An Orthographic Prediction Error as the basis for efficient Visual Word Recognition

Benjamin Gagl[1,2*], Jona Sassenhagen[1], Sophia Haan[1], Klara Gregorova[1], Fabio Richlan[3] and Christian J. Fiebach[1,2,4]

[1] Department of Psychology, Goethe University Frankfurt, Frankfurt/Main, Germany
[2] Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt/Main, Germany
[3] Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria.
[4] Brain Imaging Center, Goethe University Frankfurt, Frankfurt/Main, Germany
* Corresponding author: E-Mail gagl@psych.uni-frankfurt.de

Abstract

Most current models assume that the perceptual and cognitive processes of visual word recognition and reading operate upon neuronally coded domain-general low-level visual representations – typically oriented line representations. We here demonstrate, consistent with neurophysiological theories of Bayesian-like predictive neural computations, that prior visual knowledge of words is utilized to 'explain away' redundant and highly expected parts of the visual percept. Subsequent processing stages, accordingly, operate upon an optimized representation of the visual input, the orthographic prediction error, highlighting only information relevant for word identification. We show that this informationally optimized representation is related to orthographic word characteristics, accounts for word recognition behavior, and is processed early in the visual processing stream, i.e., in occipital cortex and before 200 ms after word-onset. Based on these findings, we propose that prior visual-orthographic knowledge is used to optimize the representation of visually presented words, which in turn allows for highly efficient reading processes.

Introduction

Written language – script – developed over the last ~8,000 years in many different variants (Haarmann, 2007). It is a symbolic representation of meaning, based on the combination of simple high contrast visual features (oriented lines) that our brains translate efficiently into linguistically meaningful units. Cognitive-psychological models of reading specify the perceptual and cognitive processes involved in activating orthographic, phonological, and lexico-semantic representations of perceived words from such low-level visual-perceptual features (for a review see Norris, 2013). While some models – consistent with other domains of perception (e.g., Riesenhuber & Poggio, 1999 for object recognition) – indeed assume oriented line representations as the lowest-level visual feature involved in visual word recognition (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Davis, 2010; Dehaene, Cohen, Sigman, & Vinckier, 2005; McClelland & Rumelhart, 1981; Perry, Ziegler, & Zorzi, 2007; Whitney & Cornelissen, 2008), other cognitive models use as starting point a more integrated, domain-specific representations, i.e., letters (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Rayner, & Pollatsek, 2003; Sibley, Kello, Plaut, & Elman, 2008).

Interestingly, this does not take into account findings from visual neuroscience indicating that already the neuronal representation of an oriented line is an abstraction of the visual input: The phenomenon of end-stopping describes that an oriented line (i.e., the frequently-assumed low-level input into the visual word recognition system) is not represented in the brain by many neurons with receptive fields along the length of the line, but by only two neurons that have their receptive fields at the beginning and end of the line (Bolz & Gilbert, 1986; D. H. Hubel & Livingstone, 1987; David H. Hubel & Wiesel, 1965). While preserving the representation of line length and angle, this neuronal representation can be more efficient by several orders of magnitude. Given these results, we hypothesized that early perceptual stages of visual word recognition should also operate upon informationally optimized representations of the visual-orthographic input.

To provide a computationally explicit account for explaining end-stopping, Rao and Ballard (Rao & Ballard, 1999) successfully adapted the computational principles of predictive coding (Srinivasan, Laughlin, & Dubs, 1982). Predictive coding postulates that perceived regularities in the world are used to build up internal models of the (hidden) causes of sensory events, and that these internal predictions are imprinted in a top-down manner upon the hierarchically lower sensory systems, thereby increasing processing efficiency by inhibiting the processing of correctly predicted input (K. Friston, 2005a; Rao & Ballard, 1999). When sensory input violates these expectations or is not fully predicted,

a prediction error signal is generated and propagated up the cortical processing hierarchy in a bottom-up fashion (e.g. Todorovic, van Ede, Maris, & de Lange, 2011), where it is used for model updating and thus serves to optimize future predictions (Clark, 2013; Rao & Ballard, 1999). In the case of line representations and end-stopping, neurons with receptive fields at the beginning and end of the line fire and this information is propagated to higher areas where they activate abstract line representations, which in turn in a recursive, top-down manner 'predict away' the activity of the receptive fields between the two endpoints of the line (Rao & Ballard, 1999). Predictive coding has by now received support in many domains of perceptual neuroscience, from retinal coding (Srinivasan et al., 1982), auditory (Todorovic et al., 2011; Wacongne, Changeux, & Dehaene, 2012) and speech perception (Arnal, Wyart, & Giraud, 2011; Gagnepain, Henson, & Davis, 2012) to object (Kersten, Mamassian, & Yuille, 2004) and face recognition (Schwiedrzik & Freiwald, 2017), indicating that this framework is likely a generalized computational principle of the brain.

Most readers can process written language at a remarkably high speed. We reasoned that the high efficiency of visual-orthographic processing necessary for efficient reading makes it likely that the visual system also optimizes the 'low-level' perceptual representations used for orthographic processing during reading. Inspired by the wide applicability of the principles of predictive coding (see previous paragraph), the present model-based study explores whether computational principles of predictive coding may contribute to the informational optimization of perceived written words already at the earliest neurocognitive stages of reading. In an influential theoretical paper, Price and Devlin (Price & Devlin, 2011) have proposed that principles of predictive coding may be involved in visual word recognition. Their 'Interactive Account' model focuses explicitly on 'intermediate-level' stages of visual word processing that are attributed to the left ventral occipito-temporal cortex (vOT; often also referred to as 'visual word form area', e.g., Dehaene & Cohen, 2011; Dehaene et al., 2005). The Interactive Account model postulates that at the level of lvOT, visual-perceptual information that is propagated bottom-up from early visual to higher areas when reading a string of letters, is integrated with phonological and semantic information fed to lvOT from higher cortical areas in a top-down manner. Empirical support for this proposal comes from a study by Kherif et al. (2011) demonstrating semantic priming effects between words and pictures of objects, in lvOT.

However, predictive coding as a general model of cortical processing should, in principle, not be restricted to a specific level of processing, but rather apply to all sensory-

perceptual stages of the reading process – including also reading-related visual processes in lower level visual areas (i.e., that take place before the integrative processes attributed to the vOT/visual word form area in the Interactive Account model; see, e.g., also Fig. 2 of Price & Devlin, 2011). We thus hypothesized here that 'higher level' linguistic expectations – either in the form of contextual constraint from preceding input or in the form of our knowledge of the orthography of a language – should be imprinted upon the earliest stages of visual-orthographic processing, thereby 'optimizing' earlier, i.e., pre-lexical processing stages that are typically associated with brain processes located posterior to the visual word form area (e.g., Dehaene et al., 2005) and temporally earlier than 250 ms (Grainger & Holcomb, 2009).

Interestingly, the feature-configurations that constitute letters and words, i.e., that are part of our orthographic knowledge of language, contain highly redundant information (Changizi, Zhang, Ye, & Shimojo, 2006) – like vertical lines often occurring at the same position (e.g., the left vertical line in E, R, N, P, B, D, F, H, K, L, M) or letters often positioned at the same location in a word (e.g., s or y as final letters in English). As such redundancies contribute very little to unique letter and word identification, using prior orthographic knowledge to subtract the redundant part of the percept is a plausible strategy of our brain to reduce the amount of to-be-processed information – and thus a plausible way of increasing the efficiency of the neuronal code that is fundamental to visual word recognition.

We here propose that during the earliest stages of visual word recognition, following the principles of predictive coding, the visual-orthographic input signal is 'optimized' on the basis of our knowledge and expectations about the redundancies of the respective script. In other words, we propose that our orthographic knowledge of language is literally used to 'predict away' the uninformative part of visual input during reading. As a result, the subsequent stages of visual word recognition (as described in several models of reading; see above and (Norris, 2013) for review) can proceed upon an informationally optimized representation of the input. As this informationally optimized input representation highlights the unexpected (and thus more informative) part of the stimulus, we termed it the orthographic prediction error (oPE). In the following, we describe one possible, computationally explicit implementation of this proposal, which we refer to as the Prediction Error Model of Reading (PEMoR), and we report quantitative evaluations of the PEMoR using lexicon-based, behavioral, EEG, und functional MRI data. To compare the PEMoR with an alternative account that implements early visual processing stages without top-down predictions and prediction errors, we also

conducted most of the reported analyses for a full pixel-based representation of the perceived stimuli (Pixel based Model of Reading; PixMoR). As a result, for most model evaluations, we compare two parameters, one reflecting strictly bottom-up visual processing without a prediction-based optimization step and one based on a top-down-/prediction-based optimization of the sensory representation of the perceived word.

The Prediction Error Model of Reading (PEMoR)

The Prediction Error Model of Reading postulates that one identifies words not on the basis of the full physical input into the visual system contained in a string of letters, but rather based on an optimized (and thus more efficient) neuronal code representing only the informative part of the percept (while redundant and expected information is cancelled out at earliest-possible processing stages; Rao & Ballard, 1999). In the predictive coding framework, this non-redundant portion of a stimulus is formalized as a prediction error; we apply this principle to visual word recognition, and propose that internal (i.e., knowledge- or context-dependent) visual-orthographic expectations are subtracted from the sensory input, so that further processing stages operate upon an orthographic prediction error (oPE) signal (Fig. 1a). It is commonly believed that higher level linguistic representations can initiate specific expectations about upcoming words (DeLong, Urbach, & Kutas, 2005; Kliegl, Nuthmann, & Engbert, 2006; Nieuwland et al., 2018; Price & Devlin, 2011) – e.g., about the class (noun or verb) and meaning of the next word in a sentence like "The scientists made an unexpected … (discovery)". The fundamental difference between these psycholinguistic assumptions about semantic and syntactic predictions and the proposed visual-orthographic prediction in PEMoR, is that we postulate predictive processes already at much earlier stages of visual processing.

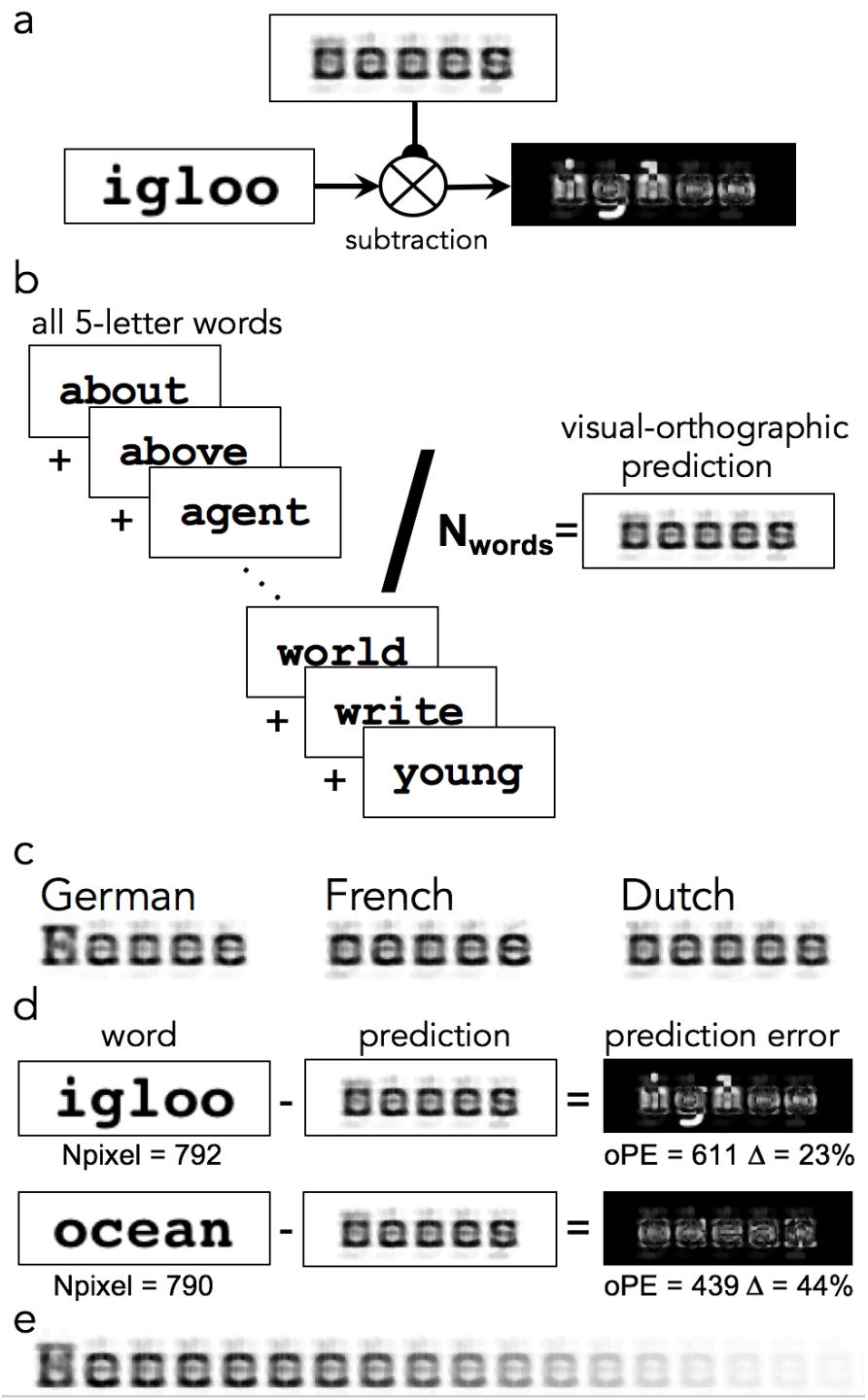

Figure 1. Prediction error model of reading (PEMoR). (a) The PEMoR assumes that during word recognition, redundant visual-orthographic information is 'explained away', thereby highlighting the informative aspects of the percept. Subsequent stages of word recognition and linguistic processing (i.e. accessing abstract letter and word representations), thus, operate upon an informationally optimized input representation. This assumption is here tested for single-word reading, i.e., independent of context, by subtracting a 'visual-orthographic prediction' from the input. (b) The visual-orthographic, knowledge-based prediction is implemented as a pixel-by-pixel mean across image representations of all known words (here approximated by all words in a psycholinguistic database; only five letters words, as in most experiments reported here; but see Supplemental figure 1b and e for a prediction including different word lengths). The resulting visual-orthographic prediction, shown on the right, contains the most redundant visual information across all words. (c) Across multiple languages, these predictions are very similar, with the exception of the upper-case initial letter that is visible in the German prediction (because experiments in German involved only nouns). (d) The orthographic prediction error (oPE) is estimated, for each word, by a pixel-by-pixel subtraction of the orthographic prediction from the input word (based on their image representations; see Methods for details). While the two example words have similar numbers of pixels, subtracting the orthographic prediction results in substantially different residual (i.e., oPE) images. The values underneath the prediction error images represent a quantitative estimate of the orthographic prediction error, the sum of all gray values from each pixel per image, and show that the amount of information reduction (Δ) can differ strongly between words. (e) Letter-length unspecific prediction for German, based on ~190.000 words.

Here, we quantitatively test the assumptions of the PEMoR for the most frequently investigated paradigm in reading research, single word recognition. In the absence of sentence context, the redundant visual information (i.e., the visual-orthographic prediction or, in Bayesian terms, the prior) is a function of our orthographic knowledge of words. We approximate this prior knowledge quantitatively as the pixel-by-pixel mean over image representations of all words derived from a psycholinguistic database (Brysbaert et al., 2011; see Fig. 1b and Methods). Interestingly, the resulting visual-orthographic predictions look similar across different languages sharing the same writing system (compare Fig. 1c) and correlate highly with each other (i.e., based on the individual gray values of the pixels from the prediction image; r ranging from .95 to .99).

We estimate the orthographic prediction error as a pixel-by-pixel subtraction of this visual-orthographic prediction (or prior in Bayesian terms) from each perceived word (Fig. 1d). This step of 'predicting away' the redundant part of written words reduces the amount of to-be-processed information by up to 51% (on average 33%, 37%, and 31% for English, French, and German, respectively; see Methods, Formula 4), thereby optimizing the visual input signal in the sense of highlighting only its informative parts (Fig. 1d). According to the PEMoR, the resulting orthographic prediction error is a critical early (i.e., pre-lexical) stage of word identification, representing (part of) the access code that our brain uses to activate word meaning.

We test this model by calculating for each stimulus item a numeric prediction error (oPE) value, which equates the sum of all gray scale values in the respective (200 by 40 pixel) stimulus image to represent the PEMoR. While this per-item summary oPE value does not take into account the spatial layout of the stimulus item, it represents an estimate of the amount of neuronal activation needed to represent the specific stimulus. Importantly, representing the oPE as a single value allows us to compare it directly to other typical word characteristics that are closely tied to different psychological models, like word frequency (Brysbaert et al., 2011) or orthographic familiarity (Coltheart, Davelaar, Jonasson, & Besner, 1977; Yarkoni, Balota, & Yap, 2008). In the following, we provide empirical support for this model by demonstrating that our orthographic prediction error (i) is correlated with orthographic familiarity of words measured as a property of lexicon statistics, (ii) accounts for response times in three languages, (iii) is represented in occipital brain regions, and (iv) electrophysiological signals from 150-250 ms after word onset.

As there exists – to the best of our knowledge – no generally accepted null model against which to compare the PEMoR, we also quantified the information contained in

each stimulus item prior to prediction-based optimization, by calculating the sum of all pixels in each original stimulus image. We used this pixel count parameter as an estimate of the full bottom-up information that would have to be processed in the absence of prediction-/top down-based optimization of the percept. For most empirical model evaluations reported in the following, we thus compare the performance of the pixel count parameter with the orthographic prediction error since current models of visual word recognition do not specify processing down to the level of individual pixels of a word image

Materials and Methods

Implementation of the PEMoR

The estimation of the orthographic prediction error as assumed in PEMoR was implemented by image-based computations. Using the EBImage package in R (Pau, Fuchs, Sklyar, Boutros, & Huber, 2010), letter-strings were transformed into gray scale images (size for, e.g., 5-letter words: 140x40 pixels) that can be represented by a 2-dimensional matrix in which white is represented as 1, black as 0, and gray as intermediate values. This matrix representation allows an easy implementation of the subtraction computation presented in Fig. 1a, i.e.,

$$(1)\quad \begin{bmatrix} SI_{1,1} & \dots & SI_{140,1} \\ \vdots & \ddots & \vdots \\ SI_{1,40} & \dots & SI_{140,40} \end{bmatrix} - \begin{bmatrix} P_{1,1} & \dots & P_{140,1} \\ \vdots & \ddots & \vdots \\ P_{1,40} & \dots & P_{140,40} \end{bmatrix} = \begin{bmatrix} oPE_{1,1} & \dots & oPE_{140,1} \\ \vdots & \ddots & \vdots \\ oPE_{1,40} & \dots & oPE_{140,40} \end{bmatrix}$$

where $SI_{x,y}$ indicates the sensory input at each pixel. $P_{x,y}$ reflects the prediction matrix which is in the present study calculated as an average across all words (or a subset thereof) in a lexical database e.g., the example shown in Fig. 1b is based on 5,896 nouns of five letters length from the English SUBTLEX database (Heuven et al., 2014). This orthographic prediction was estimated by transforming each of n words into a matrix as described above and then averaging the values included in these matrices:

$$(2)\quad \frac{\sum_1^n \begin{bmatrix} SI_{1,1} & \dots & SI_{140,1} \\ \vdots & \ddots & \vdots \\ SI_{1,40} & \dots & SI_{140,40} \end{bmatrix}}{n} = \begin{bmatrix} P_{1,1} & \dots & P_{140,1} \\ \vdots & \ddots & \vdots \\ P_{1,40} & \dots & P_{140,40} \end{bmatrix}$$

The PEMoR model postulates that during word processing, SI is reduced by the prediction matrix P, resulting in an orthographic prediction error matrix (oPE) as shown above in formula (1). The resulting orthographic predication error is therefore black (i.e. value 0) at pixels were the prediction was perfect and gray to white (i.e. value > 0) where the visual information was not predicted perfectly. As a last step, a numeric value for the orthographic prediction error of each stimulus was determined by summing all values of its prediction error matrix. This numeric representation of the prediction error is used as parameter for all empirical evaluations.

$$(3)\ \sum \begin{bmatrix} oPE_{1,1} & \dots & oPE_{140,1} \\ \vdots & \ddots & \vdots \\ oPE_{1,40} & \dots & oPE_{140,40} \end{bmatrix} = oPE_{sum}$$

The amount of information reduction ($I_{reduced}$) achieved by this predictive computation can then be calculated by relating the numeric representation of the prediction error to an analogous numeric representation of the respective word SIsum:

$$(4)\ 1 - \frac{oPE_{sum}}{SI_{sum}} * 100 = I_{reduced}$$

Participants.

35, 54, 39, 31, and 38 healthy volunteers (age from 18 to 39) participated in the two German lexical decision studies, the fMRI, the EEG, and the handwriting experiments, respectively. All had normal reading speed (reading scores above 20th percentile estimated by a standardized screening; unpublished adult version of Mayringer & Wimmer, 2014), reported absence of speech difficulties, had no history of neurological diseases, and normal or corrected-to-normal vision. Participants gave written informed consent and received student credit or financial compensation (10€/h) as incentive for participating. The research was approved by the ethics board of the University of Salzburg (EK-GZ: 20/2014; fMRI study) and Goethe University Frankfurt (#2015-229; EEG study, lexical decision studies). Behavioral results for English, and French were obtained from publicly available data sets, whose samples are described elsewhere (Ferrand et al., 2010; Keuleers et al., 2012).

Materials, experimental procedures, and statistical analyses.

Lexicon-based Characterization of the Orthographic Prediction Error. We calculated the number of pixels per word, the orthographic prediction error, and established word characteristics (Orthographic Levenshtein distance (Yarkoni et al., 2008), word frequency) for 3,110 German (Brysbaert et al., 2011) nouns (i.e., the subset used for the empirical evaluations later on; with uppercase first letters), for 5,896 English (Heuven et al., 2014) words, 5,638 French (New, Pallier, Brysbaert, & Ferrand, 2004) words, and 4,418 Dutch (Keuleers, Brysbaert, et al., 2010) words. All items had a length of five letters. For the German nouns, we additionally estimated a more comprehensive set of orthographic word characteristics, including bi-, tri-, quadirgram-frequencies (i.e., occurrences of 2, 3, 4 letter combinations), and Coltheart's N (Coltheart et al., 1977); see Fig. 2b). Orthographic Levenshtein distance and Coltheart's N were estimated with the vwr Package in R (Keuleers, 2013).

Accounting for Word Recognition Behavior. German lexical decision task 1: 800 five-letter nouns and 800 five-letter nonwords (400 pronouncable pseudowords, 400 unpronouncable non-words/consonant clusters) were presented in pseudorandomized order (Experiment Builder software, SR-Research, Ontario, Canada; black on white background; Courier-New font; .3° of visual angle per letter; 21″ LCD monitor with 1,024 × 768 resolution and 60Hz refresh rate), preceded by 10 practice trials. Participants judged for each letter string whether it was a word or not using a regular PC keyboard, with left and right arrow keys for words and non-words, respectively. Before stimulus presentation, two black vertical bars (one above and one below the vertical position of the letter string) were presented for 500 ms, and letter strings were displayed until a button was pressed. Response times were measured in relation to the stimulus onset. German lexical decision task 2 including noisy stimuli reports a replication in German with 70 five-letter words and 70 nonwords (36 pseudowords, 34 consonant clusters) with no noise with identical procedures except that data were acquired in small groups of up to 8 participants. In addition, words with 20% or 40% noise added (i.e. 20% or 40% of pixels were displaced; for details see Gagl et al., 2014) were presented in blocks of 140 (70 five-letter words and 70 nonwords).

Linear mixed model (LMM) analysis implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) of the R statistics software were used for analyzing lexical decision data as LMMs are optimized for estimating statistical models with crossed random effects for items. These analyses result in effect size estimates with confidence

intervals (SE) and a t-value. Following standard procedures, t-values larger than 2 are considered significant since this indicates that the effect size ±2 SE does not include zero (Kliegl, Wei, Dambacher, Yan, & Zhou, 2011). For the presentation in Fig. 3a,b,d,e,g,h,k,l co-varying effects were removed by the keepef function of the remef package (Hohenstein & Kliegl, 2014/2017). All response times were log-transformed, which accounts for the ex-Gaussian distribution of response times. In addition, orthographic prediction error, and number of pixels were centered and normalized by R's scale() function in order to optimize LMM analysis.

Cortical Representation of the Orthographic Prediction Error. 60 five-letter words and 180 pseudowords were presented in pseudorandom order (yellow Courier New font on gray background; 800 ms per stimulus; ISI 2,150 ms) as well as 30 catch trials consisting of the German word Taste (button), indicating participants to press the response button. Catch trials were excluded from the analyses. All items consisted of two syllables and were matched on OLD20 (Yarkoni et al., 2008) and mean bigram frequency between conditions. To facilitate estimation of the hemodynamic response, an asynchrony between the TR (2,250 ms) and stimulus presentation (onset asynchrony: 2,150 + 800 ms) was established and 60 null events were interspersed among trials; a fixation cross was shown during inter-stimulus intervals and null events. The sequence of presentation was determined by a genetic algorithm (Wager & Nichols, 2003), which optimized for maximal statistical power and psychological validity. The fMRI session was divided into 2 runs with a duration of approximately 8 min each.

A Siemens Magnetom TRIO 3-Tesla scanner (Siemens AG, Erlangen, Germany) equipped with a 32-channel head-coil was used for functional and anatomical image acquisition. The BOLD signal was acquired with a T_2*-weighted gradient echo echo-planar imaging sequence (TR = 2,250 ms; TE = 30 ms; Flip angle = 70°; 86 x 86 matrix; FoV = 192 mm). Thirty-six descending axial slices with a slice thickness of 3 mm and a slice gap of 0.3 mm were acquired within each TR. In addition, for each participant a gradient echo field map (TR = 488 ms; TE 1 = 4.49 ms; TE 2 = 6.95 ms) and a high-resolution structural scan (T_1-weighted MPRAGE sequence; 1 x 1 x 1.2 mm) were acquired. Stimuli were presented using an MR-compatible LCD screen (NordicNeuroLab, Bergen, Norway) with a refresh rate of 60 Hz and a resolution of 1,024x768 pixels.

SPM8 software (http://www.fil.ion.ucl.ac.uk/spm), running on Matlab 7.6 (Mathworks, Inc., MA, USA), was used for preprocessing and statistical analysis. Functional images were realigned, unwarped, corrected for geometric distortions by use

of the FieldMap toolbox, and slice-time corrected. The high-resolution structural image was pre-processed and normalized using the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm8). The image was segmented into gray matter, white matter, and CSF compartments, denoised, and warped into MNI space by registering it to the DARTEL template of the VBM8 toolbox using the high-dimensional DARTEL registration algorithm (Ashburner, 2007). Functional images were co-registered to the high-resolution structural image, which was normalized to the MNI T_1 template image, and resulting normalization parameters were applied to the functional data, which were then resampled to a resolution of 2×2×2 mm and smoothed with a 6 mm FWHM Gaussian kernel.

For statistical analysis, we first modeled stimulus onsets with a canonical hemodynamic response function and its temporal derivative, including movement parameters from the realignment step and catch trials as covariates of no interest, a high-pass filter with a cut off of 128 s, and an AR(1) model (K. J. Friston et al., 2002) to correct for autocorrelation. For the group level statistics, t-tests were implemented with a voxel level threshold of $p < .001$ uncorrected and a cluster level correction for multiple comparisons ($p < .05$ family-wise error corrected).

Cortical timing of the Orthographic Prediction Error. 200 five-letter words, 100 pseudowords, and 100 consonant strings (nonwords) were presented for 800 ms (black on white background; Courier-New font, .3° of visual angle per letter), followed by an 800 ms blank screen and a 1,500 ms hash mark presentation, which marked an interval in which the participants were instructed to blink if necessary. In addition, 60 catch trials (procedure as described for fMRI study) were included in the experiment. Stimuli were presented on a 19″ CRT monitor (resolution 1,024 × 768 pixels, refresh rate 150Hz), and were preceded by two black vertical bars presented for 500 - 1,000 ms to reduce stimulus onset expectancies.

EEG was recorded from 64 active Ag/Ag-Cl electrodes (extended 10-20 system) using an actiCAP system (BrainProducts, Germany). FCz served as common reference and the EOG was recorded from the outer canthus of each eye as well as from below the left eye. A 64-channel Brainamp (BrainProducts, Germany) amplifier with a 0.1–1,000 Hz band pass filter sampled the amplified signal with 500Hz. Electrode impedances were kept below 5kΩ. Offline, the EEG data were re-referenced to the average of all channels. EEG data were preprocessed using MNE-Python (Gramfort et al., 2014), including high (.1 Hz) and low pass (30 Hz) filtering and removal of ocular artifacts using ICA (Delorme, Sejnowski, & Makeig, 2007). For each subject, epochs from 0.5 s before to 0.8 s after

word onset were extracted and baselined by subtracting the pre-stimulus mean, after rejecting trials with extreme (>50 µV peak-to-peak variation) values. Multiple regression analysis, with the exact same parameters as for the behavioral evaluation (orthographic prediction error, number of pixels, word/non-word, and the interactions with the word/non-word distinction), was conducted and a cluster-based permutation test (Maris & Oostenveld, 2007) was used for significance testing. 1,024 label permutations were conducted to estimate the distribution of thresholded clusters of spatially and temporally (i.e., across electrodes and time) adjacent time points under the null hypothesis. All clusters with a probability of less than an assumed alpha value of 0.05 under this simulated null hypothesis were considered statistically significant. The presentation of effect patterns (line and box-plots) in Fig. 6 co-varying effects were removed by the keepef function of the remef package (Hohenstein & Kliegl, 2014/2017).

Application to handwriting. We obtained handwriting samples (26 upper and 26 lower case letters; 10 common German compound words, 10-24 letters long) from 10 different writers (see Fig. 6a,b for examples). The single letters were scanned and centered within a 50x50 pixels image. These images were used to estimate, for each script separately, pixel-by-pixel predictions for upper and lower-case letters (see also Fig. 6a,b), analogous to the procedures described above and in Fig. 1b. Subsequently, these predictions were subtracted from each letter of the alphabet, within the respective script sample (matrix subtraction; Formula 1). In contrast to computer fonts the correlation of the orthographic prediction error and the respective item's number of pixels was high (r = .98). To compensate this, the orthographic prediction error was normalized by a division with the respective pixel count. Readability ratings (5-point Likert scale) were obtained from 38 participants (27 females; mean age 25 years) by presenting all ten versions of all ten handwritten compound words, in addition to the identical word in computerized script.
For the handwriting data, we implemented a LMM analysis that predicted the orthographic prediction error (Fig. 6c-d) from the following parameters: mean prediction strength (i.e., mean of the values extracted from the prediction matrix), number of all non-white pixels (both scaled), and letter case. The random effect on the intercept was estimated for each script. In addition, a second LMM was estimated for readability ratings with the orthographic prediction error as the only predictor and participants as random effect on the intercept and as random effect of the orthographic prediction error slope.

Results

Lexicon-based Characterization of the Orthographic Prediction Error

Cognitive psychologists have developed several quantitative measures to characterize words (Brysbaert et al., 2011; Coltheart et al., 1977; Yarkoni et al., 2008), mostly derived from large text corpora and psycholinguistic word databases (Heuven, Mandera, Keuleers, & Brysbaert, 2014; Keuleers, Brysbaert, & New, 2010; see Fig. 2a for most essential characteristics and examples). Abundant empirical research demonstrates that these lexicon-based word characteristics are predictive of different aspects of reading behavior (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Rayner, 2009). Accordingly, understanding how the orthographic prediction error, derived from the implemented PEMoR (see Fig. 1), relates to these measures provides an essential first indication that this informationally optimized and supposedly pre-lexical perceptual signal is indeed involved in word recognition.

Across all words, the orthographic prediction error (i.e., the sum of all gray values after subtracting the knowledge-based prior from the actual stimulus image; cf. Fig. 1d and Methods) clusters with several measures that are commonly interpreted as orthographic (Fig. 2b): These widely-used (psycho-) linguistic characteristics reflect the (non-) uniqueness of words in terms of their orthographic similarity to other words (e.g., the number of Coltheart neighbors (Coltheart et al., 1977) or the orthographic distance (OLD20; (Yarkoni et al., 2008); cf. Fig. 2a) and letter co-occurrences (e.g., bi- and trigram frequencies; cf. Fig. 2a). Note that these measures describe the statistics of letters and letter combinations in all words retrieved from a lexicon database (Keuleers, Brysbaert, et al., 2010). In cognitive psychological research, one associates these measures consistently with the first, i.e., orthographic, stages of processing written words before lexical access (Coltheart et al., 2001; Grainger & Jacobs, 1996). These correlations are an impressive result as it demonstrates that a neurophysiologically inspired transformation of the visual stimulus, i.e., the here-proposed orthographic prediction error (oPE), is meaningfully related to orthographic properties of words as derived from lexicon-based statistics. Crucially, this is achieved while (a) reducing the to-be-processed information content by more than 30% and (b) at the same time retaining the ability of discriminating the word identities, as indicated by a strong correlation of $r = .87$ between the representational similarity matrices (Edelman, 1998; Kriegeskorte et al., 2008) of the word and orthographic prediction error images (Fig. 2d). This latter result indicates that the representational similarity structure, or in other words the discriminability between items, is preserved after deriving the oPE from the sensory input as proposed by the PEMoR.

a

Orthographic measure	Example: *read*	Value
Coltheart neighbors (N)	real, head, road, lead, rear …	11
Orthographic Levenshtein distance (OLD)	real: 1, ready: 1, red: 1, great: 2, bad: 2, …; Mean distance over the 20 nearest words	1
Bigram frequency	re, ea, ad; summed occurrences in lexicon of all bigrams in the word	29951
Trigram frequency	rea, ead; summed occurrences in lexicon of all trigrams in the word	4528
Quadrigram frequency	read; summed occurrences in lexicon of all quadrigrams in the word	90
Word frequency	Log. occurrences per million in a subtitles text corpus (SUBTLEX)	4.1

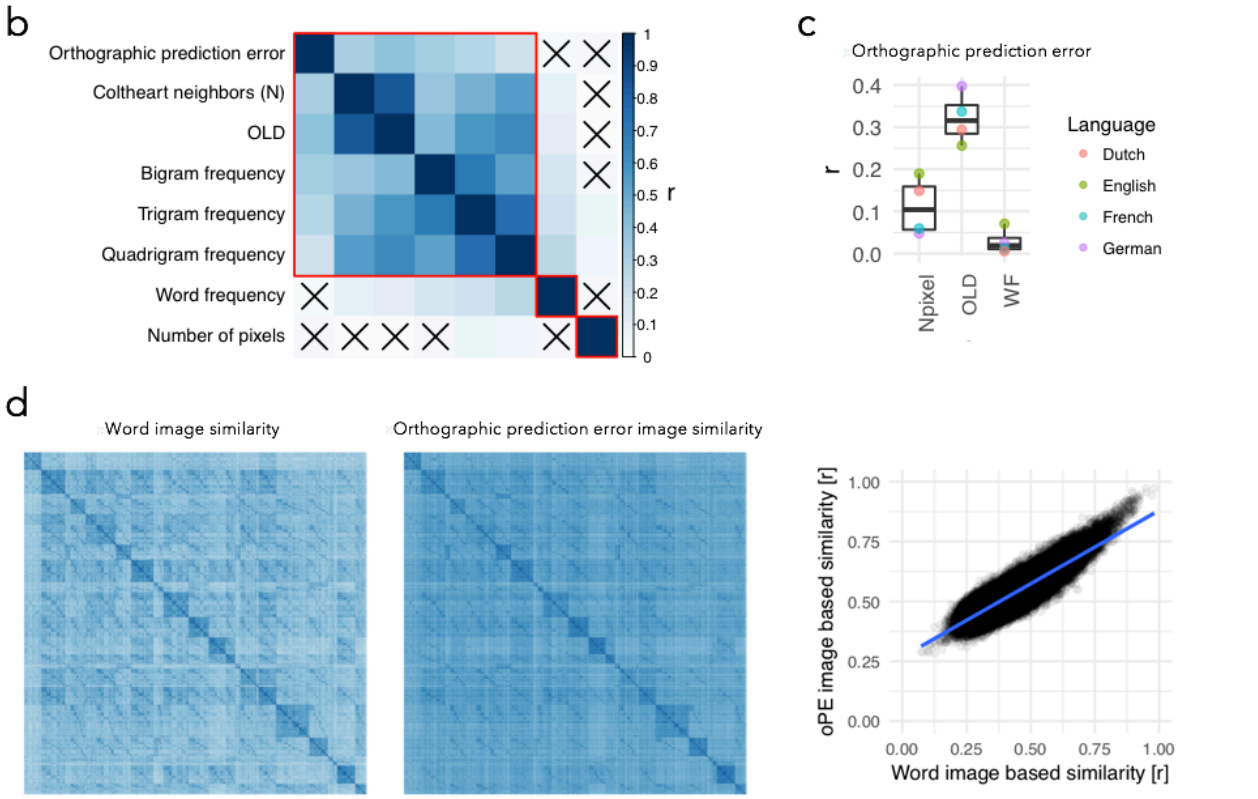

Figure 2. Comparison of orthographic prediction error to established lexicon-based word characteristics. (a) Overview of established word characteristics, exemplified for the word 'read': Coltheart's neighborhood size (Coltheart N; Coltheart et al., 1977), orthographic Levenshtein distance (OLD20; Yarkoni et al., 2008), sub-lexical frequency measures (bi-, tri-, and quadri-gram frequencies, i.e. number of occurrences of two, three, and four-letter combinations from the target word, in the lexicon), and word frequency as calculated from established linguistic corpora (see Methods for details). (b) Clustered correlation matrix between the orthographic prediction error, the number of pixels per original image, which represents an estimate of the pure amount of physical bottom-up input in the present study, and the described word characteristics (cf. panel a for explanations), applied to a set of 3,110 German nouns. Red rectangles mark clusters (obtained from a standard hierarchical clustering algorithm using the dendrogram) and black crosses mark non-significant correlations ($p < .05$; Bonferroni corrected to $p < .00179$). Number of pixels refers to the original stimulus item and is used as a simplified model of the full bottom-up physical input (PixMoR; see text). (c) Correlations between the orthographic prediction error and number of pixels per word (Npixel), orthographic similarity (OLD20), and word frequency (WF), for four different languages. (d) Representational similarity matrices (RSM; cf. Ref. Kriegeskorte, Mur, & Bandettini, 2008) for original word images (left panel) and orthographic prediction error images (central panel). Each similarity matrix reflects the correlations among the gray values of all 3,110 words (in total 9,672,100 correlations per matrix), with words sorted alphabetically (color scale equivalent to the one used in panel b). The right panel shows the correlation between word- and orthographic prediction error-based RSMs. Each dot represents a position on the similarity matrix allowing to relate the similarity values derived from either the physical input or the prediction error image. The high correlation shown here indicates that the similarity structure, or in other words the discriminability, present in the physical input is still represented in the prediction error images.

In contrast, the orthographic prediction error is not correlated with the frequency of occurrence of a word in a language (Fig. 2b). The word frequency effect typically indicates the difficulty of accessing word meaning based on an already-decoded orthographic access code (Coltheart et al., 2001). This dissociation between the orthographic prediction error and word frequency replicates across languages (Fig. 2c) and is much more pronounced for the orthographic prediction error than for the so-far dominant measures of orthographic similarity and orthographic neighborhood (Fig. 2b). With the raw pixel count of the words (Fig. 2b), as a reflection of the PixMoR, only two standard orthographic measures (tri- and quadrigram frequency) were weakly correlated. This dissociation of the correlation structure of the prediction error and the pixel count provides the first evidence that the neurophysiologically inspired orthographic prediction error is more important for a mechanistic understanding of reading than the full physical input contained in a printed word as assumed in the PixMoR.

Accounting for Word Recognition Behavior

As a next empirical test of the visual-orthographic prediction model of reading, we evaluated how well the orthographic prediction error performs in accounting for behavior in an established and widely-used word recognition task, i.e., the lexical decision task. Thirty-five human participants were asked to decide as fast as possible by button press whether written letter-strings (presented on the computer screen; 1,600 items; 5 letters length; language: German) were words or not. Remember that the orthographic prediction error represents the deviance of a given letter-string from our knowledge-based orthographic expectation, and thus how (un-)likely it is that the given letter-string is a word. Accordingly, participants should be fast in identifying letter-strings with low orthographic prediction error as words and fast in rejecting non-words with a high orthographic prediction error.

Fig. 3a shows exactly this pattern of response times, i.e., a word/non-word by orthographic prediction error interaction (linear mixed model/LMM estimate: 0.03; SE = 0.01; t = 5.0; see Methods for details on linear mixed effects modeling and Supplemental Table 1 for detailed results). No significant interaction or fixed effect of the number of pixels estimate (i.e. the sum of all pixels contained in a word), representing the PixMoR, was found (Fig. 3b; Interaction: estimate: 0.00; SE = 0.01; t = 0.0; Fixed effect: -0.01; SE = 0.00; t =1.8). To directly compare if the response times are more adequately described by the PEMoR or the PixMoR, we performed an explicit model comparison (see Methods for details) of four models. The full model, including as predictors the orthographic

prediction error and the number of pixels, a pure prediction error model, a pure number of pixels model, and a null model without any of the two predictors. Fig. 3c shows that, in contrast to the null model, the three alternative models showed higher model fits (all χ^2's > 9.9; all p's < .007; Bonferroni corrected p threshold: 0.0083), but this increase was significantly larger for the models including the orthographic prediction error. In addition, the model including only the orthographic prediction error explained substantially higher amounts of variance when compared to the model including only the number of pixels parameter (AIC difference: 34; $\chi^2(0) = 34.2$; $p < .001$) with no substantial increase for the combined model (AIC difference: 3; $\chi^2(2) = 7.2$; $p = .02$). This finding indicates that for German, the PixMoR explains substantially less variance in word recognition behavior than the PEMoR.

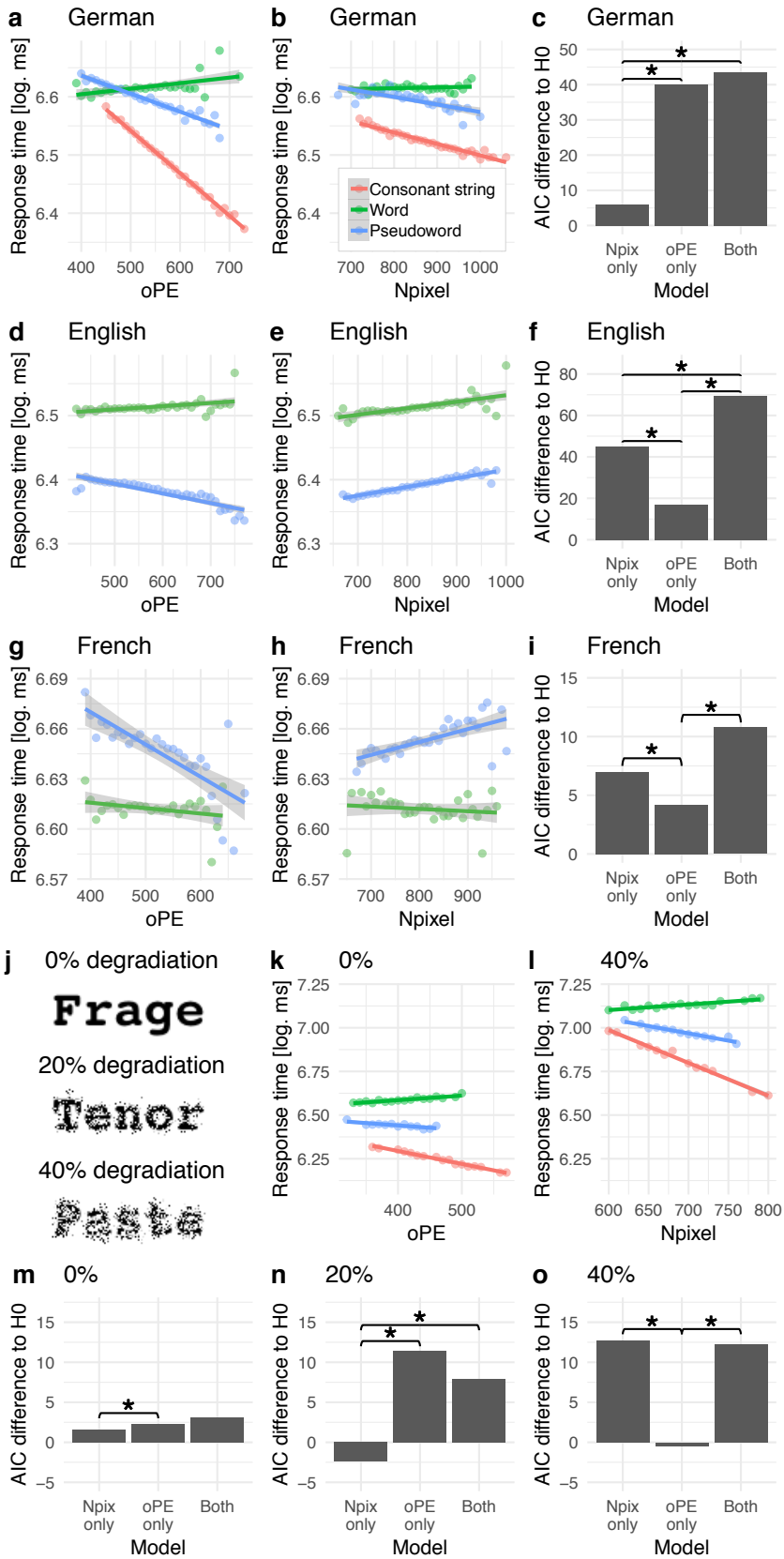

Figure 3. Word/non-word decision task behavior. (a) Orthographic prediction error (oPE) and (b) number of pixels (Npixel) effects on response times in a word/non-word decision task (German nouns, 5 letters length; overall error rate 7.4%; see Supplemental Table 1 for detailed statistical analysis). Green lines show the effects for words, blue lines for pseudowords (pronounceable non-words), and red lines for consonant strings (unpronounceable non-words). Dots represent mean reaction time estimates across all participants, separated into bins of oPE (width of 10) and stimulus category, after excluding confounding effects. (c) Results from model comparisons. First, a null model was established with only word/non-word status and word frequency as predictors. Subsequently, a model adding only the oPE predictor, a model adding only the Npixel predictor, and one model adding both predictors to the null model, were compared to the null model. Note that also the interaction terms with the word/non-word parameter were included. The Akaike Information Criterion (AIC) for difference to the null model is shown for each model. A positive value represents an increase in model fit; asterisks mark significant differences ($p < .05$ Bonferroni corrected for multiple comparisons; 6 comparisons, three in relation to the null model and three, marked with asterisks, comparing the alternative models; corrected significance threshold $p < .0083$). (d-f) Analogous results for English and (g-i) for French word/non-word decision tasks. Visual noise experiment: (j) Example stimuli representing the three visual noise levels. (k) Orthographic prediction error effect (oPE) when no noise was applied, replicating the first study presented in a (error rate: 6%). (l) Number of pixels effect (Npixel) in the condition where noise was strongest (error rate: 33%). (m-o) Model comparisons including the full models and the models with oPE and Npixels only for each of the noise levels. Note that for the noise study, AIC comparisons were Bonferroni corrected for nine comparisons (corrected significance threshold $p < .0055$).

Additionally including orthographic distance (OLD20 Yarkoni et al., 2008) as predictor improved the model fit further (AIC difference comparing the full model with and without OLD20: 104; $\chi^2(2) = 105.8$; $p < .001$) but did not affect the significance of the word/non-word-by-orthographic prediction error interaction (Interaction effect estimate after including additional parameters: 0.03; SE = 0.01; t = 5.2). This finding indicates that despite its correlation with other orthographic measures (Fig. 2b, c), the orthographic prediction error accounts for unique variance components in word recognition behavior that cannot be explained by other word characteristics.

We also replicate this interaction when calculating the orthographic prediction error using a length-unspecific visual-orthographic prediction (i.e., based on all ~190,000 German words from the SUBTLEX database (Brysbaert et al., 2011); 2-36 letters length; cf. Fig. 1e; LMM estimate of interaction effect: 0.03; SE=0.01; t=4.5; for replication in English and a more extensive investigation of the interaction effect for multiple word lengths see Supplemental figure 1a). Interestingly, length-specific and length-unspecific orthographic prediction errors are highly correlated (e.g., German: r = .97), showing that the prediction-based word recognition process proposed by the PEMoR model is independent of word length constraints. This finding is in line with evidence from natural reading, which shows that one can extract low-level visual features like the number of letters from the parafoveal vision before fixating the word (Cutter, Drieghe, & Liversedge, 2014; Gagl, Hawelka, Richlan, Schuster, & Hutzler, 2014; Schotter, Angele, & Rayner, 2012). The use of a fixed of word length in our German lexical decision experiment is therefore not necessarily artificial since in natural reading word length is known before fixation. In sum, these results demonstrate that the orthographic prediction error is meaningfully related to word recognition behavior and independent of word length.

Generalization across languages

The interaction effect between lexicality (word/non-word status) and orthographic prediction error could be replicated in two open datasets from other languages, i.e., British English (Keuleers, Lacey, Rastle, & Brysbaert, 2012; 78 participants and 8,488 words/non-words: Fig. 3d; estimate: 0.008; SE = 0.002; t = 4.2) and French (Ferrand et al., 2010; 974 participants and 5,368 words/non-words: Fig. 3g; estimate: 0.005; SE = 0.002; t = 2.0); see also Supplemental Figure 2 for two further datasets from Dutch and Supplemental Table 1 for detailed results. However, in contrast to German, in both datasets we also found a significant effect of the number of pixels parameter (Fig. 3e,h; British: fixed effect: 0.008; SE = 0.001; t = 6.7; French: interaction with word/non-word

status: -0.007; SE = 0.002; t = 3.0). In terms of model comparison, the pattern derived from German, i.e., the greatest increase in model fit when including the orthographic prediction error, could not be recovered for English and French. Rather, we found that the role of the number of pixels parameter for describing the response times was larger than in German (see Fig. 3f,i). Still, the combined model showed the best model fit in all three languages (oPE only vs. full model: AIC difference English: 52; $\chi^2(2) = 56.5$; $p < .001$; French: 6; $\chi^2(2) = 10.5$; $p = .005$; Npixel only vs. full model: AIC difference English: 24; $\chi^2(2) = 28.6$; $p < .001$; French: 3; $\chi^2(2) = 7.8$; $p = .02$; Bonferroni corrected p threshold: 0.0083) indicating that both the orthographic prediction error and the number of pixels parameter are relevant in explaining word/non-word decision behavior. To summarize, for English and French, model comparisons showed that in addition to the prediction error, the parameter reflecting more directly the physical stimulus input explained a greater amount of variance than in German. Nevertheless, in all three languages, the orthographic prediction error explained unique variance components, which further supports its relevance for understanding visual word recognition. Future research should aim at clarifying the differential reliance on the bottom-up input itself in different languages but also see the next section for a potential explanation.

Word recognition behavior under conditions of visual noise

We speculated that the more significant role for bottom-up input in the British and French datasets might result from the presence of several sources of additional perceptual variability. For example, word length changed from trial to trial (English, 2-13 letters; French, 2-19 letters) and a proportional font (Times new roman) was used in the English dataset, while we had used only five-letter words presented in a monospaced font in both the German experiment and the implementation of the PEMoR (Fig. 1). Even though such unpredictable perceptual variation, without any doubt, is not the standard case in naturalistic reading (i.e., through the integration of visual information from parafoveal vision), in a single word reading paradigm it reduces the ability to predict visual features of upcoming stimuli and thus unnaturally decreases the performance of our model. For example, using a proportional-spaced font removes structure (e.g., the letter separation) both in the sensory input and the orthographic prior, which results in less precise predictions and more substantial prediction errors. This reduction in prediction strength, in turn, increases the correlation between the number of pixels in the input image and the derived orthographic prediction error (cp. Monospace font: r = .05 vs. proportional font: r = .49, both in German; see Supplement 2). In the face of this, it is particularly

noteworthy that the orthographic prediction error, as proposed here, remained a highly relevant predictor in the English and French data set. It should also be stressed again that in natural reading, low-level visual features like word length or letter position can be picked up in parafoveal vision, so that the visual system may be able to dynamically adapt its predictions to the upcoming word (Schotter et al., 2012). Future work will, therefore, have to specify in more detail the nature of orthographic priors in naturalistic reading.

To directly test if visual word recognition relies more firmly on the bottom-up input when visual word presentation includes unpredictable perceptual variations, we conducted a second lexical decision experiment. We presented visual word stimuli with an explicit manipulation of visual noise (0% vs. 20% vs. 40% noise level) to reduce the predictability of visual features (for details see Methods section). A noise manipulation, rather than, e.g., a comparison of different fonts, was applied since noise levels can be easily manipulated and quantified (i.e., in terms of the number of displaced pixels). In contrast, a direct comparison of fonts is more difficult because the contrast of proportional vs. mono-spaced font is confounded with many other visual differences like total stimulus width (Hautala, Hyönä, & Aro, 2011; Marinus et al., 2016). In addition, the 0% noise stimuli allowed us to replicate our original behavioral finding. Figure 3j shows examples word stimuli.

We found, in general, that response times and errors increased with the amount of noise that was applied to the visual-orthographic stimuli (0%: response time/RT: 613 ms, 6% errors; 20%: RT: 739 ms, 12% errors; 40%: RT: 1,105 ms, 33% errors; compare also Fig. 3k and l). When no noise was applied we replicated our first study (cp. Fig. 3k and a) with a significant interaction between the orthographic prediction error and the word/non-word factor (estimate: 0.05; SE = 0.02; t = 2.3; see Supplemental Table 1 for detailed results). As in the first experiment, no effect or interaction was found for the number of pixels parameter. With 20% noise, we still could identify a fixed negative effect of the orthographic prediction error (estimate: -0.06; SE = 0.02; t = 3.3) however without a significant interaction pattern. Also, the fixed effect of the number of pixels was not significant. With 40% noise, however, no significant effect of the orthographic prediction error could be found but as expected from the above discussion of noise effects, we observed now a significant fixed effect of the number of pixels parameter, as well as an interaction with word/non-word status (Fig. 3l; estimate: 0.08; SE = 0.03; t = 2.9). A similar impression can be obtained from the model fit results showing that including the orthographic prediction error resulted in significantly higher model fits for 0% and 20% noise conditions compared to models in which only the number of pixels predictor was

included (see Fig. 3m-n; 0% AIC difference: 1; $\chi^2(0) = 1$; $p < .001$; 20% AIC difference: 13; $\chi^2(0) = 13.7$; $p < .001$). With 40% noise, inclusion of the number of pixels parameter resulted in a higher model fit (see Fig. 3o; AIC difference: 13; $\chi^2(0) = 13.2$; $p < .001$), but including the orthographic prediction error had essentially no effect.

In sum, the behavioral experiments reported in this section demonstrate that the orthographic prediction error contributes substantially to visual word recognition. We find the PEMoR is highly relevant when the visual information presented in the lexical decision tasks is with a restricted variability and, therefore, high predictability (i.e., all words with the same number of letters), which is typically the case in natural reading situations (i.e., sentence reading). The predictability of visual features results in greater reliance on the orthographic prediction error compared to the pure bottom-up sensory input. In contrast, when the perceptual variability (i.e., complexity) increases, i.e., due to a variation of the number of letters, proportional fonts (e.g., as in the case of the English study) or visual noise the pure bottom-up signal, i.e., as assumed in the PixMoR, becomes the adequate parameter to explain visual word recognition behavior. Thus, the behavioral evidence indicates that one implements efficient neuronal coding when one can predict the perceptual properties of the letter strings. In case the perceptual properties are highly variable, predictive processing is hampered as only weak predictions can be formed suggesting the PixMoR as a "fallback" strategy as an approximation in effortful reading conditions.

Cortical Representation of the Orthographic Prediction Error

The PEMoR assumes that the orthographic prediction error is estimated at an early stage of the word recognition processes, i.e., in the visual-perceptual system and before word meaning is accessed and one can activate higher-level linguistic representations of the word. Note that also the end-stopping phenomenon was found up to middle temporal regions. We accordingly hypothesized that brain systems involved in computing or representing the orthographic prediction error should be driven by this optimized representation of the sensory input independent of the item's word/non-word status (i.e., for words and non-words alike). Localizing the neural signature of the orthographic prediction error in the brain during word/non-word recognition, thus, is a further critical test of the PEMoR. Of note, a strict bottom-up model of word recognition (and perception in general, i.e., PixMoR) would make a different prediction, i.e., that activation in visual-sensory brain regions should be driven by the full amount of physical information in the percept (Goodyear & Menon, 1998; Henrie & Shapley, 2005). Processes that take

place after word-identification, i.e., that involve higher levels of linguistic elaboration can only operate on mental representations of words, so that brain regions involved in these later stages of word processing should distinguish between words and non-words.

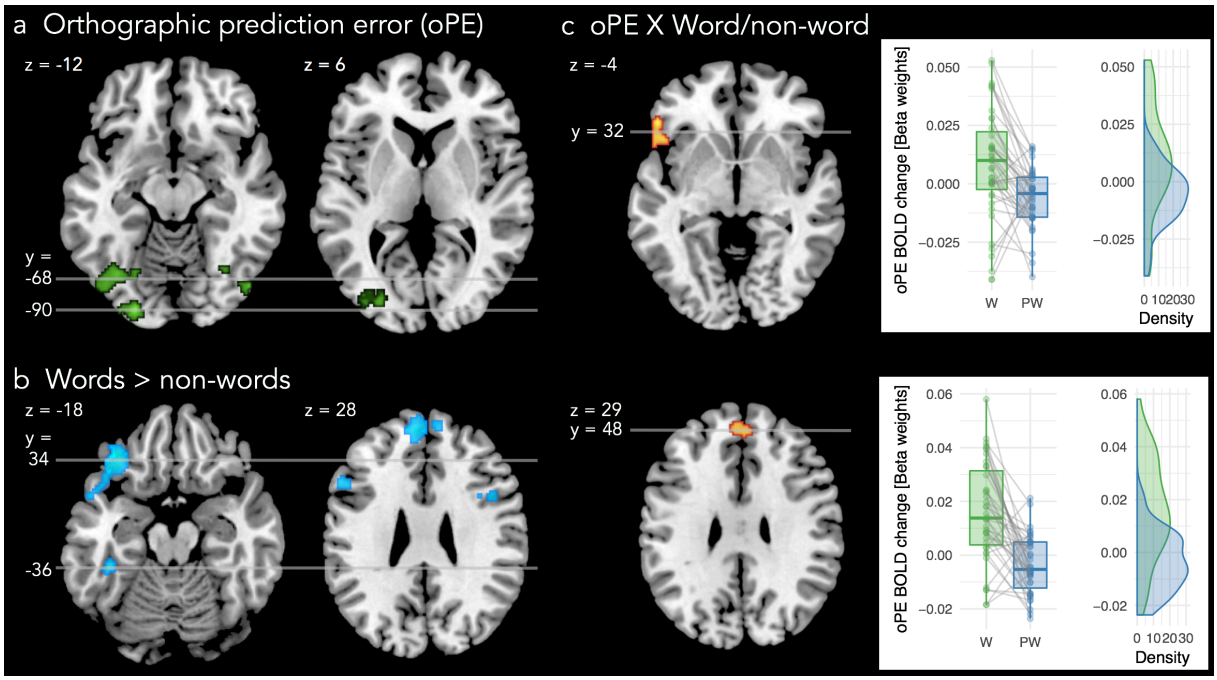

Figure 4. fMRI results demonstrating the neuroanatomical localization of orthographic prediction error effects. BOLD activation during silent reading (see Methods for further details, and Table 1 for exact locations of activation effects): (a) Analysis demonstrating a positive orthographic prediction error (oPE) effect in bilateral occipital activation-clusters. This regression analysis used item-specific oPE values as covariate, independent of stimulus condition, and shows brain regions with greater activity for letter strings characterized by a higher oPE, independent of stimulus type. (b) Clusters of higher BOLD activation for words than for non-words. (c) Two frontal activation clusters showing a oPE by word/non-word interaction, i.e. positive and negative oPE effects for words and non-words, respectively. Boxplots show individual beta weights; lines connect word and non-word betas from each individual. No effects of the number of pixels per word were found. Threshold voxel level: p<.001 uncorrected; cluster level: p<.05 family-wise error corrected. Boxplots represent the median (line), the data from the first to the third quartile (box) and ±1.5 times the interquartile range (i.e. quartile 3 minus quartile 1; whiskers).

We tested these hypotheses about the localization of the orthographic prediction error by measuring BOLD activation changes using functional MRI while 39 participants silently read words (German nouns) and pronounceable non-words (i.e., pseudowords), in randomized order (see Methods for details). We identified three left- and one right-hemispheric brain regions in the occipital cortex that showed higher levels of activation when reading items with higher orthographic prediction error – both for words and non-words (Fig. 4a and Supplemental Table 2). Prior research (Dehaene & Cohen, 2011; Dehaene et al., 2005) has identified a region in the mid-portion of the left occipito-

temporal cortex as critical for reading: the visual word form area. Consistent with our hypothesis, all four activation clusters representing the orthographic prediction error are located posterior to this so-called visual word form area (Dehaene & Cohen, 2011), which supports our claim of an 'early' role for the orthographic prediction error signal before word identification. Importantly, no brain areas showed activity dependent on the pure amount of bottom-up information in the percept (i.e., the number of pixels parameter).

Only brain regions involved in the activation of word meaning and subsequent processes should differentiate between words and non-words. We observed higher activity for words than non-words, independent of the orthographic prediction error, more anteriorly in left temporal and prefrontal cortex (Fig. 4b and Supplemental Table 2). Third, the left inferior frontal gyrus (pars triangularis) and the medial portion of the superior frontal gyrus (mSFG) mirrored the word/non-word decision behavior reported above, in that higher prediction errors lead to increased activation for words but decreasing activation for non-words (Fig. 4c and Supplemental Table 2). The fMRI experiment, thus, supports our hypothesis that during the earliest stages of visual processing, i.e., presumably before accessing word meaning, an optimized perceptual signal, the orthographic prediction error, is generated and used as a basis for efficient visual-orthographic processing of written language. Only at later processing stages (in more anterior temporal and prefrontal cortices), the brain differentiates between words and non-words.

Cortical timing of the Orthographic Prediction Error

While of the fMRI results demonstrate a representation of the orthographic prediction error in presumably 'early' visual brain regions, the temporal resolution of fMRI precludes inferences concerning the temporal sequence of cognitive processes during word recognition. The millisecond time resolution of EEG has helped to consistently attribute the extraction of meaning-from perceived words to a time window of around 300 to 600 ms post word onset (N400 component of the event-related brain potential/ERP; Kutas & Federmeier, 2011). Visual-orthographic processes associated with the orthographic prediction error should thus temporally precede this time window, most likely to occur during the N170 component of the ERP (Barber & Kutas, 2007; Carreiras, Armstrong, Perea, & Frost, 2014; Grainger & Holcomb, 2009). To test this hypothesis, we measured EEG while 31 participants silently read words and non-words (including both pseudowords and consonant-only strings). We fitted a multiple regression model (analogous to the model used for the analysis of behavioral data) to the EEG data (Linzen

& Engemann, 2017) with the orthographic prediction error, the number of pixels, word/non-word–status, and their interactions as parameters (see Methods for details).

Regression-estimated ERPs (see methods for details) show a significant effect of the orthographic prediction error on electrical brain activity between 150 and 250 ms after stimulus onset (Fig. 5a). In this early time window, letter-strings characterized by higher prediction errors elicited significantly more negative-going ERPs over posterior-occipital sensors, for both words and non-words. In line with the temporal sequence of processes inferred from their neuroanatomical localizations (i.e., fMRI results), a significant word/non-word effect then emerged between 200-570 ms (Fig. 5b), followed by an interaction between word/non-word-status and orthographic prediction error at 360-620 ms (Fig. 5c). In this interaction cluster, higher prediction errors led to more negative-going ERPs for non-words, as observed for all stimuli in the earlier time window, but showed a reverse effect for words, i.e., more positive-going ERPs for words with higher prediction errors (Fig. 5c). This pattern of opposite prediction error effects for words vs. non-words is analogous to the effects seen in word/non-word decision behavior and the frontal brain activation results obtain with fMRI.

As in the fMRI study, we found no effect of the bottom-up input as such (pixel count), even though it is well-established that manipulations of physical input contrast (as determined, e.g., by the strength of luminance Johannes, Münte, Heinze, & Mangun, 1995) can increase the amplitude of early ERP components starting at around 100 ms. We performed an explicit model comparison between statistical models including the orthographic prediction error compared to a model including the number of pixels parameter (analogous to the analysis of behavioral data), for both time windows in which the orthographic prediction error was relevant (early fixed effect and later interaction). In both time windows the model including the orthographic prediction error resulted in better fit (AIC difference: 16 at 230 ms at posterior sensors: $\chi^2(0) = 16.0$; $p < .001$; and 5 at 430 ms at frontal sensors: $\chi^2(0) = 5.0$; $p < .001$) . Even when investigating the combined models, including both parameters, we found a tendency for a better fit in the oPE only model (AIC difference: 3 at 230 ms at posterior sensors and 3 at 430 ms at frontal sensors).

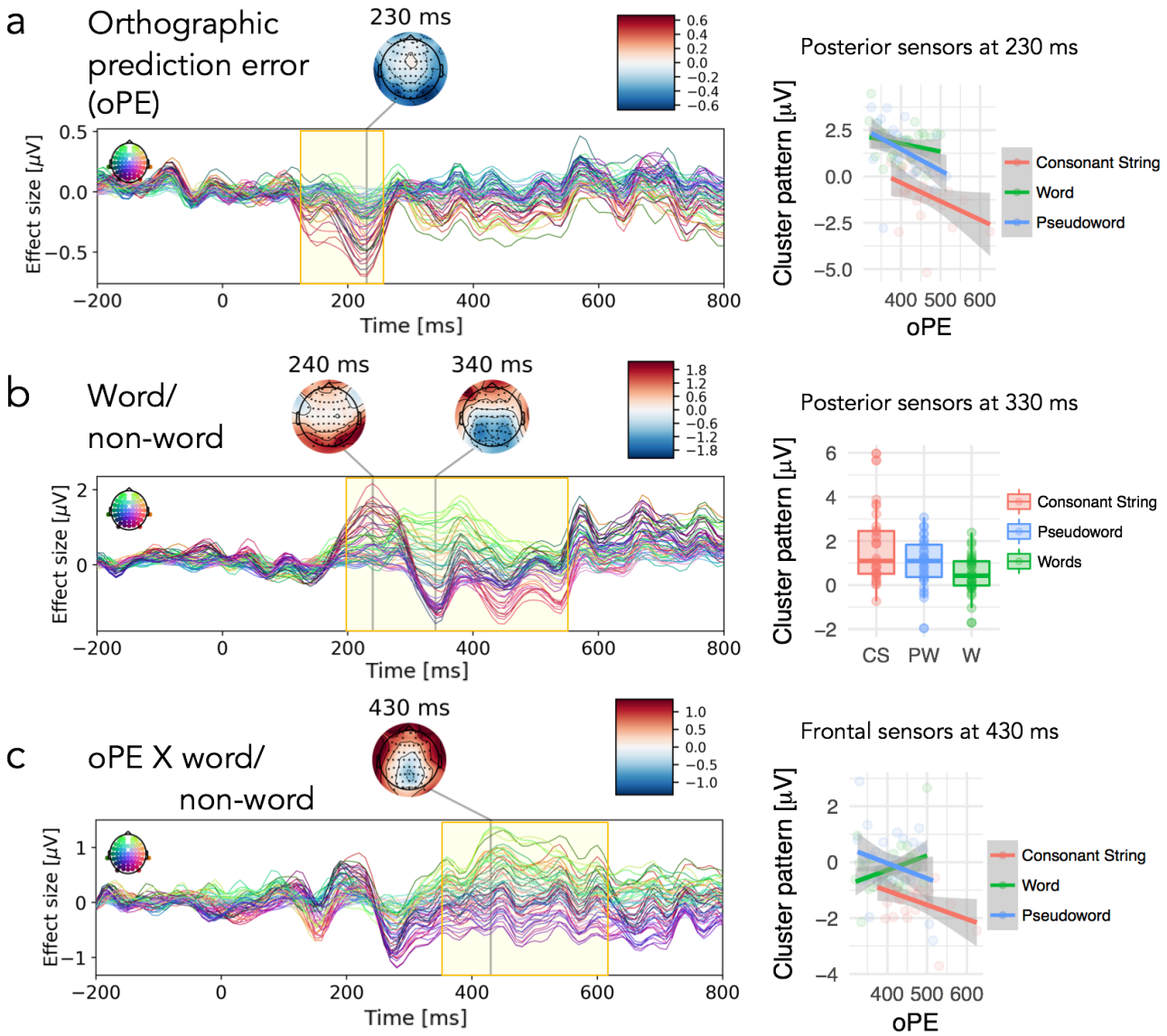

Figure 5. EEG results: Timing of orthographic prediction error effects. Effect sizes from regression ERPs are presented as time courses for each sensor and time-point (left column; color coding reflects scalp position) with yellow areas marking time windows with significant activation clusters for silent reading of 200 words and 200 non-words (100 pronouncable pseudowords, 100 consonant strings; see Supplemental Figure 3 for a more detailed visualization of the significance of spatio-temporal activation clusters). ERP results are shown for (a) the orthographic prediction error (oPE) main effect, (b) the word/non-word effect, and (c) the oPE by word/non-word interaction. Results indicate significant oPE, word/non-word, and oPE by word/non-word effects starting around, 150, 200, and 360 ms, respectively. The right panel shows the activation patterns related to the significant activation clusters (cf. Supplemental Figure 3) in more detail. Dots represent mean predicted µV across (a,c) all participants and items separated by oPE and stimulus category, and (b) all items separated by stimulus category, excluding confounding effects (see Methods). No significant activation clusters were found for the parameter representing the number of pixels. Boxplots represent the median (line), the data from the first to the third quartile (box) and ±1.5 times the interquartile range (i.e. quartile 3 minus quartile 1; whiskers). The frontal cluster includes the following sensors: AF3, AF4, AF7, AF8, F1, F2, F3, F4, F5, F6, F7, F8, SO1, SO2, FP1, FP2, Fz. The posterior cluster includes the following sensors: O2, O1, Oz, PO10, PO3, PO4, PO7, PO8, PO9, POz.

To summarize, EEG results converge with behavioral and fMRI results. They suggest that relatively early on in the cortical visual-perceptual processing cascade, the amount of perceptual processing devoted to the orthographic percept is smallest for letter-strings with highly expected visual features (i.e., low orthographic prediction error). 100 to 200 ms later, i.e., in a time window strongly associated with semantic processing (Kutas & Federmeier, 2011), the prediction error effect was selectively reversed for words, and thus started to differentiate between the two stimulus categories. This finding mirrors behavioral results and activation patterns in the anterior temporal lobe and prefrontal cortex found in the fMRI dataset. In sum, these results support the PEMoR's proposal that orthographic representations are optimized early during visual word recognition, and that the resulting orthographic prediction error is the basis for subsequent stages of word recognition.

Applying the PEMoR to handwritten script.

The electronic fonts used for all above-reported experiments introduce a highly regular structure that favors some of the PEMoR's core processes, like the calculation of the orthographic prediction error (i.e., the prior). We showed above that when reducing the high regularity of computerized script by visual noise, reading performance decreases and the orthographic prediction error becomes less relevant for describing reading behavior. To demonstrate the 'real world' validity of the PEMoR with even less regular scripts, we applied a variant of this model to the reading of naturalistic handwritings obtained from 10 different writers. The extreme variability of different handwritings strongly influences their readability (compare Figs. 6a,b). The visual-orthographic predictions which have here been implemented based on single letters and separately for each handwriting, accordingly vary substantially in the strength and precision between individual handwritings (cp. 'prediction' of Fig. 6a,b). As in PEMoR for computer script, we once more define prediction strength in terms of the darkness of gray values of the prediction image, i.e., the mean gray value across pixels, and observe lower oPEs for handwritings that allow stronger predictions (Fig. 6c; linear mixed model statistics: Estimate: -0.05; SE = 0.01; t = 7.4). The precision of the prediction is represented by the inverse of the number of gray pixels included the prediction image; more precise predictions are more focused and less distributed, and also elicit lower orthographic prediction errors (Fig. 6d; 0.02; SE = 0.01; t = 2.1; see Supplemental Table 1 for full results). Finally, we obtained the rated readability of each handwriting based on ten written words and observed that the readability is higher for handwritings that produce

lower prediction errors (Fig. 6e; 38 raters; Estimate: -5.9; SE = 1.0; t = 6.2). These results demonstrate that (variants of) the PEMoR can account for reading processes not only in highly formalized stimuli but also in more naturalistic settings.

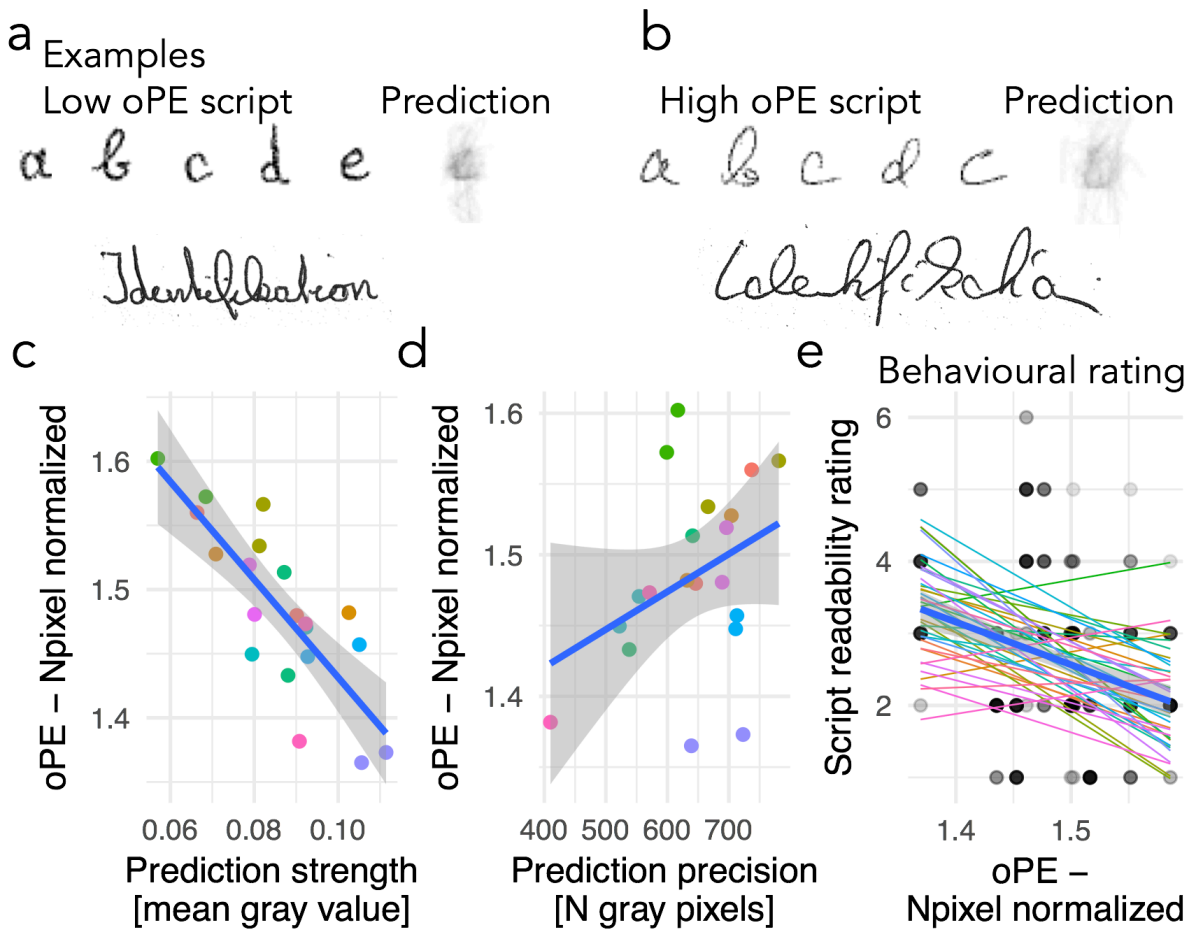

Figure 6. Applying the PEMoR to the perception of handwritings. Examples for two (out of 10 empirically obtained) different handwritings. (a) A handwriting including single letters and the respective (letter-level) orthographic prediction estimated based on all 26 lower case letters (written in isolation). In addition, the word Identifikation (identification) is presented for both handwritings, as an example (out of 10). These words were used to acquire the readability rating. (c) Relationship between prediction strength and the mean orthographic prediction error across all letters for each script. Note that the oPE estimate for handwritings was normalized (i.e. divided) by the number of pixels since the number of pixels differed drastically between scripts (e.g. compare Examples in a and b). (d) Relationship between the precision of the prediction and the orthographic prediction error. Point color reflects each of 10 individual scripts, separately for upper- and lower-case letters. (e) Script readability ratings in relation to the orthographic prediction error (lower- and upper-case prediction error combined). Blue line reflects the overall relationship and thin lines represent each rater.

Discussion

Here we investigated if an efficient neuronal code representing the visual information in words, i.e., analogous to the neural representation underlying the end stopping

phenomenon for oriented lines, is the basis of proficient reading. We found that our Prediction Error Model of Reading (PEMoR) is a plausible account. Surprisingly, since the focus of the PEMoR is visual processing, the resulting prediction error, i.e., the non-redundant and thus informative part of the visual percept, represents not only the visual but also orthographic word information. We concluded this from the correlations between various lexicon-based descriptors of words, associated with orthographic stages of visual word recognition, and the prediction error representation. Our empirical observations also support this conclusion: We found that the orthographic prediction error (i) accounts for word identification behavior, (ii) explains brain activation in visual-perceptual systems of the occipital cortex, (iii) explains brain activation as early as 150 ms after the onset of the letter-string, and (iv) is represented in high-level lexical processing in frontal regions as well as the N400 component as part of the processes that underlie behavior. We inferred the latter from the finding that both the brain activation (i.e., in frontal areas and the N400) and the behavioral evidence showed comparable interaction pattern. Also, the PEMoR provides a quantitative estimate of the amount of information reduction achieved by this mechanism (i.e., in our data between 31 and 37% on average depending on language, with an upper limit of 51% at the level of the individual word). Finally, we have provided the first evidence that the principles of predictive coding may also apply to more naturalistic reading situations, for example, to account for individual differences in the readability of handwriting. In sum, our findings indicate that the basis for fast access to the meaning of written words is an informationally optimized neuronal code representing visual-orthographic word information.

We also found evidence that the reliance on the orthographic prediction error in word recognition is related to the perceptual quality of the stimulus. We showed that in case the visual occurrence of the stimulus is less predictable, e.g., due to visual noise, the amount of visual information (i.e., the number of pixels) predicted the behavioral performance better than the prediction error. As described previously (Rao & Ballard, 1999), efficient coding in a predictive system relies on the structure present in the stimulus (i.e., more structured handwriting results in stronger and more precise predictions). If the structure is compromised, i.e., by visual noise, the predictive system breaks down as predictions become weaker and less precise (i.e., imagine when the images used as the basis for the prediction in the PEMoR would be noisy). As a consequence, word processing relies on less efficient neuronal codes under such conditions.

Most results reported here relied on experiments with fixed word lengths, while naturalistic reading involves considerably more variability at the level of the input. However, para-foveal vision provides information about word length to the visual system before the actual processing of the word (Schotter et al., 2012), so that it can dynamically implement best-fitting visual-orthographic predictions (priors) online during reading. This would, in principle, allow for optimized sensory processing as described by the PEMoR in natural reading situations. This hypothesis must be tested in future studies but fits with previous theoretical proposals which have acknowledged the integration of top-down predictions from multiple linguistic domains (for example at the phonological, semantic, or syntactic level DeLong et al., 2005; Eisenhauer, Fiebach, & Gagl, 2019; Nieuwland et al., 2018; Price & Devlin, 2011). Critically, our results go beyond these earlier models by demonstrating that top-down guided expectations are implemented already onto early visual-orthographic processing stages.

The so-far dominant model of visual word recognition in the brain (Dehaene & Cohen, 2011; Dehaene et al., 2005) postulates that words are 'assembled' bottom-up along the visual pathway, starting with symbolic representations of letter features up to successively more complex higher-order representations. In this and similar models (including computational accounts of visual word recognition, e.g., Coltheart et al., 2001) the symbolic letter representation is inferred from the visual input stimulus. Here we show that the representational spaces of our original word stimulus images (i.e., the visual input) and their derived orthographic prediction errors (Fig. 2d) correlate to a high degree, indicating that, in principle, the PEMoR does not contradict current models of visual word recognition. In contrast, PEMoR specifies explicitly and in a testable manner the neuronal code from which we infer symbolic representations like letters. Prediction-based top-down optimization of the visual-orthographic input, as proposed here, is thus not necessarily incompatible with the current models of reading and visual word recognition, but offers a specification of a previously underspecified visual input representation.

Predictive coding-based theories, in general, assume that higher-level processing is concerned with prediction error minimization as a core computation (K. Friston, 2005b; Price & Devlin, 2011). As described in the introduction, such an account would be fundamentally different from most assumptions in visual word recognition models. Accordingly, a future challenge for model development will be the incorporation of the orthographic prediction error and current model assumptions about orthographic processing. One possibility here could be the inclusion of the orthographic prediction

error as partial evidence in an evidence accumulation process with the goal of word recognition (similar as previously described in Ref. Gagl, Richlan, Ludersdorfer, Sassenhagen, & Fiebach, 2016; Ratcliff, Gomez, & McKoon, 2004; Summerfield & de Lange, 2014).

In sum, we demonstrate that during reading, visual information is optimized by 'explaining away' redundant visual information based on top-down predictions. This study provides strong evidence that reading follows domain-general mechanisms of predictive coding during perception (Clark, 2013) and is also consistent with the influential hypothesis of a Bayesian brain, which during perception continuously combines prior knowledge and new sensory evidence (K. Friston, 2005a; Knill & Pouget, 2004). We propose that the result of this optimization step, i.e., an orthographic prediction error signal, is the efficient neuronal access code to subsequent 'higher' levels of word processing, including the activation of word meaning. These data provide the basis for a new understanding of early, i.e., pre-lexical orthographic stages of visual word recognition, rooted in a strong and widely accepted, domain-general neurophysiological model –prediction-based perception (K. Friston, 2005a; Rao & Ballard, 1999). At the same time, our results provide crucial converging evidence in support of predictive coding theory.

Acknowledgments

We thank Rebekka Tenderra, Anne Hoffmann, Jan Jürges, and Kirsten Hilger for help with EEG data acquisition. In addition, we thank Mark D'Esposito, David Poeppel, Matt Davis, and Ulrike Basten for helpful comments on a previous version of the manuscript. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2013) under grant agreement n° 617891 awarded to CJF and from the European Community's Horizon 2020 Programme (2016) under grant agreement n° 707932 awarded to BG.

Author contributions

B.G. and C.F. wrote and revised the manuscript and all of the authors edited the final drafts. B.G., J.S., and C.F. conceptualized the model. B.G. and S.H. implemented the model. B.G. implemented the model simulations, conceptualized behavioral experiments, and behavioral data analysis. B.G. and K.G. implemented behavioral data acquisition, and the handwriting adaptation. B.G., J.S., and S.H. designed, measured and analyzed the EEG study. B.G. and F.R. designed, measured and analyzed the fMRI study.

References

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nature Neuroscience, 14(6), 797–801. https://doi.org/10.1038/nn.2810

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. NeuroImage, 38(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. Journal of Experimental Psychology. General, 133(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. Brain Research Reviews, 53(1), 98–123. https://doi.org/10.1016/j.brainresrev.2006.07.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bolz, J., & Gilbert, C. D. (1986). Generation of end-inhibition in the visual cortex via interlaminar connections. Nature, 320(6060), 362. https://doi.org/10.1038/320362a0

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The Word Frequency Effect. Experimental Psychology, 58(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. Journal of Experimental Psychology: Human Perception and Performance, 42(3), 441–458. https://doi.org/10.1037/xhp0000159

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. Trends in Cognitive Sciences, 18(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Changizi, M. A., Zhang, Q., Ye, H., & Shimojo, S. (2006). The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes. The American Naturalist, 167(5), E117–E139. https://doi.org/10.1086/502806

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In Attention and performance VI. Proceedings of the Sixth International Symposium on Attention and Performance, Stockholm, Sweden, July 28-August 1, 1975.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review, 108(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(6), 1778–1786. https://doi.org/10.1037/xlm0000013

Davis, C. J. (2010). The spatial coding model of visual word identification. Psychological Review, 117(3), 713–758. https://doi.org/10.1037/a0019738

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. Trends in Cognitive Sciences, 15(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. Trends in Cognitive Sciences, 9(7), 335–341. https://doi.org/10.1016/j.tics.2005.05.004

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience, 8(8), 1117–1121. https://doi.org/10.1038/nn1504

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage, 34(4), 1443–1449.

Edelman, S. (1998). Representation is representation of similarities. Behavioral and Brain Sciences, 21(4), 449–467.

Eisenhauer, S., Fiebach, C. J., & Gagl, B. (2019). Context-based facilitation in visual word recognition: Evidence for visual and lexical but not pre-lexical contributions. ENeuro, ENEURO.0321-18.2019. https://doi.org/10.1523/ENEURO.0321-18.2019

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. Psychological Review, 112(4), 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., … Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. Behavior Research Methods, 42(2), 488–496. https://doi.org/10.3758/BRM.42.2.488

Friston, K. (2005a). A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2005b). A theory of cortical responses. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian Inference in Neuroimaging: Applications. NeuroImage, 16(2), 484–512. https://doi.org/10.1006/nimg.2002.1091

Gagl, B., Hawelka, S., Richlan, F., Schuster, S., & Hutzler, F. (2014). Parafoveal preprocessing in reading revisited: Evidence from a novel preview manipulation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(2), 588–595. https://doi.org/10.1037/a0034408

Gagl, B., Richlan, F., Ludersdorfer, P., Sassenhagen, J., & Fiebach, C. J. (2016). The lexical categorization model: A computational model of left ventral occipito-temporal cortex activation in visual word recognition. BioRxiv, 085332. https://doi.org/10.1101/085332

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal Predictive Codes for Spoken Words in Auditory Cortex. Current Biology, 22(7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015

Goodyear, B. G., & Menon, R. S. (1998). Effect of Luminance Contrast on BOLD fMRI Response in Human Primary Visual Areas. Journal of Neurophysiology, 79(4), 2204–2207. https://doi.org/10.1152/jn.1998.79.4.2204

Grainger, J., & Holcomb, P. J. (2009). Watching the Word Go by: On the Time-course of Component Processes in Visual Word Recognition. Language and Linguistics Compass, 3(1), 128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. Psychological Review, 103(3), 518.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. NeuroImage, 86, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Haarmann, H. (2007, March 15). Geschichte der Schrift | Haarmann, Harald | Verlag C.H.BECK Literatur - Sachbuch - Wissenschaft. Retrieved March 27, 2017, from http://www.chbeck.de/Haarmann-Geschichte-Schrift/productview.aspx?product=20173

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. Vision Research, 51(15), 1719–1727. https://doi.org/10.1016/j.visres.2011.05.015

Henrie, J. A., & Shapley, R. (2005). LFP Power Spectra in V1 Cortex: The Graded Effect of Stimulus Contrast. Journal of Neurophysiology, 94(1), 479–490. https://doi.org/10.1152/jn.00919.2004

Heuven, W. J. B. van, 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(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Hohenstein, S., & Kliegl, R. (2017). remef: Remove Partial Effects [R]. Retrieved from https://github.com/hohenstein/remef (Original work published 2014)

Hubel, D. H., & Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. Journal of Neuroscience, 7(11), 3378–3415. https://doi.org/10.1523/JNEUROSCI.07-11-03378.1987

Hubel, David H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of Neurophysiology, 28(2), 229–289. https://doi.org/10.1152/jn.1965.28.2.229

Johannes, S., Münte, T. F., Heinze, H. J., & Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. Cognitive Brain Research, 2(3), 189–205. https://doi.org/10.1016/0926-6410(95)90008-X

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. Annual Review of Psychology, 55(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research (Version 0.3.0) [GNU R]. Retrieved from https://cran.r-project.org/web/packages/vwr/index.html

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. Behavior Research Methods, 42(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

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. https://doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. Behavior Research Methods, 44(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

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

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. Frontiers in Psychology, 1, 238. https://doi.org/10.3389/fpsyg.2010.00238

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. Trends in Neurosciences, 27(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience, 2. https://doi.org/10.3389/neuro.06.004.2008

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. https://doi.org/10.1146/annurev.psych.093008.131123

Linzen, T., & Engemann, D. (2017). Sensor space least squares regression — MNE 0.15.dev0 documentation. Retrieved June 23, 2017, from http://martinos.org/mne/dev/auto_examples/stats/plot_sensor_regression.html

Marinus, E., Mostard, M., Segers, E., Schubert, T. M., Madelaine, A., & Wheldall, K. (2016). A Special Font for People with Dyslexia: Does it Work and, if so, why? Dyslexia, 22(3), 233–244. https://doi.org/10.1002/dys.1527

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mayringer, H., & Wimmer, H. (n.d.). SLS 2-9 - Salzburger Lese-Screening für die Schulstufen 2-9 – Hogrefe Verlag. Retrieved July 29, 2019, from https://www.testzentrale.ch/shop/salzburger-lese-screening-fuer-die-schulstufen-2-9.html

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(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : A new French lexical database. Behavior Research Methods, Instruments, & Computers, 36(3), 516–524. https://doi.org/10.3758/BF03195598

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018, April 3). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. https://doi.org/10.7554/eLife.33468

Norris, D. (2013). Models of visual word recognition. Trends in Cognitive Sciences, 17(10), 517–524. https://doi.org/10.1016/j.tics.2013.08.003

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBImage—an R package for image processing with applications to cellular phenotypes. Bioinformatics, 26(7), 979–981. https://doi.org/10.1093/bioinformatics/btq046

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. Psychological Review, 114(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. Trends in Cognitive Sciences, 15(6), 246–253. https://doi.org/10.1016/j.tics.2011.04.001

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1), 79–87. https://doi.org/10.1038/4580

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. Psychological Review, 111(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. Behavioral and Brain Sciences, 26(04), 445–476.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019. https://doi.org/10.1038/14819

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. Attention, Perception, & Psychophysics, 74(1), 5–35.

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. Neuron, 96(1), 89-97.e4. https://doi.org/10.1016/j.neuron.2017.09.007

Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2008). Large-Scale Modeling of Wordform Learning and Representation. Cognitive Science, 32(4), 741–754. https://doi.org/10.1080/03640210802066964

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive Coding: A Fresh View of Inhibition in the Retina. Proceedings of the Royal Society of London. Series B, Biological Sciences, 216(1205), 427–459.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. Nature Reviews Neuroscience, 15(11), 745–756. https://doi.org/10.1038/nrn3838

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. The Journal of Neuroscience, 31(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. Journal of Neuroscience, 32(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: A general framework using a genetic algorithm. NeuroImage, 18(2), 293–309. https://doi.org/10.1016/S1053-8119(02)00046-0

Whitney, C., & Cornelissen, P. (2008). SERIOL Reading. Language and Cognitive Processes, 23(1), 143–164. https://doi.org/10.1080/01690960701579771

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. Psychonomic Bulletin & Review, 15(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Supplementary Materials

Fig. S3.1. Behavioral evaluation, including stimuli with 4-8 letters.
Fig. S3.2. Dutch lexical decision behavior and prediction using a proportional script.
Table S3.3. Results from linear mixed model regression analysis
Table S4. Reliable activation clusters from the fMRI evaluation with respective anatomical labels
Fig. S5. Detailed description of significant activation clusters in the EEG study

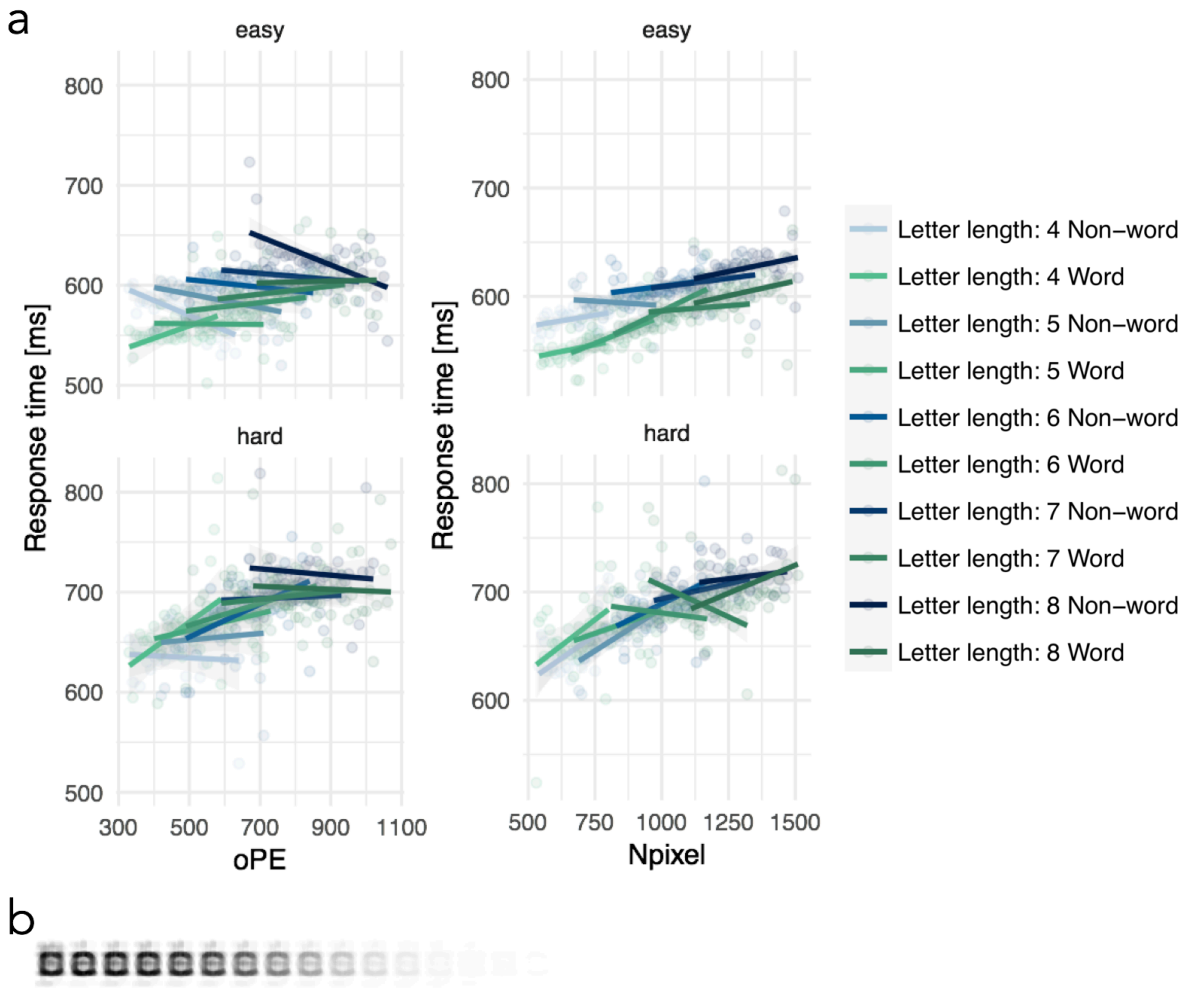

Figure S3.1. Behaviroal evaluation including multiple word lengths. (a) Response times aggregated across participants from the British lexicon (BLP) project (Keuleers et al., 2012) for the word lengths 4-8. The left panel shows the word/non-word by orthographic prediction error (oPE) interaction and the right panel shows the word/non-word by number of pixels (Npixel) interaction for each word length separately. In addition, the upper panel shows letter strings that are correctly categorized in nearly all cases (accuracy > .95) and the lower panel shows the response times to the items, which were less accurately processed (i.e., accuracy < .95). The median split resulted in a subset of the BLP (i.e., the easy words) which are roughly comparable to words used in the previous experiments (e.g. see Fig. 3), as the BLP study includes a large number of very rare words (median log. word frequency per million is .3). Bluish colors represent non-words (N) and greenish colors represent words (W), while the hue of the colors reflects word length (i.e., bright to dark reflects short to long letter strings). For both effects, we first estimated linear regression models with either the oPE or the Npixel effect and allowing interactions with word/non-word status, word length, and accuracy. Note that the oPE in this first analysis was based on length-specific predictions (i.e., for the estimation of the oPE of four-letter words, all four-letter words of the lexicon were included in the prediction). For the oPE model, a significant four-way interaction was found (estimate = -1.078e-04; SE = 4.199e-05; t = -2.567). Separating hard vs. easy words allowed us to disentangle the four-way interaction: In easy words/non-words, we found a consistent (i.e., across length levels) oPE by word/non-word interaction (estimate = 1.530e-04; SE = 4.047e-05; t = 3.780) in the same direction as previously shown (positive effect for words and a negative effect for non-words). For hard words/non-words, we found that the oPE by word/non-word interaction was inconsistent across letter length levels, which was indicated by a significant oPE and letter length interaction (estimate = -3.530e-05; SE = 8.092e-06; t = -4.363). In addition, for the hard words both the oPE by word/non-word interaction (estimate = -1.685e-04; SE = 6.905e-05; t = -2.440) and the main effect of oPE were reversed (estimate = 2.828e-04; SE =

5.802e-05; t = 4.874 compare to estimate = -1.000e-04; SE = 2.440e-05; t = -4.101, for easy words). For the Npixel model, no four-way interaction and no Npixel interaction or main effect were found. In sum, in this analysis we showed that the oPE by word/non-word interaction shown previously for word lengths of five letters (see main text) is consistent for easy-to-process English items with word lengths from 4-8 letters. Secondly, the word/non-word by orthographic prediction error interaction was also reliable when the prediction included all words of all letter lengths from the English lexicon (see part b of this Figure) and the orthographic prediction error estimation was based on this length-unspecific prediction (estimate: 0.02; SE=0.007; t=3.349). (b) Letter-length unspecific prediction for English based on ~60,000 English words from the SUBTLEX database (Heuven et al., 2014).

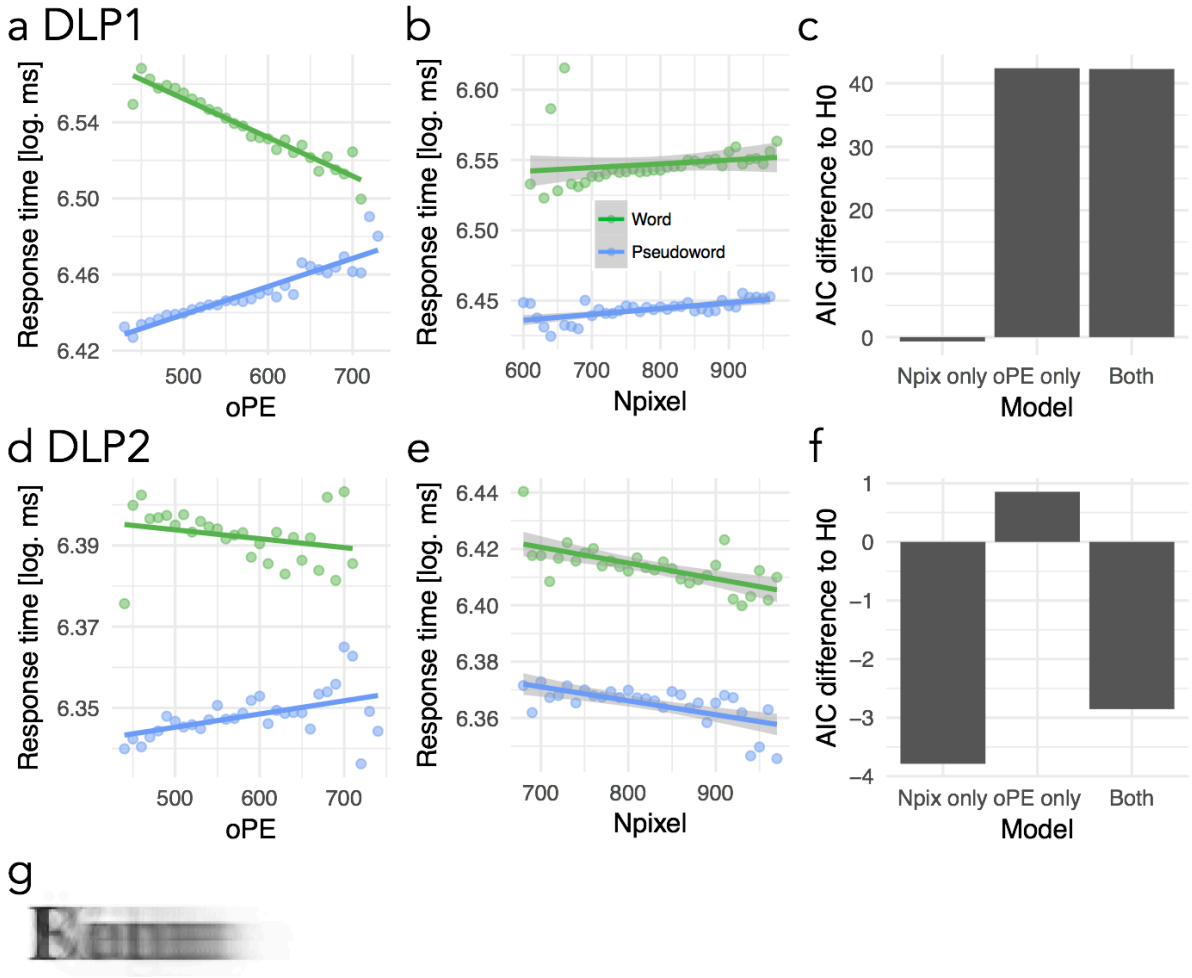

Figure S3.2. Dutch lexical decision behavior and prediction using a proportional script. (a) Effect of the orthographic prediction error parameter, (b) number of pixels parameter and (c) showing the same model comparisons as implemented in Figure 3 for the data from the first Dutch lexicon project (DLP1; (Keuleers, Diependaele, & Brysbaert, 2010); 4,305 five-letter stimuli; 39 participants) and the same effects and model comparisons for the second Dutch lexicon project (DLP2; (Brysbaert, Stevens, Mandera, & Keuleers, 2016); 3,145 five-letter stimuli; 81 participants) are presented in (d,e,f). Before going into the details of the two studies one has to note that the patterns we have found in the data in relation to our parameters of interest do not replicated within these two Dutch studies and, in addition, do not replicate with the findings from German, English, and French shown in Figure 3. In general, this is difficult for the interpretations of the results. For the DLP1 pattern we found a significant interaction of the orthographic prediction error with word/non-words and no significant effect of number of pixels. The interaction pattern in contrast to the findings in other languages (Fig. 3a), however, was qualitatively different as it showed a negative orthographic prediction error effect for words and a positive effect for non-words. The pattern is exactly the inverse from all other languages. Still model comparisons highlighted that the orthographic prediction error was relevant for the model fit since the predictor increased the model fit with no further increase of fit when the number of pixel parameter was included. None of these findings could be replicated in the DLP2 dataset, showing no significant fixed effects or interactions and no substantial changes in model fit relation to the null model. (g) Prediction image from a PEMoR implementation using five-letter words with a proportional Times New Roman script.

Table S3.3. Results from linear mixed model regression analysis (with the exception of the British data including multiple word lengths was estimated based on word aggregated data) for the behavioral lexical decision tasks (LDT) and handwriting analyses.

	E	SE	*t*
German LDT N°1: Orthographic prediction error based on word length specific prediction			
Intercept	6.49	0.023	288
Orthographic prediction error (oPE)	-0.03	0.004	6.5
Number of pixels (Npixel)	-0.007	0.004	1.8
Word/non-word (Lex)	0.33	0.009	33.1
Word frequency	-0.12	0.004	33.5
Error	-0.03	0.005	6.2
oPE X Lex	0.03	0.006	5.0
Npixel X Lex	0.000	0.006	0.1
German LDT N°1: Orthographic prediction error based on word length general prediction			
Intercept	6.48	0.023	288.3
Orthographic prediction error (oPE)	-0.03	0.004	6.3
Number of pixels (Npixel)	-0.01	0.004	1.7
Word/non-word (Lex)	0.33	0.010	33.2
Word frequency	-0.12	0.004	35.5
Error	-0.03	0.005	6.2
oPE X Lex	0.03	0.006	4.5
Npixel X Lex	-0.00	0.006	0.0
German LDT N°1: Orthographic prediction error based on word length specific prediction including orthographic Levenshtein distance and word frequency			
Intercept	6.66	0.023	237.1
Orthographic prediction error (oPE)	-0.02	0.004	4.3
Number of pixel (Npixel)	-0.00	0.004	0.2
Word/non-word (Lex)	0.29	0.011	27.0
Error	-0.03	0.005	6.2
Orthographic Levenshtein distance	-0.08	0.008	10.5
Word frequency	-0.12	0.004	35.5
oPE X Lex	0.03	0.006	5.2
Npixel X Lex	-0.00	0.005	0.6
German LDT N°2 including noise: 0%			
Intercept	6.32	0.024	263.9
Orthographic prediction error (oPE)	-0.02	0.016	1.4
Number of pixels (Npixel)	-0.00	0.015	0.2
Word/non-word (Lex)	0.27	0.05	5.4
Word frequency	-0.07	0.02	4.9

Error	-0.07	0.010	6.8
oPE X Lex	0.05	0.02	2.3
Npixel X Lex	-0.02	0.021	1.2

German LDT N°2 including noise: 20%

Intercept	6.45	0.026	245.4
Orthographic prediction error (oPE)	-0.06	0.017	3.3
Number of pixels (Npixel)	-0.00	0.013	0.3
Word/non-word (Lex)	0.37	0.049	7.5
Word frequency	-0.14	0.02	6.1
Error	-0.14	0.010	5.4
oPE X Lex	0.04	0.022	1.6
Npixel X Lex	0.02	0.022	0.7

German LDT N°2 including noise: 40%

Intercept	6.84	0.042	162.9
Orthographic prediction error (oPE)	-0.02	0.021	1.0
Number of pixels (Npixel)	-0.08	0.018	4.1
Word/non-word (Lex)	0.14	0.049	2.8
Word frequency	-0.11	0.06	1.9
Error	-0.00	0.010	0.1
oPE X Lex	-0.00	0.028	0.1
Npixel X Lex	0.08	0.026	2.9

British LDT

Intercept	6.39	0.013	507.1
Orthographic prediction error (oPE)	-0.007	0.001	5.3
Number of pixels (Npixel)	0.008	0.001	6.7
Word/non-word (Lex)	0.12	0.003	46.2
Word frequency	-0.067	0.001	58.0
oPE X Lex	0.008	0.002	4.2
Npixel X Lex	-0.003	0.002	1.9

British LDT 4-8 Letters: Length specific prediction

Intercept	6.26	0.157	39.7
Orthographic prediction error (oPE)	-0.001	0.000	5.0
Number of letters (Nletters)	0.062	0.027	2.3
Word/non-word (Lex)	0.155	0.162	0.3
Error	0.043	0.165	0.8
oPE X Lex	-0.001	0.000	4.5
oPE X Nletters	-0.001	0.000	3.3
oPE X Error	-0.002	0.000	5.1
Nletters X Lex	-0.006	0.028	0.8
Nletters X Error	-0.245	0.172	1.4

Lex X Error	-0.036	0.028	1.3
oPE X Lex X Nletters	0.001	0.000	2.4
oPE X Lex X Error	0.002	0.000	5.0
oPE X Nletters X Error	0.001	0.000	3.2
Nletters X Lex X Error	0.003	0.030	0.1
oPE X Lex X Nletters X Error	-0.001	0.000	2.6

British LDT 4-8 Letters: Length general prediction

Intercept	5.25	0.421	12.5
Orthographic prediction error (oPE)	0.002	0.000	3.7
Number of letters (Nletters)	0.250	0.061	4.1
Word/non-word (Lex)	1.064	0.438	2.4
Error	1.264	0.443	2.9
oPE X Lex	-0.002	0.001	3.1
oPE X Nletters	-0.000	0.000	3.6
oPE X Error	-0.002	0.001	4.0
Nletters X Lex	-0.183	0.065	2.9
Nletters X Error	-0.002	0.001	4.0
Lex X Error	-1.426	0.467	3.1
oPE X Lex X Nletters	0.001	0.000	2.9
oPE X Lex X Error	0.002	0.001	3.6
oPE X Nletters X Error	0.001	0.000	4.0
Nletters X Lex X Error	0.228	0.068	3.5
oPE X Lex X Nletters X Error	-0.001	0.000	3.3

British LDT 4-8 Letters: Number of pixel

Intercept	6.590	0.157	42.0

Number of pixel (Npixel)	0.000	0.001	0.3
Number of letters (Nletters)	0.092	0.028	3.2
Word/non-word (Lex)	-0.124	0.162	0.8
Error	-0.309	0.165	1.9
Npixel X Lex	0.000	0.001	0.2
Npixel X Nletters	0.000	0.001	1.4
Npixel X Error	0.000	0.001	0.4
Nletters X Lex	-0.059	0.029	2.0
Nletters X Error	-0.090	0.030	3.0
Lex X Error	0.035	0.171	0.2
Npixel X Lex X Nletters	0.000	0.001	0.9
Npixel X Lex X Error	0.000	0.001	0.1
Npixel X Nletters X Error	0.000	0.001	1.2
Nletters X Lex X Error	0.069	0.031	2.2
Npixel X Lex X Nletters X Error	0.000	0.001	1.2

French LDT

Intercept	6.63	0.005	1,333
Orthographic prediction error (oPE)	-0.002	0.001	2.0
Number of pixels (Npixel)	0.002	0.001	1.3
Word/non-word (Lex)	-0.040	0.003	11.6
Word frequency	-0.042	0.001	34.1
oPE X Lex	0.005	0.002	2.0
Npixel X Lex	-0.007	0.002	3.0

Dutch LDT

Intercept	6.45	0.019	348.1
Orthographic prediction error (oPE)	0.005	0.002	3.2
Number of pixels (Npixel)	0.001	0.002	0.6
Word/non-word (Lex)	0.101	0.004	23.8
Word frequency	-0.061	0.002	36.9
oPE X Lex	-0.016	0.002	6.6
Npixel X Lex	0.002	0.002	1.0

Dutch LDT2

Intercept	6.35	0.016	391.1
Orthographic prediction error (oPE)	0.002	0.002	1.1
Number of pixels (Npixel)	-0.001	0.002	0.6
Word/non-word (Lex)	0.048	0.005	9.4
Word frequency	-0.023	0.001	26.9
oPE X Lex	-0.003	0.003	1.3
Npixel X Lex	0.003	0.003	0.5
Handwriting: Script based orthographic prediction error			
Intercept	1.465	0.010	154.3
Mean prediction strength	0.052	0.007	7.4
Number of pixels with a prediction	0.015	0.008	2.1
Letter case	0.039	0.012	3.2
Handwriting: Readability ratings			
Intercept	11.5	1.4	8.1
Mean prediction strength	-5.9	1.0	6.2

Note. E: Estimate; SE: Standard error; *t*: t-value. All *t*'s >2 are considered a significant effect.

Table S4. Reliable activation clusters from the fMRI evaluation with respective anatomical labels (most likely regions from the Harvard-Oxford atlas; order of brain regions is relative to the order of peak components), cluster size (in voxels of size 2x2x2), and peak voxel coordinates (MNI space).

Hemisphere		Cluster extent [N voxels]	T	x	y	z
Orthographic prediction error based analysis (positive relationship)						
Occipital fusiform gyrus / Lateral occipital gyrus	L	95	6.6*	-24	-90	-12
			4.3	-34	-88	-10
Lateral occipital gyrus	L	81	4.8	-28	-84	6
			4.3	-34	-76	6
			3.8	-38	-86	4
Lateral occipital gyrus / Occipital fusiform gyrus	R	104	5.1	48	-76	-12
			4.1	44	-64	-18
			4.0	34	-64	-14
Occipital fusiform gyrus / Lateral occipital gyrus	L	170	4.9	-36	-68	-12
			4.2	-48	-76	-10
			4.0	-24	-68	-12
Words > Pseudowords						
Frontal orbital cortex / Inferior frontal gyrus, pars triangularis	L	1347	6.6	-36	34	-18
			6.3	-40	28	-8
			6.1	-54	26	-4

Superior frontal gyrus / Frontal pole	L/R	427	5.5	-6	52	28
			3.9	-10	62	22
			3.7	10	56	26
Temporal Fusiform Cortex, posterior division	L	120	5.2	-40	-36	-18
			5.2	-34	-42	-24
Middle Temporal Gyrus, posterior division / Superior Temporal Gyrus, posterior division / Middle Temporal Gyrus, temporooccipital part	R	113	4.5	60	-34	-2
			4.0	50	-26	-2
			3.8	52	-38	0
Inferior Frontal Gyrus, pars triangularis / Frontal Pole	R	164	4.3	56	32	10
			3.9	48	34	-12
			3.4	50	34	-4
Precentral Gyrus / Inferior Frontal Gyrus, pars opercularis	R	98	3.9	44	10	28
			3.9	38	4	32
			3.7	42	16	22
Orthographic prediction error by word/non-word interaction (positive relationship for words and negative for non-words)						
Inferior frontal gyrus, pars triangularis / Frontal operculum cortex	L	125	5.5	-52	32	-4
			4.6	-48	20	-4

Paracingulate gyrus / Superior frontal gyrus	L/R	90	4.4	-4	48	28
			4.2	4	48	30

Note. Cluster-level FWE-corrected at $p < .05$, peak-level uncorrected at $p < .001$; * Significant after FWE-correction on the voxel level. Order of regions presented per cluster corresponds to the order retrieved from the probabilistic Harvard-Oxford atlas.

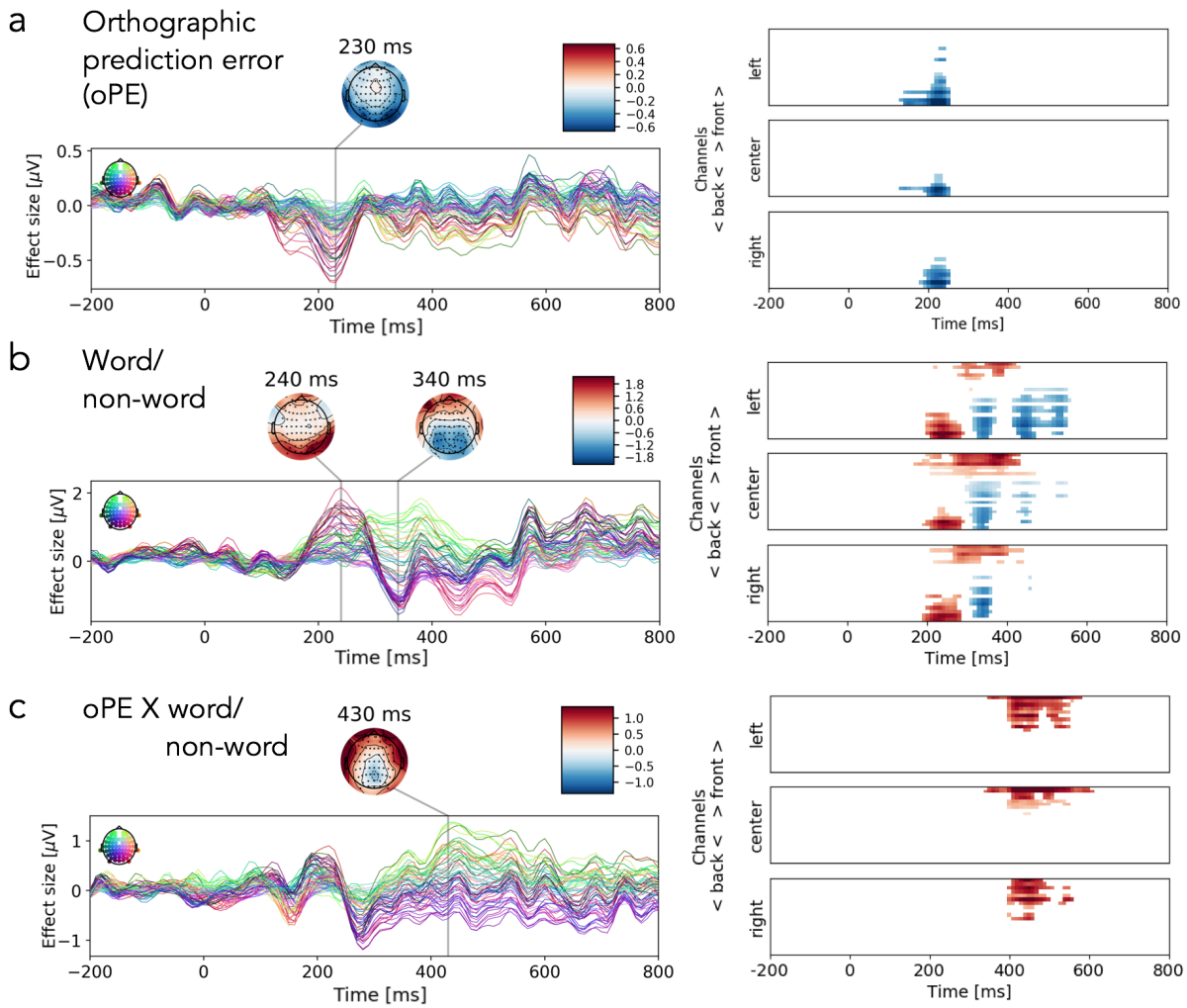

Figure S5. Detailed description of significant activation clusters in the EEG study for (a) the orthographic prediction error; (b) word/non-word effect; (c) interaction of word/non-word and the orthographic prediction error. On the left, the effect sizes from regression ERPs are presented as time courses for each sensor and time-point (color coding reflects scalp position). This part of the Figure reproduces Figure 5. The right column displays time courses with one line per channel, masked by significance using cluster statistics (see Methods for details; Maris & Oostenveld, 2007).