The What, When, Where, and How of Visual Word Recognition

Manuel Carreiras $^{1,2}$, Blair C. Armstrong $^1$, Manuel Perea $^{1,3}$, and Ram Frost $^{1,4,5}$

1. Basque Center for Cognition, Brain, and Language
2. IKERBASQUE- Basque Foundation for Science
3. Universitat de València
4. The Hebrew University of Jerusalem
5. Haskins Laboratories, New Haven, CT.

Keywords: Visual word recognition; VWFA; orthographic processing; neural connectivity; computational modelling; feedback vs. feedforward information

Corresponding Author: Carreiras, M. (m.carreiras@bcbl.eu).

Basque Center on Cognition, Brain, and Language
Paseo Mikeletegi 69, 2nd Floor
20009 - San Sebastian (Spain)
Abstract

A long-standing debate in reading research is whether printed words are perceived in a feedforward manner on the basis of orthographic information, with other representations such as semantics and phonology being activated subsequently, or whether the system is fully interactive and feedback from these representations shapes early visual word recognition. We review recent evidence from behavioural, fMRI, EEG, MEG and biologically-plausible connectionist modeling approaches, focusing on how each approach provides insight into the temporal flow of information in the lexical system. We conclude that, consistent with interactive accounts, higher-order linguistic representations modulate early orthographic processing, and discuss how biologically-plausible interactive frameworks and coordinated empirical and computational work can advance theories of visual word recognition and other domains (e.g., object recognition).
The What, When, Where, and How of Visual Word Recognition

A viable theory of visual word recognition needs to specify ‘what’ the building blocks of printed words’ internal representations are, and describe ‘how’ they are processed and assembled to give rise to word identification. These central ‘what’ and ‘how’ questions have been the focus of research (and controversy) in cognitive science since its very beginning, and have traditionally been addressed by combining inventive experimental designs and reaction time (RT) measures (see BOX 1). More recently, the availability of techniques such as fMRI have provided new opportunities to ask precise ‘where’ questions, focusing on locating the neurocircuitry involved in recognizing printed words. Given the brain’s architectural constraints, ‘where’ information often tells us something important about ‘what’ types of representations are activated during visual word recognition and ‘how’ readers eventually recognize words [1-3].

A comprehensive account of how complex stimuli such as words are processed requires, however, a detailed description of the temporal flow of information, elucidating ‘when’ words’ internal representations (e.g., letters, syllables, morphemes, lexical entries, etc.) are activated (see Figure 1 for contrasting frameworks). In that respect, ‘when’ questions constrain any theory of ‘how’, by detailing the sequence of events from stimulus presentation to word recognition. In fact, one of the oldest debates in visual word recognition concerns the demarcation line between bottom-up and top-down processing, asking whether or not the visual stimulus feeds into the lexical level in a predominantly
hierarchical manner, wherein orthographic representations feed into higher level linguistic representations, or whether higher level linguistic information such as phonological, and morphological structure exerts a top-down influence on visual-orthographic processing relatively early (see BOX 2). Cognitive neuroscience has rekindled this debate through the introduction of techniques like EEG and MEG, which have the appropriate temporal resolution to track the time course of processing. Note, however, that the ‘where,’ ‘what,’ ‘how,’ and ‘when’ questions are to a large extent interdependent. The human brain is generally constructed so that the trajectory of increased complexity, in terms of moving from relatively simple “microfeature” representations (e.g., the line-segments in a letter) to complex, higher-order representations (e.g., a representation of the whole word form) is occipital-to-frontal, whereas the trajectory of high-level modulation is frontal-to-occipital. Since ‘where’ information is correlated with the flow of processing (early/simple or late/higher-order), locations of brain activations are often taken to support claims regarding the temporal order of processing. Here we will discuss the potential danger of using evidence of ‘where’ to make inferences about ‘when’ (and ‘how’), review the findings that have been obtained with techniques having the appropriate temporal resolution for tracking the time course of printed word processing, and point to the desirable cross-fertilization between behavioural data, neuroimaging techniques, and neurobiologically-plausible computational models, for developing a mechanistically explicit theory of visual word recognition.
Many fMRI studies have investigated the brain circuits that underlie reading. Two points on which this research converges is that left hemisphere plays a major role in reading, and that the reading circuit consists of a network with two major pathways: (1) a dorsal pathway including the occipital, supramarginal and angular gyri, the premotor and the pars opercularis in the inferior frontal cortex, and (2) a ventral pathway that integrates the left fusiform, middle and anterior temporal and the pars triangularis in the inferior frontal cortex [4]. This notwithstanding, there is still a heated debate regarding the characterization of directionality of flow of information in these pathways (i.e., 'when' and 'how'). Specifically, the literature is unsettled regarding the extent to which higher-level lexical representations that are not necessarily orthographic modulate the relatively early processing of orthographic information (see BOX 3).

One of the most relevant examples of such debates is the role of the left fusiform gyrus, the putative visual word form area (VWFA) [5, 6]. From an anatomical processing perspective (i.e., a 'where' distinction), this brain region is considered to be a relatively “early” processing area. The left fusiform gyrus has been shown to be more activated for words or pseudowords than for false fonts or consonant strings [7-10]. It is thus commonly accepted that the left fusiform is
involved in orthographic processing. There is a controversy, however, regarding what specific information is represented in this brain region and how sensitive it is to top-down information. One theoretical position is that the VWFA is a prelexical hub, specific for written words, that computes and stores strictly visual and abstract prelexical orthographic representations in a primarily feedforward manner [8, 11, 12]. Another theoretical position, however, postulates that activation of the visual form area is modulated by higher-order linguistic properties of the stimuli such as phonology, morphology, and semantics [13, 14]. These two approaches provide very different views of reading: The former is compatible with the notion of feedforward temporal (and also structural) modularity (see BOX 2), wherein reading is considered to rely on a sequence of consecutive brain areas sensitive to a hierarchy of orthographic representations (e.g., letters, letter clusters of increasing size) that culminates in the recognition of a word. The latter considers reading as a fully interactive processing system wherein higher-level linguistic information that is not necessarily orthographic modulates early perceptual-orthographic processing.

Whereas proponents of the feedforward approach have relied on the argument that activation of the VWFA reflects a stage of orthographic processing that is immune to phonological and semantic influences which come into play only later on [12, 15, 16], there is mounting evidence suggesting that early print processing in the VWFA is modulated by higher-levels of lexical information. For example, sensitivity to high-level variables such as lexical frequency has been observed in the left fusiform [17]. Furthermore, the VWFA was similarly activated when target words were preceded by masked printed word primes or by masked pictures [18] (see BOX 1). Note that bidirectional flow of information
wherein higher-order levels of processing constrain feedforward assembly has also been proposed for object recognition [19-24]. Adopting the recycling hypothesis [12] (i.e., the neurocircuitry for visual object recognition has been recycled to compute the representations necessary for human reading), simple parsimony considerations would lead to the assumption of similar principles regarding the flow of information for visual object and visual word recognition.

Despite the above evidence, the debate regarding whether processing of visual word recognition is feedforward or not is still as active as ever. This is because the critical distinction between the two conceptual approaches regarding ‘how’ information flows in the circuits is mostly temporal in nature (“early” vs. “late”). However, fMRI integrates processes over a relatively long period of time. Thus, although the timing of stimuli presentation can be well-controlled (e.g., masked priming, fMRI adaptation, etc.), the temporal resolution of the BOLD response is too slow to unequivocally distinguish between activations that are feedforward versus feedback (i.e., the ‘when’ question) [25]. This leads us to the inherent advantage of magnetoencephalography (MEG) and electroencephalography (EEG).

\[\text{Box 3 about here}\]

\[\text{Box 3 about here}\]

**MEG and EEG: the dynamics of the time course**

EEG and MEG are time-sensitive methods with a temporal resolution in the range of milliseconds (ms). Combined with appropriate designs, they can reveal the temporal ordering of the neural processes involved in visual word recognition, tracing the time course from low-level visual perception to letter
perception and word meaning. It is this time course that provides important
evidence to adjudicate between different theoretical stances regarding the flow
of information (feedforward vs. feedback). Both techniques tap synchronized
neuronal activity over time triggered by some cognitive event in the brain [26].
Unlike RTs, which give us the end result of processing in the system as a whole
(and not specifically of lexical processing, see BOX 1), both MEG and EEG provide
a continuous measure of the intermediate events that have led to the final
response. MEG also provides some spatiotemporal constrains, allowing for some
synthesis of ‘when’ and ‘where’ information.

One of the earlier markers of visual word recognition is a left lateralized
N150/N170 response that has been found to differentiate orthographic stimuli
such as words and pseudowords from other stimuli such as symbols [27-29].
Selective responses to letters in this time window have also been found in the
inferior occipitotemporal cortex using intracranial recordings [30, 31] and MEG
[32, 33], particularly for normal readers but not for dyslexic children [34-37].
Thus, it has been proposed that the left-lateralized N170 could be an automatic
response related to typical visual word recognition, and that it could be
associated with the activation found using fMRI in the VWFA-left fusiform gyrus
[38](see however Rauschecker et al. [39] for bilateral effects). In fact, all this
evidence supports the claim that at approximately 150 ms from stimulus onset
the visual system responds only to the frequency of letter combinations, and that
lexical and phonological effects come into play much later [15, 16, 40]. As we will
expound below, however, higher-level linguistic information has been shown to
exert its influence already at 100 ms (that is, before 170 ms) from stimulus
onset. For instance, early ERP components in the range of 100-200 ms have been
reported to be sensitive to lexical frequency [41-45]. Thus, from a simple time-
scale perspective, an early marker of visual word recognition as revealed by ERP
measures (but not fMRI measures) seems to be susceptible to modulation from
higher-order lexical information.

Another early marker of reading is the N250, which was originally found
to be sensitive to orthographic similarity in combined masked priming and EEG
studies [46, 47]. However, subsequent studies have shown that the N250 is also
modulated by lexical factors [48, 49]. In particular, this ERP component was
found to be sensitive not just to letter identity but also to the phonological status
of the letters—that is, whether the letters are consonant or vowels [46, 50-52].

For example, Carreiras and colleagues [46] showed that masked subset priming
of consonants (e.g., mln-melon) and masked full identity priming (e.g., melon-
mln) do not significantly differ from each other already in the N250
component, whereas masked vowel subset priming (e.g., aio-amigo) and masked
identity priming (e.g., amigo-amigo) do significantly differ. Because consonants
are more lexically constraining than vowels in predicting word identity [53], this
effect demonstrates that top-down lexical information modulates the N250
component. Note that the same pattern of response is revealed in the later N400
component and in RTs in the behavioral experiments. This suggests that the
accumulated lexical information (and/or lexical competition) that is generating
the masked prime has exerted its full impact by 250 ms from stimulus onset. In
fact, the dissociation found between transposed-letter priming effects for word-
word pairs (e.g. ‘casual–causal’) and for nonword-word pairs (e.g. ‘barin–brain’)
in the N250 component [48] reinforces the hypothesis of high-order lexico-
semantic information constraining orthographic form-level processing in the N250.

Consistent with a sustained and early interactive co-activation of a network of sites contributing to reading, Thesen and colleagues [54] found a strong phase-locking from 170 to 400 ms between the left fusiform and more anterior language areas when comparing words versus false fonts using MEG and intracranial recording. Other recent reports of very early neurobiological responses to phonological information in anterior areas are also consistent with a top-down flow of information during visual word recognition. Using MEG in a masked priming paradigm, Wheat and colleagues [55-57] observed stronger responses to pseudohomophones than to orthographic control primes within 100 ms of target word onset in a cluster that included the left inferior frontal gyrus (pars opercularis) and the precentral gyrus. Note that a parallel pattern of activation found in the middle occipital gyrus suggests that these regions could be oscillating together during visual word recognition at a very early stage. Thus, the inferior frontal gyrus may exert feedback control on regions involved in lower level analysis of written words. In fact, a recent study [58] provides evidence of top-down feedback from the inferior frontal gyrus to the left ventral occipitotemporal cortex via dynamic causal modeling of MEG data. Specifically, the researchers found that words (as compared to false fonts) activated the left inferior frontal gyrus. More importantly, they showed that feedback connections from the inferior frontal gyrus onto the left ventral occipitotemporal cortex within the first 200 ms provided the best fit for the data relative to a model with only feedforward connectivity between those regions. One possible explanation for this feedback is that the inferior frontal region is sending higher-level
information (e.g., phonology) to constrain the representations computed in the left fusiform. Alternatively (or complementarily), these two brain regions may be interacting bidirectionally as part of a constraint network with as-yet underspecified graded specialization across the different contributing brain regions. Although the specific representations and dynamics of the frontal-occipototemporal areas remain to be fleshed out, these data clearly challenge the notion of temporal and structural modularity in orthographic processing.

**Biologically plausible connectionist modeling: A platform for advancing theories of visual word recognition**

From the empirical data outlined above, it is clear that substantive theoretical advance will require an integrated understanding of the contributions of a large set of distributed representations stored in different brain regions—brain regions that are accessed (at least initially) at different points in time as activity cascades throughout the brain. Connectionist modeling offers a mechanistic platform that is ideally suited for these investigations because it allows researchers to probe the ‘where’ and ‘when’ of visual word recognition and directly relate them to the questions of ‘what’ (representations) and ‘how’ (explicit computational processing mechanisms) work in concert to enable the perception of written words. Moreover, models allow researchers to explore the emergent properties of these systems and develop targeted empirical research agendas for the future.

The basic capacities of connectionist networks as outlined above were keenly demonstrated via the Interactive Activation model [59-61]. In this model,
information from low-level visual feature detectors flowed “bottom-up” to a "lexical" representation of whole words, while simultaneously being able to flow “top-down” from higher levels of representation. This model was thus able to explain and generate predictions regarding top-down influences related to word reading, such as the word superiority effect (i.e., the processing advantage for letters embedded in words relative to isolated letters) [62].

From this work and the general mechanics of constraint satisfaction systems [63], as well as the presence of bidirectional connectivity between brain regions that process different aspects of a word’s representation (e.g., letters, phonology, semantics), connectionist theories of visual word recognition will, by default, expect and generate some degree of top-down influence for maximally accurate word identification. The more critical concern, then, is if these top-down influences are substantive enough to be theoretically significant and may not be dismissed for reasons of parsimony [64-67]. Proficient reading also involves an optimization of efficiency in addition to one of accuracy; that is, of correctly reading words as quickly as possible. Thus, if a strictly feedforward system could, in principle, enable highly efficient word recognition, under what circumstances would the brain “choose to pay the price” of waiting for additional top-down constraints because of inadequacies in a strictly-feedforward signal? [68]

Connectionist models offer several avenues for exploring these possibilities. One particularly important recent advance is the ability to incorporate additional neurobiological constraints into standard connectionist models (e.g., by specifying different sub-populations of inhibitory and excitatory neurons) to simulate electrophysiological and behavioral responses (B. C.)
Armstrong, PhD thesis, Carnegie Mellon University, 2012)[69, 70]. For example, Laszlo and Plaut[69] showed how a model that instantiated these principles was able to generate and explain electrophysiological dynamics corresponding to the N400 ERP component. Additionally, they were able to advance the field by offering an account of an important discrepant finding between the behavioural and electrophysiological literatures: why the N400 ERP component is not sensitive to the lexicality of the stimulus (e.g., words and pseudowords vs. acronyms and illegal strings), whereas behavioural responses are. Specifically, they showed that the initial settling dynamics, during which the prominent deflection typically associated N400 ERP component was displayed, were primarily driven by the orthographic wordlikeness of the stimulus (e.g., in terms of its orthographic neighborhood). However, nonlinear settling dynamics in the network caused a change in these activation patterns later on in processing, such that valid lexical types (words and acronyms) were more active than nonwords (pseudowords and illegal strings), consistent with typical behavioural lexical decision data.

Relatedly, Laszlo and Armstrong [70] further extended this work to account for how simple context effects (e.g., word repetition) modulate the N400 component that has been associated with lexical-semantic access. This was accomplished via the incorporation of a neuron-specific fatigue mechanism so that neurons that fired recently would not be able to fire at their maximum rate for a brief period of time. This resulted in a reduction of the N400 component amplitudes for stimuli in the semantic representation, regardless of the lexical status of the string of character that was input to the network. Moreover, they were recently able to generate specific predictions regarding the power-
frequency spectra that should be evoked by words and nonwords (Laszlo & Armstrong, unpublished)—data that are increasingly influential in establishing the causal links between which brain regions influence one another and the temporal order (i.e., 'when') in which this occurs [58, 70]. (See Figure 2). The result of this work was targeted insights into 'what' aspects of a word's representation are modulated by related context. Furthermore, this neurally-inspired account therefore presented an alternative explanation for the long-accepted verbal account of N400 repetition effects [71], which stated that reduced N400 amplitudes resulted from an (underspecified) facilitation mechanism instead of a fatigue mechanism.

With such explicit models in hand, it is possible to add or subtract different feedback connections and evaluate which of these models best captures empirical electrophysiological data such as ERP waveform amplitudes and power over time in as parsimonious a manner as is possible. These models thus present an opportunity for an “experimental” approach to theoretical development, by allowing for the direct assessment of the impact of different theoretical assumptions. Moreover, by virtue of the domain-general nature of the framework, it is possible to naturally extend these principles to the study of other levels of representation. For instance, these principles can readily be employed to study ERP components associated with earlier processes related to visual word recognition (e.g., N170, N250) and determine 'when' and 'how' these representations shape other purportedly earlier.
Combined with a domain-general learning theory and recent advances in ‘deep’ neural networks [72], it has also recently become possible to train networks with multiple intermediate “hidden” layers whose input values or target outputs are not explicitly set by the experimenter. Consequently, researchers can step back from explicitly specifying the exact nature of the representations that are expected to mediate between a low-level visual form and higher-order lexical information. Instead, they can simply specify a more neutral input (e.g., a low-level visual input associated with a word in a retinotopically-centered coding scheme) and observe how unsupervised learning mechanics can gradually extract higher-order information such as position-insensitive visual representations and intermediate representations of different grain sizes [73]. This approach enables the empirical investigation of whether particular intermediate representations assumed to mediate between visual inputs and higher-order lexical outputs (e.g., bigrams, syllables) are in fact representations that will emerge during learning.

Taken together, biologically plausible connectionist models are showing strong initial promise in being able to provide detailed and explicit mechanistic accounts of ‘what’ and ‘where’ different types of representations are stored [72, 73], ‘when’ they interact, and ‘how’ the processing mechanics of neurobiologically-constrained computations operate the way they do. Moreover, these models can be used not only as ‘process’ models of the typical mature system, but can also trace the development of lexical skills and top-down versus bottom-up interactivity [74], as well as seamlessly enable studies of different types of reading impairments [61].
Concluding remarks

Old debates regarding the modularity—or lack thereof—of visual word recognition are back again with the advent of techniques such as fMRI and EEG/MEG. However, the rekindling of these debates via research using these techniques offers the opportunity for unique theoretical advance that was not possible via behavioural investigations alone, and also enables the investigation of important new questions (see BOX 4). No longer are researchers constrained to ‘black box’ theorizing regarding the internal mechanics of the brain that mediate between stimulus and response. Rather, these representations can now be monitored directly and used to motivate specific theoretical claims about the intermediate internal representations and processes that subserve visual word recognition. Like never before, it is therefore possible to achieve integrated theories of ‘what’, ‘when’, ‘where’, and ‘how’ visual words are represented and processed in the brain.
Although behavioural investigations are useful for understanding visual word recognition, these studies suffer from an inherent limitation: they only provide an “end-state” reflection of the state of processing via an indirect behavioural response (e.g., lexical decision time as signalled via a key-press). Thus, these techniques do not provide direct insight into the internal temporal dynamics underlying ‘how’ different representations are activated. Moreover, these techniques simultaneously run the risk of producing data that are contaminated by pre- and post-lexical processes (e.g., decision making).

These limitations notwithstanding, techniques that provide relevant indirect insight into the time-course of different processes have been developed that relate to the ‘when’ question regarding feedforward and feedback processes. In this context, the masked priming technique [75] deserves special consideration. In masked priming, a target word is preceded by a briefly-presented masked priming stimulus (e.g., mln-melon). By manipulating the structural relationships between prime and target (e.g., at the orthographic, phonological, morphological, and other levels), at different exposure durations (e.g., typically between 10 and 60 ms), researchers have observed different time courses of processing for different properties of printed words (e.g., orthographic and phonological representations) [76].

The rationale behind this experimental approach is that the minimal prime duration that is required to obtain a specific priming effect reflects the time that is necessary for activating that information (e.g., orthographic, phonological, morphological, or semantic information). Nonetheless, this procedure has limitations [77]; for instance, its lack of ecological validity. A related and more
ecologically-valid technique is to present the words in the context of normal silent reading while the participants’ eye movements are registered [78]. Here, of particular interest is the very early parafoveal preview benefit effect using the boundary technique in which the relationship between a parafoveal preview and a target word is manipulated. Specifically, the parafoveal preview is replaced by the target word once the fixation crosses an invisible “boundary” located next to the target word. The differences in the fixation durations on the target word caused by different structural manipulations of the parafoveal preview reflect ‘what’ information was already processed in the parafovea (e.g., orthography and/or phonology and/or morphology, etc.) [79].

There is ample evidence that high-level information, such as phonological [80, 81]; morphological [82, 83], and lexical information [84] influences very early aspects of the overall visual word recognition process. This evidence challenges the traditional claim of temporal and structural modularity according to which printed words are principally identified on the basis of orthographic information alone in skilled readers (the underlying logic behind some researchers’ conception of the VWFA), with phonological and semantic information being retrieved subsequently [64, 85].

**Box 2: Structural and temporal modularity, interactivity, and orthographic processing**

The main theoretical stand underlying the feedforward approach is that pure orthographic models have substantial descriptive adequacy to account for a large set of (mostly behavioral) data in visual word recognition [86]. One basic tenet in this feedforward view is that in principle “feedback cannot improve...
performance at either the lexical or prelexical level” (p. 306) and, indeed, a number of well-known phenomena (e.g., word superiority effect) that have been traditionally attributed to top-down feedback can be explained parsimoniously in a feedforward manner. By this view, the game of processing printed words is largely played in the court of orthographic processing, such that a significant part of the recognition process is determined by considering the surface structural properties of the printed stimulus alone (i.e., letters, letter sequences, etc.). Interestingly, this position is consistent with the idea that the identification of visual forms in general and letter strings in particular can be achieved through a low-level visual pattern recognition system shared by humans and baboons [87-89].

The strongest version of the feedforward view postulates structural modularity, according to which orthographic processing is in principle non-penetrable by other linguistic dimensions. Moreover, processing within the orthographic system proceeds bottom-up from low-level features to full orthographic words. The weaker version assumes temporal modularity [90], and posits that the word recognition system is simply set so that the processing of printed words proceeds until an orthographic word unit is recognized; only once this is accomplished does the orthographic representation make contact various other linguistic properties (e.g., phonology, morphology, semantics; [64]). Note that this approach may allow for top-down interactivity as long as it is constrained to occur after initial orthographic coding is complete.

The contrasting approach argues for full interactivity between lower- and higher-order representations at all processing levels. Here, the demarcation line beyond ‘when’ and ‘where’ “perceptual orthographic” processing ends and
“linguistic” processing begins is blurred [72]. By this view, high-level linguistic considerations that are not purely orthographic (e.g., how some letters correlate with phonology and meaning, and how letter clusters are constrained by lexical, morphological, and phonological structure) shape the distributional properties of letters in a given language, and the word recognition system learns these features to enable efficient (i.e., fast and accurate) reading in that language. Language-specific retinal-perceptual learning effects (i.e., cross-linguistic difference of processing letters at different retinal eccentricities) [91] suggest that reading habits stemming from the overall structure of a language indeed affect the functional structure of early stages in the visual pathway, and are thus compatible with this view. For example, frequently-encountered visual configurations result in perceptual learning that allows for rapid and efficient recognition of a word form, and these configurations are determined by the correlation of orthography with phonology and meaning that are characteristic to a language [92]. Because different languages are characterized by different relations between orthography, phonology, and semantics (amongst other representations), interactive models that allow for phonological, morphological, and semantic information to come into play early on, are argued to be better accounts for the substantial cross-linguistic differences observed in early orthographic processing [85].

Box 3: Future explorations of interactivity in fMRI and MEG

Several studies have shown intrinsic functional connections between Broca’s area and ventral occipitotemporal regions [58, 93-96] Anatomical connections between frontal and occipital regions through the superior
longitudinal fasciculus and/or the inferior fronto-occipital fasciculus have also been documented [97-105]. Taken together, these findings provide the neurobiological platform for possible top-down effects from frontal areas, and thus offer interesting avenues for future investigations. Moreover, it is important to keep in mind that readers not only activate the left fusiform and other regions of the ventral pathway when reading. They also activate the dorsal pathway including regions such as the left supramarginal gyrus, the left superior temporal cortex, and the left inferior parietal cortex—particularly, the angular gyrus, which has a modulatory effect on the visual cortex [4, 106]. Further, there are other functional pathways in the reading circuit starting in the occipital cortex that do not necessarily involve the left fusiform [107]. The reading circuit includes not only one, but multiple pathways from vision to higher-order temporal lobe language areas. Thus, the full scope of interactivity (or lack thereof) between regions spanning the different pathways of the reading circuit should be more fully established. Finally, it is important to note that the dorsal and ventral pathways are not modular systems that operate independently of each other, but exchange information during visual word recognition [108, 109]. In fact, structural connectivity between regions belonging to each of the two pathways (e.g., the posterior inferior temporal regions, including the left fusiform, and the posterior superior temporal and inferior parietal regions, including the supramarginal gyrus) has been documented [110]. Moreover, functional connectivity between these regions has been shown in skilled readers [111] but not in dyslexic individuals [112]. Further studies are required to determine how brain regions falling along the ventral and dorsal pathways
interact and cooperate during visual word recognition, as well as how these interactions relate to other similar processes such as object recognition.

Box 4. Outstanding questions and future directions

1. How anatomical and functional connectivity, and consequently the temporal flow of information, evolves from preliterate to skilled reading?

2. What are the functional/anatomical differences underlying reading disabilities, and how their understanding can help the implementation of remediation programs?

3. How are connectivity and interactivity modulated by different languages having different scripts, different orthography-to-phonology relations, and different morphological systems?

4. What do the constraints of possessing different orthographic, phonological, and semantic representations have on visual word recognition in the case of bilingual and multilingual readers?

5. How can more biologically plausible computational models interact with empirical investigations to produce theories that are mechanistically explicit, comprehensive, and parsimonious?

6. To what extent will a neurobiological theory of visual word recognition that considers the full patterns of brain connectivity and interactivity provide insights into domain-general mechanisms shared by other related abilities such as object recognition?
Acknowledgements

We acknowledge support to M.C. from the European Research Council (ERC-2011-ADG-295362) and from the Spanish Ministry of Economy and Competitiveness CONSOLIDER-INGENIO2010 CSD2008-00048 and PSI2012-31448; to B. C. A. from the Marie Curie program (PCIG13-2013-618429); to R. F. from the NICHD (RO1 HD067364); to R.F. and to M. C. (PO1HD 01994); and to M. P. from the Spanish Ministry of Economy and Competitiveness (PSI2011-26924).

We thank S. Laszlo for discussion of the modeling section.
References


17 Suppl 1: p. i125-35.


**Figure 1.** Core architectural and functional assumptions of temporally-modular feedforward versus interactive models of visual word recognition.

(a) Temporally-modular feedforward models claim that the visual-orthographic information is processed in a set of distinct, hierarchically-organized processing stages, such that each stage (e.g., the activation of letter and orthographic-lexical representations) occurs in a strictly feedforward—and in the strongest form, sequential—fashion. Critically, additional non-visual-orthographic representations (e.g., phonology, semantics) are not accessed until orthographic access is complete and/or if accessed before that point, higher-level representations never feed back to influence the orthographic computation.

(b) Interactive-Activation models [59] claim that visual information continuously cascades throughout the entire orthographic-phonological-lexical-semantic network. This enables partially-resolved phonological and lexical-semantic representations (amongst others) to feed back and provide constraint on other (lower) levels of representation in the network such as orthography. Note that additional intermediate levels of representation (e.g., letter-clusters) have been suppressed for simplicity in both figures, and that these are but two examples of each type of network (e.g., other feedforward theories suggest direct sublexical input to phonology but are nevertheless feedforward). Solid blue lines denote feedforward connections; dashed green lines denote feedback connections.
Figure 2. Integrating insights from more biologically-plausible connectionist models and neuroimaging data.

Recent connectionist models that employ large pools of excitatory neurons and small pools of inhibitory neurons (here, inhibitory sub-populations are coded by a “-”, all other neurons are excitatory), as well as sparse/weak distal connections (thin arrows) and dense/strong local connections (thick arrows), and emergent hidden representations that are not explicitly specified by the modeler (denoted by < > around the name of the approximate representation that should emerge in that pool) have enabled major advances towards explaining and predicting behavioural and neuroimaging data with a computationally explicit theory. For example, Laszlo and Plaut [69] and Laszlo and Armstrong [70] used the architecture illustrated in (a) to generate (b) simulated ERP components that explained empirical ERP data from (c) an analogous experiment involving lexical types including words (e.g., DOG), acronyms (e.g., DVD), pseudowords (e.g., GORK) and illegal strings (e.g., XFQ) both on their initial presentation (thick lines) and when the item was repeated (thin lines). Note that the units in the model data are arbitrary and are omitted; in the empirical data, the x-axis ranges from 0-900 milliseconds and negative is plotted up. All empirical and simulation data were drawn from Laszlo & Armstrong [70], used with permission. These ERP data were collected from an electrode placed on a middle parietal location. This work suggests that (d) in a more comprehensive model that also contains orthographic and semantic representations (here the <o-p>, <o-s>, and <p-s> labels denote intermediate pools of neurons that map between orthography, phonology, and semantics, respectively), it is in principle possible to study the
activation dynamics and representations that emerge in (e) brain regions associated with different representations, of which a subset of the most critical regions are shown in a lateral cross-section of the left hemisphere. The colour of the circles denotes the theoretical representations in the model that these regions might subserve. IF = Inferior Frontal Cortex; SG = Supramarginal Gyrus; AG = Angular Gyrus; AT = Anterior Temporal Cortex; FG = Fusiform Gyrus (includes Visual Word Form Area; VWFA); OC = Occipital Cortex.
(a) 

- Semantics 

- <Hidden> 

- <Orthography> 

- Visual Input 

(b) 

RECORDINGS FROM SIMULATION 

Recording Activation 

Time (unit updates) 

First Presentation 

Second Presentation 

(c) 

RECORDINGS FROM ANALOGOUS EEG/ERP EXPERIMENT 

Middle Parietal Electrode 

Recordings 

Amplitude (µV) 

N400 Window 

First Presentation 

Second Presentation 

(d) 

- Semantics 

- <O-S> 

- <O-P> 

- <Orthography> 

- <Phonology> 

- Visual Input 

(e) 

Brain regions: 

- pM 

- SG 

- AG 

- MT 

- AT 

- FG 

- OC 

- IF