BRAIN RESEARCH REVIEWS 53 (2007) 98–123

available at www.sciencedirect.com

www.elsevier.com/locate/brainresrev

BRAIN RESEARCH REVIEWS

Review

Interplay between computational models and cognitive electrophysiology in visual word recognition

Horacio A. Barber[a,*], *Marta Kutas*[a,b]

[a]*Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA*
[b]*Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA*

ARTICLE INFO

Article history:
Accepted 6 July 2006
Available online 14 August 2006

Keywords:
Language
Reading
Visual word recognition
Computational model
Event-related potential (ERP)
EEG
MEG
Cognitive electrophysiology

ABSTRACT

In this article, we discuss the relevance of electrophysiological data to the enterprise of analyzing and understanding the reading process. Specifically, we detail how the event-related brain potential (ERP) technique (and its magnetic counterpart) can aid in development of models of visual word recognition. Any viable and accurate account of reading must take into account the temporal and anatomical constraints imposed by the fact that reading is a human brain function. We believe that neurophysiological (especially, although not limited to electrophysiological) data can serve an essential reference in the development of biologically realistic models of reading. We assess just how well extant electrophysiological data comport with specific predictions of existing computational models and offer some suggestions for the kinds of research that can address some of the remaining open questions.

Contents

---

[*] *Corresponding author.*
E-mail address: hbarber@ull.es (H.A. Barber).

0165-0173/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainresrev.2006.07.002

1. Introduction

Reading is a remarkable, uniquely human eye–brain activity during which splotches of ink – constituting words, phrases, statements, demands, requests, promises, threats, among others – come to bear meaning in a reader's mind. Large-scale alphabetization is a contemporary phenomenon. Somewhere within the past 5000 years or so, the human species learned to read and thereby radically changed its world. To this day, psychologists and neuroscientists are actively seeking to decipher the neuro (and psycho)-computational processes of normal reading.

Whether or not human brains are genetically pre-wired for language processing with a language acquisition device, they are incontrovertibly not pre-wired with a hard-wired, reading acquisition device. Children *learn* to read, acquiring the requisite reading skills via substantial effort, and most often with explicit instruction. Learning to read thus is accompanied by functional reorganization of brain systems that clearly evolved for other functions. Nonetheless, by the age of five, most children have mastered the basic visual and language skills necessary for reading, thereafter merely reinforcing and refining the visual–language interface via continued exposure and practice (Carr, 2005). In this acquired neural system for reading, visual information (as in a printed word) travels from the retina of both eyes to the cerebral cortex of both hemispheres where it is analyzed at perceptual and linguistic levels.

To understand how the human brain reads is to understand the essence and order of the neurocognitive computations by which the different visual inputs to the two eyes and the two cerebral hemispheres are processed (i.e., decoded, encoded, transferred, and integrated). Furthermore, any realistic account requires an appreciation for the remarkable speed at which these processes transpire: an average adult reader can read about three to four words per second, scanning a text with fast eye movements and fixating words for durations of 200–300 ms or so. Although some words may be re-fixated, others are skipped and perceived only parafoveally (Rayner, 1998). Word recognition occurs so quickly that it may be easy to forget that it is the endproduct of a set (series or cascade) of time-consuming physical processes. More specifically, the neural computations on which successful reading relies are the result of metabolic changes that occur at various times in various brain areas after visual input impinges upon the retina. No realistic model of reading thus can ignore the fact that reading is a cognitive skill effortfully learned by areas of the human brain evolved for other functions. While models of visual word recognition (VWR) need not specify exactly which brain areas are involved, they do need to respect the nature of the specific visual representations involved, their encapsulation and/or interaction, and the time courses of their availability and deployment.

Computational models of VWR have become especially popular among cognitive researchers in the past decade or so (for excellent reviews, see Grainger and Jacobs, 1998; Randall et al., 2000). In general, computational models have considerable advantages over traditional strictly verbally descriptive models. First and foremost, computational models (i.e., computer programs) are explicit: indeed, they must be fully specified, if they are to run at all. Second, these specifications must be internally consistent: different parts of a model cannot lead to internal conflicts or produce contradictory outputs. Third, computational models provide researchers a means for testing theories via simulations affording comparisons with human performance, in terms of internal representations, mechanisms and execution, and outputs. Last, but not least, computational programs can be "lesioned" to simulate malfunctions resembling those consequent to human brain damage.

It should go without saying then that empirical data—behavioral, neuropsychological, and neurophysiological—are critical for both the developmental and refinement of VWR models with neurologically plausible architectures and greater predictive power. Empirical data not only impose processing and storage constraints on VWR models but also provide a means for assessing each model's predictions. Modeling in turn generates new questions and new hypotheses that can then be assessed via a whole gamut of available technologies that generate new data, and so on.

The benefits of such a dynamic interplay between computational models and empirical research are clearly evident in several computational models of VWR based largely on behavioral measures (reaction time and accuracy). By contrast, on the whole, there is no similar give-and-take between computational modelers and electrophysiological researchers, perhaps because computational models have been agnostic if not silent regarding the time courses of the various neurophysiological processes or the brain areas involved in VWR. However, as will become patent in the following review, this state of affairs is in rapid flux; the most recent generation of VWR models has begun to acknowledge the brain's role in reading.

After briefly reviewing the most prominent computational proposals, we discuss how the particular strengths of event-related brain potential (ERP) measures, as a direct measure of brain activity with exquisite temporal resolution, might help to redress some of the acknowledged weaknesses of present-day VWR models. In this review, we focus on three major respects in which we believe most VWR models are found lacking: (1) specification of the time courses of VWR processes, (2) instantiation of the consequences of the physical–neural space in which VWR transpires, and (3) recognition of polysyllabic and polymorphemic words.

2. Models of visual word recognition and various reading phenomena explained

The first and perhaps most influential cognitive model of VWR is Morton's (1969, 1980) logogen model. This model is based on a mental lexicon (mental dictionary) made up of individual lexical entries (word representations). The logogen model distinguishes between representations of word meanings stored in the *cognitive system* and word forms stored in *logogen systems*. Separate logogen systems exist for dealing with auditory and visual inputs and for producing oral outputs (reading comprehension, listening comprehension, and speech, respectively). Each unit (logogen) in the logogen system has an activation threshold. When the amount of incoming information exceeds the threshold for a particular word, its meaning becomes available (i.e., is accessed) via the cognitive system.

Although both the notions of an absolute threshold and a mental lexicon have been found lacking to some extent (see below), the logogen model has had an undeniable impact on models of VWR. Data from neurologically intact as well as neurologically impaired humans have led to increasingly more elaborate, empirically more accurate, and as a consequence more readily testable versions of the model (e.g., Ellis and Young, 1988; Shallice, 1988). This type of model has obvious advantages for organizing the empirical psycholinguistic evidence and for generating qualitative predictions. As already mentioned, computational models carry some additional benefits, especially the potential of generating quantitative predictions. Although a detailed review of the field is beyond the scope of this article, we proceed with a few representative computational models of VWR together with examples of how they have simulated specific aspects of human VWR performance.

2.1. *Connectionist models based on learning algorithms*

An important class of VWR models, distinctly different from the logogen model, comprises computational models based on learning algorithms. Such models are implemented in artificial networks loosely inspired by the neural architecture of the brain. Similar to real neural networks, such artificial networks are made up of a (large) number of highly interconnected functional processing units. The activity level of each unit and the strength of each connection between units are numerically determined, in such a way that information is represented as the general pattern of those values. These "neural" networks typically include three (or more) layers of units (see Fig. 1); an input unit layer in which the stimuli are represented, an intermediate or hidden unit layer which is crucial for many learning algorithms, and an output unit layer, the pattern of activity of which corresponds to the network response. Initial

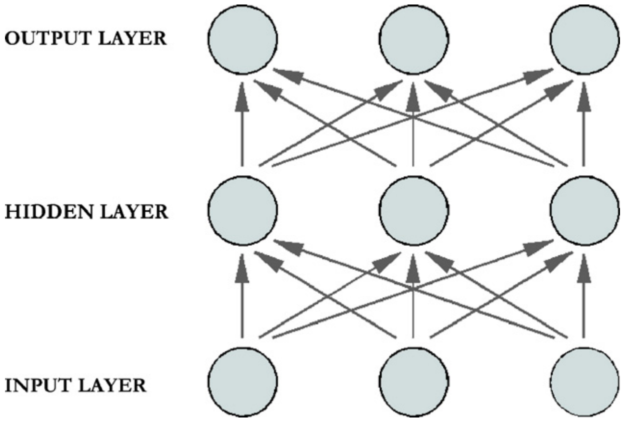

Fig. 1 – Basic architecture of a feedforward connectionist network; activity flows from the input layer to the output layer, through a hidden layer. Connections weights are gradually adjusted by mean of a learning algorithm, and information becomes represented in a distributed way across the network.

random values are assigned as connection weights between units, so any input (stimuli) introduced to the net produces a random output (response). The model is then trained with a large set of input–output vector pairs, and learning is achieved by altering the functional organization of the network. In each training trial, the connection weights between units are adjusted, bringing the next output closer to the expected one. As a result, the network's response to any given input is *statistically* predictable from previous input–output associations.

It is important to understand the parallels drawn between artificial and real neural networks as a general metaphor. Units in artificial neural networks are not really neurons but neuron-like with oversimplified features, artificial network structures usually do not even try to mimic the structure of real brain tissue, and more importantly, some of the common learning algorithms do not seem to work according to any obvious physiological principles. For example, the backpropagation algorithm, which is used in many connectionist models to adjust weights during training by propagating error (discrepancy between the desired output of a unit and its actual output) back to the hidden units, is not biologically plausible, as such. It is thus commonly assumed that the brain uses less powerful learning rules in combination with more neurons as well as more complex patterns of connectivity. Most connectionist models of VWR are thus more accurately viewed as models of learning and human performance than as models of the brain or even of brain functions *per se*. The challenge of rendering connectionist models more biologically plausible at an implementational level is a relevant issue in the field of computational neuroscience that is not directly addressed here (see Churchland and Sejnowski, 1992; Rolls and Treves, 1998). Obviously, advances on this front will have crucial implications for brain models of cognition.

A number of different connectionist models of reading have been devised sharing the common assumption that words (or any other unit intervening between the visual input and meaning) are not locally represented (Seidenberg and McClelland, 1989; Hinton and Shallice, 1991; Plaut et al., 1996; Harm and Seidenberg, 2004). According to these, so-called PDP (*Parallel Distributed Processing*) models, meaning emerges from the activity of and connections among highly distributed representations that could be the neural instantiation of mental states. On this view, there is no need for a mental lexicon built of lexical units, as meanings are represented by weights and connections across a set of units, with each unit participating in the representations of many different words. Words are thus not associated with any lists of orthographic, phonological, grammatical, and semantic features; rather, they are stimuli that operate directly on internal states (Elman, 2004).

A classic example of such a connectionist network is Seidenberg and McClelland's (1989) model of reading aloud which learns to map distributed orthographic representation inputs onto distributed phonological representation outputs. After training with a set of English words and their pronunciations, S&M's network was able to generate correct pronunciations for regular as well as irregular words (i.e., those with irregular spelling–sound correspondences, like "sew" or "bowl"). Moreover, assuming a correspondence between human reaction times and network error scores, the model was able to mimic two typical phenomena in normal readers—the word frequency effect (faster naming of frequent than infrequent words) and the word regularity effect (faster naming of low-frequency regular than irregular words). The S&M model, however, was limited in its ability to generate pronunciations for pseudowords and to discriminate pseudowords from real words.

Plaut et al. (1996) modified the S&M model to overcome these limitations: they (1) altered the inputs to be letter clusters tagged in terms of their position within a word, thereby encoding contextual information, and added (2) recurrent connections (feedback connections from the hidden unit layer to itself, i.e., its input) and (3) attractor structures to the model. Networks with recurrent connections introduce a contextual unit layer which receives activity from a unit layer in one time step ("store" it) and send it back to the same layer in the next time step. In this way, the network can maintain a "sense of history" and process sequences that unfold over time. Attractor structures allow the network to settle into one of several stable low energy states (attractors), across a number of processing cycles, depending on the inputs (input vectors remain active until the output is constant). The behavior of this type of network is dynamic, thereby adding a critical temporal dimension to the simulations. Furthermore, as attractors drive partial inputs to more familiar (stored) representations (i.e., attractors), attractor networks are better able to "read" noisy, degraded, or partial information, as occur in natural environments.

Another useful property of the attractor structures is that they permit the learning of arbitrary associations (in contrast to connectionist networks that generalize, giving similar responses to similar inputs). Such arbitrary associations are typical of the relationship between orthography (or phonology) and semantics, given that words that look alike do not necessarily correspond to similar concepts. Hinton and Shallice (1991; see also Plaut and Shallice, 1993) took advantage of this property of attractor networks to build a model that produces word meanings (a set of semantic features) in response to spelling patterns. Lesioning such a model (by removing units or connections) yielded a pattern of errors resembling those committed by patients with deep dyslexia (e.g., substituting a semantically related or visually similar word for the one actually shown). Likewise, after training, this model responds to degraded stimuli much like normal human readers do (McLeod et al., 2001).

As already noted, all these networks are essentially variants of the general model of lexical processing proposed by Seidenberg and McClelland (1989), according to which reading results from the cooperation of orthographic, phonological and semantic components. A recent large-scale model with this triangular structure is detailed in Harm and Seidenberg (2004). In this model of meaning activation, distributed semantic representations are activated via an orthographic pathway and a phonologically mediated pathway. Unlike in Morton's logogen model, these two pathways to meaning are not independent, but rather cooperate to arrive at a meaning. Although based on the same computational principles, each pathway contributes to a different degree depending on specific word characteristics (e.g., frequency or spelling–sound consistency) and context (e.g., priming effects) because the two are trained to resolve different mappings. As in humans, orthographic-to-semantic mappings require more time to be learned initially but this ultimately becomes the faster pathway. In contrast, the orthographic–phonological–semantic pathway predominates in the early stages of training but in the end it is relatively slower because it involves more intermediate steps.

Arguably, the most noteworthy virtue of the PDP type models is their parsimony. Modelers start with very few assumptions about network structure. Indeed, at first glance, these connectionist models may seem to be free of theoretical bias. Denying the existence of local representations, however, is a strong assumption, and this has been the principal criticism against this kind of VWR model. Perhaps for this reason, some computational modelers have opted instead for network structures based on local representations.

2.2. *Localist connectionist models*

Localist connectionist models of VWR differ from connectionist models described thus far in that they start with localist rather than distributed representations. Unlike in distributed representations, where the same units may be involved in representing different kinds of information, in localist models, different kinds of information are encoded in different units or group of units. Connectionist models that incorporate localist representations also take as a given the traditional view that reading involves the cooperation of different subsystems, routes, and levels of processing; these subcomponents are then further specified according to theoretical considerations (Fig. 2).

Although distributed and localist modeling constitute two very different approaches to the simulation of cognitive processes, they are not necessarily mutually exclusive. Models based on local representations, for example, do represent distributed knowledge, albeit at a different level (than the localist representations). Likewise, models based on distributed representations usually include units that can be considered as local representations, but at a lower level (e.g., the semantic features that encode distributed word meanings)[1]. In addition, localist models can incorporate learning algorithms that modify the strengths of the prespecified connections, while some types of connectionist networks can learn localist representations (see Page, 2000 for a further discussion and a defense of localist approaches).

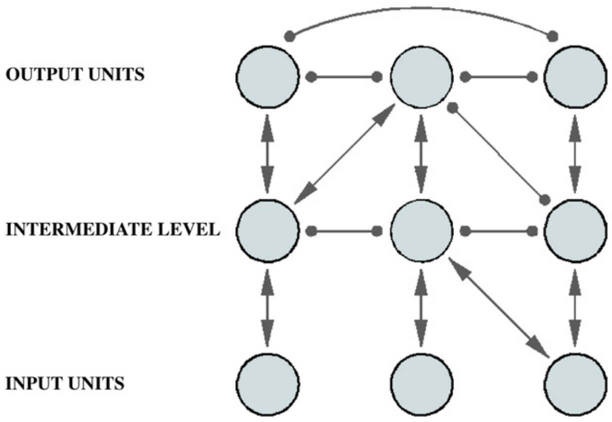

Fig. 2 – An example network with localist representations; activation can spread interactively between units (localist representations) through excitatory (arrows) or inhibitory (black circles) connections which are hand-wired according to theoretical considerations.

In fact, the first connectionist model of letter and word perception, the *Interactive Activation and Competition* (IAC) model (McClelland and Rumelhart, 1981; Rumelhart and McClelland, 1982), is a localist model whose architecture was hand-wired rather than learned. It includes units at three representational levels: a visual feature level (lines in different orientations), a letter level (letters of the alphabet in each position within the word), and an orthographic word level (word units). Activation spreads in this model interactively because it flows across levels in parallel through bidirectional inhibitory or excitatory connections. Unlike in the logogen model in which activation is discrete, with logogens activated only after some threshold activation level is reached, activation in these connectionist networks flows continuously or in a cascade-like fashion. The cornerstone of such interactive models is the *inhibition hypothesis*, according to which activation of a specific symbolic unit at any level (e.g., words) is the outcome of a competition among local representations with interconnections at the same level. Initially, activated word units compete via their lateral inhibitory connections while they send excitatory feedback to their corresponding letters. This feedback mechanism explains, for example, why letters are usually detected faster when they appear within a word than within a nonword letter string.

A further extension of the IAC model principles can be found in the *Multiple Read-Out Model* (MROM) of Grainger and Jacobs (1996); the MROM is a model of orthographic processing developed to simulate human performance in word identification and lexical decision tasks. On this model, a correct response to a word stimulus is given when activation of a word representation reaches a critical level. In the lexical decision task, for example, a "yes" response is given when some word representation exceeds some specified criterion and a "no" response is given if no representation is sufficiently activated after some (pre)specified time. The model also introduces a "fast guess" mechanism that also can lead to a "yes" response when the total activation in the orthographic lexicon is high, even before any specific word representation has reached threshold criterion. In other words, lexical decisions can be made even before lexical identification is fully complete. The model admirably reproduces a broad range of behavioral data and the "fast guess" mechanism is especially useful in simulating the orthographic neighborhood size effect,[2] and its pattern of interactions with other lexical

---

[1] However, they are still considered distributed models, viewed from the perspective of the primary process being modeled.

[2] The orthographic neighborhood size of a word is usually defined as the number of words that share all but one letter with that word. Low-frequency words with many orthographic neighbors are on average recognized faster than those with fewer neighbors.

variables in words and pseudowords (e.g., the frequency of orthographic neighbors). A more recent version of the model incorporates phonological processing as well; when letters activate word units in the orthographic lexicon, they also activate a parallel set of nodes in a phonological lexicon (Jacobs et al., 1998).

Coltheart et al. (1993, 2001) have elaborated a more general computational model of VWR and reading aloud, based on the same philosophy, called the *Dual Route Cascade* (DRC) model. In line with other classical reading models, the DRC model has two different input and output lexicons. The DRC assumes two distinct processing routes from print to sound—one lexical or purely orthographic and the other nonlexical. In this model, visual feature units activate letter units and, in the lexical route, letters activate words in the orthographic lexicon, which then activate representations in the phonological lexicon. A distinctive characteristic of the DRC model is that it incorporates a set of grapheme-to-phoneme correspondence rules into the nonlexical route. These rules convert letter strings into phoneme strings serially. Although it has not yet been implemented, the model will eventually also house a semantic system and thus possess a semantically mediated route as well. The DRC model has been quite successful in simulating important behavioral phenomena in normal readers, in both pronunciation (reading pseudowords and phonologically irregular words and producing the regularity by frequency interaction) and lexical decision tasks (where the three response mechanisms of the MROM are incorporated). Moreover, the DRC adequately models certain reading phenomena in aphasic individuals, including the dissociation between surface and phonological dyslexia.[3] Pseudowords, for instance, are pronounced by means of grapheme–phoneme correspondence rules using the nonlexical route, whereas previously learned irregular words are read out by means of the lexical route. Regular words can use both routes. Surface and phonological dyslexia are simulated by damage (i.e., turning down the activation parameters) to the lexical and nonlexical routes, respectively.

2.3. *Summary*

In this section, we described some representative examples of two major classes of computational models of VWR: those with distributed representations and learning mechanisms and those with local representations and hand-wired architectures. These two classes of models not only differ in implementation but reflect two very different philosophies about modeling in cognitive science.

PDP models stress the role of learning. Word recognition, in this computational framework, is an instantiation of a continuous and dynamic learning process that depends on the interaction between the new input, the current state of the system, and its previous history. Information provided by previous encounters with any given linguistic information is crucial in shaping the system's response. The statistical regularities of the language and its use are therefore important determinants of the system's behavior. Neural networks of this type have been able to simulate a variety of behavioral observations using homogeneous computational principles. Triangular models based on Seidenberg and McClelland (1989) seminar proposal, for example, are able to correctly pronounce both words and pseudowords and to simulate naming latencies thereof (e.g., Plaut et al., 1996), as well as to offer viable explanations for frequency and regularity effects, and the impact of these on the division of labor between the orthographic–semantic and phonologically mediated pathways. These networks also can mimic characteristic error patterns associated with phonological processing of homophones (words that look different but sound the same). These types of models likewise can account for certain semantic effects in word reading such as the imageability effect (low-frequency words with higher imageability are named faster than those words with lower imageability) and semantic priming effects when two words are presented sequentially (Harm and Seidenberg, 2004). Last but not least, these models can be "lesioned" to reproduce responses of certain aphasic patients or tweaked to parallel the performance of normal readers encountering partial or noisy information (Hinton and Shallice, 1991; McLeod et al., 2001).

Models based on localist representations, by comparison, have typically been presented as computational instantiations of classical verbally descriptive models, especially the logogen model (Grainger and Jacobs, 1996; Coltheart et al., 2001). The functional architecture of these models including their basic units, levels, and connections is pre-wired. This not only makes the structure and dynamics of the model more transparent but also allows for direct testing of specific hypotheses drawn from the empirical data. Although learning and the statistical regularities within the language under study also have prominent roles in such models, they tend to focus more on the interaction between stable subsystems operating under different processing principles. Like PDP models, these models have successfully simulated an impressive range of behavioral effects. For example, the latest version of the DRC model mimics all the naming effects previously simulated by Plaut and collaborators (1996), as well as homophone processing and priming. Likewise, both the DRC and MROM models can simulate the main findings of lexical decision task performance: faster responses to words than to pesudowords, the frequency effect, interaction between frequency and size or frequency of orthographic neighborhood, as well as homophony effects.

Despite their different theoretical underpinnings and mechanistic implementations, both classes of models of VWR cannot only adequately simulate some reading phenomena but in many cases, the same reading phenomena. Modeling success therefore in no way guarantees which model is accurate; the brain may simply perform the computations in a different way. One crucial test of any computational model is just how well it can simulate a large variety of different effects (in size and timing) including those that have yet to be discovered. Together with biological plausibility, flexibility to adapt to an ever-growing body of new empirical evidence is key.

---

[3] Surface dyslexia is characterized by relatively intact reading of regular words and pseudowords, together with poor reading of exception or phonologically irregular words (resulting in regularization errors). Phonological dyslexia, in contrast, is characterized by poor reading of nonwords with relatively intact word reading.

In summary, each class of computational models discussed herein was developed to account for a circumscribed set of reading phenomena and we have every expectation that future developments will extend their scope to encompass others. For example, there is no obvious practical or theoretical reason why the Harm and Seidenberg (2004) model cannot simulate response latencies in naming and lexical decision tasks or why the DRC model cannot incorporate semantic processes and model semantic effects in reading (Coltheart et al., 2001). Likewise, certain behavioral phenomena that currently elude one or the other type of model may ultimately be achieved either via relatively minor modifications to the existing models (Ziegler et al., 2001; Zevin and Seidenberg, 2006) or via novel combinations of computational principles, e.g., connectionist models that are able to learn localist representations (Bowers, 2002). These potential improvements notwithstanding, in the next section, we turn our attention to what are arguably greater challenges for *all* VWR models, the solutions to which we maintain would lead to qualitative reformulations of extant models, and in so doing a deeper understanding of the reading process.

3. Limitations of current visual word recognition models

3.1. *Normal visual word recognition, retinas, and cerebral hemispheres*

With a few notable exceptions, incorporating the physiological and anatomical milieu within which reading computations take place has not been a priority for computational modelers. While this may have been an effective strategy in the initial stages of theoretical and computational development, it is unlikely to be the wisest tack now just as VWR models are becoming increasingly more successful in accounting for reading phenomena. Explaining the last 2–5% of the variance so often offers the most difficult challenge for any model that it would seem to be an especially good criterion for adjudicating among alternative proposals all of which can seem to account for a large portion of the variance, albeit in fundamentally different ways.

Arguably, the most salient anatomical characteristic of the brain is its separation into two hemispheres. *Bi-hemispheric neural networks* have been used to test different proposals about hemispheric specialization and interhemispheric transfer in language processing (Reggia and Levitan, 2003; Weems and Reggia, 2004). These models are composed of input and output layers connected via two different association layers (representing the hemispheres) with connections between them. Reggia and Levitan (2003) developed a bi-hemispheric network of lexical identification (associating graphemes to lexical representations) to examine how hemispheric specialization might emerge at least in this type of task. They found that a processing advantage for one of the networks can arise due to its larger size, greater excitability, or a faster learning rate. In a related study (Weems and Reggia, 2004), they used a similar model of lexical decisions to test various hypotheses regarding interhemispheric communication. Simulations supported a model in which communication spreads in both directions and the two hemispheres cooperate to generate outputs compared to a model in which information spread is mostly unidirectional or in which the two hemispheres act independently. Despite the limited scope of these simulations, they have been useful in testing specific hypotheses about the functional organization of the brain and thus could prove to be of even greater benefit to large-scale models of VWR.

Of the few modelers who have introduced some biological constraints into their computational models of VWR, special attention has been devoted to the implications of the fact that visual information is projected to the two cerebral hemispheres, each with its own representations and processing capabilities. It is commonly assumed that receptors in the center of the fovea transmit signals to both hemispheres, creating visual field overlap around fixation. Some researchers, however, argue instead primarily from neuropsychological data for a split fovea in humans; that is, a clean bisection of the fovea, with information from each visual hemifield sent to the contralateral hemisphere (Brysbaert, 1994; Leff, 2004). On this view, when a word is presented in central vision, letters falling to the left of fixation are projected to the right hemisphere (RH), whereas letters falling to the right of fixation are projected to the left hemisphere (LH). A few recent models of reading have incorporated this proposal (Shillcock et al., 2000; Whitney, 2001).

The SERIOL model, for example, is a 5-layered model (retina, feature, letter, bigram, word) that takes letter strings as inputs and connects them to lexical units (Whitney, 2001, 2004). The retinal level reproduces the distributions of the different cones in a real human retina via an acuity gradient, with activation decreasing with distance from fixation. The feature level honors the split-fovea principle in its representation of suborthographic features, such that the LH receives information from the right visual field (RVF) and vice versa. At this level, the retinal acuity gradient is transformed into a location gradient, with weights decreasing as a function of a letter's ordinal position within a word from first (most activated) to last (least). At the subsequent letter level, this location gradient is converted into a firing pattern, funneling parallel activation into serial activation. At the bigram level, ordered pairs of letters are recognized, with the relative time of bigram firing being the basic unit for lexical access at the word level. Within this model, information in the RH is transferred to the LH at the letter level, in line with the proposal that there exists a specific (word form) area in the left cerebral hemisphere where orthographic information is initially analyzed (McCandliss et al., 2003).

Shillcock and collaborators (Shillcock et al., 2000; Shillcock and Monaghan, 2001) likewise have implemented a bi-hemispheric neural network in which the two halves of a word are processed differently after a foveal split. Their network has two input layers, one corresponding to each visual field; these are in turn connected to two corresponding interconnected hidden unit layer "hemispheres". Shillcock et al. argue that the asymmetrical distribution of information within words would naturally lead the two model "hemispheres" (and presumably in humans as well) to different optimal solutions based on the different partial (half) word inputs they receive. More specifically, faced with the need to quickly analyze the high density of information coming from word initial segments, the network's "RH" may settle on a coarse-coding scheme such that information from bigrams or trigrams (letters pairs or triplets) is jointly processed. By contrast, the lower density of word-final information may allow the network's "LH" to settle on a finer-coding scheme in which units represent individual letter positions. As semantic priming simulations show that training does indeed lead the model to develop more loosely related semantic connections in the "RH" layer than in the "LH" layer. One important consequence of these two different processing/representation strategies is that the "LH" layer responds rather accurately (at least more so than the "RH") to a given word input, whereas the "RH" layer shows some activation even to items that are orthographically and/or semantically close to the input word (Monaghan et al., 2004).

While the split-fovea hypothesis is controversial and thus in need of additional empirical support, networks embodying this hypothesis, nonetheless, provide a clear example of how computational models can contribute to a debate about the anatomical bases of reading not only in showing the feasibility of a proposal but in generating specific predictions that can be assessed via electrophysiological measures.

3.2. Orthographic systems across languages and sublexical units

3.2.1. Cross-linguistic studies of reading

For the most part, empirical work on and thus the computational models of VWR have been based on languages with alphabetic orthographies—mainly English. There is thus a relative paucity of VWR models dealing with non-alphabetic systems (e.g., syllabic Japanese Kana or logographic Chinese). A notable exception is the Lexical Constituency Model (LCM), which has been used to simulate orthographic, phonological, and semantic priming effects on naming in Chinese (Perfetti et al., 2005). The LCM includes both local and distributed representations, fixed connection weights, and activation that spreads via thresholded functions. In addition to a level of radical units (Chinese orthographic units), LCM contains phonological, orthographic, and semantic levels. While phonology is not used to identify words, it is considered part of the word to be identified. Critical for present purposes, the phonological unit's contribution to a word's identity (in addition to the contributions from orthography and meaning) is considered universal. This feature of LCM thus can explain phonological reading effects even in languages, like Chinese, where, due to the preponderance of homophones, phonological information does not unambiguously map onto meaning. At the same time, the developers of LCM are explicit in their belief that the specific nature(s) of the orthographic unit(s) for any given writing system will naturally influence the particular basic processes involved in VWR for that language. We likewise believe that other characteristics that distinguish one language family from another such as the left-to-right scanning or the absence of an explicit graphic representation of vowels in Semitic languages will impact the relative weightings of different information types during VWR, if not the actual information used. Even more subtle differences, such as the degree of consistency between orthography and phonology (e.g., low in English versus high in Spanish), also must play some non-trivial role in VWR and thus in models thereof. It should go without saying that comparative experimental studies across different orthographic scripts and languages and computational VWR models that can handle the different patterns across cross-linguistic data will be crucial to a full understanding of the structure and dynamics of VWR.

3.2.2. Most words are not monosyllabic

Presumably, for the sake of simplicity, the vast majority of VWR models (all of those reviewed thus far) have dealt exclusively with monosyllabic words, even though in many languages, polysyllabic words are far more numerous. This limitation has been acknowledged by the authors of contemporary models of VWR (e.g., Coltheart et al., 2001; Harm and Seidenberg, 2004), but only a few researchers have attempted to remedy the situation. Ans et al. (1998), for example, developed a connectionist feedforward network – the *Multiple-Trace Memory* (MTM) model—that simulates the reading of polysyllabic French words. MTM has two distinct word processing modes operating in series. Initially, all words (and pseudowords) are analyzed via a global route. However, whenever the global route is stymied, for example, by a pseudoword or a low-frequency word, an analytic mode based on word syllable segments takes over. More precisely, letter string inputs are analyzed in order, starting with the largest unit that the system recognizes as familiar (usually a syllable) and so on; in this way, the network performs automatic syllabic segmentation. While MTM offers an adequate starting point for polysyllabic word processing, a similar endeavor with languages like English, in which words have less clear syllabic boundaries, will undoubtedly raise a whole host of additional challenges (Rastle and Coltheart, 2000).

In order to capture the syllabic frequency effect reported for some languages,[4] it has been proposed that interactive-activation models should incorporate an intermediate level made up of syllabic units that compete via lateral inhibitory mechanisms (Carreiras et al., 1993). It is not obvious to us, however, how current versions of these models would be affected by such modification and if, for example, they could still account for some of the previously simulated orthographic effects, such as the neighborhood frequency effect (see Conrad and Jacobs, 2004 for a further discussion). Since PDP models are able to simulate frequency effects of units that are not locally represented (e.g., the lexical frequency effect), it is an open question whether syllable frequency effects could emerge in them from the interaction of orthographic and phonological information.

3.2.3. Not all words are monomorphemic

A related problem, but with different implications, faces models with morphemes as sublexical units. In principle, morphemes can mediate between lower levels (orthographic

---

[4] Words beginning with a high-frequency syllable are associated with longer latencies than words starting with a low-frequency syllable in lexical decision tasks.

and/or phonological) and semantic representations, so the different units could be hierarchically integrated in an interactive-activation type model (Taft, 1994). Unlike orthographic and phonological units, however, morphemic units also can serve as semantic units in their own right. However, morphemes need not necessarily be made up of some other subunits such as syllables given that in many languages there is no correlation between morphological borders and syllabic borders (e.g., the monosyllabic word “CATS” is made up of two morphemes (CAT-S) and the monomorphemic word “WINDOW” is made up of two syllables “WIN-DOW”). Integrating both syllabic and morphemic units into a unified model of VWR thus could prove to be a thorny problem.

Seidenberg and Gonnerman (2000) proposed an alternative PDP model based on the assumption that a word's morphological structure is epiphenomenal, emerging from the confluence of orthographic, phonological, and semantic factors. Likewise, some learning or neuropsychological effects related to inflectional morphology (like the past tense generation in English verbs) have been simulated without any explicit morphological level of representation (Plunkett and Juola, 1999; Joanisse and Seidenberg, 1999). However, no model of reading that has been implemented to date is able to adequately capture the mounting evidence that morphology does play some role in VWR. Clearly, extending the current generation of computational models—admittedly quite successful in dealing with monosyllabic words—to handle polysyllabic and polymorphemic words as readily is one of the major challenges that contemporary modelers of VWR face.

3.3. *Timing is critical*

A critical variable that will ultimately need to be an intrinsic part of any viable computational models of VWR is the time course with which different information types become available and are used until a letter string is recognized (or not) as a word. Of particular importance, especially from a modeling perspective, is whether the different information types are processed simultaneously and in parallel or in a particular serial or cascaded order sequence. Although none of the models described thus far has as of yet fully mimicked the dynamics of the word recognition process, the temporal dimension of current VWR models has been singled out as in need of improvement (e.g., Harm and Seidenberg, 2004), and a few tentative speculations have been tendered (Whitney, 2001). Some models have successfully simulated naming latencies (e.g., Plaut et al., 1996; Coltheart et al., 2001) or predicted entire response distribution functions (Grainger and Jacobs, 1996), but those are still measures of the total time that the whole process consumes to reach an optimal solution. Processing cycles or settling timings have not yet been correlated with partial computations to account for process dynamics, mainly due to the lack of an adequate quantitative description of the temporal dimension, which obviously is an essential prerequisite. We believe that this is an arena where electrophysiological data (event-related brain potentials or ERPs, in particular) can make important, arguably unique, contributions to the study of VWR processes. More specifically, electrophysiological data can help to delineate the various subprocesses involved in VWR by identifying qualitatively different processes (how many there are, what they are, and their relation to proposed physical or psychological distinctions), their relative time course(s) of (dis)engagement—including onset, offset, duration, and relative order, and their susceptibility to context.

4. Electrophysiology and mental chronometry of visual word recognition

Electrophysiological measures, in particular, ERPs recorded from the human scalp have proven to be especially useful tools in the study of various aspects of human cognition, including language comprehension (see reviews in Rugg and Coles, 1995; Münte et al., 2000; Rösler, 2005; Kutas and Federmeier, in press; Kutas et al., in press). ERPs are voltage fluctuations over time (usually 1–2 seconds) triggered by some sensory, motor, or cognitive event. As any given ERP is quite small relative to the background noise in the electroencephalogram, researchers typically compute a mean ERP by averaging across multiple segments of electroencephalographic recordings presumably time-locked to physically or conceptually similar events. As the average ERP at the scalp at any moment is a reflection of the sum of post-synaptic potentials at that moment (primarily from pyramidal cells in the neocortex), it provides a reasonable estimate of the neural activity supporting the neocortical neural processes that unfold as an event is analyzed. The ERP has an exquisite temporal resolution spanning the range from milliseconds to seconds—the very range during which reading processes occur. ERPs thus are now routinely combined with the well-known paradigms of experimental psychology to track the neural processes involved in stimulus evaluation, decision-making, implicit and explicit memory, response preparation, selection, and execution, among other processes.

Recognizing the printed word, for example, requires a series of mental operations (just which and in what order remain contentious issues), many if not all of which are reflected in the ERP triggered by word onset not only during the word's actual appearance but for hundreds of milliseconds thereafter. ERPs thus are an exquisite tool for tracking the mental operations of reading, particularly those that occur between the initial appearance of sensory input and the apprehension or construction of its meaning (see Fig. 3 for a representative visual ERP to a word).

4.1. *N400 component*

The N400 is the most intensively studied ERP-language-related component, although it is unequivocally not language-specific. The N400 is a negative-going component starting around 200 or so ms and peaking around 400 ms after the onset presentation of a word or any other potentially meaningful stimulus (Kutas and Hillyard, 1980; Kutas and Van Petten, 1994; Kutas and Federmeier, 2000). N400 amplitude to words within sentences is a linear function of how predictable those words are based on the sentence context up to that point (where predictability is measured by offline cloze probability measures); the more predictable a word, the smaller the N400 elicited (Kutas and Hillyard, 1984; Van Petten and Kutas, 1990).

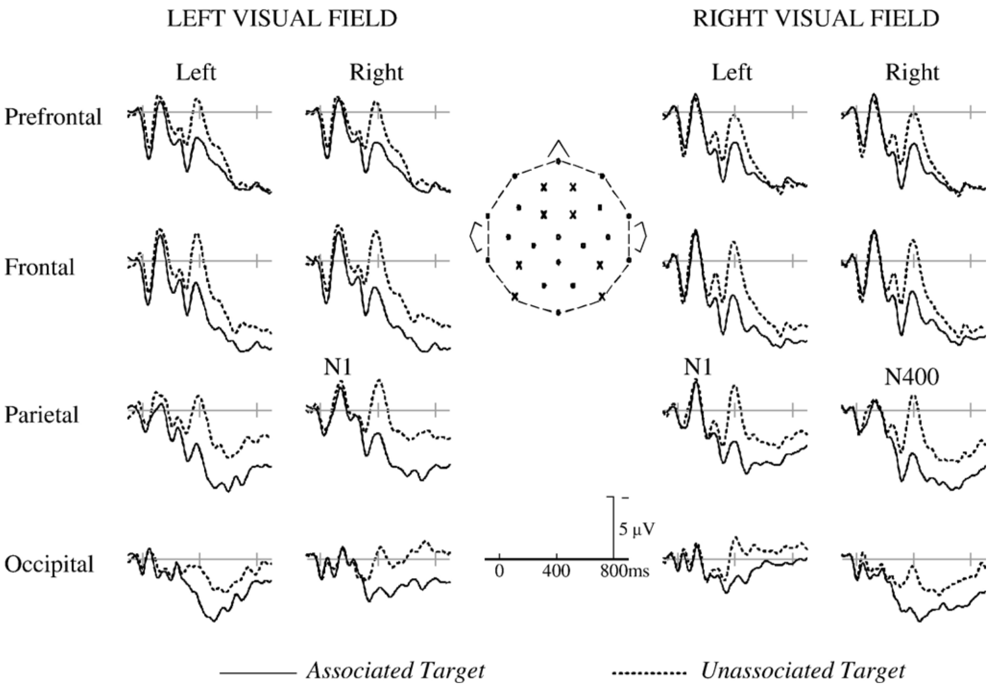

Fig. 3 – Visual event-related potential (grand average, $N = 18$, average age = 20 years old) elicited by the second of a pair of words randomly presented for 200 ms 2° lateral to fixation in the left or right visual field following a centrally presented prime word that was either semantically related or semantically unrelated to it (prime to target onset to onset interval = 500 ms). Each recording is referred to the offline average of left and right mastoids. Negative plotted up. Note the contralateral posterior N1 component and the large N400 to semantically unrelated words (data reported in Coulson et al., 2005).

A similar reduction in N400 amplitude characterizes the ERP to the second of two successive words occurring as a pair or in a running list when they are semantically or associatively related (e.g., Bentin et al., 1985) or identical (Rugg, 1990).

Under certain task instructions, the amplitude of the ERP during the N400 region is also sensitive to phonological (Radeau et al., 1998) and morphosyntactic relations (Barber and Carreiras, 2003), suggesting that these different codes are available and come together (perhaps inevitably and critically) to affect ongoing processing between 200 and 500 ms post-word onset, at least in normal young adults. Accordingly, it has been suggested that the N400 is a good index of the ease of accessing information within long-term semantic memory and its integration with the local context (Kutas and Federmeier, 2000).

N400s in response to written or spoken words can vary in onset latency, duration, and relative amplitude distribution across the scalp, as a function of sensory modality of input, among other factors. Although the N400 is in some sense modality-independent, its timing and continued sensitivity to input modality taken together suggest that the neural processes reflected in N400 activity might be the ongoing, cumulative process by which incoming information cues (and thus makes available) information stored in long-term semantic memory for meaning activation and construction. As such, N400 amplitude has been successfully used as an indirect probe of the effect of orthographic information on meaning activation. A reduction in N400 amplitude, for example, has been seen when a word (e.g., chair) was preceded by a pseudoword (e.g., wable) derived from a word semantically related to the target word (Deacon et al., 2004). This result shows that even nonword letter combinations partially activate semantic representation(s) if they resemble a real word prime to some extent and that N400 amplitude is sensitive to such activation.

Holcomb and colleagues (2002) also reported a direct relation between orthographic neighborhood size and N400 amplitude. They took this relation to reflect more widespread activation of representations in semantic memory following words with more as opposed to fewer orthographic neighbors (note: the mechanism of activation was unspecified). More recently, Braun et al. (2006) examined the relationship between ERP amplitudes and lexical activity levels in a lexical decision task using stimuli chosen according to their degree of activity in MROM (Grainger and Jacobs, 1996). Specifically, the overall lexical activity of each word and pseudoword stimulus was defined as the average summed lexical activation across the first seven cycles of processing. A clear advantage of this approach is that, rather than selecting a single variable such as orthographic neighborhood size to index lexical activation, the model-generated values are a non-linear combination of various lexical variables (e.g., the frequencies of the orthographic neighbors and bigram or trigram frequencies). ERP results showed an initial distinction (i.e., amplitude differences) between words and pseudowords around 300 to 390 ms followed between 450 and 550 ms (i.e., modulation of N400 amplitude) by a sensitivity to lexical activation levels but only for pseudowords. The authors suggest that, in line with MROM's decision mechanisms, the early categorical distinction between words and pseudowords reflects an initial identification process for "yes or word" responses, while the subsequent graded effects of lexical activity are consistent with a temporal deadline based on global activation that leads to "no or pseudoword" responses.

These kinds of studies demonstrate just how the N400 can be combined with predictions of a computational model to probe meaning activation by units smaller than a word (i.e., sublexical units). Such results open up the possibility for exploring potential activation by a large bevy of variables including but not limited to the order or position of letters within a word, bigrams, syllables, morphemes, etc. Data of this type could prove especially crucial, for example, in helping to adjudicate between contrasting proposals of letter encoding, e.g.; claims for letter position encoding (Coltheart et al., 1993; McClelland and Rumelhart, 1981) versus claims for contextual encoding via letter clusters such as bigrams or trigrams (Seidenberg and McClelland, 1989) or some hybrid position (Whitney, 2001). According to the SERIOL model (Whitney, 2001), for example, initial serial processing of letters and letter positions is followed by the activation of open bigrams (ordered but not necessarily consecutive letter pairs). Letter positions are encoded via temporal firing patterns, where the time of firing depends on 40 Hz oscillatory cycles (each letter position corresponds to a successive 25 ms subcycle).

Precise predictions of this sort from brain data necessitate a relatively detailed timeline of the different neurocomputational events that transpire from stimulus onset up to the N400, starting with the initial processes of letter encoding. ERP correlates of the effects of the bigram and trigram frequency have been reported as early as 100 ms after word presentation onset (Hauk et al., 2006a), showing the sensitivity of the scalp ERP not only to lexical and semantic processing but also to early word-form information. An electrophysiological delineation of the neural processes of VWR can in turn be used to weigh in as evidence for or against various computational VWR models. In particular, a timeline of ERP sensitivities can speak to what variables are and are not important for which particular process or stage of processing, when they have their effects, and their degree of independence or interactivity.

4.2. *Electromagnetic estimates of the time course of orthographic processing*

4.2.1. *Linguistic versus non-linguistic stimuli*

One strategy for identifying visual processes specifically involved in letter encoding has been to compare the time course of word processing directly with that of other perceptually and conceptually similar but non-letter or word stimuli and look for the point and location of initial ERP divergence.

Schendan et al. (1998), for instance, examined ERPs elicited by various word-like stimuli (words, letters strings, and pseudo-fonts), object-like stimuli (objects and pseudo-objects), and faces. Over occipital brain areas, ERPs to word-like and object-like stimuli diverged around 90 ms or so after stimulus onset, presumably reflecting a low-level analysis of the physical characteristics of the visual input. By around 125 ms, the P150 component over parietal brain areas distinguished well-learned categories of visual patterns (such as words and faces) from other less well-learned visual inputs. In a similar vein, ERPs elicited by orthographic stimuli (words, pseudo-words, and consonant strings) and non-orthographic stimuli (alphanumeric symbols and strings of forms) reliably diverged at around 140 ms and showed a characteristic lateral asymmetry; over posterior temporal and occipital regions, orthographic stimuli were associated with an N150 that was larger over the left than right hemisphere whereas non-orthographic stimuli showed a reverse asymmetry. Intracranial recordings likewise have revealed selective responses to letters in inferior occipito-temporal region between 150 and 200 ms as neurons in a nearby area selectively responded to faces (Allison et al., 1994; Nobre et al., 1994). Electrophysiological data thus indicate that the brain reacts to well-learned (versus novel) input patterns at around 125–200 ms post-stimulus onset (if not before) by first analyzing the physical features of the input. Letter strings, like faces, constitute a type of rapidly detected perceptual category, presumably due to the massive exposure to (the regularities of) orthographic stimuli that the average adult reader experiences across the course of a lifetime. The absence of any difference in the electrophysiological response to words and illegal words at this point (prior to 200 ms) implies that these early processes do not involve any meaning activation.

Similar inferences have been drawn from visual processing studies using magnetoencephalography (MEG). MEG is a method in which weak magnetic fields, generated by current flow in the brain, are noninvasively measured using SQUID (superconducting quantum interference device) detectors outside the skull. MEG and EEG both derive from synchronized neuronal activity in the brain and have a time resolution in the millisecond range. However, due to the differential sensitivity of electric and magnetic fields to the skull and the geometric orientation of intracranial current flow, the two methods offer complementary information about brain activity; with MEG recording, it is primarily the activity of the cortex in the sulci rather than gyri (in which the pyramidal cells are oriented perpendicular to the skull) that are detected (for a review, see Hämäläinen et al., 1993). MEG studies of VWR have differentiated a so-called Type I activity pattern around 100 ms, reflecting low-level analysis of visual features and a Type II activity pattern around 150 ms (greater for letter than symbol strings), originating in the inferior occipito-temporal cortex and reflecting object-level processing (Tarkiainen et al., 1999, 2002).

4.2.2. *Lexicality and word frequency*

An alternative strategy for demarcating the initial perceptual processes involved in letter encoding has been to compare the ERP responses to various types of orthographic stimuli (without any additional non-linguistic contrasts for comparison). These types of studies have generally reported a word–nonword divergence in the ERP significantly earlier than the 300–350 ms difference that is characteristic of investigations that include non-orthographic stimuli. These early ERP differences of wordness (lexicality) range from 100 to 200 ms (100 ms: Sereno et al., 1998; 125 ms: Compton et al., 1991; 150 ms: Proverbio et al., 2004; 192 ms: Dehaene, 1995).

Attempts to further delineate the time course of word processing have been based on systematic manipulations of variables known to influence word processing such as word frequency. Such investigations are potentially of great relevance to reading modelers as several VWR models assume that frequent words are recognized as a whole without any intermediate analysis whereas infrequent words undergo at least some intermediate analysis. This assumption implies not only a qualitative difference in the processing of frequent and infrequent words but also that this frequency-based processing difference occurs relatively early in the word recognition process. ERP data are clear in showing that an eliciting word's frequency of usage influences neural processing at multiple time points during the course of VWR. Word frequency, for example, modulates the ERP between 300 and 500 ms (N400); with all other factors held constant, N400 amplitude is an inverse function of a word's eliciting frequency (e.g., Bentin et al., 1985; Rugg, 1990; Barber et al., 2004). These N400 frequency effects however can be overridden by semantic context: large frequency effects on the N400, seen for open class words early in a sentence, are absent for such words by sentence end (Van Petten and Kutas, 1990). Dambacher et al. (2006) further observed that the effect of semantic context (e.g., predictability estimated from offline cloze probabilities) was larger for low- than high-frequency open class words. Word frequency effects were seen as early as 130 or so ms both in ERP (132–192 ms: Sereno et al., 1998, 2003; 135–175 ms: Proverbio et al., 2004; 150–190 ms: Hauk and Pulvermüller, 2004; 140–200 ms: Dambacher et al., 2006) and MEG investigations (Assadollahi and Pulvermüller, 2001). Assadollahi and Pulvermüller (2003) also reported that the latency of the frequency effect interacted with word length, ranging between 150 ms for short words and 240 ms for long words, suggesting that formal and lexical variable interactions might explain part of the latency variability observed across studies. An intermediate ERP component over left frontal sites with a peak latency that varies inversely with written word frequency, dubbed the Frequency Sensitive Negativity (FSN), was described by King and Kutas (1998). In their sample of words from a series of isolated sentences, the FSN peaked around 280 ms for high-frequency words and around 335 ms for low-frequency words (see also Osterhout et al., 1997; Münte et al., 2001; although not all researchers have been successful in isolating an FSN, Brown et al., 1999). A similar sensitivity to lexical frequency was described for the latency of the M350 component[5] (Embick et al., 2001). In summary, it seems that frequency impacts word processing at multiple time points, relatively early, late, as well as at points in between, suggesting that several word processes are sensitive to experience. To date, there is no strong evidence for qualitative differences in the processing of high- versus low-frequency words, as predicted by some VWR models, except fairly late (only for closed class words, Münte et al., 2001; or after repetition, Rugg, 1990). At the same time, no ERP studies have thus far capitalized on these reported sensitivities to frequency to test specific VWR proposals, such as whether there is any sign in the ERP that low-frequency words are of necessity analyzed at a sublexical level as they cannot be recognized globally (Coltheart et al., 2001; Ans et al., 1998) or whether low-frequency words do in fact place greater demands on a slow, phonologically mediated pathway (Harm and Seidenberg, 2004).

4.2.3. *Conclusion*

Electrophysiological investigations of linguistic versus nonlinguistic stimuli, different kinds of letter strings (from consonant strings to real words), and of the lexical variables such as frequency and length all suggest that, *when* the brain distinguishes between these various inputs depends, it seems not just on the nature of the input but also on the task demands. In broad stroke, designs which include orthographic as well as non-orthographic stimuli tend to find a later latency of divergence than designs limited to letter strings. This suggests the possibility that human brains might adopt a reading set or mode (based on prior materials or other expectations) that predisposes the word processing system to process incoming stimuli as orthographic, thereby speeding up (at least some) VWR processes.

4.3. *Tracking phonological processes*

In addition to the availability and use of various orthographic codes, VWR models need to be informed by the availability and use of other types of lexical information during the recognition process. As already mentioned, word phonology has been hypothesized to be one such important source of information for VWR (e.g., Frost, 1998). Consistent with this proposal, semantically incongruous words that were homophonic with the expected ending of contextually constrained sentences were found to elicit smaller N400s (indicating priming) than nonhomophonic anomalous endings (Newman and Connolly, 2004). Experiments with Chinese homophones also have revealed phonological priming effects in the N400 time window, supporting the proposition that phonological information is considered even in a language where its role could reasonably be questioned (Liu et al., 2003).

The first ERP divergence between nonpronounceable nonwords and pseudowords reportedly appears between 250 and 350 ms post-letter string onset (Bentin et al., 1999) and has been variously linked to the accessing of phonological representations (Proverbio et al., 2004) or to grapheme–phoneme conversion mechanisms (Simon et al., 2004). Similarly, the M250 and M350 components of event-related MEGs reveal sensitivities to phonotactic probability (frequency of occurrence of a sound and sequence of sounds in words) (Pylkkänen et al., 2002).

Phonological processing also has been investigated more directly by recordings taken as participants render rhyming judgments about visually presented word pairs. Rhyming effects have been observed for both the N200 and N400 components (Rugg and Barrett, 1987; Valdés-Sosa et al., 1993). Kramer and Donchin (1987), for example, found a greater negativity peaking at 350 ms for non-rhyming relative to rhyming word pairs. Based on a comparison of orthographic and phonological conditions, they concluded that a phonological mismatch can be detected ~260 ms after word onset. A similar latency has been reported for the onset of the N400 visual rhyming effect, being very consistent across different

---

[5] Detection of event-related responses in single individuals is easier in MEG than in EEG, which makes this method especially suitable for the detection of small latency differences.

groups of children and adults (Grossi et al., 2001). Bentin et al. (1999) likewise observed modulation of an N320 component by non-rhyming words, as well as a distinction between pronounceable and unpronounceable letters strings in an N350 component. As these effects were contingent on there being a phonological task, they could either reflect phonological processing *per se* or semantic activation resulting from phonological processing. All told, however, such data indicate that there is phonological mediation between orthography and semantics, at least under certain conditions. Moreover, considering that orthographic and lexical effects in the ERP have tended to appear earlier than these phonological effects, the data pattern is consistent with the hypothesis that the phonological pathway is slower than the more direct orthography-to-semantics pathway. This processing speed difference, however, is equally explicable by models in which grapheme–phoneme conversion unfolds serially (Coltheart et al., 2001) as by connectionist PDP models (Harm and Seidenberg, 2004).

4.4. *Word class distinctions*

In addition to orthographic and phonological information, words are associated with other types of information such as word class that also might be considered in early stages of VWR, especially when words appear within sentences. Word category violations (also referred to as phrase structure violations), for example, elicit an enhanced left anterior negativity between 150 and 300 ms, referred to as LAN, whose latency presumably is a function of the time at which word category information is available (e.g., Neville et al., 1991; Friederici, 1995; Hahne and Friederici, 1999). An early frontal negativity with a left hemisphere (N280) focus was initially proposed to distinguish closed class or function words (determiners, articles, conjunctions, prepositions) from open class or content words (nouns, verbs, adjectives, adverbs; Neville et al., 1992; Nobre and McCarthy, 1994). King and Kutas (1998) however showed that the N280 was not specific to closed class words in that it could be subsumed by latency variability in the same negativity—an LPN, which peaks early for closed class words and later for open class words (see also Osterhout et al., 1997; Münte et al., 2001; Brown et al., 1999; ter Keurs et al., 2002). Pulvermüller et al. (1995) proposed that differences in laterality, rather than in amplitude or latency, best distinguish open from closed word classes at around 200ms, the same latency at which amplitude differences between German verbs and nouns were observed (Preissl et al., 1995; Pulvermüller et al., 1999; see also a similar result in Dutch: Kellenbach et al., 2002). Federmeier et al. (2000) also described word class differences for nouns versus verbs in English. They presented unambiguous nouns, unambiguous verbs, and class-ambiguous words that could be a noun or a verb (e.g., "drink") in grammatically unambiguous contexts. Class-ambiguous words differed from the other items 100 ms after word onset, even when sentential context rendered them unambiguous. Moreover, at around 200 ms, unambiguous verbs elicited a left-lateralized anterior positivity. Taken together, such results suggest that some information about a word's grammatical category may become available relatively early in the VWR process, although the possibility that these word class effects are secondary to various semantic and/or syntactic processes has not yet been ruled out. Additionally, one study described an ERP response to gender agreement violations appearing at around 100–150 ms post-word onset, showing that, at least in Hebrew, morphosyntactic gender information also is available quite early (Deutsch and Bentin, 2001). In other languages (e.g., German or Spanish), the equivalent effect has been seen later, post-300ms (Gunter et al., 2000; Barber and Carreiras, 2005). Meaning activation and integration processes thus may be the outcome of more than initial orthographic decoding (as determined by familiarity and other lexical factors), including the co-activation of phonological and perhaps even some morphosyntactic word features.

4.5. *Conclusions*

All of these data have yet to be wholly integrated into an explicit timeline of visual word recognition. Nonetheless, there have been partial attempts to make order of the extant behavioral and brain data on VWR. Sereno and Rayner (2003; see also Sereno et al., 1998), for example, claim that some level of lexical identification must occur between 60 and 150 ms after word fixation. Their argument is based on evidence of (1) top–down influences on oculomotor control (eye movements) during reading which they take to support a certain level of lexical identification before the eyes leave a word to initiate a new saccade, (2) an average fixation duration during reading around 200–250 ms,[6] and (3) estimates of initial peripheral transmission and late oculomotor programming and execution times. Their claim finds some support in the few ERP experiments that have revealed lexical frequency effects in this time window (e.g., Sereno et al., 1998, 2003). This inference, however, rests on the questionable, albeit not uncommon, assumption that word frequency effects on the ERP necessarily indicate lexical access. At present, it is unclear why more ERP studies have failed to show such early effects reliably. Hauk et al. (2006b) would attribute the inconsistencies to the confounding of many lexical factors in previous studies. Along with some late effects, they too found some early effects of different lexical features (obtained from a Principal Components Analysis of corpus-based information) on ERPs to written words in a lexical decision task using a regression analysis. More specifically, they describe effects of word length and letter N-gram frequency (a measure highly correlated with bigram and trigram frequency) around 90 ms, an effect of lexical frequency around 110ms, and effects of both morpho-semantic coherence (the degree to which words sharing a root morpheme are semantically related to each other) and lexicality (in a factorial analysis) around 160 ms.

---

[6] During normal reading, words are presumed to be perceived parafoveally before being directly fixated. This preview can make available at least some information about the upcoming word (e.g., orthographic information) and thereby speed up its subsequent identification by at least in 20 ms or so (Rayner, 1998). In most ERP experiments, sentences are presented one word at a time at the same location. This procedure circumvents the artifacts due to eye movements of various sorts during natural reading, but at the same time deprives the system of such preview benefits, likely delaying some of the computations involved in word identification.

These variables also produced topographically similar effects in parts of the waveform post 200 ms. Hauk et al. maintain that this timeline is most consistent with an interactive cascade model in which word-form and lexical representations are accessed consecutively, albeit with substantial processing overlap. The similarity of the ERP effects observed in association with these different lexical variables in the later time windows is argued to reflect parallel processing of post-lexical information, in contrast to the earlier effects associated with lexical access and lexical selection processes.

Taken together, these electrophysiological studies of VWR point to parallel activations of different kinds of word-related information, all available soon after the initial orthographic analysis of a written word (Fig. 4). When different types of visual stimuli are presented, ERPs associated to well-learned versus novel stimuli differ from each other around 100 ms after presentation onset, discriminating orthographic from non-orthographic stimuli ~50 ms later. Similarly, a few studies using only orthographic stimuli and linguistic tasks have reported wordness or lexicality effects (word versus nonwords or/and pseudowords) and word frequency effects in this same (early) time range (although the impact of these variables has been more consistently reported at later time points, especially post-300 ms). Phonological information seems to become available a little later between 200 and 350 ms, while the availability of grammatical class and other morphosyntactic information seems to depend at least in part on how this information is represented within a language. Under certain circumstances (e.g., languages), some lexico-semantic information might be at least partially activated within 150–250 ms of word onset, after which other information associated with that word seems to accumulate gradually. All these information types are apparently integrated between 300 and 500 ms. Finally, ERP data clearly indicate that the temporal course of VWR processes remains highly flexible and responsive to top–down factors (such as local sentence context and task demands).

This general description of VWR as a fast although extended, cumulative, and adaptable process accords well with brain models of visual perception more generally, as described next. While the different visual pathways are loosely hierarchically organized (with higher visual areas coding more complex representations than lower visual areas), the significant coincidence of activation times across the many different striate and extrastriate areas from the outset indicates that numerous computations are being performed in distinct, but highly interconnected areas in parallel (Bullier and Nowak, 1995). Newer models of visual perception stress both the extended temporal course and cumulative nature of context-sensitive, constructive visual processing, with perceptual representations becoming progressively detailed and elaborated over time, via not only local activity but substantial feedforward, feedback, and lateral connections (e.g., Rousselet et al., 2004, Lamme, 2003). While

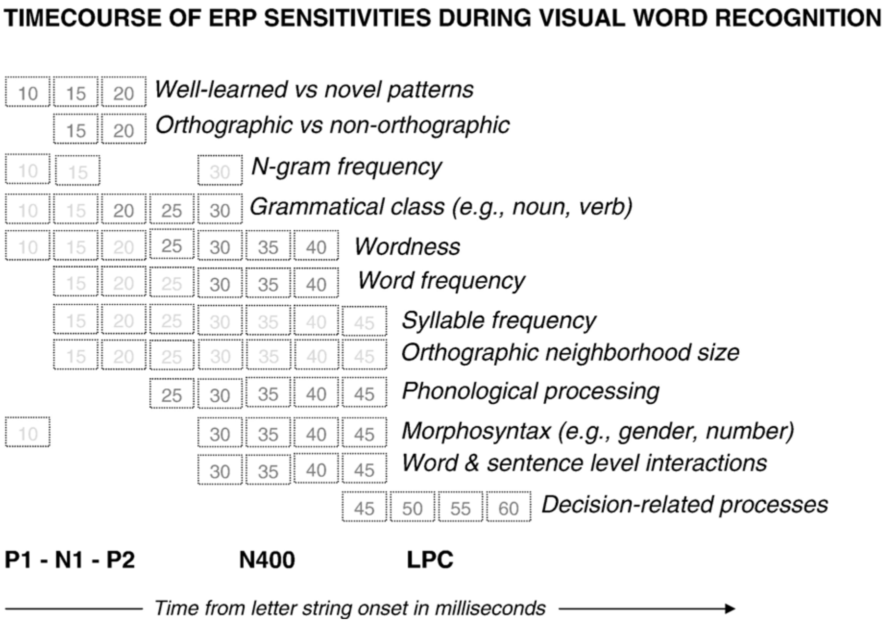

Fig. 4 – A schematic representation of the timing (in ms) of the responsivity of the amplitude of visual ERP elicited by letter strings (often although not always words) to a host of independent variables based on the published literature. These data were collected using different physical stimulus parameters, tasks, and languages; as these factors, among others, directly influence neural processing, a single table of this sort does ERP investigations of VWR a major disservice. It is also important to note that some of the effects at the same latency present with different amplitude distributions at the scalp. Nonetheless, we offer this timeline highlighting (darker shade) the latencies of reliable ERP effects which have been replicated in several laboratories but also presenting those (lighter shade) that have been reported only once or twice. Even a cursory review of the literature underscores the flexibility, adaptability, and context sensitivity of VWR processes. Accordingly, this table can only be interpreted with the accompanying text.

the primary visual cortex may at first glance seem like a bottleneck for the entry of visual information into the brain, beyond V1 visual information spreads quickly through multiple parallel pathways, each with its own representations, processes, and functions. In this way, as visually presented linguistic stimuli are being recognized, mental representations progressively emerge from the confluence of different information types and top–down influences. This general picture poses a daunting challenge to computational models of VWR—namely, incorporating the temporal constraints imposed by the empirical data together with the modulatory mechanisms that allow the VWR system to be sensitive to context and task demands for extended time epochs.

5. Electrophysiology and lateralization constraints on visual word recognition

A central issue for any cognitive model is to determine just how the anatomy of the brain constrains the functional processes it supports. One major cortical division that is unquestionably very relevant for VWR is the brain's separation into two, albeit highly interconnected hemispheres, which differ both anatomically and physiologically. VWR is thus the product of the cooperation of two somewhat similar but nonetheless different anatomical structures with different ways of processing much, if not all, of the same visual information. Neuropsychological and neuroimaging data have unequivocally demonstrated the superiority of the left hemisphere for many linguistic tasks, but there is also empirical support for right hemisphere contributions to language processing (e.g., Beeman and Chiarello, 1998). Moreover, at least some right hemispheres seem capable of word recognition, even if they are not normally responsible for or perhaps even normally involved in VWR in the intact brain (Zaidel, 2002). The specific roles of the two cerebral hemispheres as well as detailed specifications of just when and how they interact would seem to be crucial issues for any computational model of language comprehension in general and VWR in particular.

The presence of two hemispheres with different processing capabilities has largely been ignored by cognitive and computational models of VWR presumably under the assumption that, during reading, the visual information that arrives in the right hemisphere is almost immediately transferred to the left where all relevant word recognition processes are then conducted. However, neither the quick transfer of verbal information nor the superiority of the left hemisphere for some, or even all, aspects of the word recognition process necessarily nullifies parallel processing of printed words by the right hemisphere. The various split-fovea computational models that incorporate some anatomical constraints of the visual system however differ in the role they assign to each hemisphere. SERIOL, for example, assumes an immediate transfer of low level orthographic (Whitney, 2001). Shillcok and collaborators (2000), on the other hand, give the right hemisphere partial autonomy in processing lexical information. The exquisite temporal resolution of ERPs makes them especially suitable for investigating the time course of interhemispheric transfer and, in combination with other techniques, for specifying just what kinds of functions each hemisphere abets.

5.1. *Orthography*

Some of the early orthographic effects previously described appear laterally asymmetric at the scalp, most often being larger over left hemisphere than homologous right hemisphere sites. Although the relationship between the superficial topography of electrical fields and their neural generators is not straightforward given the inverse problem,[7] these kinds of ERP asymmetries have been taken in combination with other findings to reflect left hemisphere specialization for orthographic processing. Nobre and McCarthy (1994), for example, observed a larger N150 component in the left (relative to right) hemisphere in response to written words (see also Schendan et al., 1998). The N170 effect associated with orthographic (but not with non-orthographic) processing was similarly found to be larger over the left (relative to the right) posterior–temporal/occipital regions (Bentin et al., 1999; Simon et al., 2004). Additionally, the differences between orthographic and non-orthographic stimuli surface 70 ms earlier over left than right electrode recording sites, allowing for the possibility of interhemispheric transfer.

Similar early effects in MEG recordings, with their relatively better spatial resolution,[8] were localized to the left inferior temporo-occipital cortex (Salmelin et al., 1996; Kuriki et al., 1998; Tarkiainen et al., 1999, 2002). This localization is consistent with the proposed visual word form area (McCandliss et al., 2003), in which several neuroimaging studies have observed significant activations during letter processing (although its exact location and functional significance remain controversial, see for example Price and Devlin, 2003). Cohen et al. (2000), for example, recorded fMRI activity following words presented to the two visual half fields in individuals with a partial commissurotomy of the posterior portion of the corpus callosum as well as from brain intact controls. The Visual Word Form Area in the middle portion of the left fusiform gyrus was activated in the control participants regardless of which field was stimulated whereas in the posterior split-brain patients it was activated only with right visual field (left hemisphere) presentation. These activations were linked to ERP activity at around 180–200 ms over left posterior sites also present only with right visual field presentations in the patient group. Cohen et al. concluded that visual information is initially processed in temporo-occipital areas contralateral to the stimulated hemifield but is subsequently transferred to the Visual Word Form area in the

---

[7] Were the origin of an electric field inside of a sphere (namely, the head) and the characteristic conductivity of the medium (e.g., brain, cerebrospinal fluid, skull and scalp) known, it would be possible to calculate its effect over the surface. However, the calculation of the origin of electrical activity measured at the surface of a sphere (inverse problem) has no unique solution, even with an infinite number of noise-free measurements. It is thus necessary to assume specific models of the sources in order to estimate approximate solutions.

[8] Magnetic signals are less distorted than electrical fields when they propagate through the skull and other tissues although they fall off more rapidly.

left hemisphere—an area specialized for letter string processing. Applying a source analysis to their ERP data, Hauk and collaborators (2006a,b) likewise observed involvement of the occipital region in both hemispheres; at 100 ms post-word onset, an effect of word length was localized in right occipital areas and an N-gram frequency effect was localized around the Visual Word Form area in the left hemisphere.

These results are generally in accord with models that implicitly assume that visual information initially arriving in the right hemisphere is quickly transferred to the left hemisphere through the posterior portion of corpus callosum. More specifically, they support the computational architecture of SERIOL, in which suborthographic features of the left and right side of a word are initially processed in parallel in the right and left hemispheres but are soon thereafter integrated in the left hemisphere at a sublexical level (Whitney, 2001).

5.2. *Phonology*

The linguistic superiority of the left hemisphere does not seem to be restricted to early orthographic analyses. Phonological ERP effects also exhibit left lateralization at the scalp (Bentin et al., 1999; Simon et al., 2004; Proverbio et al., 2004), paralleling lateralized fMRI activation patterns during phonological tasks (Binder and Price, 2001). Taken together, these data suggest a model in which the left hemisphere is highly specialized for the sequential analysis of sublexical orthographic and phonological representations. The relative involvement of the two cerebral hemispheres in the direct processing and recognition of larger word units remains an open question (Ellis, 2004).

5.3. *Semantics*

The cerebral hemispheres also have been found to differ in their sensitivities to various types of associative and semantic relationships. Accuracy and reaction time data from visual half field studies, for example, have led to the proposal that the left hemisphere is especially sensitive to close lexical–semantic relationships (e.g., cat–dog) whereas the right hemisphere is more sensitive to a wider net of looser semantic associations (e.g., goat–dog; e.g., Chiarello, 1998). Such results are consistent with split-fovea models implying different functional organizations of semantic associations and meaning representations in the two hemispheres based on the different natures of the partial information initially available to each (Monaghan et al., 2004).

Coulson et al. (2005) examined ERPs in a priming paradigm in which the word prime first appeared in central vision and the target appeared shortly thereafter (500 ms) randomly 2° lateral to fixation in the left or right visual field (see Fig. 3). ERPs to semantically associated words elicited a greater positivity (following the early sensory potentials) relative to unassociated words with presentation to either visual field. Although a slight delay might have been predicted with left visual field presentations if the input were transferred to the left hemisphere via the callosum, the onset latencies of the priming effects in the two visual fields were statistically indistinguishable. Presentation to the two hemifields, however, did seem to unfold with a somewhat different temporal course: for targets presented in the right visual field, the priming effect was larger earlier (300–500 ms) than in the left visual field (600–800 ms). If replicable, these hemispheric differences in the degree of semantic activation over time are consistent with parallel activations and integrations of word meanings in both hemispheres, albeit likely with different processing algorithms.

5.3.1. *Conclusions*

In summary, it behooves computational modelers of VWR to attend to the processing consequences of at least certain anatomical constraints and electrophysiological researchers to provide empirical data (and ensuing inferences about functional constraints) for model testing. At minimum, these should include facts and consequences of the lateral organization of the visual system, and the relative autonomy interaction between the two hemispheres, each with its own representations and processes.

6. Electrophysiological analyses of the visual recognition of complex words

Visual word recognition is often conceptualized as a process whereby a specific set of ordered orthographic units is functionally linked to a unique conceptual representation. As language is typically construed as a combinatorial system, a critical issue has been identifying the fundamental building blocks that are to be combined. Although a lively debate regarding the basic functional units of language continues to rage among linguists, most cognitive models confer words a central role (Jackendoff, 2002). In the case of languages with alphabetic orthographies, printed words are represented by letter combinations. Letters, or letters clusters like bigrams, thus can be considered orthographic units (Grainger and Whitney, 2004). Other codes, like graphemes (usually understood as the written version of phonemes; Henderson, 1985) or sub-syllabic segments of onset, nucleus, and coda (initial consonant cluster, vowel, and final consonant cluster, respectively, e.g., “W–O–RD”; Shallice et al., 1983) also have been put forward as perceptual units. Obviously, the choice of basic units (initial and otherwise) has important implications for a model's performance. Accordingly, solid empirical evidence for the perceptual unit that is to serve as the starting point for any model of VWR is a must.

As previously noted, several computational models have a phonological level inbetween the orthographic and meaning levels, although the specific phonological implementations vary considerably in detail (Harm and Seidenberg, 2004; Coltheart et al., 2001). ERP data attest to the sensitivity of semantic analysis to phonological information during word reading and thus offer up results for testing different models and theoretical proposals. Demonstrating the psychological reality of any particular sublexical unit during reading requires evidence that words can be decomposed into these units during VWR and in turn that these units provide access to and facilitate the integration of meanings.

The bulk of computational models of VWR deal only with short (monosyllabic) words, which can be successfully modeled without any intermediate representations. The question of whether or not, and, if so, which, additional intermediate units may be needed to explain VWR more generally becomes readily apparent when models must accommodate longer and more complex multi-syllabic or polymorphemic words, the most frequent staple of most natural languages. Both syllables and morphemes have been hypothesized to be such intermediary sublexical units (Taft and Forster, 1976). Syllables are typically meaningless units of orthographic and phonological information; by contrast, morphemes are semantic units carrying both semantic and morpho-syntactic information (e.g., the morpheme "–S" in the word "CATS" gives semantic information about quantity; by contrast, the syllable "-DOW" in the word "WINDOW" only refers to its phonology). Within most PDP models, these intermediate levels of representation are not built in as a separate representation as they are considered emergent properties that arise from regularities in the input language as analyzed by the network (brain surrogate). The repeated co-occurrence of specific letter sequences in a particular language, for example, presumably makes the processor sensitive to certain orthographic and phonologic patterns. And, since the transition probabilities for letter sequences that coincide with morphologic or syllabic boundaries tend to be on average lower than transition probabilities among letters internal to a syllabic or morphological unit, the probability of bigram frequencies might account for various effects that have been attributed to intermediary sublexical units (Seidenberg, 1987).

6.1. *Syllables*

There is considerable empirical support primarily from reaction time data but also to some extent from electrophysiological data for the claim that syllables may be a critical interface between sublexical orthography and sublexical phonology during VWR (see Carreiras and Grainger, 2004 for review of behavioral evidence). Barber et al. (2004) investigated the combined effects of initial syllable frequency and word frequency on lexical decision for written Spanish words (a language with clear syllable boundaries). The probabilities of bigram transitions, orthographic neighborhood size, and syllable–morpheme overlap were controlled and thus eliminated as potential confounds. The N400 frequency effect – smaller N400 amplitudes for high- than low-frequency words – was replicated. At the same time, a reversed syllable frequency effect was observed: negative amplitudes were larger for words beginning with high than low-frequency syllables. Moreover, the syllable frequency effect was apparent earlier (150–200 ms) than the word frequency effect (300–350 ms), suggesting different time courses for these two variables during VWR. Syllable frequency effects of this type are assumed to reflect early activation of syllables which in turn activate a cadre of lexical candidates that share a common first syllable, which thereafter compete in an interactive activation fashion (e.g., McClelland and Rumelhart, 1981). The ERP effect thus could reflect either these activation–inhibition processes or attempts by the system to integrate any of those partially activated meanings into the context. This syllable frequency ERP effect was replicated in German (a language with a less clear syllabic structure); the first syllable effect at around 180 ms preceded the lexicality effect (words versus pseudowords) at around 350 ms by about 170 ms (Hutzler et al., 2004).

In a different paradigm, decisions were made about Spanish words and pseudowords made up of letters written in two different colors; for some words, the color boundary was coincident with the word's syllabic boundary (match) whereas for others the color–syllable boundaries were misaligned (mismatch). The ERP data showed color–syllable congruency effects between 180 and 260 ms for both pseudowords and low-frequency words, but not for high-frequency words (Carreiras et al., 2005). The absence of a syllable effect for high-frequency words accords well with the assumptions of the MTM model of polysyllabic word recognition (Ans et al., 1998), according to which only low-frequency words and pseudowords are decomposed in syllables if and when global word recognition processes fail. The MTM model, however, cannot simulate the inhibitory effects of high-frequency syllables reported in Barber et al. (2004) because it does not include a competition mechanism of lateral inhibition among syllable representations. This would seem to require, for instance, that a localist connectionist network include an intermediate level of representation. On the other hand, these data do not rule out the possibility that activation of word cohorts sharing the same syllable arises from orthographic redundancy, which could explain the inhibitory effect. In that case, a connectionist network without local representations could account for the frequency syllable effect. Quite reasonably, computational modelers are loath to modify their models of VWR without indisputable evidence for an actual rather than merely emergent level of syllabic representations. Such work would benefit greatly from extension to languages with deep orthographies and explicit investigations of the orthographic or phonological nature of the proposed syllabic representations.

6.2. *Morphemes*

Similar issues arise with regard to morphological decomposition during VWR. Words that are composed of more than one morpheme are likely to be represented neurally not just by single whole-word units but also decomposed into their constituent morphemes. Stems and affixes, for example, could be separately stored, which some researchers then believe would necessitate a set of combinatorial rules during word recognition. On such a view, morphemes also could serve as access points to a word representation during VWR. And, indeed, different dual-route models have been proposed to explain recognition of morphologically regular and irregular words. On such models, irregular words are listed within the mental lexicon in their full form, whereas only the stems of regular words are stored (Schreuder and Baayen, 1995; Pinker and Ullman, 2002). By contrast, many connectionist models do not build in any *a priori* role for word morphology since both morphemes and rules are presumed to become implicitly represented in the connection weights during network training (Rumelhart and McClelland, 1986; McClelland and Patterson, 2002).

Differences between the ERPs to morphosyntactic violations on regular and irregular words have fairly consistently been found. Regularization of German nouns with irregular plural forms, for example, was associated with a left anterior negativity starting at 200 ms, whereas incorrect number inflections of regular words produced a more centrally distributed effect (Weyerts et al., 1997). A similar dissociation was observed when violations produced by incorrectly inflected regular and irregular verbs were compared, in German (Penke et al., 1997) and in different Romance languages (Gross et al., 1998; Rodríguez-Fornells et al., 2001). These data have been taken to support dual process models in which irregular words are stored as a whole but regular words are decomposed into their stems and affixes. Whereas the data are consistent with this interpretation, they do not necessarily warrant it.

The representational structure of morphologically complex words also has been investigated. In one experiment, regular and irregular German participles were either presented twice (repetition priming) or were preceded by their infinitival form (morphological priming). For regular participles, both types of primes produced similar ERP priming effects; by contrast, irregular participles showed no morphological priming in the ERP (Weyerts et al., 1996). In a comparable study in English, controlling for formal priming, verb stems were preceded either by their past tense form or by the past tense form of an unrelated verb (Münte et al., 1999). Only regular pairs (e.g., start–started) were associated with a significant reduction in N400 amplitude; irregular verbs (e.g., think–thought) showed no such morphological priming. A similar pattern of results was obtained in Spanish, a language with a different verb inflectional morphology (Rodríguez-Fornells et al., 2002). Although the ERP data suggest differential treatment of regular and irregular word forms at a neural level, consistent with a dual process account, they do not yet constitute strong evidence against connectionist models which show that such processing differences between regular and irregular word forms could emerge in principle from different contributions (weightings) of the phonological and semantic components in a single-mechanism network (Seidenberg and Gonnerman, 2000).

In a different type of priming paradigm, a morphological characteristic of some Spanish words, namely, grammatical gender was used to investigate whether words are decomposed into a stem and a gender affix. In Spanish, certain noun pairs can be distinguished by their gender markings, even though they are neither semantically nor morphologically related (e.g., rata–rato: *rat–moment*). These so-called stem homographs – despite being two different morphemes – have identical orthographic and phonologic representations. Compared to unrelated words, both morphologically related word pairs (NIÑA–niño: *boy–girl*) and stem homographs (RATA–rato) showed a similar reduction in N400 amplitude between 250 and 450 ms followed by a greater negativity for stem homographs relative to morphologically related words (Barber et al., 2002). Subsequent experiments ruled out strictly orthographic similarity or rhyme mismatching as the sole causes of the priming effects for the stem homographs (Domínguez et al., 2004). Instead, the initial similarity between morphological and stem homographic ERP priming effects was taken to reflect the application of the same segmentation rule to the same item in both cases. The subsequent negativity was hypothesized to reflect the morphological parsing failure resulting from this procedure for stem homographs, i.e., a type "of lexical garden-path effect".

Two studies in English with isolated words add to the claim that written words are decomposed into their morphemes during word recognition. In a lexical decision task, pronounceable pseudowords containing no English morphemes (e.g., *flermuf*) were found to elicit large N400s in contrast to pseudowords formed by combining a prefix and a relatively non-productive bound stem (e.g., in-ceive), which produced a smaller N400 that was indistinguishable from the response to real words (McKinnon et al., 2003). Finally, an MEG experiment reported an inhibitory effect of morphological family size (number of morphological derivates associated with a root) on the peak latency of the M350 component. This delay in response to words with larger morphological families was taken to reflect competition among all the forms derived from the same root, all of which would have been activated initially (Pylkkänen et al., 2004). The authors argued for a cohort model of morphological activation similar to that invoked to account for phonologic effects (Pylkkänen et al., 2002) and the ERP syllable neighborhood effect (Barber et al., 2004).

6.3. *Conclusions*

In summary, during visual word recognition, the brain extracts various types of information presumed to be characterizing word representations. To that end, the visual (or word) processing system could segment words into a variety of sublexical units, associated with different information types (phonological, syllabic, morphological). The question is does it, and if so, into which units and under what circumstances. Various subunits could trigger the initial activation of many potential meanings followed by competition and inhibition leading to the unique activation of the correct representation of a given printed word within a specific context. Models based on distributed representations have shown that many of the behavioral phenomena that support such a vision of VWR could naturally emerge from structure in the input—regularities of various sorts (letter combinations, phonological, syllabic, and morphological), even if these information types are not explicitly (or a priori) represented as distinct sublexical level representations. This controversy raises a challenge for both computational modelers and electrophysiological researchers: how best to simulate word recognition phenomena that would seem to require various intermediate orthographic and semantic representations and how best to refine experimental designs and stimulus controls in order to eliminate possible confounds due to superficial word stimulus characteristics.

Ultimately, the resolution of questions about what types of information are used, in what combinations, when, and how at a mechanistic level in the service of VWR will require comparative studies of reading in different languages, including those with different orthographies. It is not unreasonable to expect that the brain has arrived at slightly different solutions for the different instantiations of the "same" reading problem (e.g., decoding letters versus pictograms). Indeed, when word recognition processes were examined in a group of Chinese–English bilinguals, the pattern of ERPs to various word contrasts in the two languages shared some general commonalities but were also characterized by some between language differences in latency, amplitude, and scalp distribution of the relevant components and effects (Liu and Perfetti, 2003). Moreover, across-language differences are unlikely to be limited to differences in the physical characteristics of the stimuli in the languages being compared. However, even if they were, the processing of other types of word information might depend, at least in part, on language-specific features. Thus, some of the reported effects for various intermediate level units could well be language-specific. For example, it has been suggested that, in languages (like French) where there is an inconsistent mapping between orthography and phonology, only sublexical output phonology and not sublexical input phonology may be structured syllabically, with the main functional phonological unit in recognition process being the phoneme (Ferrand et al., 1996). Clearly, investigations of the different functional units of VWR must consider different languages.

A case in point is Ziegler and Goswami's (2005) integration of findings from the literatures on phonological development, reading acquisition, and dyslexia to develop a theoretical framework of reading that cuts across languages (the psycholinguistic grain size theory of reading). Their primary claim is that the size of the fundamental reading unit (i.e., graphemes, phonemes, syllables, onsets/rimes, etc.) within any given language is determined by the phonological characteristics of that language because reading development is grounded in phonological awareness. They propose that the degree of consistency in the mapping between phonology and orthography within a language is the main determinant of the nature (specifically, grain size) of lexical representations as well as accompanying reading strategies and the nature of the associated reading problems (dyslexia) within that particular language. Moreover, since skilled reading depends on reading development, the relative importance of (or reliance on) different sublexical representations across languages is also likely to vary. This is an excellent example of how a close collaboration between experimental psychology and modeling combined with cross-linguistic research can lead to integrative models that may help to reveal universal principles of reading and their specific implementations across languages.

7. Interplay between electrophysiology and computational models

7.1. *Mutual constraints in principle and in practice*

We are not the first to suggest the mutual benefits that computational modelers and cognitive neuroscientists might derive from joining forces to analyze VWR. Indeed, in their open commentary on an early overview of neuroimaging data (Posner and Raichle, 1995), Jacobs and Carr (1995) wrote of two important ways in which functional neuroimaging can be used to constrain computational models of cognition: (1) by providing information about the anatomical loci of different subprocesses and hence system decomposability and (2) by delineating the temporal dynamics of the cognitive process (es) under investigation. Electrophysiological data have proven useful to VWR investigations in both these regards. Even with their limited spatial resolution, ERP data have placed constraints on the anatomical organization of all reading models by revealing, for example, that the temporal course of word processing and the use of sentential context in the two hemispheres differ (Coulson et al., 2005). ERP measures also have been combined with other techniques to provide evidence on localization of various VWR processes (Cohen et al., 2000). The prospects not only for more but for more spatially precise contributions of this sort seem especially good given the recent proliferation of analytic tools for localizing the sources of electromagnetic signals and for combining data across different imaging modalities (EEG, MEG, fMRI, Transcranial Magnetic Stimulation or TMS). By its nature, electrophysiological data are even better suited to contribute important information about the timing of reading process, and we have reviewed a growing body of relevant ERP data along with some specific proposals (Sereno and Rayner, 2003; Hauk et al., 2006a,b). The extant computational models are, in principle, able to simulate process dynamics and event sequences, for example, in activation levels of different processing cycles or in settling times in the case of attractor networks. This would then seem to be an especially promising avenue for future explorations.

In addition to these two ways in which neuroimaging studies can help to constrain computational models, we have highlighted some qualitative principles of neural processing derived from the electrophysiological data that we believe also should be considered in any realistic model of VWR; e.g., evidence of massive interactivity and the fast confluence of information during the VWR process, as well as the flexibility of the reading system and its dependency on context and previous states. We noted some of the conclusions that have been derived from these principles, such as a clear lack of support for the traditional distinction between lexical and post-lexical processing (Coulson and Federmeier, in press).

Jacobs and Carr (1995) also speculated about how computational models might contribute to cognitive neuroscience, for example, by generating specific predictions such as about levels of brain activation. In Section 4.1, we described two clear examples of this (Holcomb et al., 2002; Braun et al., 2006). Cross-fertilization between cognitive electrophysiology and computational modeling thus has already proven feasible in both directions. There are many levels at which one approach may influence thinking and/or research in the other, ranging from very general and qualitative to very specific and quantitative: in some cases, results in one field may only loosely constrain the other, in other cases, the mutual interplay may be very explicit and direct, as in the generation of new and testable hypotheses, or in simulation of specific electrophysiological effects. Although we cannot foresee all the fruits of such a union, we believe that all the available signs portend sufficient success to make it worth the time and effort.

Dehaene and collaborators (2005) have similarly argued that cognitive models of VWR must be rooted in what is known about the neural bases of perception and categorization. They specifically highlight some properties of the visual cortex for object recognition which they propose must likewise be relevant for any model of VWR. One such property is the hierarchical organization of the receptive fields of different neurons, which results in a progressive increase of the complexity of the preferred features of different populations along the occipito-temporal ventral pathway. This organization affords the integration of lower level features into more abstract, higher-level representations and is governed by perceptual learning rules. Dehaene et al. argue that the perception and recognition of orthographic stimuli are subject to the same principles of hierarchical organization and plasticity as the perception and recognition of other types of learned categories; these processes thus must unfold with units specialized for bars of various orientations, letter fragments, letter shapes, abstract orthographic representation (e.g., independent of the spatial location or the font type), bigrams, and other intermediate units. However, as language (and by association true reading) is a uniquely human skill, noninvasive neurophysiological techniques are unavoidable if formal proposals of reading are to be successfully integrated with neuroscientific data.

7.2. Three examples of cross-talk among methodologies

7.2.1. ERPs and cognitive modeling: localizing the bottleneck in dual task paradigms

Some excellent examples of the fruitful marriage of cognitive models and electrophysiological research can be found in other cognitive domains. Across a number of studies, different ERP components (e.g., N1/P2, P3, N400, and LRP) linked to qualitatively different processes (sensory/perceptual/attentional processes, categorization processes, semantic processes, and response preparation, respectively) have been used to test alternative hypotheses about the loci of dual task interference in the three different dual task paradigms: the psychological refractory paradigm (PRP), the encoding-speeded response (ESR) paradigm, and the attentional blink (AB) paradigm. The PRP effect refers to the increase in RT to the second of two successive stimuli requiring responses as the interval between them is decreased. The ESR effect refers to a similar effect even when the first of the two stimuli is masked and requires an unspeeded much delayed response. The attentional blink refers to an impairment in accuracy for identifying the second of two target stimuli appearing 400–600 ms after detection of the first target. Only slightly delayed P3s together with substantial and commensurate delays in the LRP and RT at short intervals compared to long ones in the PRP paradigm situate the locus of interference consequent to stimulus identification and categorization but before response selection (Osman and Moore, 1993; Luck, 1998). By contrast, P3 delays together with delayed RTs in the ESR paradigm situate the locus of interference somewhere during the processes of stimulus identification (Arnell et al., 2004). Normal early sensory potentials and semantic modulation of N400 amplitude together with completely suppressed P3 potentials during the attentional blink localize AB interference to the stage of working memory encoding (Luck et al., 1996; Vogel et al., 1998). ERP data thus have unequivocally demonstrated that dual task interference effects can arise at different stages depending on the nature of the task, and in so doing have helped to refine the cognitive models of dual task processing.

7.2.2. ERPs and computational models: error processing

ERP data have been similarly informative when combined with computational models in certain domains. Yeung et al. (2004), for example, built a model of error detection and conflict monitoring from a combination of computational simulations and ERP data. They proposed an account of the so-called error-related negativity (ERN or Ne)—a potential seen over fronto-central sites following the commission of an error (or negative feedback). Their computational simulation was not only able to integrate some apparently contradictory ERN findings, but perhaps even more importantly, predicted a previously unreported ERP effect (N2) that was subsequently confirmed. Whether or not Yeung et al.'s conflict model of the ERN turns out to be wholly or even partially correct, this study is a prime example of the type of mutual interplay between computational modeling and cognitive electrophysiology that we espouse and aim to encourage with this review.

7.2.3. Computational models of eye movements and reading

What we espouse is cross-talk similar to that between VWR modelers and researchers who record eye movements as a primary dependent measure. Certain of these models, for example, aim is to "explain" specific subparts of the reading process such as the various patterns of eye movements of readers scanning a text, including saccade generation, landing positions, fixation times, and regressive saccades; these are known to be determined by the physical and linguistic characteristics of words (reviews in Reichle et al., 2003; Reichle, 2006). Some of these computational models have incorporated relevant physiological properties of the oculomotor system. The SWIFT model (Engbert et al., 2005), for instance, embodies the separation between the "where" and "when" pathways in the brainstem via temporal and spatial control of saccades with little interaction. As in some models of VWR, anatomical constraints derived from the split-fovea proposal have also been implemented in a computational model of eye movement control (McDonald et al., 2005). Reichle et al. (2003) offer some tentative speculations of how their influential A–Z model of oculomotor control might be implemented in the brain. Eye movement control is situated within the more general process of reading, which involves the cooperation of different cortical and subcortical brain regions. Cortical areas involved in VWR presumably work together with anterior and posterior attention networks, as well as with the oculomotor system including the pulvinar nucleus, frontal eye fields, superior colliculus, and brainstem. Clearly, it is assumed that those computational models of eye movements will eventually be integrated into more general brain models of reading.

7.3. What electrophysiological measures to model?

A reasonable question that arises for modelers who are convinced that there is some value in modeling electrophysiological patterns associated with VWR is how to simulate patterns of electromagnetic activity. Most extant VWR models were developed to simulate behavioral patterns—primarily, error percentages and response latencies. There is, however, no reason in principle or practice that such models could not be adapted to simulate other dependent variables such as

EEG power (in various frequency bands) and coherence, ERP amplitudes, and especially onset, peak, or duration latencies, either alone or in combination with each other as well as coincident behavioral (and eye movement) measures, when available; in the near future, there may also be value in modeling ERP "components" derived from various data reduction and/or decomposition techniques such as principal components analysis, independent components analysis, and wavelet analysis. Naturally, the choice of the best measure will depend on the specific VWR/reading phenomena to be simulated (after piloting), but all of them are potentially useful in analyzing the nature and time course of important VWR processes. Unlike a reaction time which is an endproduct measure that reflects the sum of individual contributions from virtually every stage of processing leading up to the overt response, the ERP provides a continuous measure of processing between the input and the response. Moreover, to the extent that different ERP effects are shown to index specific processes, these different effects can be used to isolate the effects of various experimental manipulations to specific stages of processing (e.g., certainly stimulus versus response process, but even more precisely, pre- or post-perceptual, conceptual, working memory consolidation, or response selection and execution processes as discussed above). As ERPs are a direct measure of neural activity and reflect the membrane currents at that moment in time, the latency of a reliable difference between any two ERP waveforms at the same recording site can be used as an upper limit on the time by which the brain must have appreciated the difference(s) between the stimuli eliciting them; ERP onset latencies (along with peak latencies and duration latencies) thus can be especially useful in specifying when certain variables influence neural processing (whether or not they are reflected in reaction times) and in delineating the time course of VWR processes. A continuous measure also provides the flexibility needed to measure the potential flexibility of reading processes themselves in the face of different contexts.

ERPs at the scalp surface are presumed to reflect primarily the synchronous synaptic activity from a large number (at least 10,000 or so) pyramidal cells that have spatially aligned current flow acting in synchrony within patches of neocortex (see Kutas and Dale, 1997). These synaptic computations (10,000–20,000 synapses/cell generating synaptic potentials that are subject to spatial and temporal summation) are the basis of important cognitive computations. Amplitude changes without concomitant changes in spatial distribution across two or more conditions are often taken as engagement of the same neural process, albeit differing in strength or degree; these may reflect smaller post-synaptic potentials in the same set of neurons, smaller number of active neurons, or less temporal synchrony (see Rugg and Coles, 1995). Moreover, statistically reliable between-condition differences in the distribution of scalp potentials show incontrovertibly that corresponding neural generators do differ *somehow* with respect to their location, polarity, or intensity—in some combination of location, polarity, and relative or overall strength (Urbach and Kutas, 2002). Our understanding about the exact relationship between neurobiological processes and information processing is still relatively limited; indeed, specifying these relationships is a major goal of the ambitious program of cognitive electrophysiology, but past research and present models underscore the value of using electrophysiological signals as markers of different cognitive processes and their unparalleled utility in delimiting the temporal locus of processing bottlenecks, for example.

7.4. *What electrophysiological effects to believe?*

As the old adage goes, "garbage in, garbage out". Unarguably, then, a prerequisite for any model, whether cognitive or computational, is the availability of clean and reliable data to model. Skeptical modelers might find some of the ERP findings on VWR reviewed herein to be confusing and/or inconsistent as compared with the more extensive corpus of behavioral findings. We have suggested that at least part of the variability of the results across experiments is a natural consequence of the intrinsic characteristics of the brain as a cognitive system: its sensitivity to the context, as well as to previous and present states, expectancies, and goals. Viewed from this perspective, the apparent lack of any single temporally fixed set of VWR subprocesses, far from undermining electrophysiological evidence, raises, perhaps rightfully, questions about the flexibility of VWR processes—questions that then need to be addressed in a computational framework as well. Other apparent inconsistencies are better understood considering the nature of the techniques, for instance, the fact that some authors report orthographic effects in the N400 component while others found semantic effects in much earlier time windows. The likelihood of picking up electromagnetic activity on the scalp that is related to any particular cognitive event depends on both temporal and anatomical considerations (Münte et al., 2000). Indeed, many neuronal events may not be detectable at the scalp, and this could vary with circumstances and across individuals. Viewed from this perspective, it should not be so surprising that some experiments fail to detect early effects that actually transpire and have observable processing consequences in later time windows. As Hauk et al. (2006a,b) rightly pointed out, averaging per se can lead to the loss of information in some cases. Early effects are usually smaller in amplitude and of shorter duration than later effects (which can last for hundreds of milliseconds), so they are more vulnerable to variability across items and/or subjects. Finally, it is important to bear in mind that, in the scheme of things, electrophysiological data are new players in the VWR field; there are many more studies using behavioral than electrophysiological measures, although this gap is closing with the rapid increase of ERP studies of word processing in the recent past. As in any research area riddled with inconsistent data, there is a need for replication as well as a methodological and/or theoretical account of why replication fails when it does. These often include a specification of the relevant variables and take non-negligible amounts of research (often years). In this process, some ERP results will indeed fall by the wayside. Nonetheless, we feel that, far from waiting until the dust settles, computational models should (because they can) play a more central role in the integration and organization of the growing body of empirical neurophysiological data on VWR.

7.5. *Summary*

In summary, while our desire to see a bidirectional flow between empirical data to computational models is not new, the data we offer are, and we believe that the growing empirical database reviewed herein, underscores the benefits of the (relatively costly at least by comparison to behavioral methods) ERP technique in helping VWR models to meet the criterion of biological realism. We have shown how electrophysiological research can be useful in addressing questions that have been traditionally studied with behavioral techniques, such as the effect of linguistic regularities, the role of phonological processing, or the possible representation of different orthographic, phonological, and semantic units. In addition, these techniques are based on direct measures of cerebral activity, and consequently contribute information about anatomical and physiological constraints, as well as add the critical temporal dimension to the description of the reading process. On the other hand, electrophysiological researchers clearly will benefit from using computational models in the generation and design of future experiments and as a tool for organizing and interpreting the resulting data.

8. General conclusions

Reading is a very complex, integrated set of perceptual, cognitive, and motor skills that most (although not all) literate adults carry out with relative ease. Successful reading relies on attentional mechanisms, sensory–perceptual analyses, working memory and long-term memory processes, as well as eye movements (Coltheart, 1987; Klein and McMullen, 1999). A central issue in the study of reading is to understand how written words activate mental representations in the brain via the visual system, as a result of the orchestration of these various processes. Clearly, a full understanding will require due consideration to the biological structures in which the requisite computations transpire and their temporal courses.

Several computational models have been remarkably successful in simulating various aspects of human VWR. These models can be divided in two main types depending primarily upon whether they use local versus distributed representations and the number of intermediate levels represented. These models are now faced with incorporating reading-relevant biological constraints and of recognizing not just simple words but the more common complex (polymorphemic or polysyllabic) words in languages.

Although reading is not a genetically transmitted cognitive ability, it seems to involve the progressive specialization of specific brain areas. It has thus been possible to associate activity in different brain regions during reading with initial perceptual analysis, orthographic decoding, and phonological processing. Electrophysiological data in particular suggest that VWR is not an encapsulated, fixed process but rather a highly flexible, dynamic system of processes that can adapt its temporal course to the available context and task demands at hand. We have argued that this time course and its flexibility have relevant inescapable implications for any computational model of VWR. Moreover, some of the brain areas essential for reading are strongly left lateralized, although there is also support for some right hemisphere involvement. The functional consequences of these anatomical constraints need to be simulated by any viable model of VWR in processing words with one or more syllables and/or morphemes. To the extent that the ERP data continue to evidence the decomposition of complex words during the recognition process, modelers will be compelled to modify existing computational models so as to simulate such intermediate representations.

Naturally, before modelers can be expected to substantially alter their VWR models, psychologists and neuroscientists – especially (although not exclusively) those using electrophysiological and electromagnetic measures – will need to amass a solid corpus of reliable empirical data that modelers can use as specifications for their models. As outlined herein, neurophysiological measures have the temporal resolution to help resolve outstanding issues regarding the temporal course of the various reading-related processes, the nature of the various sublexical units involved, the differential roles of the two cerebral hemispheres, and the various dual route accounts, among others. We suggest that it is time that computational modelers and neurophysiologists come together in practice and in theory to unravel the mysteries of reading.

Acknowledgments

This work was supported by the postdoctoral fellowship EX2003-012 (Secretaría de Estado de Educación y Universidades of the Spanish Government and the European Social Fund) to H.A.B.; and the grants HD22614 and AG08313 to M.K. M. Kutas was a Lady Davis Fellow at Hebrew University during the writing of this article. We thank Jeff Elman and two anonymous reviewers for their helpful comments.

REFERENCES

Allison, T., McCarthy, G., Nobre, A., Puce, A., Belger, A., 1994. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. Cerebral Cortex 4, 544–554.

Ans, B., Carbonnel, S., Valdois, S., 1998. A connectionist multiple-trace memory model for polysyllabic word reading. Psychological Review 105, 678–723.

Arnell, K.M., Helion, A.M., Hurdelbrink, J.A., Pasieka, B., 2004. Dissociating sources of dual-task interference using human electrophysiology. Psychonomic Bulletin and Review 11 (1), 77–83.

Assadollahi, R., Pulvermüller, F., 2001. Neuromagnetic evidence for early access to cognitive representations. NeuroReport 12 (2), 207–213.

Assadollahi, R., Pulvermüller, F., 2003. Early influences of word length and frequency: a group study using MEG. NeuroReport 14, 1183–1187.

Barber, H., Carreiras, M., 2003. Integrating gender and number information in Spanish word pairs: an ERP study. Cortex 39, 465–482.

Barber, H., Carreiras, M., 2005. Grammatical gender and number agreement in Spanish: an ERP comparison. Journal of Cognitive Neuroscience 17 (1), 137–153.

Barber, H., Domínguez, A., De Vega, M., 2002. Human brain potentials indicate morphological decomposition in visual word recognition. Neuroscience Letters 318, 149–152.

Barber, H., Vergara, M., Carreiras, M., 2004. Syllable-frequency effects in visual word recognition: evidence from ERPs. NeuroReport 15 (3), 545–548.

Beeman, M., Chiarello, C. (Eds.), 1998. Right Hemisphere Language Comprehension: Perspectives from Cognitive Neuroscience. Lawrence Erlbaum Associates, Mahwah, NJ.

Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials, lexical decision and semantic priming. Electroencephalography and Clinical Neurophysiology 60, 343–355.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M.H., Echallier, J.F., Pernier, J., 1999. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. Journal of Cognitive Neuroscience 11, 35–60.

Binder, J.R., Price, C.J., 2001. Functional imaging of language. In: Cabeza, R., Kingstone, A. (Eds.), Handbook of Functional Neuroimaging of Cognition. MIT Press, Cambridge, pp. 187–251.

Bowers, J.S., 2002. Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. Cognitive Psychology 45, 413–445.

Braun, M., Jacobs, A.M., Hahne, H., Ricker, B., Hofmann, M., Hutzler, F., 2006. Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. Brain Research 1073–74, 431–439.

Brown, C.M., Hagoort, P., ter Keurs, M., 1999. Electrophysiological signatures of visual lexical processing: open- and closed-class words. Journal of Cognitive Neuroscience 11 (3), 261–281.

Brysbaert, M., 1994. Interhemispheric transfer and the processing of foveally presented stimuli. Behavioral Brain Research 64, 151–161.

Bullier, J., Nowak, L.G., 1995. Parallel versus serial processing: new vistas on the distributed organization of the visual system. Current Opinion in Neurobiology 5, 497–503.

Carr, T.H., 2005. On the functional architecture of language, and reading: trade-off between biological preparation, and cultural engineering. In: Mayr, U., Awh, E., Keele, S.W. (Eds.), Developing Individuality in the Human Brain: A Tribute To Michael I. Posner. American Psychological Association.

Carreiras, M., Grainger, J., 2004. Sublexical representations and the 'front end' of visual word recognition. Language and Cognitive Processes 19 (3), 321–331.

Carreiras, M., Álvarez, C.J., de Vega, M., 1993. Syllable frequency and visual word recognition in Spanish. Journal of Memory and Language 32, 766–780.

Carreiras, M., Vergara, M., Barber, H., 2005. Early ERP effects of syllabic processing during visual word recognition. Journal of Cognitive Neuroscience 17 (11), 1803–1817.

Chiarello, C., 1998. On codes of meaning and the meaning of codes: semantic access and retrieval within and between hemispheres. In: Beeman, C.C.M., Chiarello, C. (Eds.), Right Hemisphere Language Comprehension: Perspectives from Cognitive Neuroscience. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 141–160.

Churchland, P., Sejnowski, T.J., 1992. The Computational Brain (Computational Neuroscience). The MIT Press.

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

Coltheart, M. (Ed.), 1987. Attention and Performance XII: Psychology of Reading. Erlbaum, London.

Coltheart, M., Curtis, B., Atkins, P., Haller, M., 1993. Models of reading aloud: dual-route and parallel-distributed-processing approaches. Psychological Review 100, 589–608.

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, 204–256.

Compton, P.E., Grossenbacher, P., Posner, M.I., Tucker, D.M., 1991. A cognitive–anatomical approach to attention in lexical access. Journal of Cognitive Neuroscience 3, 304–312.

Conrad, M., Jacobs, A.M., 2004. Replicating syllable frequency effects in Spanish in German: one more challenge to computational models of visual word recognition. Language and Cognitive Processes 19, 369–390.

Coulson, S., Federmeier, K.D., in press. Words in context: ERPs and the lexical/postlexical distinction. Journal of Psycholinguistic Research.

Coulson, S., Federmeier, K.D., Van Petten, C., Kutas, M., 2005. Right hemispheric sensitivity to word- and sentence-level context: evidence from event-related brain potentials. Journal of Experimental Psychology: Learning, Memory and Cognition 31 (1), 129–147.

Dambacher, M., Kliegl, R., Hofmann, M., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. Brain Research 1084 (1), 89–103.

Deacon, D., Dynowska, A., Ritter, W., Grose-Fifer, J., 2004. Repetition and semantic priming of nonwords: implications for theories of N400 and word recognition. Psychophysiology 41, 60–74.

Dehaene, S., 1995. Electrophysiological evidence for category-specific word processing in the normal human brain. NeuroReport 6, 2153–2157.

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.

Deutsch, A., Bentin, S., 2001. Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and Eyes Movements. Journal of Memory and Language 45, 200–224.

Domínguez, A., de Vega, M., Barber, H., 2004. Event related potentials elicited by morphological, homographic and semantic priming. Journal of Cognitive Neuroscience 16 (4), 598–608.

Ellis, A.W., 2004. Length, formats, neighbours, hemispheres and the processing of words presented laterally or at fixation. 88, 355–366.

Ellis, A.W., Young, A., 1988. Human Cognitive Neuropsychology. Erlbaum, London.

Elman, J.L., 2004. An alternative view of the mental lexicon. Trends in Cognitive Sciences 8 (7), 301–306.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., Marantz, A., 2001. A magnetoencephalographic component whose latency reflects lexical frequency. Cognitive Brain Research 10 (3), 345–348.

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.

Federmeier, K.D., Segal, J.B., Lombozo, T., Kutas, M., 2000. Brain responses to nouns, verbs and class-ambiguous words in context. Brain 123, 2552–2566.

Ferrand, L., Segui, J., Grainger, J., 1996. Masked priming of word and picture naming: the role of syllabic units. Journal of Memory and Language 35, 708–723.

Friederici, A.D., 1995. The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. Brain and Language 50, 259–281.

Frost, R., 1998. Toward a strong phonological theory of visual word recognition: true issues and false trials. Psychological Bulletin 123, 71–99.

Grainger, J., Jacobs, A.M., 1996. Orthographic processing in visual word recognition: a multiple read-out model. Psychological Review 103, 518–565.

Grainger, J., Jacobs, A.M. (Eds.), 1998. Localist Connectionist Approaches to Human Cognition. Erlbaum, Mahwah, NJ.

Grainger, J., Whitney, C., 2004. Does the huamn mnid raed wrods as a wlohe? Trends in Cognitive Sciences 8 (2), 58–59.

Gross, M., Say, T., Kleingers, M., Clahsen, H., Münte, T.F., 1998. Human brain potentials to violations in morphologically complex Italian words. Neuroscience Letters 241, 83–86.

Grossi, G., Coch, D., Coffey-Corina, S., Holcomb, P.J., Neville, H.J., 2001. Phonological processing in visual rhyming: a developmental ERP study. Journal of Cognitive Neuroscience 13, 610–625.

Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. Journal of Cognitive Neuroscience 12 (4), 556–568.

Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. Journal of Cognitive Neuroscience 11, 194–205.

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography: theory, instrumentation, and applications to non-invasive studies of the working human brain. Reviews of Modern Physics 65, 413–497.

Harm, M.W., Seidenberg, M.S., 2004. Computing the meanings of words in reading: division of labor between visual and phonological processes. Psychological Review 111 (3), 662–720.

Hauk, O., Pulvermüller, F., 2004. Effects of word length and frequency on the human event-related potential. Clinical Neurophysiology 115, 1090–1103.

Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F., Rogers, T.T., 2006a. [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 ms. ERP correlates of orthographic typicality and lexicality in written word recognition. Journal of Cognitive Neuroscience 18 (5), 818–832.

Hauk, O., Davis, M.H., Ford, M., Pulvermüller, F., Marslen-Wilson, W.D., 2006b. The time course of visual word recognition as revealed by linear regression analysis of ERP data. NeuroImage 30 (4), 1383–1400.

Henderson, L., 1985. On the use of the term “grapheme”. Language and Cognitive Processes 2, 135–148.

Hinton, G.E., Shallice, T., 1991. Lesioning an attractor network: investigations of acquired dyslexia. Psychological Review 98, 74–95.

Holcomb, P.J., Grainger, J., O’Rourke, T., 2002. An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. Journal of Cognitive Neuroscience 14, 938–950.

Hutzler, F., Bergmann, J., Conrad, M., Kronbichler, M., Stenneken, P., Jacobs, A.M., 2004. Inhibitory effects of first syllable-frequency in lexical decision: an ERP study. Neuroscience Letters 372, 179–184.

Jackendoff, R., 2002. Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press.

Jacobs, A.M., Carr, T.H., 1995. Mind mappers and cognitive modelers: Toward cross-fertililzation. Behavioral and Brain Sciences 18 (2), 362–363.

Jacobs, A.M., Rey, A., Ziegler, J.C., Grainger, J., 1998. MROM-P: an interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In: Grainger, J., Jacobs, A.M. (Eds.), Localist Connectionist Approaches to Human Cognition. Lawrence Erlbaum Associates Inc., Mahwah, NJ.

Joanisse, M.F., Seidenberg, M.S., 1999. Impairments in verb morphology after brain injury: a connectionist model. Proceedings of the National Academy of Sciences of the United States of America 96, 7592–7597.

Kellenbach, M.L., Wijers, A.A., Hovius, M., Mulder, J., Mulder, G., 2002. Neural differentiation of lexico-syntactic categories or semantic features? Event-related potential evidence for both. Journal of Cognitive Neuroscience 14 (4), 561–577.

King, J.W., Kutas, M., 1998. Neural plasticity in the dynamics of human visual word recognition. Neuroscience Letters 244, 61–64.

Klein, R.M., McMullen, P.A. (Eds.), 1999. Converging Methods for Understanding Reading and Dyslexia. MIT Press, Cambridge.

Kramer, A.F., Donchin, E., 1987. Brain potentials as indices of orthographic and phonological interaction during word matching. Journal of Experimental Psychology: Learning, Memory and Cognition 13, 76–86.

Kuriki, S., Takeuchi, F., Hirata, Y., 1998. Neural processing of words in human extrastriate visual cortex. Cognitive Brain Research 6, 193–202.

Kutas, M., Dale, A., 1997. Electrical and magnetic readings of mental functions. In: Rugg, M.D. (Ed.), Cognitive Neuroscience. Psychology Press, UK.

Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences 4 (12), 463–470.

Kutas, M., Federmeier, K.D., in press. Event-related potential (ERP) studies of sentence processing. In: G. Gaskell (Ed.), Oxford Handbook of Psycholinguistics.

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207, 203–205.

Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307 (12), 161–163.

Kutas, M., Van Petten, C., 1994. Psycholinguistics electrified. In: Gernsbacher, M.A. (Ed.), Handbook of Psycholinguistics. Academic Press, San Diego, pp. 83–143.

Kutas, M., Van Petten, C., Kluender, R., in press. Psycholinguistics electrified II (1994–2005). In M.A. Gernsbacher and M. Traxler (Eds.), Handbook of Psycholinguistics, 2nd edition. New York: Elsevier Press.

Lamme, V.A.F., 2003. Why visual attention and awareness are different. Trends in Cognitive Sciences 7, 12–18.

Leff, A., 2004. A historical review of the representation of the visual field in primary visual cortex with special reference to the neural mechanisms underlying macular sparing. Brain and Language 88, 278–286.

Liu, Y., Perfetti, C.A., 2003. The time course of brain activity in reading English and Chinese: an ERP study of Chinese bilinguals. Human Brain Mapping 18, 167–175.

Liu, Y., Perfetti, C.A., Hart, L., 2003. ERP evidence for the time course of graphic, phonological, and semantic information in Chinese meaning and pronunciation decisions. Journal of Experimental Psychology: Learning, Memory, and Cognition 29, 1231–1247.

Luck, S.J., 1998. Sources of dual-task interference: evidence from human electrophysiology. Psychological Science 9 (3), 223–227.

Luck, S.J., Vogel, E.K., Shapiro, K.L., 1996. Word meanings can be accessed but not reported during the attentional blink. Nature 383 (6601), 616–618.

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

McClelland, J.L., Patterson, K., 2002. Rules or connections in past-tense inflections: what does the evidence rule out? Trends in Cognitive Sciences 6 (11), 456–465.

McClelland, J.L., Rumelhart, D.E., 1981. An interactive activation model of context effects in letter perception: I. An account of basic findings. Psychological Review 88, 375–407.

McDonald, S.A., Carpenter, R.H.S., Shillcock, R.C., 2005. An anatomically constrained, stochastic model of eye movement control in reading. Psychological Review 112 (4), 814–840.

McKinnon, R., Allen, M., Osterhout, L., 2003. Morphological decomposition involving non-productive morphemes: ERP evidence. NeuroReport 14, 883–886.

Mcleod, P., Plaut, D.C., Shallice, T., 2001. Connectionist modelling of word recognition. Synthese 129, 173–183.

Monaghan, P., Shillcock, R.C., McDonald, S., 2004. Hemispheric

asymmetries in the split-fovea model of semantic processing. Brain and Language 88, 339–354.
Morton, J., 1969. Interaction of information in word recognition. Psychological Review 76, 165–178.
Morton, J., 1980. The logogen model and orthographic structure. In: Frith, U. (Ed.), Cognitive Processes in Spelling. Academic Press, London.
Münte, T.F., Say, T., Clahsen, H., Schiltz, K., Kutas, M., 1999. Decomposition of morphologically complex words in English: evidence from event-related brain potentials. Cognitive Brain Research 7, 241–253.
Münte, T.F., Urbach, T.P., Düzel, E., Kutas, M., 2000. Event-related brain potentials in the study of human cognition and neuropsychology, In: Boller, F., Grafman, J., Rizzolatti, G. (Eds.), 2nd ed. Handbook of Neuropsychology, vol. 1. Elsevier Science Publishers B.V.
Münte, T.F., Wieringa, B.M., Weyerts, H., Szentkuti, A., Matzke, M., Johannes, S., 2001. Differences in brain potentials to open and closed class words: class and frequency effects. Neuropsychologia 39, 91–102.
Neville, H., Nicol, J.L., Barss, A., Forster, K.I., Garrett, M.F., 1991. Syntactically based sentence processing classes: evidence from event-related brain potentials. Journal of Cognitive Neuroscience 3, 151–165.
Neville, H., Mills, D.L., Lawson, D., 1992. Fractionating language: different neural subsystems with different sensitive periods. Cerebral Cortex 2, 244–258.
Newman, R.L., Connolly, J.F., 2004. Determining the role of phonology in silent reading using event-related brain potentials. Cognitive Brain Research 21, 94–105.
Nobre, A.C., McCarthy, G., 1994. Language-related ERPs: scalp distributions and modulations by word type and semantic priming. Journal of Cognitive Neuroscience 6, 233–255.
Nobre, A.C., Allison, T., McCarthy, G., 1994. Word recognition in the human inferior temporal lobe. Nature 372, 260–263.
Osman, A., Moore, C.M., 1993. The locus of dual-task interference: psychological refractory effects on movement-related brain potentials. Journal of Experimental Psychology: Human Perception and Performance 19 (6), 1292–1312.
Osterhout, L., Bersick, M., McKinnon, R., 1997. Brain potentials elicited by words: word length and frequency predict the latency of an early negativity. Biological Psychology 46, 143–168.
Page, M.P.A., 2000. Connectionist modeling in psychology: a localist manifesto. Behavioral and Brain Sciences 23, 443–512.
Penke, M., Weyerts, H., Gross, M., Zander, E., Münte, T.F., Clahsen, H., 1997. How the brain processes complex words: an event-related potential study of German verb inflections. Cognitive Brain Research 6, 37–52.
Perfetti, C.A., Liu, Y., Tan, L.H., 2005. The Lexical Constituency Model: some implications of research on Chinese for general theories of reading. Psychological Review 112 (1), 43–59.
Pinker, S., Ullman, M.T., 2002. The past and future of the past tense. Trends in Cognitive Sciences 6 (11), 465–472.
Plaut, D.C., Shallice, T., 1993. Deep dyslexia: a case study of connectionist neuropsychology. Cognitive Neuropsychology 10, 377–500.
Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Patterson, K.E., 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychological Review 103, 56–115.
Plunkett, K., Juola, P., 1999. A connectionist model of English past tense and plural morphology. Cognitive Science 23, 463–490.
Posner, M.I., Raichle, M.E., 1995. Precis of images of mind. Behavioral and Brain Sciences 18 (2), 327–339.
Preissl, H., Pulvermüller, F., Lutzenberger, W., Birbaumer, N., 1995. Evoked potentials distinguish between nouns and verbs. Neuroscienee Letters 197, 81–83.
Price, C.J., Devlin, J.T., 2003. The myth of the visual word form area. NeuroImage 19, 473–481.
Proverbio, A.M., Vecchi, L., Zani, A., 2004. From orthography to phonetics: ERP Measures of grapheme-to-phoneme conversion mechanisms in reading. Journal of Cognitive Neuroscience 16 (2), 301–317.
Pulvermüller, F., Lutzenberger, W., Birbaumer, N., 1995. Electrocortical distinction of vocabulary types. Electroencephalography and Clinical Neurophysiology 94, 357–370.
Pulvermüller, F., Lutzenberger, W., Preissl, H., 1999. Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. Cerebral Cortex 9, 497–506.
Pylkkänen, L., Stringfellow, A., Marantz, A., 2002. Neuromagnetic evidence for the timing of lexical activation: an MEG component sensitive to phonotactic probability but not to neighborhood density. Brain and Language 81, 666–678.
Pylkkänen, L., Feintuch, S., Hopkins, E., Marantz, A., 2004. Neural correlates of the effects of morphological family frequency and family size: an MEG study. Cognition 91, B35–B45.
Radeau, M., Besson, M., Fonteneau, E., Luis Castro, S., 1998. Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. Biological Psychology 48, 183–204.
Randall, O'Reilly, R., Munakata, Y., 2000. Computational Explorations in Cognitive Neuroscience. MIT Press, Cambridge, MA.
Rastle, K., Coltheart, M., 2000. Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. Journal of Memory and Language 42, 342–364.
Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124 (3), 372–422.
Reggia, J.A., Levitan, S., 2003. Hemispheric interaction and specialization. In: Arbib, M. (Ed.), Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge, MA, pp. 525–528.
Reichle, E.D., 2006. Computational models of eye-movement control during reading: Theories of the "eye-mind" link. Cognitive System Research 7 (1), 2–3.
Reichle, E.D., Rayner, K., Pollatsek, A., 2003. The E–Z reader model of eye-movement control in reading: comparisons to other models. Behavioral and Brain Sciences 26, 445–526.
Rodríguez-Fornells, A., Clahsen, H., Lleo, C., Zaake, W., Münte, T.F., 2001. Event-related brain responses to morphological violations in Catalan. Cognitive Brain Research 11, 47–58.
Rodríguez-Fornells, A., Münte, T.F., Clahsen, H., 2002. Morphological priming in Spanish verb forms: an ERP repetition priming study. Journal of Cognitive Neuroscience 14 (3), 443–454.
Rolls, E.T., Treves, A., 1998. Neural Networks and Brain Function. Oxford University Press, USA.
Rösler, F., 2005. From single-channel recordings to brain-mapping devices: the impact of electroencephalography on experimental psychology. History of Psychology 8 (1), 95–117.
Rousselet, G.A., Thorpe, S.J., Fabre-Thorpe, M., 2004. How parallel is visual processing in the ventral pathway? Trends in Cognitive Sciences 8 (8), 363–370.
Rugg, M.D., 1990. Event related potentials dissociate repetition effects of high and low frequency words. Memory and Cognition 18, 367–379.
Rugg, M.D., Barrett, S.E., 1987. Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgement task. Brain and Language 32, 336–361.
Rugg, M.D., Coles, M. (Eds.), 1995. Electrophysiology of Mind: Event-related Brain Potentials and Cognition. Oxford University Press, Oxford, England.
Rumelhart, D.E., McClelland, J.L., 1982. An interactive activation

model of context effects in letter perception: Part 2. The contextual enhancement effects and some tests and enhancements of the model. Psychological Review 89, 60–94.

Rumelhart, D.E., McClelland, J.L., 1986. On learning the past tenses of English verbs. In: McClelland, J.L., Rumelhart, D.E., PDP Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. . Psychological and Biological Models, vol. 2. MIT Press, Cambridge, MA, pp. 216–271.

Salmelin, R., Service, E., Kiesilä, P., Uutela, K., Salonen, O., 1996. Impaired visual word processing in dyslexia revealed with magnetoencephalography. Annals of Neurology 40, 157–162.

Schendan, H.E., Giorgio, G., Kutas, M., 1998. Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. Psychophysiology 35, 240–251.

Schreuder, R., Baayen, H., 1995. Modeling morphological processing. In: Feldman, L.B. (Ed.), Morphological Aspects of Language Processing. Erlbaum Press, Hillsdale, NJ, pp. 131–154.

Seidenberg, M., 1987. Sublexical structures in visual word recognition: access units or orthographic redundancy? In: Coltheart, M. (Ed.), Attention and Performance XII: The Psychology of Reading. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Seidenberg, M.S., Gonnerman, L.M., 2000. Explaining derivational morphology as the convergence of codes. Trends in Cognitive Science 4, 353–361.

Seidenberg, M.S., McClelland, J.L., 1989. A distributed, developmental model of word recognition and naming. Psychological Review 96, 523–568.

Sereno, S.C., Rayner, K., 2003. Measuring word recognition in reading: eye movements and event-related potentials. Trends in Cognitive Sciences 7 (11), 489–493.

Sereno, S.C., Rayner, K., Posner, M.I., 1998. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. NeuroReport 9, 2195–2200.

Sereno, S.C., Brewer, C.C., O'Donnell, P.J., 2003. Context effects in word recognition: evidence for early interactive processing. Psychological Science 14, 328–333.

Shallice, T., 1988. From Neuropsychology to Mental Structure. Cambridge University Press, Cambridge, England.

Shallice, T., Warrington, E.K., McCarthy, R., 1983. Reading without semantics. Quarterly Journal of Experimental Psychology 35, 111–138.

Shillcock, R.C., Monaghan, P., 2001. The computational exploration of visual word recognition in a split model. Neural Computation 13, 824–851.

Shillcock, R., Ellison, M.T., Monaghan, P., 2000. Eye-fixation behavior, lexical storage and visual word recognition in a split processing model. Psychological Review 107, 824–851.

Simon, G., Bernard, C., Largy, P., Lalonde, R., Rebai, M., 2004. Chronometry of visual word recognition during passive and lexical decision tasks: an ERP investigation. International Journal of Neuroscience 114, 1293–1324.

Taft, M., 1994. Interactive-activation as a framework for understanding morphological processing. Language and Cognitive Processes 9 (3), 271–294.

Taft, M., Forster, K.I., 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. Journal of Verbal Learning and Verbal Behavior 15, 607–620.

Tarkiainen, A., Helenius, P., Hansen, P.C., Cornelissen, P.L., Salmelin, R., 1999. Dynamics of letter string perception in the human occipito-temporal cortex. Brain 122, 2119–2132.

Tarkiainen, A., Cornelissen, P.L., Salmelin, R., 2002. Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. Brain 125, 1125–1136.

ter Keurs, M., Brown, C.M., Hagoort, P., 2002. Lexical processing of vocabulary class in patients with Broca's aphasia: an event-related brain potential study on grammatic comprehension. Neuropsychologia 40, 1547–1561.

Urbach, T.P., Kutas, M., 2002. The intractability of scaling scalp distributions to infer neuroelectric sources. Psychophysiology 39 (6), 791–808.

Valdés-Sosa, M., Gonzalez, A., Xiang, L., Zhang, X.L., Yi, H., Bobes, M.A., 1993. Brain potentials in a phonological matching task using Chinese characters. Neuropsychologia 31 (8), 853–864.

Van Petten, C., Kutas, M., 1990. Interactions between sentence context and word frequency in event-related potentials. Memory and Cognition 18 (4), 380–393.

Vogel, E.K., Luck, S.J., Shapiro, K.L., 1998. Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. Journal of Experimental Psychology: Human Perception and Performance 24 (6), 1656–1674.

Weems, S.A., Reggia, J.A., 2004. Hemispheric specialization and independence for word recognition: a comparison of three computational models. Brain and Language 89, 554–568.

Weyerts, H., Münte, T.F., Smid, H.G., Heinze, H.J., 1996. Mental representations of morphologically complex words: an event-related potential study with adult humans. Neuroscience Letters 206, 125–128.

Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., Münte, T.F., 1997. Brain potentials indicate differences between regular and irregular German noun plurals. NeuroReport 8, 957–962.

Whitney, C., 2001. How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. Psychonomic Bulletin and Review 8, 221–243.

Whitney, C., 2004. Hemisphere-specific effects in word recognition do not require hemisphere-specific modes of access. Brain and Language 88, 279–293.

Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The neural basis of error detection: Conflict monitoring and the Error-Related Negativity. Psychological Review 111 (4), 931–959.

Zaidel, E., 2002. Hemispheric specialization for language in the split brain. In: Boller, F., Grafman, J. (Eds.), Handbook of Neuropsychology: Language and Aphasia.

Zevin, J.D., Seidenberg, M.S., 2006. Simulating consistency effects and individual differences in nonword naming: a comparison of current models. Journal of Memory and Language 54, 145–160.

Ziegler, J.C., Goswami, U., 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic Grain Size Theory. Psychological Bulletin 131 (1), 3–29.

Ziegler, J.C., Klüppel, D., Jacobs, A.M., 2001. Pseudohomophone effects in lexical decision: still a challenge for current word recognition models. Journal of Experimental Psychology: Human Perception and Performance 27 (3), 547–559.