Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI

Tal Yarkoni [a,*], Nicole K. Speer [b], David A. Balota [a,c], Mark P. McAvoy [d], Jeffrey M. Zacks [a,d]

[a] *Department of Psychology, Washington University, One Brookings Drive, Saint Louis, MO 63130, USA*
[b] *WICHE Mental Health Program 3035 Center Green Drive, Suite 200 Boulder, CO 80301, USA*
[c] *Department of Neurology, Washington University School of Medicine, One Brookings Drive, Saint Louis, MO, 63130, USA*
[d] *Department of Radiology, Washington University School of Medicine, One Brookings Drive, Saint Louis, MO 63130, USA*

ARTICLE INFO

ABSTRACT

Reading is one of the most important skills human beings can acquire, but has proven difficult to study naturalistically using functional magnetic resonance imaging (fMRI). We introduce a novel Event-Related Reading (ERR) fMRI approach that enables reliable estimation of the neural correlates of single-word processing during reading of rapidly presented narrative text (200–300ms /word). Application to an fMRI experiment in which subjects read coherent narratives and made no overt responses revealed widespread effects of orthographic, phonological, contextual, and semantic variables on brain activation. Word-level variables predicted activity in classical language areas as well as the inferotemporal visual word form area, specifically supporting a role for the latter in mapping visual forms onto articulatory or acoustic representations. Additional analyses demonstrated that ERR results replicate across experiments and predict reading comprehension. The ERR approach represents a powerful and extremely flexible new approach for studying reading and language behavior with fMRI.

© 2008 Elsevier Inc. All rights reserved.

Introduction

In recent years, functional magnetic resonance imaging (fMRI) has emerged as an important methodological tool for studying the neural mechanisms of reading and language behavior. FMRI studies have added considerable anatomical detail to classic neuropsychological models of left-hemisphere language function (Vigneau et al., 2006), and have investigated a broad range of orthographic, phonological, semantic, and syntactic processes involved in reading and language (Bookheimer, 2002; Fiez and Petersen, 1998; Gernsbacher and Kaschak, 2003; McCandliss et al., 2003; Vigneau et al., 2006). However, the successes of fMRI have not been distributed equally across the spectrum of reading and language function. The vast majority of fMRI studies of reading have employed single-word factorial designs that use small groups of individually presented, unrelated words as stimuli. Although such designs have led to substantial advances in understanding the neural mechanisms supporting single-word reading (for reviews, see Fiez and Petersen, 1998; Jobard et al., 2003; Price, 1998; Turkeltaub et al., 2002; Vigneau et al., 2006) they may be of limited utility in studying the many semantic, syntactic, and phonological processes that are arguably essential to natural reading but depend on the presence of contextual factors (Rayner, 1998). Moreover, most single-word studies require subjects to make overt responses to isolated words, raising the possibility that the neural responses observed in such studies could be contaminated by secondary executive and motor task demands.

In the behavioral literature, psycholinguists have developed several methodological approaches to overcome the limitations of traditional single-word factorial studies. One approach is to use more naturalistic measures during reading, e.g., monitoring processing of individual words within coherent text using eye-tracking or event-related potentials (Rayner, 1998; Sereno and Rayner, 2003). This approach increases ecological validity and enables researchers to quantify the effects of variables that are impossible to investigate in single-word studies, e.g., position of a word within a sentence, probability of skipping a word, etc. A second, 'mega-study', approach is to retain single-word paradigms, but to rely on large datasets containing thousands of words instead of relatively small factorial studies (Balota et al., 2004; Seidenberg and Waters, 1989). The chief

---

* Corresponding author. Department of Psychology, Campus Box 1125, Washington University, St. Louis, MO 63130, USA.
E-mail address: tyarkoni@wustl.edu (T. Yarkoni).

advantage of the latter approach is its flexibility to model a wide range of variables and discriminate amongst alternative theoretical accounts of isolated word recognition.

Despite the demonstrable benefits of naturalistic or mega-study approaches in behavioral research, no comparable approaches to reading have emerged in the fMRI literature. Most fMRI studies continue to rely on relatively non-naturalistic single-word designs, with small sets of stimuli. The few studies that have employed more naturalistic reading paradigms—e.g., presenting subjects with coherent sentences or narratives (Ferstl and von Cramon, 2001; Keller et al., 2001; Yarkoni et al., in press-b)—have provided little information about word-level processes. Anecdotal evidence suggests two possible reasons for the absence of fMRI studies investigating word-level processes within the context of rapid reading. One potential concern is that the relatively slow sampling rate of fMRI (typically 1–3s) may preclude the study of higher-frequency experimental events that occur multiple times within each scanning frame. However, existing empirical evidence weighs against this concern. Hemodynamic responses to visual stimulation have been shown to sum approximately linearly even at very short (~ 250ms) durations (Burock et al., 1998), and brain activation in response to subliminally-presented stimuli has been observed at even shorter latencies (e.g., Dehaene et al., 2001). While there is no guarantee that the hemodynamic response to complex stimuli such as words is as linear as the response to simple visual stimuli, such an assumption seems warranted in the absence of empirical evidence to the contrary. Moreover, the same assumption implicitly underlies many other fMRI studies of higher cognition in which the neural and cognitive processes of interest occur much more rapidly than the nominal fMRI sampling rate.

A second and more important concern is that narrative stimuli may be too complex to decompose into constituent word-level processes. In a coherent reading context, individual word-level variables (e.g., frequency or length) are likely to be confounded with many other variables (e.g., grammatical class, position within sentence, number of repetitions, etc.), because the selection and order of stimuli is not experimentally constrained. Although this concern is a valid one, it can be addressed to some degree by statistically modeling potential confound. Numerous eye-tracking and EEG studies have taken this approach to model single-word processing within a sentential or narrative context (Sereno and Rayner, 2003). Moreover, the standard fMRI approach of using highly controlled factorial designs with relatively few stimuli is also subject to confounding. Most fMRI studies match word stimuli on only a handful of dimensions (e.g., length, frequency, orthographic neighborhood, etc.), despite the fact that dozens of variables are known to exert unique influences on lexical processing (Balota et al., 2006). In factorial-design fMRI studies potential confounds in uncontrolled extraneous variables are not usually modeled.

The present study used a novel *Event-Related Reading (ERR)* approach to demonstrate the viability of studying word-level processes with fMRI within a narrative reading context. The ERR method is an extension of standard rapid event-related fMRI to reading paradigms in which words are presented at a rate exceeding the nominal sampling rate. Application to an fMRI dataset demonstrated that the approach was able to flexibly replicate and extend a broad range of findings from previous single-word fMRI studies. Moreover, additional analyses ruled out alternative explanations for the ERR results while providing further evidence of their reliability and validity.

Materials and methods

Subjects

Twenty-eight right-handed, native English speakers (ages 19–34, 20 women) were given informed consent according to the guidelines set forth by the Washington University Medical School. Five subjects had data from only two ($n=1$) or three ($n=4$) stories due to equipment malfunction or subject fatigue. Non-overlapping analyses from this sample have been reported in a previous study (Speer et al., 2007).

Reading task

Subjects read 4 narratives from the book *One Boy's Day* (Barker and Wright, 1951; stimuli are available online at http://dcl.wustl.edu/stimuli.html). Narratives ranged from 8.5 to 10.9 min in length. Order was counterbalanced across subjects. Each narrative was presented one word at a time, with each word remaining on-screen for 200ms followed by a 150ms/syllable blank delay. Subjects practiced this reading method on a separate narrative prior to scanning until they reported being comfortable with word-by-word reading. Subjects were told in advance that they would be given a comprehension test at the end of the session. No overt responses were made during scanning.

Comprehension test

To ensure that subjects complied with instructions and were attentive to the narratives, a comprehension test consisting of 20 multiple-choice items (5 items per narrative, 4 possible choices per question) was administered at the end of the session. Each question assessed subjects' understanding of a portion of one of the narratives. Mean accuracy was 82.74% ($SEM=1.98\%$; range=60%–100%), and all subjects performed significantly above chance (binomial probability test, $p<.001$ for all subjects).

fMRI data acquisition

Images were acquired on a 3-T Siemens Allegra MRI scanner (Erlangen, Germany). High resolution ($1\times1\times1.25$ mm) structural images were acquired using a sagittal MP-RAGE T1-weighted sequence. Functional images were acquired using a T2*-weighted asymmetric spin-echo echo-planar sequence, with 32 slices (4.0×4.0 mm in-plane resolution) acquired every 2.048s. An additional T2-weighted fast turbo spin-echo scan acquired structural data in the same planes as the functional scans, in order to map the functional data to the structural data. The functional data were pre-processed and warped to a standard stereotactic space with isotropic voxels ($3\times3\times3$ mm) (Talairach and Tournoux, 1988). Timing offsets between slices were corrected using cubic spline interpolation. All data were realigned within and across runs for each subject, and image intensity was normalized to a whole-brain mode value of 1000. Data were then smoothed with a Gaussian filter (6 mm full-width half-maximum).

fMRI data analysis

In a typical single-word fMRI study designed to allow estimation of responses to individual word presentations, stimuli might be presented at a rate of one every 2–3 s, often with variable delays (or "jittering") inserted between presentations. However, neither the relatively wide spacing between stimuli nor

the use of a variable delay are necessary in order to obtain reliable word-level estimates. Spacing is unnecessary because the hemodynamic response is roughly additive even at very short intervals (Burock et al., 1998). Random delays are unnecessary provided that there is sufficient word-by-word variability in the variables of interest. For example, in a coherent sentence such as "Raymond walked toward the corner", the constituent words vary greatly in their relative frequency of occurrence. This natural variability should, in theory, inject sufficient variability to enable a word-by-word correlation to be performed between word frequency and BOLD signal (cf. Maccotta et al., 2001). However, the fact that the rate of word presentation (several hundred milliseconds) exceeded the sampling rate or TR (2.048 s) necessitates aggregation across words prior to GLM estimation. In the present study, this was accomplished by averaging across all words that fell within each scanning frame (see Fig. 1).

Estimates of reading-related brain activity were computed for each subject using the general linear model (GLM; Friston et al., 1995). In addition to a varying number of lexical variables, each model included regressors coding for the contrast between word reading and a fixation baseline, as well as 8 regressors coding for effects of no interest (separate regressors for trend and baseline for each of the 4 BOLD runs). Estimation of the word vs. fixation effect was facilitated by the inclusion of 6–12 fixation frames prior to and following each narrative, as well as the use of variable delays between words (see above). All variables were modeled using a finite impulse response basis set that spanned 8 frames (2.048 s each, total = 16.38 s). The shape of the HRF was estimated rather than assumed in order to validate the present methodological approach and ensure that the timecourses it produced resembled a standard HRF. Because an initial inspection of several regions confirmed that this was the case, all subsequent analyses were conducted using a summary estimate of each voxel's overall magnitude (see Statistical Analysis section below).

Lexical variables

Several different multiple regression models were tested (see Results). These models included different combinations of 12 lexical variables, defined as follows:

Length was defined as the number of letters in the string.
Frequency refers to log-transformed norms from the HAL corpus (Lund and Burgess, 1996). The HAL frequency of a word represents the number of times that word appeared in a large corpus of text culled from internet newsgroups, and is a strong predictor of lexical processing latencies (see Balota et al., 2004).

Lexical decision time and *word naming* behavioral norms. Lexical decision time refers to the time needed to determine whether a string of letters is a valid word or not. Naming latency refers to the time needed to initiate (correct) pronunciation of the word. Both estimates are taken from the English Lexicon Project (ELP; http://elexicon.wustl.edu), which is a multi-institution study of naming and lexical decision performance (Balota et al., 2007). These norms are based on over 1200 subjects, and include latencies for over 40,000 words. Each item z-score mean for naming and lexical decision is based on approximately 30 observations. These means were standardized based on each individual's mean latency and standard deviation, and then averaged across subjects who received that word.

Word position was defined as the serial position of a word within its enclosing sentence. The number assigned was absolute and not relative to the total length of a sentence. That is, the 8th word in an 8-word sentence was coded 8, and the 21st word in a 21-word sentence was coded 21.

Word repetition: the number of times a word had been presented (across narratives) to the subject prior to the current occurrence.

Quadratic length: the square of word length.

Number of syllables and number of morphemes: the number of syllables and morphemes in a word, respectively.

Orthographic Neighborhood Size (ON) refers to the number of words that can be obtained by changing exactly one letter in a word's orthographic representation (Coltheart et al., 1977). The count does not include words that may be obtained by additional insertion, deletion or transposition operations. ON values were computed using the full ELP dataset of over 40,000 words.

Phonological Neighborhood Size is the phonological analog of Orthographic Neighborhood Size. It is defined as the number of words that can be obtained by changing exactly one phoneme in a word's phonological representation. PN values were computed using the full ELP dataset of over 40,000 words.

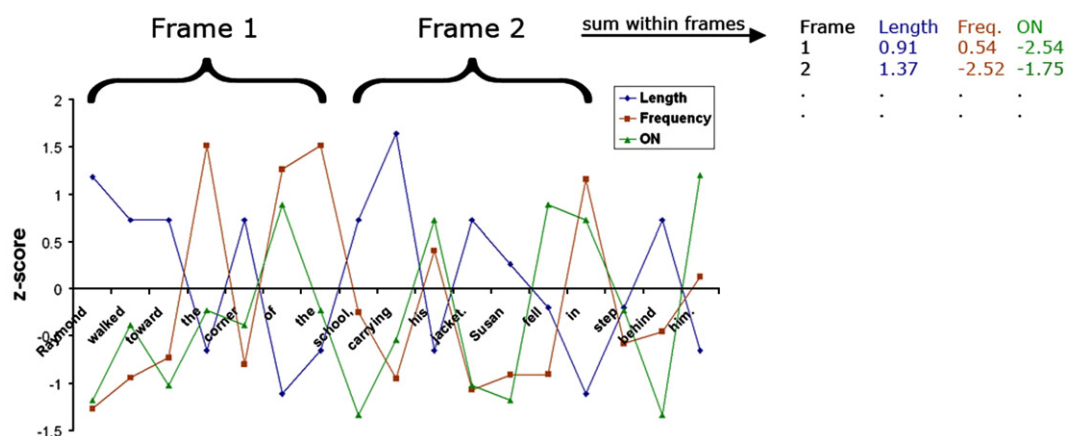**Fig. 1.** Schematic illustration of ERR methodology. First, all lexical variables in the model are standardized across all frames. Standardized values for three variables (length, frequency, and orthographic neighborhood size [ON]) are displayed for 2 sample frames (each TR = 2.048s). The standardized values for each variable are then summed within each frame to produce the frame-by-frame values entered into the GLM as predictors.

Imageability: values from the Toglia and Battig norms for how readily and vividly the referent of a word can be imaged (Toglia and Battig, 1978).

Parametric modeling of lexical variables

For the 7 parametric variables (all but the reading vs. baseline contrast), we employed the following approach (Fig. 1). First, values for each variable were standardized across all words (i.e., the mean was subtracted and the result divided by the standard deviation). This step orthogonalized the parametric variables with respect to the non-parametric regressors in the model, ensuring that the parametric estimates from the GLM would not be confounded with the word vs. fixation estimate. For the contextual variables (word repetitions and position within sentence), values were log-transformed prior to standardization, because (a) the distribution for each variable was highly skewed (i.e., a small number of words were repeated many more times than most words, and a few sentences were much longer than most), and (b) we reasoned that changes in activation related to repetition or sentence complexity should asymptote relatively quickly rather than showing a linear relationship. However, subsequent reanalysis with linear predictors produced virtually identical results.

Following standardization, values for each variable were summed within each scanning frame (Fig. 1). For example, if 6 words were presented within a given scanning frame, the standardized frequencies of all 6 words would be summed to produce the frequency value for that frame, and similarly for each of the other parametric variables. Stimuli were summed within their frame of onset, regardless of whether their presentation overlapped with the next frame (e.g., if a word was presented in the last 100 ms of frame N and first 100 ms of frame N + 1, it was summed within frame N).

Note that some words could not be modeled because they lacked values for the behavioral measures (i.e., LDT and naming). Additionally, some words were deliberately eliminated because they were outliers (defined as >3 standard deviations from the mean on any lexical variable) that could potentially bias estimates. Based on these criteria, a total of 265 out of 5058 word tokens (5.2%) were eliminated from all analyses. Words were excluded from analysis on a list-wise basis, i.e., if a word lacked a value for at least one variable, no value was assigned to *any* of the variables for that word in order to prevent misattribution of variance across correlated variables. To ensure that the presence of non-normed words would not bias the estimates for normed words, a separate set of dummy regressors was used

to model the effect of reading non-normed words vs. baseline. These dummy regressors were not analyzed any further.

Because multicollinearity (i.e., excessive correlation between predictor variables) can present a problem in multiple regression, we sought to ensure that the lexical predictors we used all contained sufficient unique variance to produce reliable estimates. In the present study, correlations between variables could be assessed at 3 different levels: word type (the set of distinct words occurring once or more in the stimuli), word token (the set of total words occurring in the stimuli, including repetitions), or design matrix (the set of regressors representing values for word tokens, standardized and summed across all frames). The correlation matrices between variables at the word type and design matrix levels are presented in Table 1. Note that only the design matrix correlations have direct implications for the reliability of the neuroimaging estimates. Table 1 demonstrates that multicollinearity was not a serious problem in the present data. In the first model tested, the largest bivariate correlation was 0.69 (between length and orthographic neighborhood), and most correlations were substantially lower. More importantly, no variable had a variance inflation factor (VIF) greater than 3.1 (multiple $R^2 = .67$ when regressing length on all other variables of interest; Table 1, Model 1). When phonological variables were added to the model, collinearity increased to a potentially problematic level for only one variable (length; multiple $R^2 = .86$; Table 1, Model 2), and was considerably lower for other variables. Critically, all variables (including length) continued to show significant effects in multiple regions in the more conservative model, suggesting that they possessed sufficient unique variance to allow reliable estimation.

Note that because the parametric regressors in the GLM were standardized prior to estimation, the resulting coefficients denote the amount of change in MRI signal associated with an increase of one standard deviation in a given variable for a single word (holding all other factors constant). These standardized quantities are generally smaller than those typically reported in fMRI studies because they are in units of effect size rather than in terms of comparisons between two discrete conditions; however, they facilitate comparison of the effects of different variables (see Results).

Statistical analysis

Statistical analysis of fMRI data was performed using a random effects model. At the first level (i.e., within subject), magnitude estimates for each effect of interest were computed on a voxel-wise basis by cross-correlating the timecourse of

Table 1
Correlations between lexical variables (Pearson's *r*). Values above the diagonal denote type-level correlations; values below the diagonal denote design matrix-level correlations (see text for explanation)

	Length	Freq	ON	LDT	Naming	PN	Syll.	Morph.	Position	M1 R²	M2 R²
Length	–	− 0.57	− 0.62	0.65	0.60	− 0.65	0.81	0.75	–	0.67	0.86
Freq	− 0.68	–	0.33	− 0.64	− 0.54	0.39	− 0.42	− 0.51	–	0.58	0.59
ON	− 0.69	0.43	–	− 0.39	− 0.39	0.76	− 0.54	− 0.43	–	0.50	0.65
LDT	0.45	− 0.52	− 0.37	–	0.69	− 0.42	0.60	0.58	–	0.38	0.43
Naming	0.53	− 0.42	− 0.48	0.52	–	− 0.34	0.48	0.47	–	0.40	0.41
PN	− 0.57	0.30	0.75	− 0.31	− 0.34	–	− 0.59	− 0.49	–	–	0.59
Syll.	0.84	− 0.55	− 0.61	0.48	0.43	− 0.56	–	0.67	–	–	0.74
Morph.	0.79	− 0.59	− 0.50	0.46	0.39	− 0.45	0.72	–	–	–	0.66
Positi	− 0.18	0.22	0.11	− 0.13	− 0.12	0.06	− 0.12	− 0.16	–	0.07	0.07
Repeat	− 0.19	0.36	0.13	− 0.19	− 0.10	0.05	− 0.19	− 0.10	− 0.05	0.16	0.18

The two rightmost columns indicate the total proportion of variance explained in each variable by all the other variables at the design matrix level for models with (M1) and without (M2) phonological variables.

activation estimated for the 8 finite impulse response regressor set with a model hemodynamic response function (HRF) convolved with a boxcar (Boynton et al., 1996). For each variable, the resulting maps were then submitted to a one-sample t-test at the second level (vs. a null hypothesis of no effect). The output of the statistical test was then transformed to a z-map and thresholded to correct for multiple comparisons.

For whole-brain analyses, we used a voxel-wise (intensity) threshold of $|z|>=3.5$ ($p<.0006$) and a cluster-wise (extent) threshold of $k>8$ voxels. This combination of thresholds has been demonstrated using Monte Carlo simulations to provide an overall whole-brain Type I error rate of $p=.05$ given the present level of smoothing (McAvoy et al., 2001). To identify regions sensitive to multiple lexical variables (see Results), each whole-brain map was thresholded at a more liberal voxel-wise threshold of $p<.05$ (two-tailed, uncorrected) and all clusters of 9 or more voxels that showed at least 6 out of 8 lexical effects were considered significant. Note that a criterion of 6/8 tests is relatively conservative, as some regions could potentially show meaningful effects for only 3 or 4 variables, and would be missed by this analysis. Nonetheless, because our primary aim was to validate the present methodological approach, we chose to emphasize Type I error protection over Type II error protection. Finally, for all ROI-based analyses, we employed an uncorrected threshold of $p<.05$.

Results

Application to reading

The ERR approach permitted analysis of any combination of variables that could be quantified on a word-by-word basis. Here we focused on variables that have been extensively studied in previous fMRI and psycholinguistic studies of word reading. Thus, an initial probe set of 8 lexical, contextual, and behavioral predictors was used to identify candidate regions-of-interest (ROIs) broadly involved in lexical processing. Subsequent statistical models that included phonological and semantic variables further elaborated on the functions supported by these regions. Particular attention was focused on the VWFA, a region of considerable interest in recent fMRI studies (Cohen et al., 2000; Gaillard et al., 2006; McCandliss et al., 2003; Vigneau et al., 2005; Vinckier et al., 2007), and in which the present results identified a strikingly
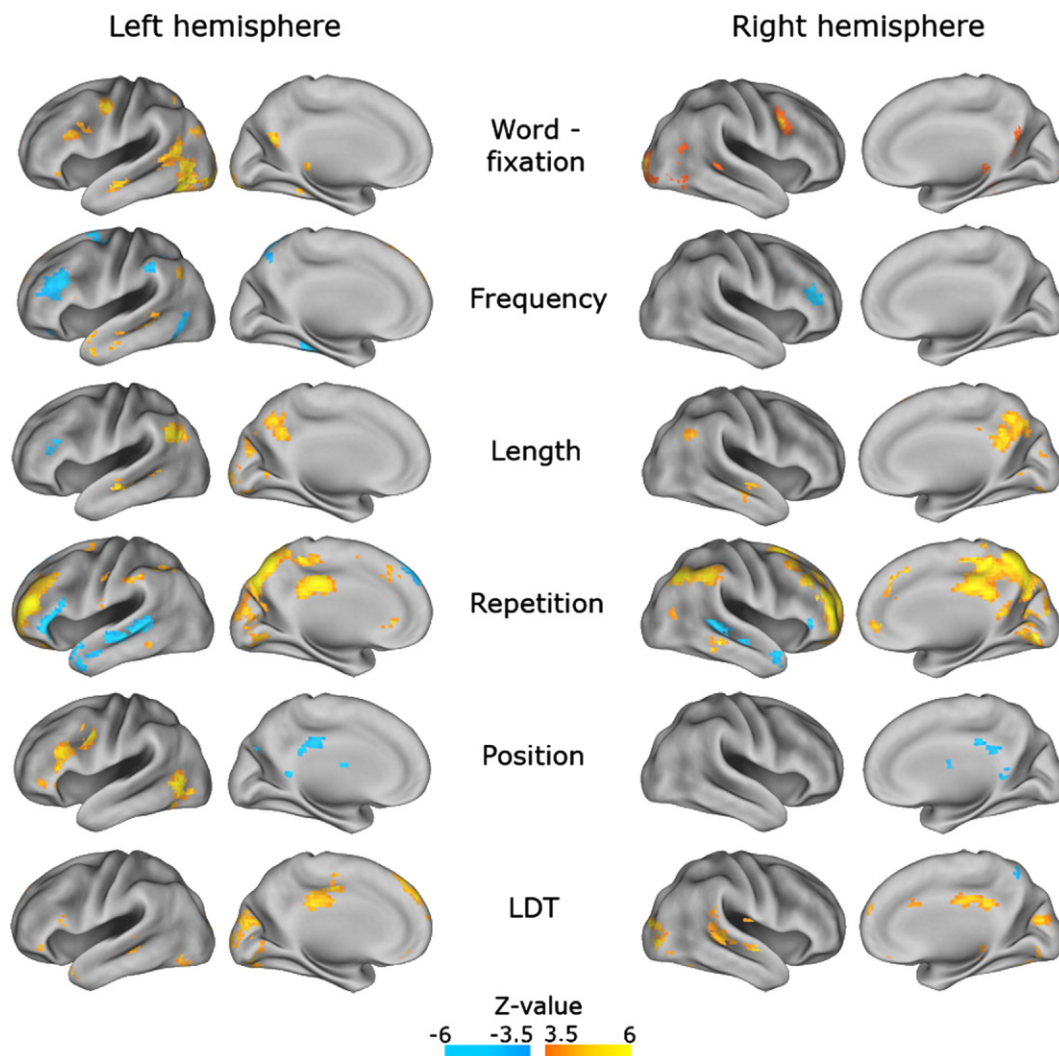

Fig. 2. Regions that showed a significant association with selected lexical variables. All regions consist of at least 9 contiguous voxels, each significant at $p<.0006$ ($|z|>3.5$), uncorrected. Orange/yellow activations denote positive effects, blue/turquoise activations denote negative effects. For full details see Table 2.

Table 2
Regions in which activation correlated with word reading relative to baseline or word-level differences in lexical variables in the first model tested (i.e., excluding quadratic length, phonological, or imageability regressors)

Variable/region	Hem.	BA	x	y	z	mm³	Word – fix (t)	t	t(rep.)	Comp. (r)
Word – fixation										
Inferior occipital gyrus	L	19	−36	−73	−3	23679	12.30	–	**10.83**	**0.45**
Precentral gyrus	L	6	−47	−3	43	3564	8.04	–	**7.44**	0.09
Parahippocampal gyrus	L		−14	−40	2	594	7.78	–	**2.88**	0.24
Middle temporal gyrus	L	21	−54	−10	−17	729	7.64	–	**5.71**	0.16
Inferior frontal gyrus	L	47	−46	31	−8	297	6.34	–	**3.79**	-0.32
Superior parietal lobule	L	7	−25	−58	43	1026	5.94	–	**4.28**	−0.09
Middle frontal gyrus	L	46	−44	20	21	648	5.64	–	**4.54**	0.23
Inferior frontal gyrus	L	9	−41	9	26	621	2.94	–	**4.52**	0.18
Superior frontal gyrus	L	10	−26	53	23	243	−5.18	–	**−3.90**	0.03
Posterior cingulate	M	23	3	−58	17	2484	7.30	–	**4.28**	0.24
Middle occipital gyrus	R	18	27	−91	4	10395	9.74	–	**10.79**	0.16
Hippocampus	R		30	−33	2	2727	9.71	–	**6.17**	−0.05
Fusiform gyrus	R	37	44	−61	−8	2619	9.36	–	**6.05**	0.16
Inferior frontal gyrus	R	9	38	5	31	4833	8.70	–	**6.95**	−0.01
Middle temporal gyrus	R	19	43	−61	13	1350	7.76	–	**4.74**	0.47
Middle temporal gyrus	R	21	51	−12	−20	810	6.74	–	**4.22**	−0.07
Fusiform gyrus	R	37	28	−39	−15	729	6.64	–	**2.90**	0.48
Superior parietal lobule	R	7	26	−57	44	270	6.03	–	**4.72**	0.42
Superior frontal gyrus	R	6	18	25	56	567	5.67	–	1.52	0.45
Caudate	R		14	14	−3	378	−5.26	–	**−3.00**	−0.09
Superior temporal gyrus	R	22	49	−9	4	837	−6.29	–	**−3.63**	−0.18
Frequency										
Middle Temporal Gyrus	L	21	−49	8	−23	1296	1.01	**8.32**	2.33	**0.41**
Superior Frontal Gyrus	L	6	−11	33	54	2889	1.93	**7.21**	4.62	**0.42**
Middle Temporal Gyrus	L	39	−52	−63	28	1458	0.82	**6.60**	4.71	0.28
Inferior Temporal Gyrus	L	20	−52	−9	−25	243	−0.29	**5.92**	2.88	0.32
Middle Temporal Gyrus	L	21	−58	−37	−4	2079	1.05	**5.86**	4.82	0.22
Inferior Frontal Gyrus	L	44	−56	18	11	648	**2.26**	**5.79**	0.14	−0.09
Middle Temporal Gyrus	L	21	−55	−16	−13	297	**3.76**	**5.30**	2.93	0.19
Superior Frontal Gyrus	L	9	−13	56	32	1134	1.84	**3.25**	3.81	0.10
Inferior Frontal Gyrus	L	47	−24	29	−8	567	1.00	**−4.57**	−1.28	−0.21
Precuneus	L	7	−28	−63	34	405	**2.63**	**−4.89**	−1.60	−0.19
Inferior Parietal Lobule	L	40	−38	−49	41	756	0.50	**−5.54**	−3.33	−0.28
Inferior Parietal Lobule	L	40	−59	−35	39	2052	−1.38	**−6.00**	−1.08	−0.06
Superior Frontal Gyrus	L	6	−26	−2	58	3051	**2.56**	**−7.06**	−1.96	−0.19
Middle Frontal Gyrus	L	46	−43	32	22	4455	0.14	**−7.68**	**−5.00**	**−0.37**
Precuneus	L	7	−16	−70	51	1998	0.56	**−7.79**	−2.49	0.00
Fusiform Gyrus	L	37	−42	−49	−14	4644	**3.91**	**−9.03**	**−3.90**	**−0.42**
Posterior cerebellum	R		25	−82	−34	1026	−0.33	**6.55**	3.17	**0.41**
Middle Frontal Gyrus	R	9	45	12	35	270	**3.30**	**−6.12**	−0.95	−0.04
Middle Frontal Gyrus	R	46	45	33	18	2430	0.34	**−9.92**	**−3.07**	−0.09
Length										
Middle Temporal Gyrus	L	39	−51	−60	25	4833	1.06	**7.80**	**4.55**	0.18
Middle Temporal Gyrus	L	21	−54	−20	−13	1134	**−2.07**	**6.75**	1.73	0.06
Lingual Gyrus	L	18	−12	−87	−14	2241	**−3.85**	**6.36**	**5.95**	0.23
Lingual Gyrus	L	18	0	−64	−1	702	**2.20**	**6.05**	1.17	0.17
Thalamus	L		−24	−25	0	243	**3.36**	**5.62**	0.58	−0.10
Middle Occipital Gyrus	L	18	−19	−99	14	486	2.81	**5.09**	**3.75**	**0.41**
Middle Temporal Gyrus	L	21	−61	−45	0	891	1.34	**4.96**	2.82	0.08
Superior Frontal Gyrus	L	9	−12	56	32	297	−0.72	**4.32**	1.91	0.20
Fusiform Gyrus	L	37	−43	−50	−17	243	1.35	**−5.66**	−1.41	−0.13
Middle Frontal Gyrus	L	46	−44	34	17	1296	0.75	**−5.95**	**−3.28**	−0.28
Precuneus	M	7	0	−61	30	15309	−0.17	**7.14**	**4.10**	0.24
Middle Temporal Gyrus	R	21	57	−18	−11	2430	−0.76	**8.85**	**5.14**	0.28
Lingual Gyrus	R	18	4	−97	−2	1404	0.36	**6.89**	1.97	0.23
Angular Gyrus	R	39	53	−59	32	1566	0.97	**5.96**	2.40	−0.06
Middle Temporal Gyrus	R	21	41	−40	4	351	0.51	**5.80**	1.80	−0.10
Lingual Gyrus	R	18	17	−80	−16	1566	**−2.43**	**5.70**	**3.10**	0.24
Middle Occipital Gyrus	R	18	19	−98	15	405	**3.71**	**5.25**	1.48	−0.10
Superior Frontal Gyrus	R	6	11	27	57	297	**2.52**	**5.03**	2.27	0.26
ON										
Superior Temporal Gyrus	L	38	−56	−60	25	270	0.06	**4.97**	0.40	−0.01
Fusiform Gyrus	L		−29	−37	−17	594	**2.09**	**−5.70**	**−2.20**	−0.24
Middle Temporal Gyrus	R	29	46	−34	0	702	**2.70**	**6.40**	1.09	0.25
LDT										
Thalamus	L		−25	−28	6	1215	0.63	**7.73**	1.07	0.32
Superior Temporal Gyrus	L	38	−46	11	−27	459	1.60	**6.80**	0.21	0.30
Medial cerebellum	L		−6	−60	−33	324	0.17	**6.57**	1.46	0.14
Medial Frontal Gyrus	L	9	−6	50	36	5184	1.38	**6.25**	1.17	0.08

Table 2 (continued)

Variable/region	Hem.	BA	x	y	z	mm³	Word – fix (t)	t	t(rep.)	Comp. (r)
LDT										
Inferior Frontal Gyrus	L	47	−49	25	−7	1296	1.20	**5.80**	1.56	0.06
Inferior Frontal Gyrus	L	44	−51	16	12	675	**2.03**	**5.77**	0.79	−0.08
Inferior Occipital Gyrus	L	18	−46	−83	−1	891	**2.97**	**5.76**	0.41	0.21
Fusiform Gyrus	L	19	−31	−74	−17	1755	1.57	**5.73**	1.13	0.12
Inferior Parietal Lobule	L	40	−50	−57	45	540	−0.76	**5.70**	0.78	−0.05
Precentral Gyrus	L	6	−59	4	21	405	1.90	**5.51**	−0.53	0.00
Caudate	L		−13	8	8	459	−1.17	**5.50**	−0.06	0.27
Middle Temporal Gyrus	L	21	−60	−38	0	1998	**3.65**	**5.31**	2.56	0.08
Middle Occipital Gyrus	L	18	−23	−89	12	513	1.67	**5.31**	1.11	0.19
Medial Frontal Gyrus	M	11	0	40	−11	756	0.42	**7.73**	−0.35	0.27
Cingulate Gyrus	M	23	−1	−26	33	4887	**−3.85**	**7.18**	2.49	0.23
Cingulate Gyrus	M	24	0	16	30	729	**−2.43**	**5.82**	0.16	0.22
Precuneus	M	7	5	−63	51	270	**−2.07**	**−5.22**	−0.25	0.25
Middle Occipital Gyrus	R	19	37	−77	3	3321	**5.03**	**8.29**	0.48	0.30
Insula	R	13	34	−6	19	675	0.37	**7.94**	0.89	0.09
Lingual Gyrus	R	18	4	−80	−2	12042	0.39	**7.83**	1.02	0.15
Middle Temporal Gyrus	R	22	53	−32	4	4671	**2.11**	**7.35**	2.65	0.08
Thalamus	R		20	−31	1	540	**3.59**	**5.72**	0.98	−0.04
Insula	R	13	32	−37	18	297	0.70	**5.25**	−0.73	0.02
Sub-Gyral	R		29	−27	−4	351	**4.33**	**5.07**	−0.30	0.33
Naming										
Precuneus	M	23	5	−60	21	1323	**2.65**	**−6.68**	**−2.84**	−0.16
Cuneus	R	19	9	−84	36	378	**−2.24**	**−5.29**	0.01	0.07
Lingual Gyrus	R	19	17	−56	−9	270	−1.77	**−5.63**	−0.50	0.04
Position										
Repetition										
Inferior Parietal Lobule	L	40	−39	−57	46	5859	−1.26	**9.44**	1.77	0.21
Caudate	L		−29	30	14	31104	**−2.43**	**9.41**	**3.56**	0.32
Lateral cerebellum	L		−33	−48	−46	9666	−1.29	**9.23**	**2.91**	0.31
Middle Occipital Gyrus	L	37	−48	−62	−5	6345	**6.11**	**8.07**	**3.89**	0.37
Inferior Frontal Gyrus	L	9	−47	17	22	5157	**4.30**	**7.91**	**4.39**	0.25
Inferior Parietal Lobule	L	40	−60	−35	39	2349	−0.98	**7.18**	0.97	0.28
Middle Frontal Gyrus	L	8	−25	−2	59	1998	**2.13**	**5.59**	0.57	0.20
Fusiform Gyrus	L	20	−40	−36	−13	378	1.25	**5.58**	−0.51	0.28
Sub-Gyral	L	10	−44	41	−3	432	0.15	**5.56**	−1.17	−0.21
Lateral cerebellum	L		−43	−73	−36	324	−0.65	**5.54**	−0.37	0.08
Cingulate Gyrus	L	31	−19	−56	25	351	0.03	**5.53**	0.11	−0.17
Inferior Frontal Gyrus	L	47	−46	26	−4	486	**3.10**	**5.32**	−2.18	0.02
Inferior Temporal Gyrus	L	37	−57	−47	−16	297	0.66	**5.26**	−0.59	0.05
Medial cerebellum	L		−8	−78	−35	243	1.30	**5.22**	0.02	0.06
Transverse Temporal Gyrus	L	41	−32	−46	15	243	0.03	**5.15**	−0.02	0.05
Transverse Temporal Gyrus	L	42	−63	−11	13	648	−0.80	**5.10**	0.46	0.17
Superior Occipital Gyrus	L	19	−31	−78	31	405	**2.69**	**4.98**	1.41	0.24
Precuneus	L	19	−35	−76	43	297	−0.24	**4.98**	0.99	0.18
Precentral Gyrus	L	6	−43	−14	30	405	−0.03	**4.79**	0.23	0.27
Medial cerebellum	L		−6	−56	−11	270	0.44	**4.23**	−1.29	0.14
Superior Temporal Gyrus	L	22	−57	−58	21	432	1.57	**−2.33**	**−2.82**	0.09
Cuneus	L	18	−11	−72	23	351	−1.27	**−5.95**	−0.49	0.04
Superior Frontal Gyrus	L	8	−11	42	46	4644	**2.13**	**−7.03**	**−3.55**	−0.17
Middle Temporal Gyrus	L	21	−55	−32	−4	5967	1.60	**−8.30**	**−5.94**	−0.05
Inferior Frontal Gyrus	L	45	−50	24	5	4482	**2.41**	**−8.36**	**−4.03**	−0.05
Middle Temporal Gyrus	L	21	−50	6	−24	3915	**2.07**	**−9.31**	**−6.18**	−0.15
Precuneus	M	7	1	−62	34	70956	**−2.23**	**8.80**	**5.68**	0.41
Cingulate Gyrus	M	32	0	25	35	4671	**−2.63**	**6.77**	**3.52**	0.34
Posterior cerebellum	M		0	−83	−36	702	−1.16	**5.35**	**2.91**	−0.07
Cingulate Gyrus	M	24	−2	−1	47	243	−1.93	**4.72**	0.32	0.28
Cuneus	M	19	0	−84	33	243	**−3.20**	**−4.23**	−0.40	−0.04
Posterior Cingulate	M	29	1	−48	4	945	−0.87	**−5.53**	−1.80	0.00
Cingulate Gyrus	M	23	0	−34	28	1728	**−4.20**	**−6.45**	−1.22	−0.02
Middle Frontal Gyrus	R	9	29	38	24	31077	−1.51	**9.84**	**3.82**	**0.41**
Inferior Parietal Lobule	R	40	44	−55	44	14769	−0.13	**8.68**	**2.69**	0.24
Middle Temporal Gyrus	R	21	62	−37	−13	837	−0.24	**7.56**	−1.59	0.01
Lateral cerebellum	R		41	−59	−44	1971	**−2.03**	**7.04**	1.12	0.01
Medial cerebellum	R		30	−38	−39	540	−1.28	**6.79**	0.62	0.00
Caudate	R		18	6	24	2349	0.38	**6.58**	1.08	**0.44**
Superior Temporal Gyrus	R	22	54	−10	6	594	**−3.31**	**6.53**	0.99	0.22
Transverse Temporal Gyrus	R	41	34	−39	14	864	0.71	**6.40**	1.19	0.25
Middle Temporal Gyrus	R	37	44	−66	11	324	**2.99**	**6.12**	0.00	0.20
Posterior cerebellum	R		2	−72	−49	378	−1.45	**6.08**	−0.34	0.22
Medial cerebellum	R		24	−67	−33	405	0.74	**5.77**	1.11	**0.43**
Medial cerebellum	R		22	−43	−52	1593	**−2.48**	**5.13**	0.02	0.11
Lentiform Nucleus	R		15	13	−5	324	**−4.37**	**4.97**	1.58	−0.02

Table 2 (continued)

Variable/region	Hem.	BA	x	y	z	mm³	Word – fix (t)	t	t(rep.)	Comp. (r)
Repetition										
Thalamus	R		1	−25	10	486	−1.21	**−5.31**	−1.89	−0.18
Middle Temporal Gyrus	R	21	50	−18	−11	297	0.48	**−5.48**	**−2.21**	0.07
Thalamus	R		1	−7	13	513	−0.13	**−5.58**	**−2.07**	−0.18
Middle Temporal Gyrus	R	21	48	8	−23	864	0.97	**−5.95**	**−5.94**	−0.18
Middle Temporal Gyrus	R		50	−35	0	1998	1.35	**−7.00**	**−4.50**	0.23
Inferior Frontal Gyrus	R	45	54	30	5	864	1.54	**−7.51**	**−4.54**	0.16

All regions were defined at a whole-brain level. Word – fix. (t) column indicates t-value for the reading – baseline contrast, in order to provide a general sense of whether each region was activated or deactivated overall. For the lexical effects, t values are displayed separately for the initial study and the replication sample (t(rep.)). Comp. (r): correlation between regional activation and individual differences in performance on the post-scan reading comprehension test. Hem: hemisphere; BA: Brodmann Area; x,y,z: Talairach co-ordinates of regional center of mass; mm³: regional volume in cubic millimeters. Bold values denote significant replications of lexical effects, or significant correlations with comprehension test performance ($p < .05$, uncorrected).

broad range of orthographic, phonological, contextual, and semantic effects.

Identification of ROIs sensitive to lexical variables

To identify a set of candidate brain regions involved in lexical access processes, we searched for regions in which activation was modulated by multiple word-level variables known to influence behavioral measures of lexical processing. The search was conducted at a whole-brain level so as not to bias the analysis in favor of regions previously identified by single-word reading studies. The probe set of variables was deliberately selected to span multiple levels of language processing, and included eight different orthographic, lexical, and contextual variables. The first variable coded for the word reading vs. fixation contrast. The next three variables coded for differences in word frequency, length and orthographic neighborhood size. These variables are strong predictors of a wide range of reading and language behaviors, including lexical decision latencies, pronunciation latencies, and eye fixation durations during reading, and reflect fundamental components of virtually all extant models of visual word recognition (Coltheart et al., 2001; Plaut et al., 1996). Another pair of variables modeled the influence of behavioral norms for lexical decision task (LDT) and word naming latencies from an independent set of subjects. These behavioral measures remain the most widely-used indices of lexical access processes, yet existing models are capable of explaining only about half of the variance in these measures (e.g., Balota et al., 2004; Yarkoni et al., in press-a). They were included in the present model in order to determine whether behavioral norms are capable of predicting brain activation above and beyond the contributions of standard lexical predictors. Such a finding would suggest that much of the variance in these behavioral variables reliably reflects neural processes that are not currently understood. Finally, two variables were used to model contextual influences on word reading: The position of a word within its enclosing sentence, and the number of times a word had previously been presented within the narrative.

Each individual variable produced reliable and replicable changes in brain activation. (Fig. 2, Table 2). For present purposes, we focus on those regions that showed concerted influences of multiple variables. To identify such ROIs, we searched for brain regions that showed at least a weak association ($p < .05$, two-tailed, uncorrected) with 6 or more lexical variables (Fig. 3, Table 3). The resulting set of ROIs included several brain regions canonically implicated in language comprehension and production in neuropsychological studies, e.g., left inferior frontal gyrus (IFG), middle temporal gyrus (MTG), and temporoparietal cortex, as well as a region of left inferotemporal cortex dubbed the *visual word form area* (VWFA) in recent lesion and neuroimaging studies (Cohen et al., 2000; McCandliss et al., 2003). Additionally, several regions that have received relatively little attention in previous neuroimaging studies of language were identified, including ROIs in the precuneus, posterior cingulate, and dorsomedial prefrontal cortex (Table 3); however, because the role of these in language is not well understood, the present analyses focus primarily on canonical language regions and the VWFA.

Inspection of the correlations between the lexical variables and brain activation suggested a qualitative dissociation between the response properties of the VWFA and those of regions classically implicated in amodal language comprehension and production (Table 3). In both the VWFA and canonical language regions (including left temporoparietal cortex and bilateral middle temporal gyrus and inferior frontal gyrus), activation increased during reading relative to fixation, decreased with increasing number of word repetitions, and correlated positively with LDT and word position. However, differential responses were observed to the three standard lexical variables of length, frequency, and ON. Whereas activation in canonical language regions correlated positively with all three variables, activation in the VWFA correlated negatively in all three cases (Fig. 4).

Optimized word form processing in the VWFA

The fact that the VWFA response to visually presented words diverged from other language regions specifically for the variables of length, frequency, and ON is instructive given that these three variables were defined on the basis of orthographic properties, in contrast to the behavioral and contextual variables. Several models of VWFA function ascribe it a primarily visual/orthographic role in language processing (Cohen et al., 2000; McCandliss et al., 2003). An influential hypothesis is that the VWFA contains neuronal assemblies optimized to detect and identify orthographically regular letter strings (McCandliss et al., 2003). The present finding that VWFA activation correlated negatively with orthographic neighborhood size and word frequency supports such a hypothesis. Words with many neighbors are, by definition, more similar to other words in the lexicon, and the visual system should presumably be optimized to detect such words more rapidly. Similarly, the negative relationship between VWFA activation and word frequency has been interpreted as evidence that VWFA neurons process frequently encountered words more fluently than infrequent words (Kronbichler et al., 2004; Kuo et al., 2003).
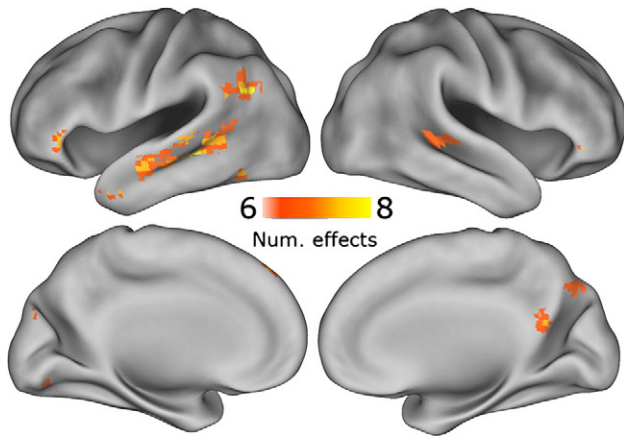
Fig. 3. Brain regions in which activation correlated with at least 6 out of 8 lexical effects (each at $p<.05$, uncorrected). See Table 2 for details.

More problematic for an orthographic fluency account of VWFA function was the inverse correlation between VWFA activation and word length. In behavioral studies, longer words typically produce slower responses (Balota et al., 2004; Weekes, 1997), suggesting that long words should be processed *less* efficiently by the VWFA. However, recent work suggests the effects of word length are non-linear: Lexical decision latencies decrease with length for words 2–7 letters in length, and increase for words longer than 8–9 letters (New et al., 2006). A powerful but non-intuitive prediction is that if the VWFA is indeed optimized for rapid detection of visual word forms of an optimal or average length, it should show a curvilinear activation profile that parallels behavioral RTs. Conversely, amodal language regions that support post-orthographic processes should not.

To test this prediction, we computed a new multiple regression model that added a quadratic length regressor to the existing variables. A significant quadratic trend was indeed found in the VWFA. Length correlated most negatively with activation for words 7–9 letters in length ($t(27)=2.37$, $p<.03$; Fig. 5). Only one other ROI showed a similar U-shaped relationship with word length: In the precuneus, activation was the mirror inverse of the VWFA, increasing with length to a peak around 7–9 letters and subsequently decreasing (Fig. 5; $t(27)=-3.91$, $p<.001$). Three other ROIs showed a significant but qualitatively different quadratic trend, with activation increasing exponentially as a function of length (Fig. 5; dorsomedial PFC ($t(27)=3.34$, $p<.002$; left inferior frontal

gyrus ($t(27)=2.37$, $p<.04$; right medial temporal gyrus ($t(27)=2.49$, $p<.02$).

Orthographic versus phonological coding

The finding that the VWFA appears to be optimized for processing orthographically valid letter strings does not require that the representational coding scheme supported in the VWFA be a purely visual/orthographic one. VWFA activation is modulated not only by visual properties such as orthographic regularity but also by phonological processing demands (Dietz et al., 2005; Sandak et al., 2004b; Xue et al., 2006). Recent behavioral evidence similarly suggests that some effects conventionally attributed to orthography, e.g., neighborhood size effects, may have a phonological basis (Yates, 2005; Yates et al., 2004). Thus, it remains unclear at what stage of processing letter strings are first parsed into higher-order units such as phonemes, syllables, or morphemes. An important question is whether the VWFA's role in visual word recognition is restricted to providing orthographic inputs to other brain regions (e.g., left temporoparietal cortex) for phonological decoding, or whether it partly implements the computations involved in grapheme-to-phoneme conversion, as several researchers have suggested (e.g., Sandak et al., 2004a).

To address this question, a second statistical model was computed that added three additional variables to the existing set of predictors. One variable coded for phonological neighborhood size (PN; Yates, 2005), calculated in a manner analogous to the standard definition of orthographic neighborhood size (Coltheart et al., 1977). Each word's PN value was equal to the number of words in the lexicon that could be produced by changing exactly one phoneme. By modeling ON and PN simultaneously, we could quantify the extent of orthographic versus phonological neighborhood effects in the VWFA. The other two variables coded for the number of syllables and number of morphemes in each word. Syllable and morpheme counts are highly intercorrelated with both word length and each other (Table 1), but when the three variables are estimated simultaneously, their parameter estimates should provide relatively uncontaminated indices of phonological, morphological, and orthographic processes, respectively.

The results suggested that there were both orthographic and phonological effects in the VWFA. The length effect in VWFA remained statistically significant when controlling for number of syllables and morphemes ($t(27)=-3.53$, $p<.002$;

Table 3
Lexical effects in regions in which activation correlated with at least 6 lexical variables in the probe set

Region	Hem	BA	x	y	z	mm³	Word – fix.	Length	Freq.	ON	LDT	Naming	Position	Repetition
Middle temporal gyrus	L	21	−56	−34	−4	6480	**5.50**	**6.03**	**5.97**	**3.17**	**4.86**	−1.94	**4.25**	**−7.30**
Superior temporal gyrus	L	39	−53	−60	23	2538	**3.17**	**7.51**	**5.37**	**4.21**	**3.86**	−2.88	**3.07**	**−4.69**
Superior frontal gyrus	L	8	−12	47	42	1512	**3.18**	**4.89**	**7.08**	**2.44**	**5.33**	−1.92	**3.37**	**−6.14**
Inferior frontal gyrus	L	47	−50	27	−1	1377	**2.69**	**2.52**	**3.99**	**3.17**	**4.10**	2.06	**4.49**	**−6.49**
Middle temporal gyrus	L	21	−48	7	−29	567	**2.30**	**5.05**	**4.94**	**3.26**	**3.50**	−2.61	**2.88**	**−5.31**
Fusiform gyrus	L	37	−42	−57	−15	459	**7.65**	**−2.96**	**−3.72**	**−2.18**	**2.68**	1.51	**5.19**	**−4.10**
Lingual gyrus	L	18	−12	−75	−13	351	**−3.15**	**3.93**	**2.16**	**3.83**	**2.87**	0.22	**−2.51**	**3.69**
Cuneus	M	19	6	−80	34	2025	**−3.33**	**4.38**	0.80	1.19	**4.50**	**−4.80**	**−3.73**	**6.22**
Posterior cingulate	M	23	4	−55	18	729	**4.63**	**3.38**	0.34	**−2.71**	**3.09**	**−3.40**	**−2.29**	**3.60**
Precuneus	M	31	1	−69	29	243	**3.14**	**4.69**	**−2.05**	−0.71	**2.22**	**−3.27**	**−2.56**	**3.93**
Superior temporal gyrus	R	22	49	−34	2	1404	**5.29**	**4.02**	**3.49**	**4.08**	**5.17**	0.67	0.65	**−6.25**
Inferior frontal gyrus	R	47	52	30	−4	324	1.25	0.26	**2.19**	1.61	1.43	**4.21**	1.36	**−3.31**
Posterior cerebellum	R		23	−79	−36	243	1.79	**4.33**	**4.86**	**2.25**	**3.05**	−1.96	**2.73**	**−3.89**

Columns for the 8 tested effects represent *t*-values. Bold values are significant at $p<.05$, two-tailed, uncorrected. ON = Orthographic Neighborhood Size; LDT= Lexical Decision Task response latencies.

Table 4), suggesting a genuine visual effect of string length rather than correlated higher-order variables. In contrast, the putative ON effect appeared to have a phonological basis: VWFA activation correlated negatively with PN ($t(27)=-2.52$, $p<.02$) and positively with syllable count ($t(27)=2.78$, $p<.005$), but

showed no remaining association with ON ($t(27)=-.7$, ns). No unique effect of morpheme count was found in the VWFA.

Like the VWFA, most other canonical language regions, including the left temporoparietal area and bilateral IFG and MTG, also showed significant effects of syllable count and PN
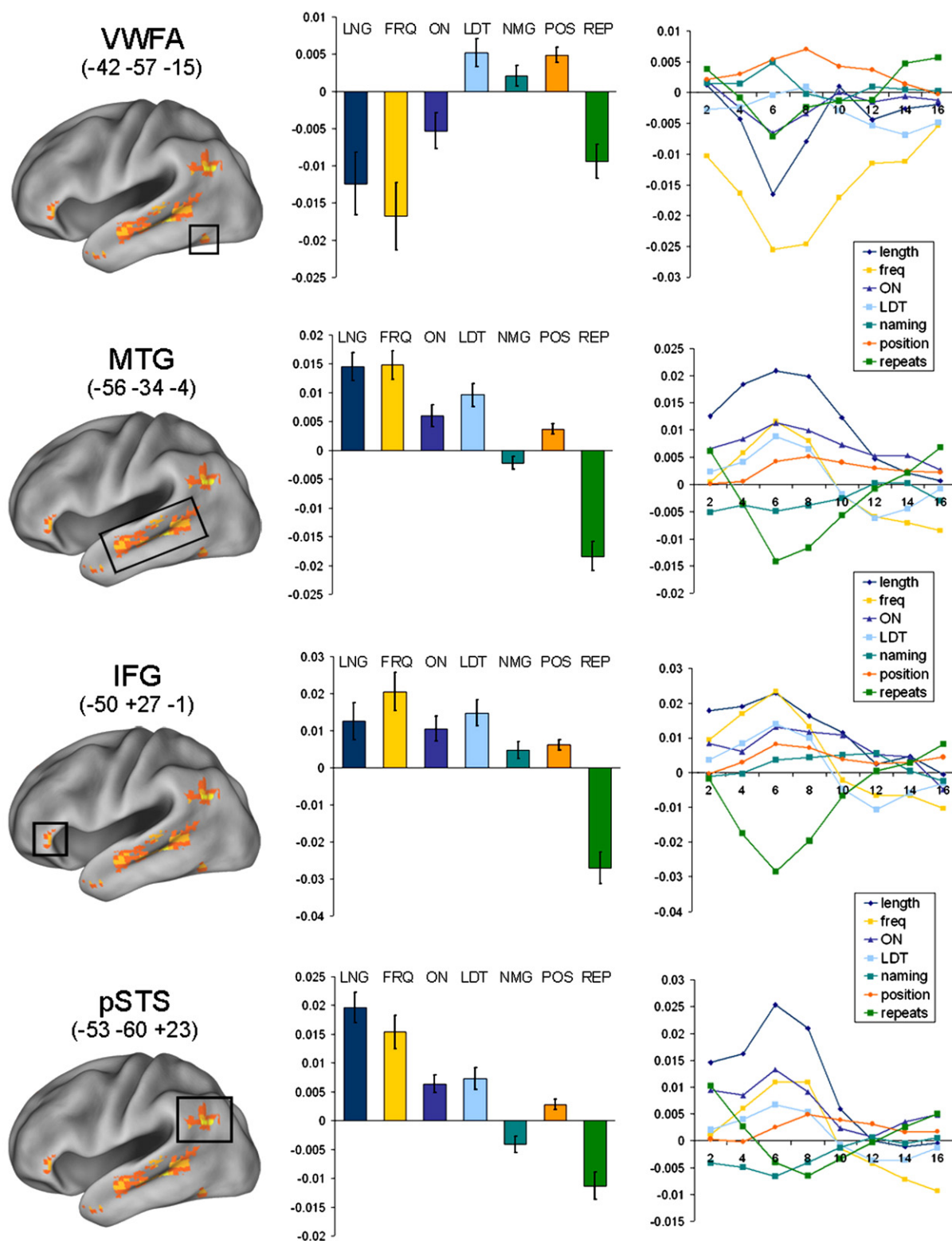
Fig. 4. Correlations between word-level variables and activation in selected functionally-defined ROIs (see Table 3 for details). Each row displays summary magnitude estimates and detailed timecourses for a single region. Error bars represent one standard error of the mean. Units on the *y*-axis represent % change in BOLD signal per standard deviation in each lexical variable. For timecourses, units on the *x*-axis represent time (in seconds) elapsed since word onset.
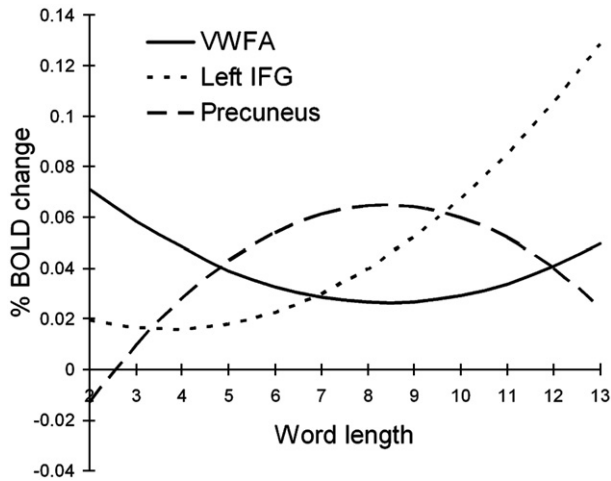
Fig. 5. Estimated % BOLD signal change (*y*-axis) as a function of word length in ROIs that showed a significant quadratic effect of length. Activation in the right MTG and dorsomedial PFC ROIs (not illustrated) showed a quadratic trend virtually identical to the left IFG.

but not of morpheme count or ON, though the direction of the PN effect was positive rather than negative (Table 4). Interestingly, word length effects were substantially reduced in all five regions, and remained significant only in the right MTG and left temporoparietal area. Of the five canonical language regions, only right MTG showed a significant correlation with number of morphemes ($t(27)=2.29$, $p<.03$).

Effects of word imageability

A final question addressed was whether the VWFA supports semantic processing of words in addition to its role in orthography and phonology. Rapid processing of semantic attributes might be expected based on the close spatial proximity of the VWFA to other inferotemporal areas associated with object perception (e.g., the fusiform face area; Kanwisher et al., 1997), as well as behavioral evidence indicating that semantic variables such as imageability exert a sizeable influence on lexical decision latencies (Balota et al., 2004). Several neuroimaging studies have found a positive correlation between VWFA activation and word imageability or concreteness (Bedny and Thompson-Schill, 2006; Sabsevitz et al., 2005; Wise et al., 2000). However, these studies allowed several seconds of processing per trial and used experimental tasks that explicitly directed attention to semantic properties of the stimuli. The

increased VWFA activation observed for highly imageable words might therefore reflect top-down reactivation of object-specific IT cortex during conscious construction of mental images rather than relatively automatic semantic processing (for discussion, see Balota et al., 2006; Dehaene et al., 2002).

To test whether VWFA activation correlated with word imageability in the present paradigm, we added imageability norms (Toglia and Battig, 1978) to the multiple regression model, retaining all earlier variables (including the phonological variables). (Norms were available for only 68.2% of the words in the narratives; the others were left unmodeled; see supplemental analyses below.) A positive correlation was observed between VWFA activation and imageability ($t(27)=3.35$, $p<.002$). No other ROI showed a positive correlation with imageability. Interestingly, when imageability was included in the model, the association between VWFA activation and frequency was eliminated ($t(27)=-.05$, *ns*), likely due to the strong correlation between the two variables ($r=-.74$ across stimulus words). The effects of other variables on VWFA activation generally remained significant ($p<.06$ for length, $p<.05$ for PN, naming, repetition, and position). Thus, even in the most conservative model tested, VWFA activation was independently influenced by a broad range of orthographic, phonological, contextual and semantic variables.

Additional validation analyses

The above results demonstrated that the ERR approach produces robust, theoretically informative results that are broadly consistent with previous single-word fMRI studies. However, several issues remained unaddressed. First, the inclusion of incomplete imageability norms in the above model assumed that the ERR approach is largely unaffected by data omission, which requires empirical validation. Second, none of the models tested above controlled for potential confounding effects of grammatical class. Finally, while the presence of statistically significant results is compelling, a demonstration that the ERR effects identified are replicable and predict meaningful behavioral criteria would be still more convincing. To address these concerns, we conducted a series of additional analyses.

Model resilience to word omissions

For the multiple regression model that included imageability, a large proportion of words (31.8%) could not be modeled

Table 4
Lexical effects for regions in Table 3 following inclusion of phonological and morphological variables

Region	Hem	BA	*x*	*y*	*z*	mm³	Word – fix.	Length	Freq.	ON	PN	LDT	Nmng.	Syll.	Morph.	Pos.	Rep.
Middle temporal gyrus	L	21	−56	−34	−4	6480	**6.06**	0.69	**5.80**	1.48	**2.15**	**3.95**	0.13	**4.83**	1.84	**4.20**	**−7.16**
Superior temporal gyrus	L	39	−53	−60	23	2538	**3.45**	**2.96**	**5.21**	0.63	**3.26**	**3.28**	−1.47	**3.46**	0.61	**2.82**	**−4.07**
Superior frontal gyrus	L	8	−12	47	42	1512	**3.47**	−0.50	**6.91**	−0.82	**3.68**	**4.02**	−0.94	**5.28**	1.90	**3.59**	**−5.17**
Inferior frontal gyrus	L	47	−50	27	−1	1377	**3.14**	−1.61	**3.89**	1.73	**2.82**	**3.12**	**3.53**	**3.49**	0.89	**4.58**	**−6.06**
Middle temporal gyrus	L	21	−48	7	−29	567	**3.49**	−0.33	**3.66**	0.61	**2.25**	**2.73**	−1.47	**4.01**	1.29	**2.72**	**−5.22**
Fusiform gyrus	L	37	−42	−57	−15	459	**7.27**	**−3.53**	**−3.98**	−0.70	**−2.52**	1.20	**3.18**	**2.78**	0.61	**5.10**	**−3.93**
Lingual gyrus	L	18	−12	−75	−13	351	**−2.98**	**2.14**	**2.16**	**3.20**	−1.78	1.87	0.54	−0.86	1.55	**−2.62**	**3.67**
Cuneus	M	19	6	−80	34	2025	**−3.31**	**2.33**	1.10	**2.25**	**−2.78**	**3.25**	**−4.62**	**−2.00**	**2.61**	**−3.71**	**5.69**
Posterior cingulate	M	23	4	−55	18	729	**4.39**	0.63	0.35	−0.53	**2.13**	**2.13**	**−2.51**	−0.28	**3.29**	**−2.30**	**2.61**
Precuneus	M	31	1	−69	29	243	**3.15**	**2.85**	**−2.07**	0.24	−1.76	1.96	**−2.92**	−0.85	0.32	**−2.00**	**3.64**
Superior temporal gyrus	R	22	49	−34	2	1404	**5.16**	0.84	**3.93**	1.54	**2.38**	**4.59**	**2.36**	**3.39**	**2.29**	0.96	**−6.24**
Inferior frontal gyrus	R	47	52	30	−4	324	1.16	**−2.18**	1.92	0.41	1.40	0.89	**5.14**	**2.20**	1.63	1.38	**−3.04**
Posterior cerebellum	R		23	−79	−36	243	1.24	**3.64**	**4.48**	−0.11	**2.59**	**2.91**	−1.74	0.82	−0.47	**2.91**	**−2.59**

In addition to the 8 columns represented in Table 3, there are columns for number of syllables (Syll.), number of morphemes (Morph.), and Phonological Neighborhood Size (PN). Bold values are significant at $p<.05$, two-tailed, uncorrected.

because they lacked imageability norms. The ability to successfully estimate the influence of imageability on brain activation was therefore contingent on the model's robustness to item loss. Does omitting substantial numbers of words catastrophically affect the models' detection power? To address this, we computed a new model that randomly eliminated approximately 50% of all words for each subject but was otherwise identical to our first regression model (which included 7 lexical variables). Words were randomly and independently omitted for each participant; each word had a 50% change of being eliminated. Word variables were eliminated in a list-wise manner: either all variables (length, freq, etc.) were included for a word, or none were.

The model showed substantial robustness to word omission. Effects of the lexical variables were, unsurprisingly, somewhat weaker; however, detection power was still good and there was no evidence of a systematic increase in false positives. At an ROI level, 128 of 140 ROIs (91%) continued to show a significant effect at $p < .05$, uncorrected. Importantly, the majority of ROIs identified in Table 2 would have been detected even if the same whole-brain analysis had been conducted on the 50%-omission model. Fig. S1 compares the results of the two models for a sample variable (word frequency). Thus, far from requiring that every word in the stimuli be coded for, the present methodology appears to provide reliable estimates even when half of the words are omitted.

Controlling for grammatical class

A second potential problem associated with modeling word-level variables within the context of narrative text is that the grammatical class of the stimulus words cannot be constrained a priori, and may be naturally correlated with lexical variables. For example, function words tend to be shorter and more frequent than content words. There is ongoing debate as to whether words of different grammatical class recruit different neural systems (e.g., Bedny and Thompson-Schill, 2006; Longe et al., 2007; e.g., Shapiro et al., 2005). Might the correlations observed between lexical variables and neural activity reflect the confounding influence of grammatical class? To rule out this concern, we computed a new model that included dummy-coded regressors for different parts of speech (nouns, verbs, adjectives, adverbs, and closed-class words). Words were classified using an automated part-of-speech tagger (Toutanova et al., 2003). The inclusion of the syntactic variables had a negligible influence on the observed assocations between the lexical variables of interest and neural activity: 118 of 119 ROIs continued to show a significant effect at $p < .05$; the vast majority (112/119) remained significant at $p < .001$; and significance levels actually *increased* in 23 ROIs. (The 21 ROIs associated with the word reading vs. fixation contrast were not tested because this contrast was replaced by the 5 sets of syntax regressors.) Thus, the lexical results could not be attributed to confounding grammatical effects.

Replication of effects in an independent sample

As a stringent test of the reliability of the ERR approach, we assessed the degree to which the results of the present study (Experiment 1) could be replicated in an independent sample drawn from a different study (Experiment 2; Yarkoni et al., in press-b). In Experiment 2, subjects ($n = 29$) read 48 shorter narratives (each approximately 49 s long) describing the everyday activities of a seven year old boy. The design and

analytic approach in Experiment 2 were similar in most respects to Experiment 1. The main difference was that power to detect word-level effects in Experiment 2 was substantially lower than in Experiment 1 due to exclusion of data from a different reading condition conceptually dissimilar to Experiment 1. In brief, Experiment 2 included both a coherent reading condition similar to Experiment 1, and a 'scrambled' condition in which materials were coherent at the sentential but not at the narrative level (for details, see Yarkoni et al., in press-b). For present purposes, lexical regressors were estimated separately for each condition, and only the coherent condition was tested for replication of Experiment 1 effects. Statistical analysis in Experiment 2 was performed only at an ROI level; thus, all effects were tested at $p < .05$, uncorrected.

We attempted to replicate effects by testing each of the 140 ROIs identified in the first regression analysis of Experiment 1 (Table 2) for the same effect in the replication sample. Across all variables, 87 out of 140 ROIs showed a statistically significant effect ($p < .05$) in the same direction in Experiment 2 as in Experiment 1, a rate grossly exceeding chance expectation (binomial probability test, $p < 10^{-75}$). An additional 38 ROIs showed a non-significant effect in the same direction (Table 2), and only one ROI showed a significant effect in the opposite direction (versus the chance expectation of 7 false positives). Moreover, the rate of replication exceeded chance for five of the eight individual word variables (word vs. fixation: 20/21 ROIs; length: 10/18 ROIs; frequency: 13/19 ROIs; position: 3/13 ROIs; repetition, 16/39 ROIs; $p < .05$ for position, $p < .001$ for all other variables), and at least one ROI replicated for each of the remaining three variables. Thus, item-level differences in rapidly presented words exerted highly consistent effects on brain activity across the two experiments despite the differences in task structure.

Lexical activations predict individual differences in reading comprehension

A final validation analysis sought to demonstrate the external validity of the lexical estimates produced by the present approach. We reasoned that if the activation changes associated with lexical variation played a functional role in supporting reading, they should also predict individual differences in reading comprehension, as assessed on the post-scan multiple-choice comprehension test. To test this hypothesis, we correlated individual differences in comprehension performance with activation in the ROIs identified by the group-level analyses (Table 2). The analysis was performed separately for each of the lexical variables (e.g., comprehension was correlated with frequency-related activation in frequency-defined ROIs, length-related activation in length-defined ROIs, etc.). Collapsing across all variables, significant correlations ($p < .05$) were observed in 15 out of 140 ROIs (Table 2), a rate significantly different from chance (binomial probability test, $p < .006$). Moreover, in all 15 cases, the direction of the correlation between comprehension and brain activation mirrored the direction of the group-level effect for the sample as a whole (e.g., in regions where brain activation was greater for low-frequency words than high-frequency words, better-performing subjects showed even stronger responses to low-frequency words). This finding generalized to non-significantly correlated ROIs: better comprehension was associated with amplification of the group-level effect in 112 of 140 ROIs (binomial probability: $p < 4.2 \times 10^{-13}$). When very small effect sizes ($r < .1$) were excluded, the effect even stronger:

correlations in the group-average direction were observed in 85 of 92 ROIs (binomial probability: $p < 3.8 \times 10^{-18}$). Thus, individuals with better comprehension showed an amplified pattern of reading-related activation that spanned multiple lexical variables and brain regions.

Discussion

Methodological implications

The results of the present study demonstrate the viability of studying word-level processes using fMRI even when coherent narratives are presented at a relatively rapid rate. The ERR approach introduced in this study parallels methodological trends in the psycholinguistic literature (Balota et al., 2004; Sereno and Rayner, 2003), and offers several important advantages over traditional single-word fMRI studies. One advantage is that, because the experimental paradigm more closely approximates natural reading rates and does not interrupt reading with an overt task, the resulting estimates are likely to provide more accurate insights into the neural mechanisms of reading and language processing. Another advantage is the reduced need for a priori constraint of one's stimuli. Although the proposal to exercise *less* experimental control over stimulus selection may seem counter-productive, it actually provides several benefits. First, the effective range of variation in most dimensions increases. Second, avoiding dichotomization of one's variables allows non-linear effects to be modeled (e.g., length in the present study). Third, researchers can control for a much broader range of potential confounds, including variables that may not have been considered initially. For example, most single-word neuroimaging studies match stimuli on ON, but very few match on PN (e.g., Binder et al., 2003). The fact that PN effects dwarfed ON effects in the present study raises concerns about the interpretation of previous results. Might some effects ascribed to orthographic processing really be phonological in nature? Such a possibility was easily tested in the present study by adding a variable to the model, but would be more difficult to assess in a factorial study with relatively few stimuli.

Finally, perhaps the most important advantage of the Event-Related Reading Paradigm over conventional single-word designs is the considerable flexibility it provides to study almost any word-level variable that might influence reading. Whereas use of single-word designs often restricts the scope of investigation to one or two variables, the ERR approach affords investigation of multiple variables simultaneously and in naturally coherent combinations. For validation purposes, the scope of the present investigation was restricted to a set of well-studied lexical and semantic variables. However, the range of possible applications is virtually limitless. Researchers could, for example, investigate the relationship between neural activation and latent semantic content, deep syntactic structure, or interactions between different lexical variables. Norms for variables ranging from narrative complexity to level of emotional arousal could be regressed on brain activation on a word-by-word basis. Novel variables could be validated by demonstrating unique modulation of brain activation. Importantly, such investigations could be highly cost-effective, because a single dataset could potentially be used to test a multitude of hypotheses. Ideally, therefore, increased use of an ERR approach might be accompanied by the development of large, open-access fMRI datasets that allow for open-ended

investigations of language function, mirroring the flexibility of large-scale behavioral databases.

Of course, the benefits of the ERR approach must be weighed against several potential limitations of the method. First, the absence of behavioral measurement during scanning may make it difficult to ensure that subjects are actively engaged in the task and following instructions appropriately. This absence may be a small price to pay if one's goal is to study reading processes in as natural a context as possible; nevertheless, care should be taken to ensure participant compliance whenever possible. For example, in the present study, a post-scanning comprehension test was used to ensure that all subjects were actively reading the narratives during scanning.

Second, multicollinearity between variables may result in unreliable estimates if too many predictor variables are included in the regression model. As in any other multiple regression or GLM analysis, care should be taken to ensure that correlations between variables are not excessive, and that each variable carries sufficient unique variability to serve as a viable predictor of brain activation. A failure to take such considerations into account could lead to unreliable GLM estimates that reduce the probability of detecting significant activations and make interpretation of any significant results more difficult.

Third, presenting words within coherent narratives rather than in isolation could potentially confound the effects of word-level variables with those of intercorrelated phrase-level, sentence-level, or narrative-level variables due to a relative lack of control over the order in which words are presented. While this concern is valid, we would argue that its severity depends largely on one's objectives. If the goal of a study is to characterize the neural correlates of word processing in as context-free a manner as possible, an ERR approach may indeed be inadvisable. However, if one is interested in studying word-level processing in the context of reading, the ERR approach is clearly preferable to standard single-word designs, because there is little doubt that contextual factors play an important role in word-level processing under typical circumstances. Moreover, as noted above, hypotheses about potential confounds can be directly tested using the ERR approach simply by adding variables to the multiple regression model and assessing their impact.

Finally, one might argue that the ERR approach does not go far *enough* toward approximating natural reading processes, because it does not allow subjects to control the rate of presentation, skip or refixate words, etc. We agree with this concern, and view the ERR approach as only an incremental step in the broader effort to develop more naturalistic reading paradigms for fMRI. While considerable challenges remain before fMRI can be combined with high-precision eye-tracking techniques in such a way as to allow subjects full control over stimulus presentation and eye movement, the continued development of MR-compatible eye-tracking techniques provides reason for optimism (e.g., Gitelman et al., 2000; Kanowski et al., 2007; Kimmig et al., 1999; e.g., Oeltermann et al., 2007).

Theoretical implications

The present results also have theoretical implications for existing models of visual word form analysis. Supporting the notion that the VWFA contains neuron assemblies optimized for rapid processing of valid word forms (McCandliss et al., 2003), we found that VWFA activation varied systematically

with several orthographically-defined variables, including word frequency (Chee et al., 2003a; Kronbichler et al., 2004), orthographic neighborhood size, and word length. Of particular interest in this respect was the finding that the VWFA, but not other language regions, showed a U-shaped relationship with length, consistent with recent findings of a similar relationship between length and behavioral response latencies (New et al., 2006). Importantly, however, the results also provided strong evidence for a VWFA role in integrating orthography with word sound and meaning. VWFA activation was independently modulated by several different phonological, semantic, and contextual variables. Moreover, the ON effect in this area appeared to be better explained as an effect of *phonological* neighborhood size (PN), a novel and informative finding given recent discussion over whether neighborhood size effects in behavioral tasks reflect orthographic or phonological processes (Yates et al., 2004). The fact that the VWFA was sensitive to unique effects of PN and number of syllables, but not to ON, suggests that the process of grapheme-to-phoneme conversion may initiate earlier in the processing stream than is often assumed, and is consistent with recent studies demonstrating converging orthographic, phonological, and/or semantic effects in the VWFA (e.g., Sandak et al., 2004b; Xue et al., 2006). Sandak and colleagues have interpreted such findings as evidence that the left inferotemporal cortex serves as a general "skill zone" rather than supporting orthography-specific processes, a contention the present results support (Sandak et al., 2004a).

Although we have focused primarily on the role of the VWFA, the response profiles observed in other brain regions also inform previous work on the neural mechanisms of reading. Several of the effects observed in the present study directly replicate previous results, e.g., the finding that all 5 canonical language ROIs demonstrated an effect of priming (reduced activation with increasing repetition of a word; Demb et al., 1995; Wagner et al., 2000). Other effects are novel but follow naturally from existing models, e.g., the relatively selective time-on-task effect of naming latencies on bilateral IFG activation, consistent with a role for this region in subvocal articulation or rehearsal (Henson et al., 2000; Paulesu et al., 1993).

Also important, some of the current observations are inconsistent with previous results. For example, we observed strong positive effects of frequency in IFG and MTG, where previous fMRI and studies have typically observed negative effects (Chee et al., 2003a,b; Kronbichler et al., 2004). One possible interpretation of this discrepancy is that the present paradigm, in which words were presented at a relatively natural reading pace, emphasizes the latency of neural responses rather than their magnitude. That is, high-frequency (HF) words may elicit greater activation in anterior language sites because their lexical representations are accessed earlier in the 200ms window of word presentation than low-frequency (LF) words, allowing more time for the neural response to accumulate. If true, this account would suggest that the LF>HF effects seen in previous fMRI studies are epiphenomenal with respect to reading comprehension, because natural reading rates are much closer to 200 ms than the trial durations typically used in single-word studies.

Finally, an intriguing and unexpected finding of the present study was the presence of consistent associations between lexical variables and activation in regions outside of the canonical language network. In particular, these regions included several that have been identified with the brain's 'default mode' of processing in the absence of a focal cognitive task (Raichle et al., 2001). These included areas of DMPFC, precuneus, and posterior cingulate cortex .The fact that the response profiles of these regions differed markedly from one another (Table 2) suggests that their involvement in reading cannot be attributed to a general association between processing difficulty and default network activation. For example, DMPFC activation showed a profile that closely resembled activation in bilateral IFG and MTG, including an exponentially-increasing relationship with word length. In contrast, precuneus activation showed an inverse U-shaped relationship with word length. Such dissociations suggest that default network regions may play secondary roles in lexical processing that are currently not well understood. Alternatively, it is possible that effects in default network regions reflect the confounding influences of multiple unmodeled higher-level variables. For example, DMPFC is reliably activated during theory of mind tasks and situation model processing (Ferstl et al., 2007; Ferstl and von Cramon, 2001; Yarkoni et al., in press-b); perhaps the demand for such processes covaries naturally with lexical variables such as frequency and length when presented in the context of coherent narratives. Such hypotheses could be tested in future studies using a similar multiple regression approach.

Conclusions

The present study validated a new Event-Related Reading approach to studying word-level processes with fMRI during a coherent reading context. Application to a set of well-studied lexical variables replicated several findings of previous single-word fMRI studies while also identifying several novel and theoretically informative effects. The flexibility and power of the ERR approach thus render it an ideal complement or alternative to conventional single-word fMRI designs when studying the neural mechanisms of reading and language.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2008.04.258.

References

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H., Yap, M., 2004. Visual word recognition of single-syllable words. J. Exp. Psychol. Gen. 133, 283–316.

Balota, D.A., Yap, M.J., Cortese, M.J., 2006. Visual word recognition: the journey from features to meaning (a travel update). In: Traxler, M.J., Gernsbacher, M.A. (Eds.), Handbook of Psycholinguistics. Academic Press, Amsterdam, The Netherlands, pp. 285–376.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., Treiman, R., 2007. The English Lexicon Project. Behavior. Res. Methods 39, 445–459.

Barker, R.G., Wright, H.F., 1951. One Boy's Day: A Specimen Record of Behavior. Harper.

Bedny, M., Thompson-Schill, S.L., 2006. Neuroanatomically separable effects of image-ability and grammatical class during single-word comprehension. Brain. Lang.

Binder, J.R., McKiernan, K.A., Parsons, M.E., Westbury, C.F., Possing, E.T., Kaufman, J.N., Buchanan, L., 2003. Neural correlates of lexical access during visual word recognition. J. Cogn. Neurosci. 15, 372–393.

Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu. Rev. Neurosci. 25, 151–188.

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. 16, 4207.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., Dale, A.M., 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. Neuroreport 9, 3735–3739.

Chee, M.W., Westphal, C., Goh, J., Graham, S., Song, A.W., 2003a. Word frequency and subsequent memory effects studied using event-related fMRI. Neuroimage 20, 1042–1051.

Chee, M.W.L., Venkatraman, V., Westphal, C., Siong, S.C., 2003b. Comparison of block and event-related fMRI designs in evaluating the word-frequency effect. Hum. Brain Mapp. 18, 186–193.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M.A., Michel, F., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123, 291–307.

Coltheart, M., Davelaar, E., Jonasson, J.T., Besner, D., 1977. Access to the internal lexicon. Atten. Perform. VI, 535–555.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol. Rev. 108, 204–256.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.F., Poline, J.B., Rivière, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat. Neurosci. 4, 752–758.

Dehaene, S., Le Clec'H, G., Poline, J.B., Le Bihan, D., Cohen, L., 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport 13, 1–5.

Demb, J.B., Desmond, J.E., Wagner, A.D., Vaidya, C.J., Glover, G.H., Gabrieli, J.D., 1995. Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. J. Neurosci. 15, 5870–5878.

Dietz, N.A.E., Jones, K.M., Gareau, L., Zeffiro, T.A., Eden, G.F., 2005. Phonological Decoding Involves Left Posterior Fusiform Gyrus. Hum. Brain Mapp. 26, 81–93.

Ferstl, E.C., von Cramon, D.Y., 2001. The role of coherence and cohesion in text comprehension: An event-related fMRI study. Cogn. Brain Res. 11, 325–340.

Ferstl, E.C., Neumann, J., Bogler, C., Yves von Cramon, D., 2007. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. Hum. Brain Mapp. 29, 581–593.

Fiez, J.A., Petersen, S.E., 1998. Neuroimaging studies of word reading. Proc. Natl. Acad. Sci. U. S. A. 95, 914–921.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Gaillard, R., Naccache, L., Pinel, P., Clémenceau, S., Volle, E., Hasboun, D., Dupont, S., Baulac, M., Dehaene, S., Adam, C., 2006. Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. Neuron 50, 191–204.

Gernsbacher, M.A., Kaschak, M.P., 2003. Neuroimaging studies of language production and comprehension. Annu. Rev. Psychol. 91–115.

Gitelman, D.R., Parrish, T.B., LaBar, K.S., Mesulam, M.M., 2000. Real-time monitoring of eye movements using infrared video-oculography during functional magnetic resonance imaging of the frontal eye fields. Neuroimage 11, 58–65.

Henson, R.N.A., Burgess, N., Frith, C.D., 2000. Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. Neuropsychologia 38, 426–440.

Jobard, G., Crivello, F., Tzourio-Mazoyer, N., 2003. Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. Neuroimage 20, 693–712.

Kanowski, R., Rieger, J.W., Noesselt, T., Tempelmann, C., Hinrichs, H., 2007. Endoscopic eye tracking system for fMRI. J. Neurosci. Methods 160, 10–15.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Keller, T.A., Carpenter, P.A., Just, M.A., 2001. The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. Cereb. Cortex 11, 223–237.

Kimmig, H., Greenlee, M.W., Huethe, F., Mergner, T., 1999. MR-Eyetracker: a new method for eye movement recording in functional magnetic resonance imaging. Exp. Brain Res. 126, 443–449.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., Ladurner, G., 2004. The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. Neuroimage 21, 946–953.

Kuo, W.J., Yeh, T.C., Lee, C.Y., Wu, Y.T., Chou, C.C., Ho, L.T., Hung, D.L., Tzeng, O.J., Hsieh, J.C., 2003. Frequency effects of Chinese character processing in the brain: an event-related fMRI study. Neuroimage 18, 720–730.

Longe, O., Randall, B., Stamatakis, E.A., Tyler, L.K., 2007. Grammatical categories in the brain: the role of morphological structure. Cereb. Cortex 17, 1812.

Lund, K., Burgess, C., 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. Behav. Res. Methods Instrum. Comput. 28, 203–208.

Maccotta, L., Zacks, J.M., Buckner, R.L., 2001. Rapid self-paced event-related functional MRI: feasibility and implications of stimulus- versus response-locked timing. Neuroimage 14, 1105–1121.

McAvoy, M.P., Ollinger, J.M., Buckner, R.L., 2001. Cluster size thresholds for assessment of significant activation in fMRI. Neuroimage 13, S198.

McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. 7, 293–299.

New, B., Ferrand, L., Pallier, C., Brysbaert, M., 2006. Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. Psychon. Bull. Rev. 13, 45–52.

Oeltermann, A., Ku, S.P., Logothetis, N.K., 2007. A novel functional magnetic resonance imaging compatible search-coil eye-tracking system. Magn. Reson. Imaging 25, 913–922.

Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. Nature 362, 342–345.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Patterson, K., 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychol. Rev. 103, 56–115.

Price, C.J., 1998. The functional anatomy of word comprehension and production. Trends Cogn. Sci. 2, 281–288.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98, 676–682.

Rayner, K., 1998. Eye movements in reading and information processing: 20years of research. Psychol. Bull. 124, 372–422.

Sabsevitz, D.S., Medler, D.A., Seidenberg, M., Binder, J.R., 2005. Modulation of the semantic system by word imageability. Neuroimage 27, 188–200.

Sandak, R., Mencl, W.E., Frost, S.J., Pugh, K.R., 2004a. The Neurobiological Basis of Skilled and Impaired Reading: Recent Findings and New Directions. The Cognitive Neuroscience of Reading: A Special Issue of Scientific Studies of Reading, 8, pp. 273–292.

Sandak, R., Mencl, W.E., Frost, S.J., Rueckl, J.G., Katz, L., Moore, D.L., Mason, S.A., Fulbright, R.K., Constable, R.T., Pugh, K.R., 2004b. The neurobiology of adaptive learning in reading: a contrast of different training conditions. Cogn. Affect. Behav. Neurosci. 4, 67–88.

Seidenberg, M.S., Waters, G.S., 1989. Word recognition and naming: a mega study. Bull. Psychon. Soc. 27, 489.

Sereno, S.C., Rayner, K., 2003. Measuring word recognition in reading: eye movements and event-related potentials. Trends Cogn. Sci. 7, 489–493.

Shapiro, K.A., Mottaghy, F.M., Schiller, N.O., Poeppel, T.D., Flüβ, M.O., Müller-Gärtner, H.W., 2005. Dissociating neural correlates for nouns and verbs. Neuroimage 24, 1058–1067.

Speer, N.K., Zacks, J.M., Reynolds, J.R., 2007. Human brain activity time-locked to narrative event boundaries. Psychol. Sci. 18, 449–455.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging. Thieme.

Toglia, M.P., Battig, W.F., 1978. Handbook of Semantic Word Norms.

Toutanova, K., Klein, D., Manning, C.D., Singer, Y., 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume, 1, pp. 173–180.

Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage 16, 765–780.

Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. Neuroimage 30, 1414–1432.

Vigneau, M., Jobard, G., Mazoyer, B., Tzourio-Mazoyer, N., 2005. Word and non-word reading: what role for the visual word form area. Neuroimage 27, 694–705.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55, 143–156.

Wagner, A.D., Koutstaal, W., Maril, A., Schacter, D.L., Buckner, R.L., 2000. Task-specific repetition priming in left inferior prefrontal cortex. Cereb. Cortex 10, 1176–1184.

Weekes, B.S., 1997. Differential effects of number of letters on word and nonword naming latency. Q. J. Exp. Psychol., Sec. A. 50, 439–456.

Wise, R.J.S., Howard, D., Mummery, C.J., Fletcher, P., Leff, A., Büchel, C., Scott, S.K., 2000. Noun imageability and the temporal lobes. Neuropsychologia 38, 985–994.

Xue, G., Chen, C., Jin, Z., Dong, Q., 2006. Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. Neuroimage 31, 1315–1326.

Yarkoni, T., Balota, D.A., Yap, M., in press-a. Moving beyond Coltheart's N: A new measure of orthographic similarity. NeuroImage.

Yarkoni, T., Speer, N.K., Zacks, J.M., in press-b. Neural substrates of narrative comprehension and memory. NeuroImage.

Yates, M., 2005. Phonological neighbors speed visual word processing: evidence from multiple tasks. Learn. Mem. 31, 1385–1397.

Yates, M., Locker, L., Simpson, G.B., 2004. The influence of phonological neighborhood on visual word perception. Psychon. Bull. Rev. 11, 452–457.