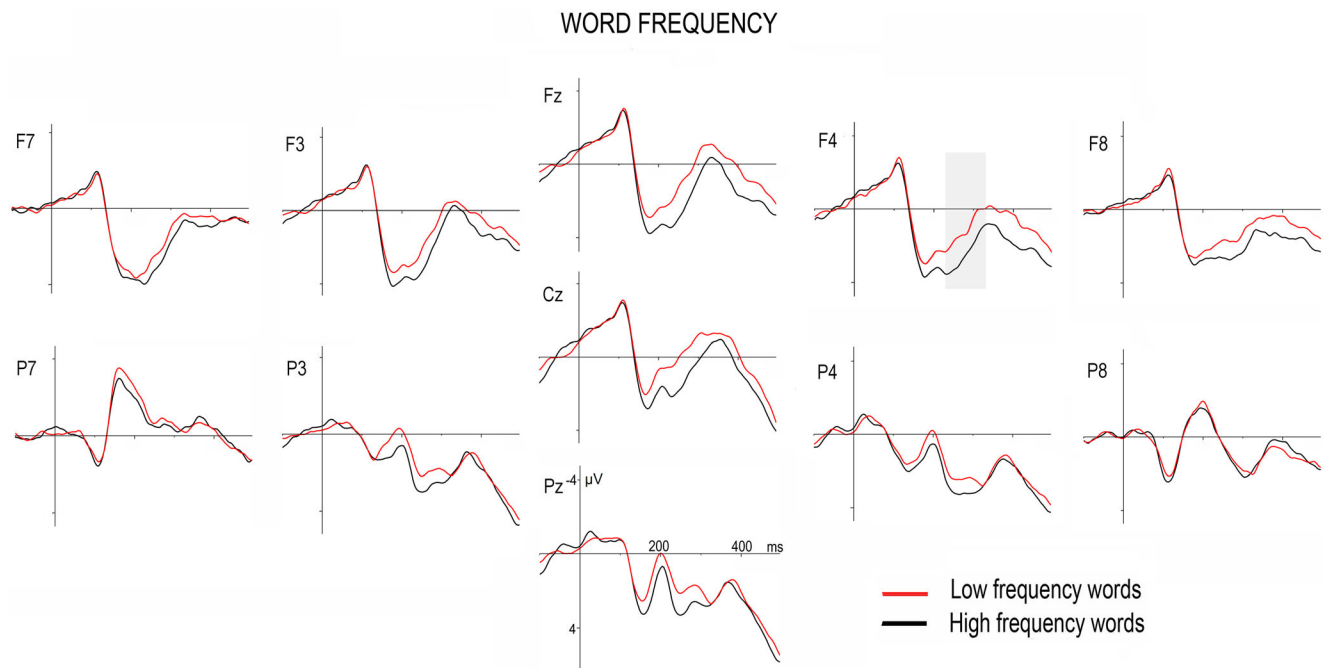


Fig. 6 Grand average ERPs to words in the two WF conditions (low and high) in 11 representative electrodes. The 225–325 ms time epoch is highlighted by the colored bar (Color figure online)

Figure 6 shows that the WF effect expands through a larger time window (150–500 ms) than the current epoch under analysis (225–325 ms). Therefore, we conducted follow up analyses on an earlier (Epoch 1: 150–225 ms) and a later (Epoch 3: 325–500 ms) epoch. These analyses showed main effects of WF in Epoch 1, midline: $F(1, 22) = 7.58, p = .012, \eta^2_p = .256$; Col. 1: $F(1, 22) = 7.35, p = .013, \eta^2_p = .250$; Col. 2: $F(1, 22) = 6.32, p = .020, \eta^2_p = .223$; Col. 3: $F(1, 22) = 4.59, p = .043, \eta^2_p = .173$. In Epoch 3, the WF effect was significant in midline and Column 1, midline: $F(1, 22) = 5.31, p = .031, \eta^2_p = .195$; Col. 1: $F(1, 22) = 5.23, p = .032, \eta^2_p = .192$. In Column 1, the main effect of WF was modulated by a significant interaction between WF, HEM, and AP distribution, $F(1, 22) = 3.8, p = .03, \eta^2_p = .150$; FC1: $F(1, 22) = 5.6; p = .026$; FC2: $F(1, 22) = 8.04; p = .01$; C3: $F < 1$; C4: $F(1, 22) = 8.8; p = .007$; CP1: $F < 1$; CP2: $F(1, 22) = 3.6; p = .072$. This interaction revealed that by this time window, WF effects were mainly obtained over central-anterior electrodes of the right hemisphere.

Appendix B: Stimulus list

High-CD (high-WF) words

frente; secreto; orden; teléfono; mente; plan; fuerza; fuego; oficina; escuela; atención; ropa; palabra; perdón; llamada; pena; chico; daño; mitad; vista; aire; viaje; cuarto; edad; boca;

cama; minuto; café; cielo; jugar; dormir; cambiar; poner; perder; extraño

Low-CD (high-WF) words

general; blanco; maldito; ganar; profesor; nave; ejército; avión; maestro; capitán; abogado; agente; asesino; carta; película; sexo; caja; sistema; música; perro; libro; cita; campo; bebé; señorita; pueblo; negocio; novia; médico; sueño; futuro; arma; país; niña; tren

Low-CD (low-WF) words

subir; deseo; tarjeta; paseo; brazo; público; cerveza; pista; piel; golpe; cerebro; vestido; cocina; carne; especie; ventana; carrera; cárcel; calma; piso; señal; papel; causa; regalo; tema; basura; sala; éxito; sorpresa; vieja; ridículo; relación; aprender; caer; echar;

Pseudowords

zubar; neseras; dieba; gornir; plango; mivétulo; cactiar; marlato; pelariad; pomir; vadar; afrandir; pendor; clovelor; mair; empliño; gabe; iglar; gronte; enertiso; tujir; tederro; afial; senea; ortín; peantra; borbeta; velíboco; cacinal; satea; muste; amisada; praño; plaz; afiste; mullaco; fierva; acimico; cismeza; viego; marda; musta; odenica; selenuza; ciel; enciosa; sefo; gurpe; amardiño; pama; cececha; soña; dastega;

lescado; malidra; sudisa; corata; murdón; pibro; canve; drasida; fitro; envecio; peva; cuda; fantina; cheno; calgo; cadreno; lajo; leme; carvol; metaz; mejonisa; casma; linta; muegro; pemo; aude; feposio; refal; vaipe; gopia; magel; muanto; semaco; ceuda; enaz; ruevo; decala; rona; cunuso; bima; pafa; anza; tavuro; siduno; maes; bama; mave; geña; ebato; ceuso; trel; salbresa

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*, 814–823. doi:10.1111/j.1467-9280.2006.01787.x
- Adorni, R., & Proverbio, A. M. (2012). The neural manifestation of the word concreteness effect: An electrical neuroimaging study. *Neuropsychologia*, *50*, 880–891. doi:10.1016/j.neuropsychologia.2012.01.028
- Alonso, M. A., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavioral Research Methods*, *47*, 268–274. doi:10.3758/s13428-014-0454-2
- Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*, 970–983. doi:10.1016/j.neuropsychologia.2011.01.003
- Amsel, B. D., & Cree, G. S. (2013). Semantic richness, concreteness, and object domain: An electrophysiological study. *Canadian Journal of Experimental Psychology*, *67*, 117–129. doi:10.1037/a0029807
- Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*, 47–53. doi:10.1016/j.bandl.2013.01.005
- Barber, H. A., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: Evidence from ERPs. *Neuroreport*, *15*, 545–548. doi:10.1097/00001756-200403010-00032
- Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience*, *4*, 132–149. doi:10.1162/jocn.1992.4.2.132
- Borowsky, R., & Masson, M. E. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 63–85. doi:10.1037/0278-7393.22.1.63
- Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*, 531–544. doi:10.3758/bf03196189
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), e10729. doi:10.1371/journal.pone.0010729
- Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*, 713–758. doi:10.1037/a0019738
- Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behaviour: The case of Greek. *Frontiers in Psychology*, *1*, 218. doi:10.3389/fpsyg.2010.00218
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*, 1246–1258. doi:10.3758/s13428-013-0326-1
- Duñabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's Ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin & Review*, *15*, 1072–1077. doi:10.3758/pbr.15.6.1072
- Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, memory, and cognition*, *26*(5), 1103–1123. doi:10.1037/0278-7393.26.5.1103
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211. doi:10.1016/0364-0213(90)90002-E
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813. doi:10.1037/0033-295x.112.4.777
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635. doi:10.1016/s0022-5371(73)80042-8
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565. doi:10.1037/0033-295x.103.3.518
- Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*, 240–244. doi:10.1111/j.1469-8986.1991.tb00417.x
- Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, *30*, 1383–1400. doi:10.1016/j.neuroimage.2005.11.048
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*, 1090–1103. doi:10.1016/j.clinph.2003.12.020
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language*, *53*, 430–444. doi:10.1016/j.jml.2005.04.001
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*, 259–273. doi:10.1016/j.jml.2010.06.002
- Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (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*, 2432–2446. doi:10.1162/jocn.2011.21614
- Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of contextual variation in word meaning based on latent semantic analysis. *Behavior Research Methods*, *45*, 718–730. doi:10.3758/s13428-012-0278-x
- Holcomb, P., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*, 938–950. doi:10.1162/089892902760191153
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 721–742. doi:10.1037/0278-7393.25.3.721
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*, 431–439. doi:10.3758/bf03208203
- 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. doi:10.1037/a0026727
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, *1148*, 138–148. doi:10.1016/j.brainres.2007.02.044

- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42, 627–633. doi:10.3758/brm.42.3.627
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1, 174. doi:10.3389/fpsyg.2010.00174
- Kounios, J., Green, D. L., Payne, L., Fleck, J. I., Grondin, R., & McRae, K. (2009). Semantic richness and the activation of concepts in semantic memory: Evidence from event-related potentials. *Brain Research*, 1282, 95–102. doi:10.1016/j.brainres.2009.05.092
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-encoding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 804–823. doi:10.1037/0278-7393.20.4.804
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463–470. doi:10.1016/S1364-6613(00)01560-6
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. doi:10.1146/annurev.psych.093008.131123
- Landauer, T. K. (2001). Single representations of multiple meanings in latent semantic analysis. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 217–232). Washington, DC: American Psychological Association. doi:10.1037/10459-012
- Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships in sentence context. *Journal of Memory and Language*, 61, 326–338. doi:10.1016/j.jml.2009.06.004
- Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, 48, 176–186. doi:10.1016/s1364-6613(00)01560-6
- Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: Evaluating the physiological time course of visual word recognition with regression analysis of single item ERPs. *Language, Cognition, and Neuroscience*, 29, 642–661. doi:10.1080/01690965.2013.866259
- Massol, S., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). When less is more: Feedback, priming, and the pseudoword superiority effect. *Brain Research*, 1386, 153–164. doi:10.1016/j.brainres.2011.02.050
- McClelland, J. L. (1994). The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. E. P. Bertelson & G. d'Ydewalle (Eds.), *Current advances in psychological science: Ongoing research* (pp. 57–88). Hillsdale: Erlbaum.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88, 375–407. doi:10.1037/0033-295x.88.5.375
- Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics*, 22, 281–300. doi:10.1016/j.jneuroling.2008.08.001
- Müller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: What is shared, what is different? *Psychophysiology*, 47, 455–466. doi:10.1111/j.1469-8986.2009.00960.x
- Nagy, M. E., & Rugg, M. D. (1989). Modulation of event-related potentials by word repetition: The effects of inter-item lag. *Psychophysiology*, 26, 431–436. doi:10.1111/j.1469-8986.1989.tb01946.x
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113. doi:10.1016/0028-3932(71)90067-4
- Parmentier, F. B., Comesaña, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2015.1105268
- Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word-identification times in young readers. *Journal of Experimental Child Psychology*, 116, 37–44. doi:10.1016/j.jecp.2012.10.014
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15, 161–167. doi:10.1037/e527342012-696
- Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, 9, 542–549. doi:10.3758/bf03196311
- 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. doi:10.1037/e633262013-687
- Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *Journal of General Psychology*, 13, 199–203. doi:10.1080/00221309.1935.9917878
- Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, 132, 68–89. doi:10.1016/j.cognition.2014.03.010
- Rabovsky, M., Sommer, W., & Abdel Rahman, R. A. (2012). The time course of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, 6, 11. doi:10.3389/fnhum.2012.00011
- Räling, R., Holzgrefe-Lang, J., Schröder, A., & Wartenburger, I. (2015). On the influence of typicality and age of acquisition on semantic processing: Diverging evidence from behavioural and ERP responses. *Neuropsychologia*, 75, 186–200. doi:10.1016/j.neuropsychologia.2015.05.031
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201. doi:10.3758/bf03197692
- Recchia, G., Johns, B. T., & Jones, M. N. (2008). *Context repetition benefits are dependent on context redundancy* (pp. 23–26). Paper presented at the Proceedings of the 30th Annual Cognitive Science Society, Washington, DC.
- Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 294–320. doi:10.1037/0278-7393.26.2.294
- 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. doi:10.1037/0033-295x.105.1.125
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46, 245–266. doi:10.1006/jmla.2001.2810
- Rodd, J. M. (2004). When do leotards get their spots? Semantic activation of lexical neighbors in visual word recognition. *Psychonomic Bulletin & Review*, 11, 434–439. doi:10.3758/bf03196591
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9, 487–494. doi:10.1016/s0022-5371(70)80091-3
- Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? In P. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223–250). Hillsdale: Erlbaum.

- Soares, A. P., Machado, J., Costa, A., Iriarte, A., Simões, A., de Almeida, J. J., & Perea, M. (2015). On the advantages of word-frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *Quarterly Journal of Experimental Psychology*, *68*, 680–696. doi:10.1080/17470218.2014.964271
- Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, *43*, 195–201. doi:10.1037/h0054636
- Smith, M. E., & Halgren, E. (1987). Event-related potentials during lexical decision: Effects of repetition, word frequency, pronounceability, and concreteness. *Electroencephalography & Clinical Neurophysiology: Supplement*, *40*, 417–421.
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. doi:10.1080/17470218.2013.850521
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*, 380–393. doi:10.3758/bf03197127
- Vergara-Martínez, M., Perea, M., Gómez, P., & Swaab, T. Y. (2013). ERP correlates of letter identity and letter position are modulated by lexical frequency. *Brain and Language*, *125*, 11–27. doi:10.1016/j.bandl.2012.12.009
- Vergara-Martínez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology*, *49*, 1277–1289. doi:10.1111/j.1469-8986.2012.01410.x
- West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*, 1024–1037. doi:10.1162/08989290051137558
- Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 878–890. doi:10.1037/0278-7393.31.5.878
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 72. doi:10.3389/fnhum.2012.00072
- Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, *18*, 742–750. doi:10.3758/s13423-011-0092-y