Review

# 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, San Sebastian, Spain
[2] IKERBASQUE Basque Foundation for Science, San Sebastian, Spain
[3] Universitat de València, València, Spain
[4] The Hebrew University of Jerusalem, Jerusalem, Israel
[5] Haskins Laboratories, New Haven, CT, USA

**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 activated subsequently, or whether the system is fully interactive and feedback from these representations shapes early visual word recognition. We review recent evidence from behavioral, functional magnetic resonance imaging, electroencephalography, magnetoencephalography, 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. We also 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 a printed word 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 (Box 1). More recently, the availability of techniques such as functional magnetic resonance imaging (fMRI) have provided new opportunities to ask precise 'where' questions, focusing on locating the neurocircuitry involved in recognizing printed words. Given the architectural constraints of the brain, '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].

However, a comprehensive account of how complex stimuli such as words are processed requires a detailed description of the temporal flow of information and elucidation of 'when' the internal representations of words (e.g., letters, syllables, morphemes, lexical entries) are activated. Figure 1 presents contrasting frameworks. In this 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 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 (Box 2). Cognitive neuroscience has rekindled this debate through the introduction of techniques such as electroencephalography (EEG) and magnetoencephalography (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. Because '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 discuss the potential danger of using evidence of 'where' to make inferences about 'when' (and 'how'), review the findings obtained using techniques with the appropriate temporal resolution for tracking the time course of printed word processing, and point to desirable cross-fertilization between behavioral data, neuroimaging techniques, and neurobiologically plausible computational

*Corresponding author:* Carreiras, M. (m.carreiras@bcbl.eu).
*Keywords:* visual word recognition; visual word form area; orthographic processing; neural connectivity; computational modeling; feedback versus feedforward information.

1364-6613/$ – see front matter
 http://dx.doi.org/10.1016/j.tics.2013.11.005

**Box 1. Measuring time courses in behavioral research**

Although behavioral 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 behavioral response (e.g., lexical decision time as signaled 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 approaches 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) for 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 required to obtain a specific priming effect reflects the time necessary for activation of that information (e.g., orthographic, phonological, morphological, or semantic information). Nonetheless, this procedure has limitations [77], such as a 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]. 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. Differences in fixation duration 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) [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' concept of the VWFA), with phonological and semantic information retrieved subsequently [64,85].

models for the development of a mechanistically explicit theory of visual word recognition.

## fMRI evidence suitable for 'where' but not for 'when'

Many fMRI studies have investigated the brain circuits that underlie reading. Two points on which this research converge is that the left hemisphere plays a major role in reading and the reading circuit consists of a network with two major pathways: (i) a dorsal pathway including the occipital, supramarginal, and angular gyri, and the premotor and pars opercularis in the inferior frontal cortex; and (ii) 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 (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

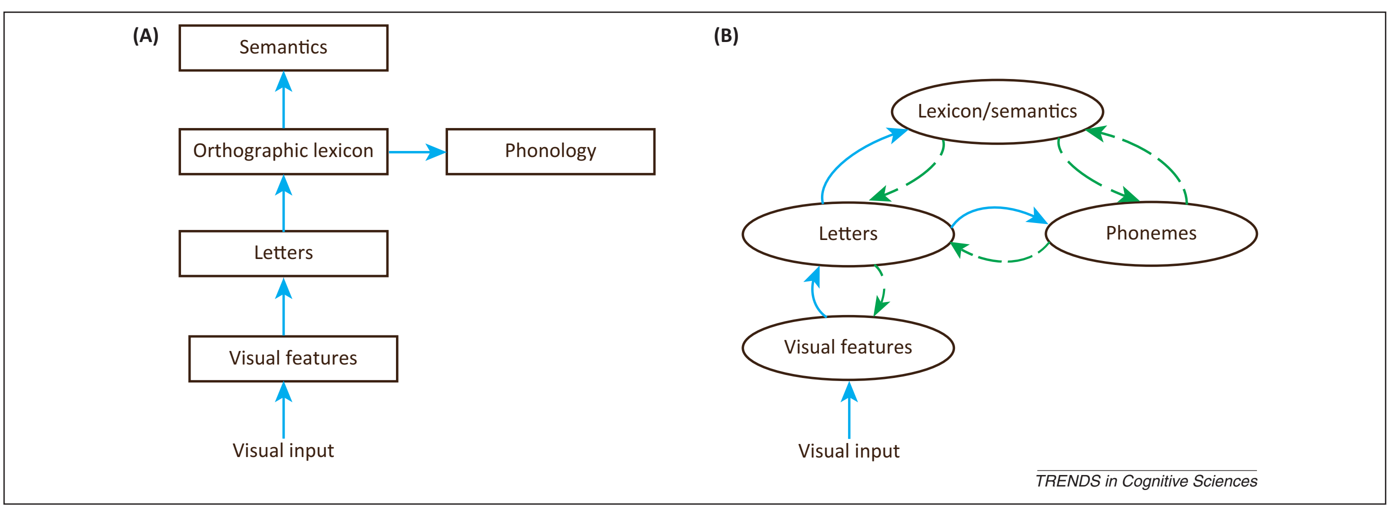


**Figure 1**. Core architectural and functional assumptions of temporally modular feedforward versus interactive models of visual word recognition. **(A)** According to temporally modular feedforward models, visual orthographic information is processed in a set of distinct, hierarchically organized processing stages, such that each stage (e.g., 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)** According to interactive activation models [59], visual information continuously cascades throughout the entire orthographic–phonological–lexical–semantic network. This enables partially resolved phonological and lexical–semantic representations (among others) to feed back and provide constraints 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 schematics, and that these are just two examples of each type of network (e.g., other feedforward theories suggest a direct sublexical input to phonology but are nevertheless feedforward). Unbroken blue lines denote feedforward connections; broken green lines denote feedback connections.