Published in final edited form as:
*Behav Brain Sci*. 2012 October ; 35(5): 263–279. doi:10.1017/S0140525X11001841.

# Towards a Universal Model of Reading

**Ram Frost**
Department of Psychology The Hebrew University And Haskins Laboratories, New Haven CT.

## Abstract

In the last decade, reading research has seen a paradigmatic shift. A new wave of computational models of orthographic processing that offer various forms of noisy position or context-sensitive coding, have revolutionized the field of visual word recognition. The influx of such models stems mainly from consistent findings, coming mostly from European languages, regarding an apparent insensitivity of skilled readers to letter-order. Underlying the current revolution is the theoretical assumption that the insensitivity of readers to letter order reflects the special way in which the human brain encodes the position of letters in printed words. The present paper discusses the theoretical shortcomings and misconceptions of this approach to visual word recognition. A systematic review of data obtained from a variety of languages demonstrates that letter-order insensitivity is *not* a general property of the cognitive system, neither it is a property of the brain in encoding letters. Rather, it is a *variant* and idiosyncratic characteristic of some languages, mostly European, reflecting a strategy of optimizing encoding resources, given the specific structure of words. Since the main goal of reading research is to develop theories that describe the *fundamental and invariant* phenomena of reading across orthographies, an alternative approach to model visual word recognition is offered. The dimensions of a possible universal model of reading, which outlines the common cognitive operations involved in orthographic processing in all writing systems, are discussed.

## 1. Introduction

The business of modeling visual word recognition has never been better. In the last decade, computational models of reading[1] have been produced at an impressive rate. However, whereas the previous generation of reading models of the 80s and 90s (e.g., The Serial Search Model, Forster, 1976; The Interactive Activation Model (IAM), McClelland & Rumelhart, 1981; The distributed developmental model, Seidenberg & McClelland, 1989; and The Dual Route Cascaded Model (DRC), Coltheart, Rastle, Perry, Ziegler& Langdon, 2001) aimed at providing a *general* framework of lexical structure and lexical processing, addressing a relatively wide range of reading phenomena (e.g., word superiority effect, context effects, phonological computation, consistency by regularity interaction, reading aloud and reading disabilities, etc.), the new wave of modeling seems to have focused mostly on the front-end of visual word recognition. The influx of such models, which center on orthographic processing, stems mainly from consistent findings coming from a variety of languages, such as English, French and Spanish, regarding an apparent insensitivity of skilled readers to letter-order. Typically, these findings have demonstrated a surprisingly small cost of letter-transpositions in terms of reading time, along with robust priming effects when primes and targets share all of their letters but in a different order (e.g., Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004; Perea & Carreiras, 2006a,b, 2008;

---

[1] In the context of the present paper, the term “models of reading” refers mainly to the modeling of visual word recognition, rather than the complex operations involved in full text reading.

Rayner, White, Johnson, & Liversedge, 2006; Duñabeitia, Perea, & Carreiras, 2007; Johnson, Perea, & Rayner, 2007; Kinoshita & Norris, 2009).

The important role of registering letter position during the process of visual word recognition and reading seems almost self-evident. Printed letters are visual objects, and the fast saccades that characterize text reading necessarily involve some level of uncertainty regarding their exact identity and location. Indeed, general concerns regarding letter position coding have already been acknowledged in the seminal discussion of the Interactive-Activation Model (Rumelhart and McClelland, 1982), and some proposals for alternative coding schemes have been subsequently offered (e.g., Seidenberg & McClelland, 1989; and see also Bruner and O'Dowd, 1958).

In this context, the apparent indifference of readers to letter-order, reported in many studies, has revolutionized the modeling of visual word recognition. Underlying this revolution is the theoretical assumption that insensitivity to letter order reflects the special way in which the human brain encodes the position of letters in printed words (e.g., Grainger & Whitney, 2004; Whitney, 2001; Whitney & Cornelissen, 2005). As a consequence, the old computational models that encoded letter positions in rigid and absolute terms (e.g., IAM, McClelland & Rumelhart, 1981; DRC, Coltheart et al., 2001; or CDP, Zorzi, Houghton, & Butterworth, 1998) were out, to be replaced by models involving letter position uncertainty, either through various forms of context-sensitive coding or by introducing noisy letter-positions (e.g., The SERIOL model, Whitney, 2001; the SOLAR and the recent Spatial Coding model, Davis, 1999; 2010; The Bayesian Reader model, Norris, Kinoshita, & van Casteren, 2010; the Overlap model, Gomez, Ratcliff, & Perea, 2008; the dual-route model of orthographic processing, Grainger & Ziegler, 2011; and see also Grainger & van Heuven, 2003; Grainger, Granier, Farioli, van Assche, & van Heuven, 2006). This constitutes a dramatic paradigm shift, since the fuzzy encoding of letter-order has become a primary component of modeling visual word recognition and reading. More importantly, by focusing almost exclusively on issues of processing letter sequences and on letter-position, reading and visual word recognition research has shifted to produce theories of orthographic processing per-se, some with the explicit aim of "cracking the orthographic code" (see Grainger, 2008, for a detailed discussion).

Admittedly, some of the extensive empirical work regarding indifference of readers to letter order has focused on whether the locus of the effect is morphological (e.g., Christianson, Johnson, & Rayner, 2005; Dunabeitia et al. 2007), or phonological (e.g., Perea & Carreiras, 2006a, b, 2008; Acha & Perea, 2010), and some experiments examined the interaction of letter-position coding with consonant vs. vowel processing (e.g., Perea & Lupker, 2004). Nevertheless, the main conclusion of these studies was that transposed-letter (TL) effects are orthographic in nature (e.g., Perea & Carreiras, 2006a, b). Purely orthographic models were considered, therefore, to have substantial descriptive adequacy, thereby accounting for a large set of data. Consequently, they were taken to represent a viable approach to visual word recognition without the need to revert to phonological or morphological considerations (see for example, Davis, 2010, for a discussion). This, inevitably narrows the array of phenomena that can be explained by the models, limiting them mainly to effects related to various aspects of orthographic form.

Paradigmatic shifts, however, should emerge only following extensive theoretical debates. If not, they may reflect only occasional fluctuations of trends and fashion, to which even scientific inquiry is not immune. The present paper takes the recent wave of modeling visual word recognition as an example of how interesting findings can eventually lead to a generation of narrow, and, therefore, ill-advised models. Through a comprehensive discussion of the theoretical shortcomings underlying the basic approach of current trends of modeling reading, it aims to outline alternative directions. These directions emerge from the following claims: orthographic effects in visual word recognition, such as sensitivity or insensitivity to letter order or any other phenomena, are the product of the *full* linguistic environment of the reader (phonology, morphology and semantic meaning), not just the structure of orthographic letter sequences. Orthographic processing cannot be researched, explicated or understood, without considering the manner in which orthographic structure represents phonological, semantic and morphological information in a given writing system. Therefore, only models that are tuned, one way or another, to the full linguistic environment of the reader can offer a viable approach to modeling reading. A word of caution though: the following discussions are not aimed at proposing specific blueprints for a new computational model of orthographic processing, neither do they point to specific modeling implementations. Their goal is to set the principles for understanding, researching and consequently modeling the processing of printed information.

## The “new age” of orthographic processing

The recent wave of modeling visual word recognition has focused on a series of findings all related to the manner by which readers treat the constituent letters of printed words. The original demonstration that letter transposition in the prime results in significant facilitation in recognizing the target, was reported by Forster, Davis, Schoknecht, & Carter (1987), who showed that TL primes (anwser-ANSWER) produce priming as large as identity primes (answer-ANSWER). This surprising effect was followed-up by Perea and Lupker who systematically examined whether and how this effect varies as a function of letter-position (Perea & Lupker, 2003, 2004). Subsequent research on eye movements argued that letter-transpositions result in some cost in terms of fixation time measures on target words during reading (Rayner et al., 2006; Johnson et al., 2007). This cost, however, seemed relatively small in magnitude. In parallel, in 2003, a demonstration of how reading is resilient to letter-transposition became well known via a text composed entirely of jumbled letters which was circulating over the Internet. This demonstration, labeled “the Cambridge University effect” (reporting a fictitious study allegedly conducted at the University of Cambridge) was translated into dozens of languages and quickly became an urban legend. Consecutive follow-up studies reported that insensitivity to letter transpositions in reading can be revealed in a variety of languages such as French (Schoonbaert & Grainger, 2004), Spanish (Perea & Lupker, 2004; Perea & Carreiras, 2006a,b), Basque (Perea & Carreiras, 2006c; Duñabeitia et al., 2007) and Japanese Kana (Perea & Perez, 2009). The facilitation caused by TL primes was shown even with extreme distortions when several letters are jumbled (snawdcih-SANDWICH, Guerrera & Forster, 2008). The abundant evidence regarding TL priming converged with other forms of priming that suggested non-rigidity of letter-position coding. For example, in an extensive investigation, Humphreys, Evett, & Quinlan (1990) showed that primes consisting of a subset of the target’s constituent letters, which kept the *relative* but not the absolute position of letters (blck-BLACK), produce significant priming. Similar effects of relative-position priming were reported by Peresotti and Grainger (1999), and by Grainger et al. (2006). Relative-position priming has also been demonstrated with superset priming where primes contain more letters than the target (juastice-JUSTICE, Van Assche & Grainger, 2006).

The “new age of orthographic processing” reflects an increased interest and preoccupation with how the cognitive system encodes and registers letter sequences. Underlying this paradigmatic approach is an implicit and almost self-evident assumption that the game of visual word recognition is played mainly in the court of constituent letter recovery. The main focus of the new age approach is, therefore, the level of processing where letter position coding is approximate rather than specific (see Grainger 2008, and Grainger & Ziegler, 2011, for a discussion). Given the abundant evidence regarding relative insensitivity to letter-position, the emergent new models of reading focused on finding creative solutions to produce what seemed to be the main characteristic of reading: letter position flexibility.

For example, both the SOLAR model offered by Davis (1999), and the recent Spatial Coding model (Davis, 2010) adopt the idea of spatial coding, and encode relative letter position by measures of relative pattern of activity across letters in a word (see also Davis & Bowers, 2004; 2006). The SERIOL model (Whitney, 2001; Grainger & Whitney, 2004; Whitney, 2008; Whitney & Cornelissen, 2008) is based on a serial activation of letter detectors that fire serially in a rapid sequence (but see Adelman, Marquis, & Sabatos-DeVito, 2011, for counter-evidence regarding serial processing). This firing sequence serves as input to a layer of “open bigram” units, which do not contain precise information about letter contiguity, but preserve information regarding relative position. For example, the word *FORM*, would be represented by activation of the bigram units #F, *FO*, *OR, RM*, but also *FR*, *OM*, and *M#*, where # represents a word boundary. A transposition prime, such as FROM, would then share all but one of these units, namely *#F, FR, FO, RM, OM* and *M#*, resulting in substantial priming. Other models obtain letter position flexibility by assuming noisy slot-based coding. For example, The Overlap model (Gomez et al., 2008) posits a noisy letter order scheme in which information regarding order of letters becomes available more slowly than information about letter identity. Similarly, to accommodate TL effects, Kinoshita and Norris (2009), Norris and Kinoshita (2008), and Norris et al. (2010), have implemented as part of their computational model a noisy letter-position scheme in which, in the limited time available for which the prime is presented briefly, information regarding order of letters as well as information about letter identity is ambiguous. In a similar vein, a combination of noisy retinotopic letter coding with either contiguous bigram detectors (Dehaene, Cohen, Sigman, & Vinckier, 2005), or with location-specific letter detectors (Grainger et al., 2006), was suggested as well to account for letter-position flexibility. Note that although all of the above models deal in one way or another with letter position flexibility, they naturally differ in the scope of phenomena they describe. Hence, while context-sensitive coding models such as SERIOL focus on finding inventive solutions for representing a string of letters, models like the Bayesian Reader model (Norris, et al., 2010), or the Spatial Coding model (Davis, 2010) offer a rather broad and comprehensive view of visual word recognition and reading. Nevertheless, discussions regarding the descriptive adequacy of all of these models have centered mainly on their relative ability to predict effects of TL priming and to reproduce a continuum of TL priming effects, given different types of distortion in the sequence of letters. For example, almost all of the 20 simulations offered to validate the recent Spatial Coding model (Davis, 2010) deal in some way with transposed-letter priming effects.

Interestingly, most of the above models have argued that letter-position flexibility reflects *general and basic brain mechanisms* (e.g., neural temporal firing patterns across letter units, Whitney, 2001; noisy retinotopic firing, Dehaene et al., 2005; split of foveal vision and interhemispheric transfer costs, Shillcock, Ellison, & Monaghan, 2000; Hunter & Brysbaert, 2008). This claim, in the context of modeling visual word recognition, does not merely aim to make the models neurologically plausible, or to extend reading research to include in it also a description of the neuro-circuitry of the visual system --the claim is deeply theoretical in terms of reading theory, because it is based on a general argument regarding *the brain and lexical processing*.

The present paper attempts to discuss the theoretical shortcomings of this approach to visual word recognition. As will be argued, sensitivity to letter-order per-se does not tell us anything interesting about how the brain encodes letter position, *from the perspective of a theory of reading*. Instead, it tells us something very interesting about how the cognitive system treats letters *in specific linguistic environments*. To reiterate, it is not the neurological claims about noisy retinotopic firing or about neural temporal firing which are being contested. Obviously, the architecture of neural circuitry determines the way visual information is encoded and consequently processed in the cortex. What is being challenged is the implication of these facts for lexical architecture and understanding reading. As a corollary claim, I will argue that as a general strategy, focusing only on orthographic coding in a model, by mapping various types of input structure of letter units to an output structure of word units, while disregarding the contribution of phonological, semantic, or morphological factors to the process, can perhaps produce a desired behaviour in terms of letter-flexibility, but it misses the complexity and interactivity of the reading process.

## 2. Preliminary assumptions

The main goal of reading research is to develop theories that describe and explicate the *fundamental* phenomena of reading. Our models are major tools in developing such theories, so that they have descriptive and explanatory adequacy. I propose, therefore, two main criteria for assessing their potential contribution. Since the merits or shortcomings of approaches for modeling reading are a major focus of the present paper, a brief exposition of these criteria will set common ground for the following discussion.

### The universality constraint

Our first criterion is that models of reading should be universal in the sense that they should aim to reflect the *common* cognitive operations involved in treating printed language across different writing systems. Languages naturally differ in their scripts and orthographic principles. A good theory of reading should be able to describe and explicate, as a first step, the cognitive procedures that are implicated in processing printed words in *any* orthography, focusing on 1) what characterizes human writing systems, and 2) what characterizes the human cognitive system that processes them. I will label these common cognitive procedures “reading universals”. Only when the reading universals are well defined and well understood, diverging predictions in cross-linguistic research can be formulated a-priori (see Perfetti, 2011, for a discussion of a universal reading science). Consider for example language X that shows some consistent pattern of behavior across readers, and language Y that consistently does not. Our theory of reading should be able to suggest a higher-level principle that simultaneously accounts for both phenomena. If it cannot, then our set of reading universals is probably incomplete or possibly wrong. Naturally, the set of reading universals ought to be quite small, general, and abstract, to fit all writing systems and their significant inter-differences.

Reading universals are empirically established, and can be supported or falsified only through cross-linguistic research. Obviously, models of reading could be locally formulated to describe the idiosyncratic properties of reading one specific language or even a group of languages, thereby not satisfying the universality constraint. However, in this case, their narrower aim should be stated explicitly. The impact of these models would be greatly reduced since they are not actually part of a general reading theory. As will be subsequently argued and demonstrated, most current models of orthographic processing are not universal.

### Linguistic plausibility

Models of visual word recognition deal with words, and words have orthographic, phonological, semantic, and morphological characteristics. Although any given model may focus on any one of these properties and not on others, it should be nevertheless constrained by findings related to the other linguistic properties of words. A model of orthographic processing, therefore, could exclusively describe the processes involved in the perception and analysis of printed letters, and consequently, the model may not derive predictions regarding, say, phonological, semantic or morphological priming. The requirement of linguistic plausibility simply states that the model should, in principle, be able to accommodate the established findings related to all linguistic dimensions of printed words, or at least it should not be structured in a way that goes counter to the established findings for other linguistic dimensions. Thus, if for example, a model of orthographic processing accurately predicts letter-transposition effects, but its architecture runs counter to what we know about morphological or phonological processing, then the model does not maintain linguistic plausibility. The requirement regarding linguistic plausibility is based on a simple argument: orthographic processing by itself is not an independent autonomous process in cognition, separable from other aspects of language, because its output must be consistent with other linguistic dimensions it is supposed to represent or feed into. Printed words were designed from the outset in any writing system to represent spoken forms, which bear meaning. Hence, orthographic structure represents one single dimension of a complex lexical architecture. Our theory of orthographic processing should, therefore, in principle, fit into a general theory of meaning recovery.

Having set the two main criteria for assessing models of visual word recognition, I will first proceed with a detailed exposition of the nature of writing systems in general and orthographic structure in particular. What drives this exposition is the claim that the orthography of any given language has evolved as a result of the linguistic environment specific to that language, and naturally it cannot be treated as independent of it. Since in every language a different solution for representing phonology and meaning by print has evolved, the process of extracting linguistic information from the graphemic array in one language may be quite different than in another language, already at the early phases of print processing. The aim of the following section is then, to set the grounds for explicating why readers in different writing systems extract from similar sequences of letter strings different types of information, and why orthographic coding in one language may be quite different than in another language.

## 3. Every language gets the writing system it deserves [2]

Humans speak about 3000 languages,[3] and a significant number of these languages have their own writing system. At first blush, what determines the large variance of spoken languages and writing systems seems arbitrary, in the sense that it reflects mainly historical events or chance occurrences such as emerging local inventions, or diffusion due to tribal migration. However, although such chance events lie at the origin of many writing systems, close scrutiny suggests that, to a large extent, the way they have eventually evolved is not arbitrary. Rather, orthographies are structured so that they optimally represent the languages' phonological spaces, and their mapping into semantic meaning, and simple principles related to *optimization of information* can account for the variety of human writing systems and their different characteristics. Outlining these principles is important from the perspective of modeling reading because they provide interesting insight regarding how the cognitive system picks up the information conveyed by print. Here I promote a view that has the flavor of a Gibsonian ecological approach (Gibson, 1986), and assumes that, to be efficient, the cognitive system that processes language must be tuned to the structure of the linguistic environment in which it operates.

The common taxonomy of writing systems focuses on the way the orthographic units represent the phonology of the language. This is the origin of the orthographic depth hypothesis (the ODH, Frost, Katz, & Bentin, 1987; Katz & Frost, 1992), and of the grain-

---

[2] I am indebted to Ignatius Mattingly who coined this expression.

[3] The number of human languages reflects an approximation since the distinction between a language and a dialect often remains in question.

size theory (Ziegler & Goswami, 2005), which classify orthographic systems according to their letter-to-phoneme transparency. This approach to reading originates from extensive research in European languages where the main differences in reading performance, the speed of reading acquisition, and the prevalence of reading disabilities was taken to result mainly from the opaque or transparent relations of spelling to sound in a given language (e.g., Seymour, Aro and Erskine, 2003; Ziegler et al, 2010). The view that the recovery of phonological information is the main target of reading (see Frost, 1998, for an extensive review and discussion of the strong phonological theory) underlies the common characterization of writing systems as differing mainly on the dimension of phonological transparency. The aim of the present section of our discussion is to widen the perspective of what writing systems are, beyond the typical (and probably simplistic) distinctions regarding phonological transparency. The focus on this factor alone for characterizing writing systems is heavily influenced by research in European languages, mostly English (see Share, 2008, for a discussion of extensive Anglocentricities in reading research). Thus, by describing the writing system of five distinctive languages, the way they have evolved, and the manner by which the orthographic information optimally represents the phonological space of the language *and* its semantic and morphological structure, possible inferences regarding how print is processed will be drawn.

## Print as an optimal representation of speech and meaning

One important fact guides this part of the discussion: although print was invented to represent speech, spoken communication is much richer than its representing writing system. Specific lexical choices of words and meaning are conveyed in spoken communication by a wide array of signals, such as stress, intonation, timing of spoken units, or even hand movements, which do not usually exist in print. This means that, as a rule of thumb, print is underspecified in any language relative to the speech it is supposed to represent. The evolution of writing systems reflects, therefore, some level of optimization aimed at providing their readers with maximal phonological and semantic information using minimal orthographic units for processing. However, given the specific characteristics of a particular language, what constitutes “optimization” in language A may be quite different than what constitutes optimization in language B. This, as I will show, is crucial for understanding orthographic processing.

Considering the variety of human languages, they differ firstly, in the structure of their phonological space, and in the way that phonological units represent morphemes and meaning. This structure represents the “ecological environment” in which orthographies have evolved in a process similar to natural selection, to allow native speakers the most efficient representational system. Putting the conclusion of this section first: to be efficient, the cognitive operations that readers launch in processing their print, that is, the “code” they generate for lexical processing, must be tuned to the idiosyncratic characteristics of their own representational system. I label this *linguistic coherence.* Thus, to process orthographic structure, the system must be sensitive to the optimal representation of *several* linguistic dimensions, in order to extract from the print maximal information. Hence, a model of reading that is linguistically coherent, must likewise include a level of description that contains *all* aspects of the language in which reading occurs. This means that a theory of visual word recognition cannot be simply “orthographic”, because the information that is extracted from print concerns complex interactions of orthography, phonology, morphology, and meaning. A model of orthographic processing, therefore, cannot be blind to this factor.

In the following exposition, five contrasting languages will be described: Chinese -- a Sino-Tibetan language, Japanese -- an Altaic language, Finnish -- a Finno-Ugric language, English -- an Indo-European language, and Hebrew -- a Semitic language. These five languages have distinct phonological, grammatical, and orthographic features, providing good coverage of the linguistic diversity in the world. The aim of this brief exposition is to demonstrate that the evolution of writing systems is not arbitrary, but mirrors a process of optimization, which is determined by constraints of the cognitive system (see Gelb, 1952, for similar arguments). These constraints concern efficiency of processing, where a substantial amount of information needs to be packed in a way so that readers of the language are provided with maximal semantic, morphological, and phonological cues via minimal orthographic units. I should emphasize then, that the purpose of the following description of writing systems is not to provide a theory of structural linguistics. What underlies this description is a deep theoretical claim regarding reading universals. For if there are common principles by which writing systems have evolved to represent orthographic information in all languages, then this must tell us something interesting about how the cognitive system processes orthographic information. If writing systems in different languages all share common strategies to provide their readers with optimal linguistic information, then it must be that the processing system of readers is tuned to efficiently pick-up and extract from print this optimal level of linguistic information. From this perspective, finding commonalities in the “logic” behind the evolution of different orthographies should have consequences for our theory of orthographic processing.

## Five contrasting languages

**Chinese—**In Chinese, words are in most cases mono-morphemic without much affixation, and the morphemic units (the words) are also monosyllabic. The permissible syllable in Chinese has no more than four phonemes (relative to seven in English). This basic structure of the language can be considered “arbitrary”, in the sense that the phonological space of Chinese could, in principle, have been different. However, once this phonological space has been established in the way it has, all resulting linguistic developments were to some extent entirely predetermined. For example, if all words in a language are monosyllabic, and if the syllabic structure is constrained to no more than four phonemes, then the number of possible Chinese words is necessarily small because the number of permissible syllables is small. This determines extensive homophony to represent the large variety of semantic meanings for meaningful complex communication (see for example, Chao, 1968). Homophony is indeed a main feature of Chinese: sometimes up to 20 different words (different meanings) are associated with a given syllable. In the spoken language, some of this homophony is resolved and disambiguated by the tones added to the syllable: high, low, rising or falling. Print, however, is underspecified relative to speech. Hence, a solution for disambiguating the extensive syllabic homophony in print had to evolve for accurate communication. It is in this perspective that the logographic writing system of Chinese should be considered -- a writing system in which different semantic radicals accompany similar phonetic radicals, to denote and differentiate between the large number of morphemes that share a given syllabic structure (see Wang, Tsai, & Wang, 2009, for a detailed review).

The structure of Chinese characters provides an additional insight. Most characters are lexically determined by a semantic radical appearing in most cases on the left side, and then the phonetic radical suggesting how to pronounce the character is added on the right side. Looking up characters in the Chinese dictionary follows the same principle: the semantic component of a character determines the initial entry, and then the cues regarding how to pronounce it consist of necessary complementary information. Semantic information comes first then, phonetic information comes second. Indeed, in his taxonomy of languages, DeFrancis (1989) categorized Chinese as a “meaning-plus-sound” syllabic system (and see Wang et al., 2009, for a similar characterization).

The lesson to be learnt from Chinese is that writing systems are not set to simply provide their readers with a means for retrieving, as quickly as possible, a phonological structure. If that were the case, an alphabet that represents the syllables of Chinese would have been employed. Writing systems evolve to provide *optimal* information by weighting the need for maximal cues about the spoken words and their specific meanings while using minimal orthographic load.

**Japanese—**In comparison to Chinese, the permissible syllabic structure of Japanese is even more constrained, and consists mostly of CV or V units.[4] However, in contrast to Chinese, words are not monosyllabic. Since words in Japanese can be composed of several syllables, tones were not necessary for constructing the sufficient number of lexical units that are required for a viable language, and indeed Japanese is not a tonal language. With about 20 consonants and 5 vowels, Japanese has 105 permissible phonological units, named mora, which consist of the basic sublinguistic phonological units of the language. A mora is a *temporal* unit of oral Japanese, the unit with which speakers control the length of word segments, and eventually determine the length of spoken words (see Kubozono, 2006). This is a very brief and rudimentary description of Japanese phonological space; however, the consequent implications for their writing system can be explicated using the same form of evolutionary arguments as in Chinese.

Historically, the first writing system of Japanese was kanji, a logographic script whose characters were imported from Mainland China (see Wang, 1981, for a review). About 2130 characters represent the current kanji script of Japanese. Since the Chinese characters along with their phonetic radicals were imported and used to represent Japanese words, which have an entirely different phonological structure, an interesting question is why kanji characters were at all appealing to Japanese speakers. Although any kind of answer would be speculative at best, a probable account seems to lie again in the restricted phonological units (words) that can be formed in Japanese, given the strict constraints on syllabic (mora) structure. If the number of permissible morae is relatively small given their very constrained structure, in order to create a sufficient number of words necessary for rich communication, the only solution is to allow for relatively large strings of morae. This, however, is not an optimal solution in terms of spoken communication (see Piantadosi, Tily, & Gibson, 2011, for a discussion of word length and efficient communication). In Japanese, words consist then in most cases of two to four morae. This again inevitably leads to a significant level of homophony, as a relatively small number of phonological forms denote a large amount of semantic meanings that are necessary for rich communication. Japanese indeed has significant homophony (although to a much lesser extent than Chinese). The use of kanji served the purpose of resolving the semantic ambiguity underlying a high-level of homophony. Indeed, the kanji characters in Japanese help in denoting specific meanings of homophones (see Seki, 2011, for a detailed description).

However, following the introduction of Chinese characters to Japan, another phonographic writing system, (hiragana and katakana), has evolved to represent the spoken language, and this evolution was to some extent inevitable as well. Note that in contrast to Chinese, Japanese is not a mono-morphemic language. The kanji characters could not be used to denote morphological inflections. Since writing systems are primarily designed to represent meaning to readers often through morphological information (e.g., Mattingly, 1992), some phonetic symbols had to be inserted to convey inflections and derivations. This is the origin of the phonographic hiragana, a script that emerged given the morphological structure of the language. In addition to hiragana, katakana graphemes were also added to denote loan words, which obviously cannot be represented by the kanji characters. It could be argued that the use of two phonographic scripts, hiragana and katakana, is a luxury of dubious

---

[4] Some Japanese syllables are non-open syllables such as geminate consonants (/Q/), or nasal N, which is often lengthened. Japanese has also two semi-vowels (or glides) /y/ and /w/.

utility; however, the advantage of this notation is that it emphasizes morphological internal structure, by assigning a separate writing system to denote morphological information. This “choice” of writing systems to emphasize morphemic constituents is consistent, and can also be demonstrated in English or in Hebrew, although with the use of different principles.

The manner with which Japanese phonograms represent the spoken subunits of the language is also not arbitrary. From a perspective of information efficiency, the optimal solution for representing the sub-linguistic units of Japanese is graphemes representing the relatively small number of morae. An alphabet where letters represent single phonemes would not do, since Japanese is a *morae-timed* language,[5] and phoneme representation would not be optimal. Memorizing 105 graphemes for decoding is pretty easy. Not surprisingly then, in the phonographic kana, letters represent morae, and Japanese children easily master the kana writing system at the beginning of the first grade, with a relatively low rate of reading disabilities in this writing system[6] (Yamada & Banks, 1994; Wydell & Kondo, 2003). More important, given the perfect match between Japanese phonological space and its representing writing system, children acquire a meta-awareness of morae about the time they learn to read (Inagaki, Hatano, & Otake, 2000) similar to the development of phonemic awareness following reading acquisition in alphabetic orthographies (e.g., Bertelson, Morais, Alegria, & Content, 1985; Cossu, Schankwiler, Liberman, Katz, & Tola, 1988; Bentin, Hammer, & Cahan, 1991; Goswami, 2000). The lesson to be learnt from Japanese is, again, that the phonological space, the manner by which it conveys meaning, and the morphological structure of the language predetermine the main characteristics of the writing system. Indeed, using kana-kanji cross-script priming, Bowers and Michita (1998) have elegantly demonstrated how the orthographic system of Japanese must interact with phonology and semantics to learn abstract letter and morphological representations.

**Finnish—**Finnish is considered a “pure phonemic system” like Greek or Latin (DeFrancis, 1989). There are 24 Finnish phonemes, eight vowels, and 16 consonants (one consonant conveyed by a two-letter grapheme NG). The consistency of grapheme-to-phoneme is perfect, with 24 correspondences to be learnt only. Some phonemes are long, but doubling the corresponding letters conveys these. Finnish thus represents the perfect example of an orthography that is fully transparent phonologically (Borgwaldt, Hellwig, & De Groot, 2004, 2005; Ziegler et al., 2010).

In the context of the present discussion, the interesting aspect of Finnish is its morphological structure as reflected by its agglutinative character (Richardson, Mikko, & Lyytinen, 2011). In Finnish compounding is very common, and printed words often have even 18-20 letters, so that printed entities such as *liikenneturvallisuusasiantuntija* (expert in travel safety) are not rare (see, for example, Bertram & Hyona, 2003; Kuperman, Bertram & Baayen, 2008). Here again, the question of interest concerns the relation between the excessively agglutinative aspect of the language and the complete transparency of the orthographic system. This relation can be explicated using the same principles of optimization of information. Condensing information in a letter string has the advantage of providing more semantic features in a lexical unit. The question is whether the writing system can support this information density, considering the decoding demands imposed by very long letter-strings. If only 24 letter-phoneme correspondences need to be used, then the reading of very long words is easy enough. Thus, the combination of extreme transparency and compounding provides speakers of Finnish with an optimal ratio of semantic information to

---

[5] Spoken Japanese is based on equitemporal (or isochronous) rhythmical sequences of syllables having the same temporal length, whether it is a CV segment, a V segment, etc.

[6] Reading disabilities in kanji are more prevalent in Japanese, making the overall rate of reading disabilities in Japan similar to the typical 7% found elsewhere.

processing demands. The lesson to be learnt from Finnish is, again, that nothing is arbitrary when it comes to orthographic structure. If entropy of letter to sound is zero, orthographic structure becomes denser, to pack maximal morphological information. The orthographic processor of Finnish readers must be tuned to this.

**English—**English is an Indo-European language with an alphabetic writing system which is morpho-phonemic. Two main features characterize the very complex English phonological space. First, English has about 22 vowels,[7] 24 consonants, and the permissible syllable can be composed of one to seven phonemes (Gimson, 1981). This brings the number of English syllables to about 8000 (DeFrancis, 1989). Obviously, this huge number does not permit any syllabic notation such as in Japanese, and, not surprisingly, English is alphabetic-phonemic. There has been abundant discussion of the extreme inconsistencies in the representation of individual phonemes in English (e.g., Borgwaldt et al., 2004; 2005; Frost & Ziegler, 2007; Ziegler et al., 2010). Some of these inconsistencies stem from simple historical reasons -- mainly influences from German or Dutch (e.g., *knight – knecht*), some have to do with the dramatic disproportion of number of vowels and vowel letters, but the main source of the English writing system inconsistency is its morpho-phonemic structure.

In English, unlike most, if not all, Indo-European languages, morphological variations are characterized by extensive phonological variations. Thus, derivations and inflections, addition of suffixes, changes in stress due to affixation, etc., very often result in changes of pronunciation (e.g., heal-health, courage-courageous, cats-dogs). Given this unique aspect of spoken English, the evolution of its writing system could have theoretically taken two possible courses. The first was to follow closely the phonological forms of the language, and convey to the reader the different pronunciations of different morphological variations (e.g., heal-helth). The second was to represent the morphological (and thereby semantic) information, irrespective of phonological form. Not surprisingly, the writing system of English has taken the second path of morphophonemic spelling, and given the excessive variations of phonological structure following morphological variations, English orthography has evolved to be the most inconsistent writing system of the Indo-European linguistic family. The lesson to be learnt from English is that writing systems, whenever faced with such contrasting options, necessarily evolve to provide readers with the meaning of the printed forms by denoting their morphological origin, rather than simplifying phonological decoding. Hence, recent suggestions that English spelling should be reformed [8] “and made consistent” stem from a deep misunderstanding of the evolution of writing systems. As stated above, every language gets the writing system it deserves. The inconsistent writing system of English is inevitable, given the characteristics of the language’s phonological space. In spite of its excessive inconsistency, it still reflects an optimization of information by providing *maximal* morphological (hence semantic) cues along with relatively impoverished phonological notations, using *minimal* orthographic symbols. Again, as will be explicated, this has immediate implications to lexical structure and lexical processing.

**Hebrew—**In the context of the present discussion, Hebrew provides the most interesting insights regarding the rules that govern the logic of evolution of writing systems. Its description, therefore, will be, slightly extended. Hebrew is a Semitic language, as are Arabic, Amharic, and Maltese. Semitic languages are all root-derived, so that the word’s base is a root morpheme, usually consisting of three consonants, and it conveys the core

---

[7] The exact number of English vowels and consonants varies across dialects.

[8] For example, the Serbo-Croatian spelling was reformed at the end of the 19th century, and was made entirely transparent. This successful change, however, stems from the fact that in Serbo-Croatian, in contrast to English, morphological variations never result in phonological variations; therefore, as a rule, the phonological structure of base words does not change with derivational morphology.

meaning of the word. Semitic words are always composed by intertwining root morphemes with word-pattern morphemes-- abstract phonological structures, consisting of vowels, or of vowels and consonants, in which there are “open slots” for the root’s consonants to fit into. For example, the root Z.M.R. that conveys the general notion of “singing”, and the word-pattern /ti--o-et/, that is mostly used to denote feminine nouns, form the word /tizmoret/ meaning “an orchestra”. Thus, the root consonants can be dispersed within the word in many possible positions. There are about 3000 roots in Hebrew, 100 nominal word-patterns, and 7 verbal patterns (see Shimron, 2006, for a review).

Since Semitic words are generally derived from word-patterns, they have a recognizable and well-defined internal structure. Word-patterns can begin with a very restricted number of consonants (mainly /h/, /m/, /t/, /n/, /l/), and these determine a set of transitional probabilities regarding the order and identity of subsequent consonants and vowels. To cast it again in evolutionary terms, this bi-morphemic structure is obligatory. If the language is root-based, and the roots convey the core meaning of words, they need to be easily extracted and recognized by the speaker. Because there are no a-priori constraints regarding the location of root consonants in the word, the only clue regarding their identity is a well-defined phonological structure of the word, which allows the root consonants to stand out. This, as will be shown has important implications to orthographic processing in Hebrew.

The major characteristic of Hebrew writing system is its extreme under-specification relative to the spoken language. Hebrew print (22 letters) was originally designed to represent mostly consonantal (root) information. Most vowel information is not conveyed by the print (Bentin, & Frost, 1987). There are two letters that, in certain contexts, convey vowels: one for both /o/ and /u/, and one for /i/; however, these letters also convey the consonants /v/ and /y/, respectively. Hebrew print consists then of a perfect example of optimization of information, where crucial morphological (and therefore semantic) features are provided, along with sufficient phonological cues, using minimal orthographic symbols. The minimalism of orthographic notation serves an important purpose: it enables an efficient and very fast extraction of root letters from the letter string; the smaller the number of letters, the easier the differentiation between root letters and word-pattern letters. This obviously comes with a heavier load on the reader when it comes to phonological decoding demands, since a substantial part of the phonological information is missing (see Frost, 1994; 1995). However, since the structure of spoken words is highly constrained by the permissible Semitic word-patterns, these decoding demands are significantly alleviated. Readers can converge on a given word-pattern quite easily with minimal orthographic cues, especially during text reading when the context determines a given word-pattern with relative high reliability, and once a word-pattern has been recognized, the full vowel information is available to the reader, even if it is not specified by the print (Frost, 2006).

The logic of this flexible evolutionary system can be demonstrated in considering the changes introduced into the Hebrew writing system throughout history. As long as biblical Hebrew was a live spoken language, its writing system was mainly consonantal as described so far. However, following the historical destruction of the Hebrew speaking national community by the Romans, Hebrew became a non-living language. If the language ceases to be spoken on a regular basis, the missing vowel information is not available to the reader at the same ease and speed as it is with spoken languages, and this increases the load of phonological decoding demands. Since the balance of optimization of information has shifted, around the 8th century vowel signs were introduced into Hebrew using diacritical marks in the form of points and dashes under the letters (“pointed Hebrew”), to alleviate the problem of phonological opacity. This served the purpose of reading religious scripts and prayer books fluently enough, without the need for semantic feedback or morphological analysis (i.e., decoding without understanding)[9]. The move from consonantal to pointed

Hebrew demonstrates how the weight of orthographic, semantic and phonological information in writing systems can dramatically shift due to changes in the linguistic environment. From the moment that phonological decoding could no longer rely on semantic feedback, orthographic structure had to change to become more complex and overburdened to supply the missing phonological information. At the end of the 19th century, Hebrew began to be reinstated as a spoken language. Without any formal decision regarding reforms in writing, and in less than a few decades, the vowel marks were naturally dropped from the Hebrew writing system, as was the case in ancient times. Thus, from the moment that the Hebrew language was revived, the balance of optimization shifted as well, and naturally reverted towards using minimal orthographic symbols. Today, Hebrew vowel marks are taught in the first grade, assisting teachers in developing decoding skills in their pupils during reading acquisition. However, starting from the end of the second grade, printed and written Hebrew does not normally include diacritical marks.

To summarize this section, we have examined five writing systems that evolved in five languages, demonstrating that orthographic structure provides readers with different types of information, pending on the language's writing system. The question at hand is whether this description is at all relevant to orthographic processing. The crucial debate then centers on whether the fact that different orthographies consist of different optimization of phonological, morphological and orthographic information, has behavioral implications in terms of processing orthographic form. If it does, then any model of orthographic processing should be somehow tuned to the structure of the language. This, however, is a purely empirical question, and the following review will examine the relevant evidence.

A large part of the findings reported will be from Hebrew or Arabic, for two main reasons. First, visual word recognition in Semitic languages has been examined extensively, and a large database is available from these languages. Second, and more important, in the present context language is considered as a factor akin to an experimental manipulation, in which important variables are held constant by the experimenter, and only a few are manipulated to pinpoint their impact on orthographic processing. For example, recent studies from Korean (Lee & Taft, 2009, 2011) suggest that letter-transposition effects are not obtained in the alphabetic Hangul[10] as they are in European languages. However, because the Korean Hangul is printed as blocks, where phonemes are spatially clustered both horizontally and vertically, the characteristics of the visual array are different than those of European languages. The significant advantage of Semitic languages is that they have an alphabetic system like English, Spanish, Dutch, or French, and from a purely orthographic perspective, they are based on the processing of letter strings just like European languages are. Hence, what is held constant is the superficial form of the distal stimulus on which the processing system operates. What is "manipulated" are the underlying or "hidden" linguistic characteristics of the orthography, which determine the ecological valence of the constituent letters. The following review will center on whether the underlying linguistic characteristics of the orthography affect the basic processing of orthographic structure.

## 4. Orthographic processing in Semitic languages

Although Hebrew writing system is not different than any alphabetic orthography, the striking finding is that the benchmark effects of orthographic processing which are revealed in European languages, such as form-orthographic priming, and most important to our discussion, letter-position flexibility, are not obtained in Hebrew, nor are they in Arabic.

---

[9] The phenomenon of reading Hebrew without understanding it is still common in religious Jewish communities outside of Israel, and indeed, this is still done with pointed text.

[10] The Korean orthography has two writing systems: one that is alphabetic -- *Hangul*, and another that is logographic (with many characters imported from Chinese) --*Hanza*.

**Form orthographic priming—**In a series of eight experiments in Hebrew and one in Arabic, Frost, Kugler, Deutsch, & Forster (2005) examined whether almost full orthographic overlap between primes and targets in Hebrew results in masked orthographic priming, as it does in English (e.g. Forster et al., 1987, Forster & Davis, 1991), French (e.g., Ferrand & Grainger, 1992), Dutch (Brysbaert, 2001), and Spanish (Perea & Rosa, 2000). The results were negative. None of the experiments produced significant priming effects either by subjects or by items. Especially revealing were two experiments that involved bilingual participants. In these two experiments, Hebrew-English and English-Hebrew bilinguals were presented with form-related primes and targets in Hebrew and in English. When tested in English, these bilingual speakers indeed demonstrated robust form priming. However, in both experiments, no such effect was obtained when these same subjects were tested with Hebrew material (and see Velan & Frost, 2011, for a replication). These findings lead to two dependent conclusions. First, the lexical architecture of Hebrew probably does not align, store, or connect words by virtue of their full sequence of letters. Second, the orthographic code generated for an alphabetic language such as Hebrew does not seem to consider all of the constituent letters (Frost, 2009). Indeed, considering the overall body of research using masked priming in Semitic languages, reliable facilitation is consistently obtained whenever primes consist of the *root letters*, irrespective of what the other letters are (e.g., Frost, Forster, & Deutsch, 1997; 2000; Velan, Frost, & Plaut, 2005; Perea, Abu Mallouh, & Carreiras, 2010). This clearly suggests that the orthographic coding scheme of Hebrew print focuses mainly on the few letters that carry morphological information, whereas the other letters of the word do not serve for lexical access, at least not initially.

**Letter-position flexibility—**This is the crux of the present discussion, since letter-position flexibility is supposed to reflect the manner by which the brain encodes letters for the reading process. A large body of research has examined letter position effects in Semitic languages, reaching unequivocal conclusions: the coding of Hebrew or Arabic letter position is as rigid as can be, as long as words are root-derived.

The first demonstration of letter-coding rigidity was reported by Velan and Frost (2007). In this study, Hebrew-English balanced bilinguals were presented with sentences in English and in Hebrew, half of which had transposed-letter words (three jumbled words in each sentence) and half of which were intact. The sentences were presented on the screen word-by-word, via RSVP (Rapid Serial Visual Presentation), so that each word appeared for 200 ms. Following the final word, subjects had to vocally produce the entire sentence. The results showed a marked difference in the effect of letter-transposition in Hebrew compared to English. For English materials, the report of words was virtually unaltered when sentences included words with transposed letters, and reading performance in sentence with and without jumbled letters was quite similar. This outcome concurs with all recent findings regarding letter-position flexibility reported in English or other European languages (e.g., Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004; Duñabeitia et al., 2007; Perea & Carreiras, 2006a,b, 2008). For Hebrew materials, on the other hand, letter-transpositions were detrimental to reading, and performance in reading sentences that included words with jumbled letters dropped dramatically.

Perhaps the most revealing finding of the Velan & Frost (2007) study concerns subjects' ability to *perceptually detect* the transposition of letters in Hebrew vs. English, as revealed by the sensitivity measure *d'*. At the rate of presentation of 200 ms per word in RSVP, subjects' sensitivity to detection of transposition with English material was particularly low (*d'*=0.86). Moreover, about one third of the subjects were at chance level in perceiving even one of the three transpositions in the sentence. In contrast, subjects' sensitivity to detecting the transposition with Hebrew material was exceedingly high (*d'*= 2.51), and not a single subject was at chance level in the perceptual task. Since *d'* taps the early perceptual level of processing, this outcome suggests a genuine difference in the characteristics of orthographic processing in Hebrew vs. English.

The substantial sensitivity of Hebrew readers to letter-transpositions raises the question whether the typical TL priming effects obtained in European languages are obtained in Hebrew. The answer seems, again, straightforward. Hebrew TL primes do not result in faster target recognition relative to letter substitution, as is the case for English, Dutch, French, and Spanish. More important, if jumbling the order of letters in the prime results in a letter-order that alludes to a different root than that embedded in the target, significant inhibition rather than facilitation is observed (Velan & Frost, 2009). This double dissociation between Hebrew and European languages regarding the effect of letter-transposition clearly suggests that letter-position encoding in Hebrew is far from being flexible. Rather, Hebrew readers display remarkable rigidity regarding letter order (and see Perea et al., 2010, for similar results in Arabic).

The extreme rigidity of letter encoding for Semitic words stems from the characteristics of their word structure. Hebrew has about 3000 roots (Ornan, 2003), which form the derivational space of Hebrew words. Since these tri-consonantal entities are conveyed by the 22 letters of the alphabet, for simple combinatorial reasons, it is inevitable that several roots share the same set of three letters. To avoid the complications of homophony, Semitic languages alter the order of consonants to create different roots. Typically, three or four different roots can share a cluster of three consonants (and thereby three letters), so that it is rare for a set of three consonants to represent a single root. For example, the consonants of the root S.L.X (“to send”) can be altered to produce the root X.L.S (“to dominate”), X.S.L (“to toughen”), L.X.S (“to whisper”), and S.X.L (”lion”). If the orthographic processing system has to pick-up the root information from the distal letter-sequence, letter-order cannot be flexible, but has to be extremely rigid. Moreover, for a system to efficiently differentiate between roots sharing the same letters but in a different order, inhibitory connections must be set between different iterations of letters, each of which represents a different meaning.

The results from Hebrew and Arabic have major implications for any theory of orthographic processing. Findings from Semitic languages presented so far demonstrate that the cognitive system may perform on a distal stimulus comprising a sequence of letters, very different types of processing, depending on factors which are unrelated to peripheral orthographic characteristics, but related to the deep structural properties of the printed stimuli. These concern firstly the morphological (and therefore the semantic) contribution of individual letters to word recognition. For Indo-European languages, individual letters of *base-words* do not have any semantic value. Since models of reading today are exclusively Anglo-centric, not surprisingly, this factor has never really been taken into account. A theory of reading, however, that is linguistically coherent, must include some parameters that consider the semantic valence of individual letters in order to satisfy the universality constraint.

We should note that these conclusions by no means imply that the *neuro-circuitry of the visual system* operates on different principles for Hebrew than it does for English or Spanish. Temporal firing patterns due to the sequential array of letters (Whitney, 2001), or noisy retinotopic firing (Dehaene et al., 2005), are probably shared by all printed forms in all languages. The conclusion so far is that these characteristics of the neural system are independent of lexical processing and do not come into play during the coding of orthographic information. Hence, they should not be a component of our theory of reading.

Perhaps the most convincing demonstration that orthographic processing and the coding of letter position in alphabetic orthographies is entirely dependent on the type of information carried by individual letters can be shown, again, in Semitic languages. Both Hebrew and Arabic have a large set of base words that are morphologically simple; meaning, they do not have the typical Semitic structure since they are not root-derived and thus resemble words in European languages. Such words have infiltrated Hebrew and Arabic throughout history from adjacent linguistic systems such as Persian or Greek, but native speakers of Hebrew or Arabic are unfamiliar with their historical origin. The question at hand is what is the nature of their orthographic processing? From the present perspective, the different types of words (Semitic root-derived vs. non-Semitic, non-root derived) are taken as an experimental factor, where both the alphabetic principle and the language are held constant, and only the internal structure of the distal stimulus is manipulated.

In a recent study, Velan & Frost (2011) examined the benchmark effects of orthographic processing when these two types of words are presented to native speakers of Hebrew. The results were unequivocal: morphologically simple words revealed the typical form priming and TL priming effects reported in European languages. In fact, Hebrew-English bilinguals did not display any differences in processing these words, and processing English words. In contrast, whenever Semitic words were presented to the participants, the typical letter-coding rigidity emerged. For these words, form priming could not be obtained, and transpositions resulted in inhibition rather than in facilitation. These findings demonstrate that flexible letter-position coding is not a general property of the cognitive system, nor is it a property of a given language. In other words, it is not the coding of letter position that is flexible, but the reader’s strategy in processing them. Therefore, structuring a model of reading so that it produces flexible letter position coding across-the-board does not advance us in any way towards understanding orthographic processing or understanding reading. The property that has to be modeled, therefore, is not letter-position flexibility, but rather flexibility in coding letter position, so that in certain linguistic contexts it would be very rigid and in others it would be less rigid. Only this approach would satisfy the universality constraint.

So far, I have established an evident flexibility of *readers* in terms of whether or not to be flexible about letter-position coding. Two questions, however, remains to be discussed so that our theoretical approach can maintain both descriptive and explanatory adequacy. First, what in the distal stimulus determines *a-priory* flexibility or rigidity in coding its letter positions? Second, why is flexible or rigid coding advantageous in different linguistic contexts?

## 5. Word structure determines orthographic processing

After demonstrating that even within language, the cognitive system performs different operations on a sequence of letters, given the deep structural properties of the printed stimuli, what remains to be explicated is what cues trigger one type of orthographic processing or another. The answer seems to lie in the structural properties of the sequence of letters that form base words.

European languages impose very few constraints on the internal structure of base words. For example, word onsets can consist of any consonant or any vowel, and since the permissible syllables are numerous, in principle, phonemes could be located in any position within the spoken word and at equal probability. There are very few phonotactic and articulatory constrains on the alignments of phonemes (such as no /p/ after /k/, in English). Although onset-rime structure determines word structure to some extent, at least in English (see Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Kessler & Treiman, 1997), the predictive value of a given phoneme regarding the identity of the subsequent one is relatively low. To exemplify, *comet* is a word in English, and *bomet* is not, but it could have been otherwise, and the word for *comet* could have been, in theory, “*temoc”, “tomec”,* *"motec"*, "*omtec*", or "*cetom*", and so on. Since letters in European languages represent phonemes, all of the above applies to written forms as well.

Semitic words, spoken or printed, are very different than European-based ones because they are always structured with a relatively small number of word-patterns. Word-patterns in Hebrew or Arabic have very skewed probabilities regarding phoneme sequences, so that Semitic words present to speakers and readers a set of *transitional probabilities*, where the probability of a phoneme or letter in a given slot depends on the identity of the previous phoneme or letter. Using the above example of the word /comet/, the game of possible theoretical iterations of phonemes in Hebrew is constrained mainly to the root-consonantal slots because all base words with the same word-pattern differ only in the sequence of root-consonants (e.g., tizmoret, tiksoret, tisroket, tifzoret, tikrovet, tirkovet, and so on, where root letters are underlined). The substantial difference in the structural properties of words in Semitic and European languages has immediate implications for orthographic processing because the uptake of information from the distal stimulus is necessarily shaped by stimulus complexity. For Semitic words, the most relevant information is the three letters of the root, and the other letters of the word assist in locating these. For English, the game of "cracking" the distal stimulus is quite different. All letters have a more or less similar contribution to word identity, the function of the individual weight of each letter to correct identification is more or less flat, and the significance of each letter to the process varies with the number of letters, depending on the position of those letters within onset-rime linguistic units.

To account for the difference in orthographic coding of "English-like" and "Hebrew-like" words, our question thus concerned the possible cues that could govern one type of orthographic processing or another: the "English-like" coding system which considers all letters equally and is flexible regarding their position, and the "Hebrew-like" coding system, which focuses on a specific subset of letters and is rigid regarding their position. I suggest that the primary cue that determines the orthographic code is whether the distribution of letter frequency is skewed or not. In linguistic systems with skewed letter-frequency, such as Hebrew, the highly repeated word-pattern letters flash-out the few letters which carry distinctive information regarding root identity and meaning. In contrast, in linguistic systems in which letters do not predict other letters, and the distribution of transitional probabilities of letters is more or less flat, orthographic coding considers all letters, focusing on their identity rather than on their position.

This account suggests that for efficient reading, the statistical properties of letter distributions of the language, and their relative contribution to meaning, have to be picked-up, and the transitional probabilities of letter sequences have to be implicitly assimilated. In the case of Hebrew, this is achieved following the repeated exposure to Semitic words that are root-derived, vs. non-Semitic words that do not have the same internal structure. These implicit learning procedures are entirely contingent on the exposure to the spoken language, and possible suggestions of how this is done will be outlined later on. At this point, however, a cardinal conclusion regarding the main characteristic of a universal model of orthographic processing is already emerging. For a model to produce different behavior as a function of the statistical properties of the language, the model has to be able to pick up the statistical properties of the language.

## 5. Advantages and disadvantages of flexible and rigid letter coding

The new age of modeling orthographic processing has focused mainly on the question of "how", that is, how does the cognitive system produce letter-position flexibility? The relative high number of such models of visual word recognition shows that there are many computational solutions to the problem. The "why" question is deeper and more fundamental because our aim in modeling is to eventually understand reading rather than to simply describe it. To a large extent, the explanation offered by the current wave of models for letter-position flexibility is, in simple terms, that *this is the way the brain works*. However, once we have established that it is only the way the brain works for languages like English, French, or Spanish, the question at hand is what does letter-position flexibility buy in terms of processing efficiency. Operating within an ecological approach, the answer to this question will focus again on the specific interaction of the reader and the linguistic environment.

For Semitic languages such as Hebrew, orthographic lexical space is exceedingly dense. If all words are derived using a relatively small number of phonological patterns, and all words sharing a word-pattern share a skeletal structure of phonemes (and therefore letters), words are differentiated only by the three root consonants (or letters). This necessarily results in a dense lexical space, in the sense that the large variation of words that are necessary for meaningful communication is created mainly by manipulating *the order* of few constituent phonemes. Thus, in Hebrew, spoonerism very often results in other lexical candidates, and words sharing the same set of letters but in a different order are the rule, rather than the exception.

The interesting question, therefore, is not why orthographic processing in alphabetic languages such as Hebrew is exceedingly rigid in terms of letter position. It could not have been otherwise, as it must fit the structure of Hebrew lexical space. The interesting question is why it is flexible in English. Why wouldn't the brain rigidly encode letters in all orthographies? What is gained by letter-position flexibility? The answer is probably two-fold. First, languages such as English, which are not constrained to a small set of phonological word-patterns, create a variation of words by aligning, adding or substituting phonemes, and not by changing their relative position, as Semitic languages do. Thus, anagrams such as "calm-clam", "lion-loin", which are the rule for Hebrew and occur in Hebrew words of any length, exist mostly for very short words of three to five letters in English (e.g., Shilcock et al., 2000). This feature is not exclusive to English, but is shared by other European languages, because unnecessary density of lexical space is not advantageous for fast discrimination between lexical candidates. Thus, the option of assigning to different base-words a different set of phonemes rather than changing their order seems to reflect a natural selection. Considering reading, this feature of the language has an obvious advantage when letter sequences have to be recognized, because most words are uniquely specified by their specific set of letters, irrespective of letter-order. Hence, a transposed-letter word such as *JUGDE* can be easily recognized since there are no word competitors that share the same set of letters. This type of linguistic environment can naturally allow for noisy letter-position processing. In general, the longer the words are, the higher the probability that they would have a unique set of letters. Consistent with this assertion, TL priming effects in European languages are indeed largely modulated by word length, with large benefit effects for long words and small effects for short words (e.g., Humphreys et al., 1990; Schoonabert & Grainger, 2004). In the same vein, Guerrera and Forster (2008) have shown that with relatively long words, even extreme jumbling of letters in the prime (only two out of eight letters are correctly positioned), targets recognition is facilitated, so that the prime SNAWDCIH primes SANDWICH. Since not a single word shares with SANDWICH this specific set of letters, SNAWDCIH can produce a priming effect.

An illustration of the behavioral implications of the major differences in the structure of lexical space between Hebrew and English is shown in patients with deficits in registering the position of letters within the word. According to Ellis & Young (1988), three distinct functions are relevant to peripheral disorders related to visual analysis of print: letter identification (or letter agnosia; e.g., Marshal & Newcomb, 1973), letter-to-word binding (letter migration problems; e.g., Shallice & Warrington, 1977), and encoding of letter position. In Letter Position Dyslexia (LPD) patients demonstrate deficits in registering the position of each letter within the word. Friedmann and her colleagues (Friedmann & Gvion, 2001, 2005; Rahamin & Friedmann, 2008) have reported several cases of Hebrew speaking patients with both acquired and developmental LPD. Interestingly, Friedmann and her colleagues have argued that pure cases of LPD are rarely reported in Indo-European languages, whereas in Hebrew they are much more prevalent. Obviously, this does not have to do with differences in brain architecture of Hebrew vs. English speakers. Rather, this has to do with the characteristics of lexical space. A case study reported by Friedmann (Shetreet & Friedmann, 2011) demonstrates this elegantly. The patient, a native speaker of English, complained about reading difficulties following an ischemic infarct. Reading tests in English could not reveal why, because the patient's reading performance was close to normal. Only when tested with Hebrew material, was a diagnosis of LPD confirmed. Since in Hebrew errors in letter position mostly result in another word, the patient had significant difficulties in reading Hebrew. In English, however, errors in letter-position seldom result in another word; LPD, therefore, did not hinder his reading significantly. Recently, Friedmann & Haddad-Hanna (in press) have reported four cases of LPD in adolescent Arabic speakers and have also described the characteristics of developmental LPD in young Hebrew readers (Friedmann, Dotan, & Rahamim, 2010), demonstrating again how LPD significantly and consistently hinders reading in Semitic languages.

If words in a language, in most cases, do not share their set of letters, and changes in letter-order do not typically produce new words, then orthographic processing can relax the requirement for rigid letter position coding without running the risk of making excessive lexicalization errors. This relaxation has a major advantage in terms of cognitive resources. Given the fast saccades during text reading, some noise must exist in registering the exact sequence of letters. A fine-grained coding system that requires overriding such natural noise is more costly in terms of cognitive resources than a coarse-grained system that is indifferent to noise (see Grainger & Ziegler, 2011, for a similar taxonomy). Note, however, that the "noise" described in the present context is not hardwired within the perceptual system (i.e., noisy retinotopic firing, etc.). Rather, it is an environmental noise, tied to the characteristics of print, where long sequences of letters are aligned one next to the other and are scanned and registered at a very fast rate. Relaxing the requirement for accurate letter-position registering without consequent damage to lexical access has a clear advantage, and is, therefore, an emergent property of skilled reading. By this view, beginning readers who are learning to spell, and have difficulties in letter decoding should not display such flexibility. I will further expand on the feature of cognitive resources later on.

The sum of these arguments leads us then to the same conclusion: letter-coding flexibility in reading is not a characteristic of the brain hardware, as current models of orthographic processing seem to suggest. It is a cognitive resource-saving strategy that characterizes reading in European languages, given the characteristics of their lexical space. Models of orthographic processing have to account for this.

The following discussions will thus center on a new approach to modeling visual word recognition. The aim of these discussions is to outline the blueprint principles for a universal model of reading that would be linguistically plausible and linguistically coherent.

## 7. Structured models vs. learning models

A common feature of most current models of orthographic processing is that they are explicitly structured to predict certain behaviors. Thus, the modelers shape their model's architecture so that its output will generate a desired outcome, in the present context, form priming or letter-position flexibility. This can be done, for example, by introducing open bigram units into the model (e.g., Grainger & van Heuven, 2003; Dehaene et al., 2005; Whitney, 2001), or by lagging the information about letter order relative to letter identity (e.g., Gomez et al., 2008). Although this approach has substantial merits in generating hypotheses regarding how specific behaviors can be produced, almost inevitably it leads modelers to focus on a narrow set of phenomena, constraining their models to deal with one dimension of processing. As I have argued at length so far, within the domain of language, this approach can be detrimental. Since reading behavior is shaped and determined by the complex linguistic environment of the reader, focusing on specific computations within the system would most probably miss the perspective of the context of these computations, thereby leading to possible misunderstandings regarding the origin of behavior which is being modeled.

To exemplify, any of the current models of orthographic processing could easily re-produce the obtained effects of Hebrew Semitic words by implementing slight modifications. Suffice to introduce tri-literal root units and set inhibitory connections between roots sharing sets of letters in a different order, and rigid letter position would show-up. This approach may yield eventually a computational model of reading Hebrew Semitic words with a relatively good fit for the Hebrew data, but the explanatory benefit of the model remains questionable. It should be emphasized that the theoretical value of a model is independent of the prevalence of the language that is being modeled. In the present context, the theoretical contribution of a similar model of reading English would be exactly the same. The conclusion that emerges from the present discussion is that “structured” models that are explicitly set to produce effects of orthographic processing, in all probability, will not be universal. Their chances of satisfying the universality constraint and achieving a full linguistic coherence are low. Following the example above, if, say, letter-order rigidity would be hardwired into a model of reading Hebrew, and the model would then indeed display the desired behavior in terms of root priming, this model would then not display the opposite effects with English-like words. That is, the model will be language-specific, or even worse, it would be “sub-language” specific, in the sense that it would simulate reading of a subset of words in Hebrew -- not even all types of words.

In contrast to structured models, learning models are set to pick-up the characteristics of the linguistic environment by themselves. A typical example is the influential model of past-tense production offered by Rumelhart and McClelland (1986). The dramatic impact of this model was in its demonstration that, both rule-like behavior (regular forms of past-tense) and non rule-like behavior (exception forms of past-tense) can be produced by training a network on a representative corpus of the language environment of children. In learning models, behavior *emerges* rather then being structured. The approach of these models focuses on the statistical regularities in the environment and on the way that these are captured by the model and shape its behavior, either through supervised or through unsupervised learning (See Rueckl, 2010, for an detailed review).

The aim of the present discussion is not to promote a connectionist approach. Connectionist models have been criticized, and rightly so, for simulating only monosyllabic words in English, but a debate regarding the merits and shortcomings of connectionism is beyond the scope of the paper. There is, however, an important analogy between the approach that lies at the heart of the architecture of learning models and what we know about learning language in general, and learning to read in particular. The internal structure of words, which determines orthographic processing, is not explicitly taught to native speakers. Similarly, lexical organization of words in any language implicitly emerges so that, for any language, orthographic codes would be optimal given the language’s phonological space, and how it represents meaning and morphological structure. This is a reading universal, and therefore it is an emergent property of the reading environment, which is picked up by readers through simple implicit statistical learning and by explicitly learning to spell. Considering Hebrew for example, Frost, Narkiss, Velan, & Deutsch (2010) have shown that sensitivity to root structure in reading, as revealed by robust cross-modal morphological priming effects, can be demonstrated already in the first grade, when children have no clue regarding the formal morphological taxonomies of their language and what governs the internal structure of printed letters. Moreover, Frost et al., (2010) have shown that English speakers who learn to read Hebrew as L2 display at the onset of learning the typical characteristics of orthographic processing in European languages. With a time course of less than a year, they assimilate the statistical properties of the language and capture the root structure of Semitic languages showing the same effects as Hebrew readers.

The earlier description of the five languages that opened the present discussion demonstrates that languages differ in a rich array of statistical properties. These concern the distribution of orthographic and phonological sublinguistic units, their adjacent and non-adjacent dependencies, the correlations between graphemes and phonemes (or syllables), the correlation between form and meaning, etc. Native speakers pick up these dependencies and correlations implicitly, without any need for formal instruction, presumably through pure procedures of statistical learning. An illuminating example of how the statistical properties of the language are implicitly assimilated comes from the logographic Chinese. Although Chinese is not an alphabetic language, and reading Chinese characters involves the recognition of a complex visual pattern, by reviewing a large corpus of behavioral and ERP studies, Lee (2011) demonstrates that Chinese readers are sensitive to the statistical mapping of orthography to phonology in their language. Lee offers a statistical learning perspective to account for Chinese reading and reading disabilities, by considering the distributional properties of phonetic radicals. Overall, the findings reported by Lee suggest that Chinese readers extract from their linguistic environment information regarding the consistency of character and sound and use it in the reading process (see also, Hsu, Tsai, Lee, & Tzeng, 2009).

The robust power of statistical learning has been demonstrated in a large number of studies, both with linguistic and nonlinguistic stimuli (e.g., Saffran, Aslin, & Newport, 1996; Gebhart, Newport, & Aslin, 2009; Endress & Mehler, 2009; Evans, Saffran, & Robe-Torres, 2009; Perruchet & Pacton, 2006). Typically these studies show that adults, young children, or even infants rapidly detect and learn consistent relationships between speech sounds, tones, or graphic symbols. These relationships can involve adjacent as well as non-adjacent dependencies (e.g., Gebhart et al, 2009; Newport & Aslin, 2004). Our discussion so far has shown that writing systems have evolved to condense maximal phonological and semantic information about their language using minimal orthographic units, and that the cognitive system learns to pick-up from the distal stimulus this optimal level of information, given the structure of the language. A model of reading that is set to operate on similar principles has, therefore, the potential to satisfy the universality constraint and be linguistically coherent, but mostly, it has the important advantage of having ecological validity. Our theory of skilled reading cannot be divorced from our theory of how this skill is learnt, and skilled visual word recognition reflects a long learning process of complex linguistic properties. If the model indeed picks-up the statistical regularities of the language and the expected reading behavior emerges, it most probably reflects the actual learning procedures of readers.

It should be emphasized that the aim of the present discussion is not to contend that learning models provide simple solutions for testing hypotheses regarding the statistical regularities that are picked up by readers in the course of literacy acquisition. Proponents of structured models would rightly argue that learning models are structured as well, in the sense that they posit an input-coding scheme, which determines to a great extent what will actually be learnt by the model. From this perspective, learning models, like structured models, also hardwire distinct hypotheses regarding the form of input that serves for the learning process. Although this argument has some merit, some critical distinctions regarding the utility of the two modeling approaches in the context of understanding reading can, and should, be outlined.

First, structured models and learning models differ in the ratio of the scope of phenomena that are produced by the model versus the amount of “intelligence” that is put into the model - narrow scope with maximum intelligence for structured models vs. large scope with minimum intelligence for learning models. This has important implications regarding what the modeling enterprise can actually teach us. When the model’s behavior is too closely related to the intelligence that is put into it, little can be learnt about the source of the behavior that is produced by the model. Second, at least in the context of visual word recognition, the modeling approaches differ in the scope of the linguistic theory that determines the *content* of intelligence that is put into the model. As explicated at length, the phenomena that constrain the architecture of current structured models of visual word recognition are by definition narrow in scope, as they concern only orthographic effects. Third, and perhaps most important, there is a major difference in the *goal* of the modeling enterprise. The aim of learning models is to learn something about the possible behaviors emerging from the model, given specific input-coding schemes. The visible representations do not do all the work. A learning model would not work right if the environment did not have the right structure and if the learning process could not pick up this structure. More broadly, the emphasis on learning requires that our theory pay attention to the structure of the linguistic environment, and a failure in the modeling enterprise has, therefore, the potential of teaching us something deeply theoretical (see Rueckl & Seidenberg, 2009, for a discussion). In contrast, the aim of structured models is often to produce a specific behavior, such as letter-position flexibility. This approach almost inevitably results in circularity. The hardwired behavior in the model is taken to reflect the mind’s, or the brain’s, circuitry, and consequently, the model’s organization is taken as behavioral explanation. Thus, as a rule of thumb, the distance between organization and explanation is by far narrower in structured models than in learning models. To exemplify, once a level of open bigrams that is hardwired in the model is shown to reproduce the desired typical effects of letter position flexibility, letter position flexibility is suggested to emerge from a lexical architecture that is based on open bigrams (e.g., Whitney, 2001; Grainger & Whitney, 2004). This circularity between organization and explanation may have detrimental consequences for understanding the fundamental phenomena of reading.

## 8. The linguistic dimensions of a universal model

As explicated at length in the initial part of our discussion, the assumption underlying most current computational models of orthographic processing is that the game of visual word recognition is played in the orthographic court, in the sense that an adequate description of the cognitive operations involved in recognizing printed words are constrained solely by the properties of orthographic structure (letters or characters, letter identity, letter location, letter sequences, etc.). However, as we have shown in the description of various writing systems, orthographic structure is determined by the phonological space of the language and the way it represents morphology and meaning. This means that letters in different languages might provide different type of information, and this must be part of a universal model of reading. Conceptual models that offer a generic lexical architecture (e.g., the bi-modal interactive activation model, Grainger & Ferrand, 1994; Diependael, Ziegler, & Grainger, 2010) do include phonological and semantic representations, but when it comes to produce a *computational* model, the focus is on orthographic entities per-se. The inherent limitation of this approach is that it inevitably leads to linguistic implausibility. This problem was recently outlined by Grainger & Ziegler (2011), who argued that orthographic processing should also be constrained by correlations of letter clusters with phonological units, and by the fact that prefixes and suffixes are attached to base words and need to be correctly detected as affixes in the process of word recognition. Thus, mapping of letter clusters such as “sh”, or “th” into phonemes, as well as affix stripping (i.e. teach-er), both required for the morphological decomposition that is necessary for base word recognition, cannot allow flexibility of letter order. The solution offered by Grainger & Ziegler (2011) was to include in a model of orthographic processing two types of orthographic codes, differing in their level of precision: one that is coarse-grained and allows for fast word recognition, and one that is fine-grained and precise and allows for correct print-to-sound conversion, and correct morpho-orthographic segmentation.

Grainger & Ziegler (2011) were right on target in realizing the severe limitation of the present approach to modeling orthographic processing. This approach has inevitably led to impoverished and narrow theories of correctly recognizing orthographic forms of English morphologically-simple base-words that are more than four letters long. However “patching-up” models by appending to them parallel computational procedures, that do whatever the original computational procedures did not do, is hardly a solution. It may conveniently fix the model’s inevitable problems, but more than fixing them, it demonstrates the basic conceptual weakness of the modeling approach. Such a strategy eventually leads to explaining *any* possible finding simply by arguing *post-hoc* that one route rather than the other was probably used, and this would hardly advance our understanding of reading. The question to be asked is why current computational approaches result in impoverished solutions, describing only a very limited set of phenomena. This leads our present discussion to the linguistic dimensions, which are necessary for a universal model of orthographic processing; necessary, in the sense that the model would satisfy the requirement for linguistic plausibility.

## Three basic dimensions: orthography, phonology, and semantics

Our above description of the logic underlying the evolution of the five contrasting writing systems -- Chinese, Japanese, Finnish, English, and Hebrew -- suggests an intricate weighting of *both* phonological and semantic factors affecting the structure of *orthographic* forms in a language, so that these convey optimal phonological and morphological information to the reader. Assuming that readers are tuned to pick-up and extract from the print this linguistic information, phonological and semantic considerations must be part of a universal model of orthographic processing. It should be emphasized that in the present context, semantic features and phonological structure are not taken as higher levels of representation into which orthographic letters are mapped. This view is common to *all* current models of reading. The claim here goes deeper, suggesting that the actual computation of an orthographic code in a given language is determined *on-line* by the transparency of mapping of graphemes into phonemes on the one hand, and by morphological and semantic considerations on the other hand, given the language properties in which reading occurs. To some extent, a similar approach is advocated by “triangular models” which describe reading in terms of a division of labor between the mapping of orthography to both phonology and semantics, and propose that such mappings are learned via associative mechanisms sensitive to the statistical properties of the language (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Harm, & Seidenberg, 1999, and see Rueckl, 2010 for a review).

The claim that phonological and morphological considerations affect orthographic processing is not only theoretical, but also has empirical support. Hebrew, for example, clearly demonstrates why morphological and semantic considerations must be part of a theory of orthographic processing. The different orthographic codes that were shown to be involved in processing Semitic vs. “English-like” base words (Velan & Frost, 2011) are entirely predetermined by the semantic value of the individual letters, that is, whether they belong to a root morpheme or not. Thus, the initial cognitive operation that Hebrew readers launch when presented with a letter-string is a search for meaningful letters that are dispersed within the word, meaningful in the sense that they convey root information. This process of searching for noncontiguous meaningful letters is early, prelexical, and can be easily demonstrated by monitoring eye-movements. For example, the optimal viewing position (OVP), the position in a word where word identification is maximal, is entirely modulated by the location of the first root letter within the word (Deutsch & Rayner, 1999). More important, Hebrew readers have been shown to search for the root letters already parafoveally. Thus, presenting the root information in the parafovea (see Rayner, 1998, 2009, for a review of parafoveal presentation and the boundary technique) results in robust parafoveal preview benefit effects, either with single word reading (Deutsch, Frost, Pollatsek, & Rayner, 2000), or during sentence reading (Deutsch, Frost, Peleg, Pollatsek, & Rayner, 2003). Interestingly, similar morphological manipulations conducted with English readers did not result in any parafoveal preview benefit (Rayner, Juhasz, White, & Liversedge, 2007, and see Rayner, 2009, for a discussion of the lack of morphological preview benefit in English). The evidence from eye-movements in Hebrew is especially compelling, since it reflects the initial phases of orthographic processing that are below the level of awareness and not governed by any conscious strategy. The contrasting findings of English vs. Hebrew regarding eye-movements demonstrate that 1) the prelexical search for letters carrying morphemic information in the parafovea is language specific, that 2) different orthographic operations are performed on a distal stimulus composed of letter-sequences in different languages, and that 3) *individual* letters are differentially processed across the visual array, given their morphological status and their contribution to recovering semantic meaning.

However, from an even more general perspective, the results from Hebrew seem to reveal an important reading universal. They exemplify the perfect fit between the optimization of information that the writing system has evolved to convey, and the cognitive operations that are launched to pick-up that information. Recall that this was the starting point of the present discussion. As explicated, the Hebrew orthography was designed to be severely phonologically underspecified in order to emphasize morphological (and thereby semantic) information. The behavior of readers mimics this evolutionary design to perfection. Already in the parafovea, orthographic processing zooms in on the letters which are root letters, that is, letters that will lead as fast as possible to meaning, although this may be only a vague meaning. Thus, the main target of orthographic processing are letters that carry the highest level of diagnosticity, but not in terms of word-form, as in English, but in terms of the morphemic units from which the word is derived. This account also provides a coherent explication of why phonological computation in Hebrew is underspecified. As Frost and his colleagues have repeatedly shown, the prelexical phonological code computed in Hebrew is indeed impoverished (Gronau & Frost, 1997; Frost & Yogev, 2001; Frost, Ahissar, Gottesman & Tayeb, 2003). Although Frost and his colleagues focused on the depth of Hebrew orthography in discussing their results, it seems that their arguments should be expanded to include what seems to be a reading universal: The representation of morphological information takes precedence over the representation of detailed phonological information when it comes to the evolution of writing systems. Given that, it also takes precedence when it comes to the cognitive processing of orthographic structure by skilled readers. Thus, a universal model of reading that is a learning model must include, one way or another, an architecture which considers the intricate relations of orthography, phonology and morphology (and therefore of meaning) in the language.

### Incorporating parameters of cognitive resources

The notion of optimization of information and the allocation of optimal cognitive resources to orthographic processing provides an important explanatory dimension to the present theoretical approach. However, if as argued, an important universal principle of processing orthographic structure is a flexibility of the processing system, that is, whether to be flexible or not about letter coding, and if this flexibility (or the lack of it) is constrained by the cognitive resources that have to be allocated for processing (more resources for rigid slot coding, less for flexible), and if these constraints are determined by the statistical properties of the language, then cognitive resources should be an integral part of the model.

How to go about that is not evident, however recent computational work by Tishby and colleagues provides challenging potential conceptual solutions. In the past few years, Tishby, Bialek and colleagues have developed a general theoretical framework for calculating optimal representations for predictability of a stimulus, given the complexity of the environment, as a function of the resources that a system allocates (Tishby, Pereira, & Bialek, 1999; Bialek, Nemenman, & Tishby, 2001; Shamir, Sabato & Tishby, 2009). In a nutshell, the computational approach developed by Tishby and his colleagues considers, in parallel, information capacity, information rate, and limitation in overall resources, and consequently computes the optimization procedures for allocating minimal resources for a given processing event, to obtain best performance in terms of predicting stimuli, given the complexity of the environment. Tishby et al., (1999), and Tishby and Polani (2010), demonstrate how the precision of representations depends, among other things, on the complexity of the environment as characterized by its predictive information regarding future events. The parallel of this computational approach to orthographic processing seems compelling. Applying this framework to visual word recognition will require including a parameter of cognitive resources necessary for precise slot coding as compared with a coarse-grained one. This choice can then be optimized given the complexity of the linguistic environment, as reflected by the statistical properties of the language. How exactly to implement this form of computations in a universal learning model of reading obviously requires extensive investigation. However, since flexibility shlould be part of the model, then allocation of cognitive resources to modulate it, is a possible solution.

## 9. Summary and conclusions

The present paper discusses the recent paradigmatic shift in reading research, and the resulting new wave of computational models of orthographic processing that center on letter-position flexibility. The main claim is, that the extensive focus on insensitivity to letter order has led to a generation of models that are non-universal, lack linguistic plausibility, and miss the complexity of the reading process. My critique is based on a set of inter-related arguments as follows: the first step in formulating a theory of orthographic processing is to provide an accurate and full description of what type of information is provided by the orthographic structure. This information is not necessarily transparent and goes way beyond a surface description of letters, letter sequences, or letter location. It reflects the phonological space of the language and the way it represents meaning through morphological structure. The cognitive system is tuned to pick up this information in an optimal way by implicitly capturing the statistical regularities of the language, and registering the inter-correlations of orthography, phonology and morphology. This necessarily results in lexical organization principles that are language-dependent and allow readers optimal and differential performance in different linguistic environments. As a consequence, orthographic processing in one language may be quite different than in another language; moreover, qualitatively different computations may be found even within a single language. Thus, a universal theory of reading should focus on what is *invariant* in orthographic processing across writing systems.

This set of claims leads us to suggest what is fundamentally wrong with the current wave of modeling orthographic processing. I argue that most recent models examine the *surface form* of orthographic structure focusing on a *variant* characteristic, which is idiosyncratic to skilled reading in European languages: flexibility of letter-position coding. I suggest that, being a variant characteristic, this specific feature of processing letter sequences does not reflect in any way the manner by which the brain encodes orthographic information for lexical processing. Rather, it reflects a strategy of optimizing encoding resources in a highly developed or skilled system, given the specific structure of words in English, French, or Spanish.

This line of criticism brings us to a set of criteria for a universal model of reading that has linguistic plausibility and is linguistically coherent. Within this context, I outline the advantage of learning models, in that they have ecological validity. Learning models are set to pick up the statistical regularities underlying the *full linguistic environment* of the reader through implicit learning; similar to the way the cognitive system implicitly picks up the relevant information from the orthographic array in the language. In parallel, I outline the pitfalls of “structured” models which hardwire a given behavior in the model. Structured models almost inevitably result in circularity when the model’s organization is taken as behavioral explanation. Structured models of orthographic processing also run the risk of mistaking a variant behavior for an invariant one thereby hardwiring it into the model. The only way of allowing additional flexibility in processing in a structured model is to assume a duality of processing in the form of parallel routes that permit opposing computations. This strategy of “patching-up” structured models does not advance us in any way in understanding human behavior. Rather, it sets us back, leading inevitably to explaining any possible finding by reverting to post-hoc argumentation.

Regarding the dimensions that need to be part of a model of orthographic processing, they should mirror the dimensions that determine orthographic structure in a language. The logic underlying this claim is that writing systems do not evolve arbitrarily, and their manner of packing and optimizing information must reflect the cognitive procedures by which readers unpack and decode this information. I have shown that only by considering phonological space and morphological structure, can a full account of orthographic structure be provided. Similarly, orthographic processing must be tuned to both the phonological and the morphological information that graphemes carry. Thus, the only viable approach to modeling visual word recognition is an approach that considers, simultaneously, the full statistical properties of the language, in terms of covariations between orthographic, phonological, semantic, and morphological sublinguistic units. Our cognitive system is, first of all, a correlation-seeking device. Hence, universal models of reading should be structured to “pick up” covariations and conditioned probabilities that exist between all of the language components.

## Acknowledgments

This paper was supported in part by The Israel Science Foundation (Grant 563/07 awarded to Ram Frost), and by the National Institute of Child Health and Human Development (Grant HD-01994 awarded to Haskins Laboratories). I am indebted to Jay Rueckl, and Kathy Rastle for their suggestions and advice, and to Steve Frost, Marcus Taft, Asher Cohen, Keith Rayner, Jeff Bowers, Sachiko Kinoshita and two anonymous reviewers, for their insightful comments on previous versions of this manuscript.

## References

Adelman JS, Marquis SJ, Sabatos-DeVito MG. Letters in words are read simultaneously, not in left-to-right sequence. Psychological Science. 2011; 21:1799–1801. [PubMed: 21030682]

Bentin S, Frost R. Processing lexical ambiguity and visual word recognition in a deep orthography. Memory & Cognition. 1987; 15:13–23.

Bentin S, Hammer R, Cahan S. The effects of aging and first grade schooling on the development of phonological awareness. American Psychological Science. 1991; 2:271–274.

Bertelson P, Morais J, Alegria J, Content A. Phonetic analysis capacity and learning to read. Nature. 1985; 313:73–74.

Bertram R, Hyönä J. The length of a complex word modifies the role of morphological structure: Evidence from reading short and long Finnish compounds. Journal of Memory & Language. 2003; 48:615–634.

Bialek W, Nemenman I, Tishby N. Predictability, complexity, and learning. Neural Computation. 2001; 13:2409–2463. [PubMed: 11674845]

Borgwaldt SR, Hellwig FM, De Groot AMB. Word-initial entropy in five languages: Letter to sound, and sound to letter. Written Language & Literacy. 2004; 7:165–184.

Borgwaldt SR, Hellwig FM, De Groot AMB. Onset entropy matters: letter-to-phoneme mappings in seven languages. Reading and Writing. 2005; 18:211–229.

Bowers JS, Michita Y. An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script Kanji-Hiragana priming. Psychonomic Bulletin and Review. 1998; 5:259–264.

Bruner JS, O'Dowd D. A note on the informativeness of parts of words. Language and Speech. 1958; 1:98–101.

Brysbaert M. Prelexical phonological coding of visual words in Dutch: Automatic after all. Memory & Cognition. 2001; 29:765–773.

Chao, YR. Language and Symbolic Systems. Cambridge University Press; London: 1968.

Christianson K, Johnson RL, Rayner KI. Letter transpositions within and across morphemes. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2005; 31:1327–1339.

Coltheart M, Rastle K, Perry C, Ziegler J, Langdon R. DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review. 2001; 108:204–256. [PubMed: 11212628]

Cossu J, Shankweiler D, Liberman IY, Katz L, Tola G. Awareness of phonological segments and reading ability in Italian children. Applied Psycholinguistics. 1988; 9:1–16.

Davis, CJ. Unpublished doctoral dissertation. University of New South Wales: 1999. The Self-Organizing Lexical Acquisition and Recognition (SOLAR) model of visual word recognition.

Davis CJ. The spatial coding model of visual word identification. Psychological Review. 2010; 117:713–758. [PubMed: 20658851]

Davis CJ, Bowers J. What do letter migration errors reveal about letter position coding in visual word recognition? Journal of Experimental Psychology: Human Perception and Performance. 2004; 30:923–941. [PubMed: 15462631]

Davis CJ, Bowers J. Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. Journal of Experimental Psychology: Human Perception and Performance. 2006; 32:535–557. [PubMed: 16822123]

DeFrancis, J. Visible speech: The diverse oneness of writing systems. University of Hawaii Press; Honolulu: 1989.

Dehaene S, Cohen L, Sigman M, Vinckier F. The neural code for written words: a proposal. Trends in Cognitive Sciences. 2005; 9:335–341. [PubMed: 15951224]

Deutsch A, Rayner K. Initial fixation location effects in reading Hebrew words. 1999; 14:393–421.

Deutsch A, Frost R, Pollatsek A, Rayner K. Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. Language and Cognitive Processes. 2000; 15:487–506.

Deutsch A, Frost R, Peleg S, Pollatsek A, Rayner K. Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. Psychonomic Bulletin & Review. 2003; 10:415–422. [PubMed: 12921418]

Diependaele K, Ziegler J, Grainger J. Fast phonology and the bi-modal interactive activation model. European Journal of Cognitive Psychology. 2010; 22:764–778.

Dishon-Berkovitz M, Elgom D. The Stroop effect: It is not the robust phenomenon that you have thought it to be. Memory & Cognition. 2000; 28:1437–1439.

Duñabeitia JA, Perea M, Carreiras M. Do transposed-letter similarity effects occur at a morpheme level? Evidence for ortho-morphological decomposition. Cognition. 2007; 105:691–703. [PubMed: 17217942]

Ellis, AW.; Young, AW. Human cognitive neuropsychology. Lawrence Erlbaum Associates Ltd; Hove,UK: 1988.

Endress AD, Mehler J. The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. Journal of Memory and Language. 2009; 60:351–367.

Evans J, Saffran J, Robe-Torres K. Statistical learning in children with specific language impairment. Journal of Speech, Language and Hearing Research. 2009; 52:321–36.

Ferrand L, Grainger J. Phonology and orthography in visual word recognition: Evidence from masked nonword priming. Quarterly Journal of Experimental Psychology. 1992; 45A:353–372. [PubMed: 1308733]

Forster, KI. Accessing the mental lexicon. In: Wales, RJ.; Walk, E., editors. New Approaches to Language Mechanisms. North-Holland; Amsterdam: 1976. p. 257-287.

Forster, KI. Form-priming with masked primes: The best-match hypothesis. In: Coltheart, M., editor. Attention & Performance XII. Lawrence Erlbaum Associates; Hillsdale, N.J.: 1987. p. 127-146.

Forster KI, Davis C. The density constraint on form-priming in the naming task: Interference effects from a masked prime. Journal of Memory and Language. 1991; 30:1–25.

Forster KI, Davis C, Schocknecht C, Carter R. Masked priming with graphemically related forms: Repetition or partial activation? Quarterly Journal of Experimental Psychology. 1987; 39A:211–251.

Friedmann N, Gvion A. Letter position dyslexia. Cognitive Neuropsychology. 2001; 18:673–696. [PubMed: 20945233]

Friedmann N, Gvion A. Letter form as a constraint for errors in neglect dyslexia and letter position dyslexia. Behavioral Neurology. 2005; 16:145–158. [PubMed: 16410630]

Friedmann N, Dotan D, Rahamim E. Is the visual analyzer orthographic-specific? Reading words and numbers in letter position dyslexia. Cortex. 2010; 46:982–1004. [PubMed: 19853849]

Friedmann N, Haddad-Hanna M. Letter position dyslexia in Arabic: From form to position. Behavioural Neurology. (in press).

Frost R. Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1994; 20:1–16.

Frost R. Phonological computation and missing vowels: Mapping lexical involvement in reading. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1995; 21:398–408.

Frost R. Towards a strong phonological theory of visual word recognition: True issues and false trails. Psychological Bulletin. 1998; 123:71–99. [PubMed: 9461854]

Frost R. Becoming Literate in Hebrew: The Grain-Size Hypothesis and Semitic Orthographic Systems. Developmental Science. 2006; 9:439–440. [PubMed: 16911440]

Frost, R. Reading in Hebrew vs. Reading in English: Is There a Qualitative Difference?. In: Pugh, K.; McCradle, P., editors. How Children Learn To Read: Current Issues and New Directions in the Integration of Cognition, Neurobiology and Genetics of Reading and Dyslexia Research and Practice. Psychology Press; 2009. p. 235-254.

Frost R, Ahissar M, Gottesman R, Tayeb S. Are phonological effects fragile? The effect of luminance and exposure duration on form priming and phonological priming. Journal of Memory and Language. 2003; 48:346–378.

Frost R, Deutsch A, Forster KI. Decomposing Morphologically Complex Words in a Nonlinear Morphology. Journal of Experimental Psychology Learning Memory, & Cognition. 2000; 26:751–765.

Frost R, Deutsch A, Gilboa O, Tannenbaum M, Marslen-Wilson W. Morphological priming: Dissociation of phonological, semantic, and morphological factors. Memory & Cognition. 2000; 28:1277–1288.

Frost R, Forster KI, Deutsch A. What can we learn from the morphology of Hebrew: a masked priming investigation of morphological representation. Journal of Experimental Psychology: Learning Memory, & Cognition. 1997; 23:829–856.

Frost R, Katz L, Bentin S. Strategies for visual word recognition and orthographical depth: A multilingual comparison. Journal of Experimental Psychology: Human Perception and Performance. 1987; 13:104–115. [PubMed: 2951484]

Frost R, Kugler T, Deutsch A, Forster KI. Orthographic structure versus morphological structure: principles of lexical organization in a given language. Journal of Experimental Psychology: Learning, Memory & Cognition. 2005; 31:1293–1326.

Frost, R.; Narkiss, A.; Velan, H.; Deutsch, A. Learning to read Hebrew as a second language: Acquiring “Semitic” markers of reading. Paper presented at the 51th Annual Meeting of the Psychonomic Society; St. Louis, Missouri. November 2010; 2010.

Frost R, Yogev O. Orthographic and phonological computation in visual word recognition: Evidence from backward masking in Hebrew. Psychonomic Bulletin & Review. 2001; 8:524–530. [PubMed: 11700904]

Frost, R.; Ziegler, J. Speech and Spelling Interaction: The Interdependence of Visual and Auditory Word Recognition. In: Gaskell, G., editor. Oxford Handbook of Psycholinguistics. Oxford University Press; Oxford: 2007.

Gebhart AL, Newport EL, Aslin RN. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. Psychonomic Bulletin & Review. 2009; 16:486–490. [PubMed: 19451373]

Gelb, IJ. A study of writing: The foundations of grammatology. University of Chicago Press; Chicago: 1952.

Gibson, JJ. The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates; Hillsdale, NJ: 1986.

Gimson, AC. An Introduction to the Pronunciation of English. Edward Arnold; Hong Kong: 1981.

Gomez P, Ratcliff R, Perea M. The overlap model: A model of letter position coding. Psychological Review. 2008; 115:577–601. [PubMed: 18729592]

Goswami U. Phonological representations, reading development and dyslexia: towards a cross-linguistic theoretical framework. Dyslexia. 2000; 6:133–151. [PubMed: 10840513]

Grainger J. Cracking the orthographic code: An introduction. Language and Cognitive Processes. 2008; 23:1–35.

Grainger J, Ferrand L. Phonology and orthography in visual word recognition: Effects of masked homophone primes. Journal of Memory and Language. 1994; 33:218–233.

Grainger J, Granier JP, Farioli F, Van Assche E, van Heuven W. Letter position information and printed word perception: The relative-position priming constraint. Journal of Experimental Psychology: Human Perception and Performance. 2006; 32(4):865–884. [PubMed: 16846285]

Grainger, J.; Van Heuven, W. Modeling letter position coding in printed word perception. In: Bonin, P., editor. The mental lexicon. Nova Science; New York: 2003. p. 1-24.

Grainger J, Whitney C. Does the huamn mnid raed wrods as a wlohe? Trends in Cognitive Sciences. 2004; 8:58–59. Grainger, J., & Whitney, C. (2004). [PubMed: 15588808]

Grainger J, Ziegler JC. A dual-route approach to orthographic processing. Frontiers in Language Sciences. 2011

Gronau N, Frost R. Prelexical Phonologic Computation in a Deep Orthography: Evidence from Backward Masking in Hebrew. Psychonomic Bulletin & Review. 1997; 4:107–112.

Guerrera C, Forster KI. Masked form priming with extreme transposition. Language and Cognitive Processes. 2008; 23:117–142.

Harm MW, Seidenberg MS. Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. Psychological Review. 1999; 106:491–528. [PubMed: 10467896]

Hsu CH, Tsai JL, Lee CY, Tzeng OJ. Orthographic combinability and phonological consistency effects in reading Chinese phonograms: an event-related potential study. Brain & Language. 2009; 108:56–66. [PubMed: 18951624]

Hunter ZR, Brysbaert M. Theoretical analysis of interhemispheric transfer costs in visual word recognition. Language and Cognitive Processes. 2008; 23:165–182.

Humphreys GW, Evett LJ, Quinlan PT. Orthographic processing in visual word recognition. Cognitive Psychology. 1990; 22:517–560. [PubMed: 2253455]

Inagaki K, Hatano G, Otake T. The effect of Kana literacy acquisition on the speech segmentation unit used by Japanese young children. Journal of Experimental Child Psychology. 2000; 75:70–91. [PubMed: 10660904]

Johnson RL, Perea M, Rayner K. Transposed-letter effects in reading. Evidence from eye-movements and parafoveal preview. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33:209–229.

Katz, L.; Frost, R. Reading in different orthographies: the orthographic depth hypothesis. In: Frost, R.; Katz, L., editors. Orthography, Phonology, Morphology, and Meaning. Advances in Psychology. Elsevier; Holland: 1992. p. 67-84.

Kessler B, Treiman R. Syllable structure and the distribution of phonemes in English syllables. Journal of Memory and Language. 1997; 37:295–311.

Kinoshita S, Norris D. Transposed-Letter Priming of Prelexical Orthographic Representations. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2009; 35:1–18.

Kubozono, H. The phonetic and phonological organization of speech in Japanese. In: Nakayama, M.; Mazuka, R.; Shirai, Y., editors. The Handbook of East Asian Psycholinguistics. Cambridge University Press; New York: 2006. p. 191-200.

Kuperman V, Bertram R, Baayen RH. Morphological dynamics in compound processing. Language and Cognitive Processes. 2008; 23:1089–1132.

Lee, CY. The statistical learning perspective on Chinese reading. In: McCardle, P.; Lee, JR.; Tzeng, OJL.; Miller, B., editors. Dyslexia across languages: Orthography and the brain-gene-behavior link. Brookes Publishing; Baltimore, MD: 2011. p. 44-61.

Lee CH, Taft M. Are onsets and codas important in processing letter position? A comparison of TL effects in English and Korean. Journal of Memory and Language. 2009; 60:530–542.

Lee CH, Taft M. Subsyllabic structure reflected in letter confusability effects in Korean word recognition. Psychonomic Bulletin & Review. 2011; 18:129–134. [PubMed: 21327354]

Marshall JC, Newcombe F. Patterns of paralexia: A psycholinguistic approach. Journal of Psycholinguistic Research. 1973; 2:179–199.

Mattingly, IG. Linguistic awareness and orthographic form. In: Frost, R.; Katz, L., editors. Orthography, Phonology, Morphology, and Meaning. Advances in Psychology. Elsevier; Holland: 1992. p. 11-26.

McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review. 1981; 88:375–407.

Newport EL, Aslin RN. Learning at a distance: I. Statistical learning of non-adjacent dependencies. Cognitive Psychology. 2004; 48:127–162.

Norris D, Kinoshita S. Perception as evidence accumulation and Bayesian inference: Insights from masked priming. Journal of Experimental Psychology: General. 2008; 137:433–455.

Norris D, Kinoshita S, van Casteren M. A stimulus sampling theory of letter identity and order. Journal of Memory and Language. 2010; 62:254–271.

Ornan, U. The final word: Mechanism for Hebrew word generation. Haifa University Press; Haifa: 2003. (In Hebrew)

Perea M, Abu Mallouh R, Carreiras M. The search of an input coding scheme: Transposed-letter priming in Arabic. Psychonomic Bulletin and Review. 2010; 17:375–380. [PubMed: 20551361]

Perea M. Carreiras M. Do transposed-letter similarity effects occur at a prelexical phonological level? Quarterly Journal of Experimental Psychology. 2006a; 59:1600–1613.

Perea M, Carreiras M. Do transposed-letter similarity effects occur at a syllable level? Experimental Psychology. 2006b; 53:308–315. [PubMed: 17176663]

Perea M, Carreiras M. Do transposed-letter effects occur across lexeme boundaries? Psychonomic Bulletin and Review. 2006c; 13:418–422. [PubMed: 17048724]

Perea M, Carreiras M. Do orthotactics and phonology constrain the transposed-letter effect? Language and Cognitive Processes. 2008; 23:69–92.

Perea, M.; Lupker, SJ. Transposed-Letter Confusability Effects in Masked Form Priming. In: Kinoshita, S.; Lupker, SJ., editors. Masked priming: the state of the art. Psychology Press; New-York: 2003. p. 97-120.

Perea M, Lupker SJ. Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. Journal of Memory and Language. 2004; 51:231–246.

Perea M, Perez E. Beyond alphabetic orthographies: The role of form and phonology in transposition effects in Katakana. Language and Cognitive Processes. 2009; 24:67–88.

Perea M, Rosa E. Repetition and form priming interact with neighborhood density at a short stimulus-onset asynchrony. Psychonomic Bulletin and Review. 2000; 7:668–677. [PubMed: 11206208]

Peressotti F, Grainger J. The role of letter identity and letter position in orthographic priming. Perception & Psychophysics. 1999; 61:691–706. [PubMed: 10370337]

Perfetti, CA. Reading Processes and Reading Problems: Progress toward a Universal Reading Science. In: McCardle, P.; Lee, JR.; Tzeng, OJL.; Miller, B., editors. Dyslexia across languages: Orthography and the brain-gene-behavior link. Brookes Publishing; Baltimore, MD: 2011. p. 18-32.

Perruchet P, Pacton S. Implicit learning and statistical learning: One phenomenon, two approaches. Trends in Cognitive Sciences. 2006; 10:233–238. [PubMed: 16616590]

Piantadosi ST, Tily H, Gibson E. Word lengths are optimized for efficient communication. Proceedings of the National Academy of Science. 2011; 108:3526–3529.

Plaut DC, McClelland JL, Seidenberg MS, Patterson K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. Psychological Review. 1996; 103:56–115. [PubMed: 8650300]

Rahamim E, Friedmann N. Developmental letter position dyslexia. Literacy and Language. 2009; 2:79–109. (In Hebrew).

Rastle K, Davis MH, New B. The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. Psychonomic Bulletin and Review. 2004; 11:1090–1098. [PubMed: 15875981]

Rastle K, Davis MH. Morphological decomposition based on the analysis of orthography. Language & Cognitive Processes. 2008; 23:942–971.

Rayner K. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin. 1998; 124:372–422. [PubMed: 9849112]

Rayner K. The Thirty Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology. 2009; 62:1457–1506.

Rayner, K.; Juhasz, BJ.; White, SJ.; Liversedge, SP. Does morphological processing in the parafovea influence eye movements in reading?. Paper presented at the 5th workshop on morphological processing; Marseille. June, 2007; 2007.

Rayner K, White SJ, Johnson RL, Liversedge SP. Raeding wrods with jubmled lettrs: There's a cost. Psychological Science. 2006; 17:192–193. [PubMed: 16507057]

Richardson, U.; Aro, M.; Lyytinen, H. Prevention of reading difficulties in highly transparent Finnish. In: McCardle, P.; Lee, JR.; Tzeng, OJL.; Miller, B., editors. Dyslexia across languages: Orthography and the brain-gene-behavior link. Brookes Publishing; Baltimore, MD: 2011. p. 62-75.

Rueckl JG. Connectionism and the role of morphology in visual word recognition. The Mental Lexicon. 2010; 5:371–400. [PubMed: 22934123]

Rueckl, JG.; Seidenberg, MS. Computational Modeling and the neural bases of reading and reading disorders. In: Pugh, K.; McCardle, P., editors. How Children Learn To Read: Current Issues and New Directions in the Integration of Cognition, Neurobiology and Genetics of Reading and Dyslexia Research and Practice. Taylor & Francis; New York: 2009. p. 101-134.

Rumelhart, DE.; McClelland, JL. On learning the past tenses of English verbs. In: McClelland, JL.; Rumelhart, DE.; the PDP Research Group. , editors. Parallel distributed processing: Explorations in the microstructure of cognition. Vol. Vol. 2. Psychological and biological models. MIT Press; Cambridge, MA: 1986. p. 216-271.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–1928. [PubMed: 8943209]

Schoonbaert S, Grainger J. Letter position coding in printed word perception: Effects of repeated and transposed letters. Language and Cognitive Processes. 2004; 19:333–367.

Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. Psychological Review. 1989; 96:523–568. [PubMed: 2798649]

Seki, A. Functional MRI Studies on Japanese orthographies: Studies in Reading Development and Reading Difficulties. In: McCardle, P.; Lee, JR.; Tzeng, OJL.; Miller, B., editors. Dyslexia across languages: Orthography and the brain-gene-behavior link. Brookes Publishing; Baltimore, MD: 2011. p. 117-132.

Seymour PHK, Aro M, Erskine JM. Foundation literacy acquisition in European orthographies. British Journal of Psychology. 2003; 94:143–174. [PubMed: 12803812]

Shallice T, Warrington EK. The possible role of selective attention in acquired dyslexia. Neuropsychologia. 1977; 15:31–41. [PubMed: 831151]

Shamir O, Sabato S, Tishby N. Learning and generalization with the Information Bottleneck. Theoretical Computer Science. 2009; 410:2696–2711.

Share DL. On the Anglocentricities of current reading research and practice: The perils of over reliance on an “outlier” orthography. Psychological Bulletin. 2008; 134:584–615. [PubMed: 18605821]

Shetreet E, Friedmann N. Induced letter migrations between words and what they reveal about the orthographic-visual analyzer. Neuropsychologia. 2011; 49:339–351. [PubMed: 21126530]

Shillcock R, Ellison TM, Monaghan P. Eye-fixation behavior, lexical storage, and visual word recognition in a split processing model. Psychological Review. 2000; 107:824–851. [PubMed: 11089408]

Shimron, J. Reading Hebrew: the language and the psychology of reading it. Lawrence Erlbaum; Mahwah, N.J.: 2006.

Tishby, N.; Pereira, FC.; Bialek, W. The Information Bottleneck Method. Proceedings of 37th Annual Allerton Conference on Communication, Control and Computing; 1999. p. 368-377.

Tishby, N.; Polani, D. Information Theory of Decisions and Actions. In: Vassilis; Hussain; Taylor, editors. Perception-reason-action cycle: Models, algorithms and systems. Springer; 2010.

Treiman R, Mullennix J, Bijeljac-Babic R, Richmond-Welty ED. The special role of rimes in the description use and acquisition of English orthography. Journal of Experimental Psychology: General. 1995; 124:107–136. [PubMed: 7782735]

Van Assche E, Grainger J. A study of relative-position priming with superset primes. Journal of Experimental Psychology: Learning, Memory and Cognition. 2006; 32:399–415.

Velan H, Frost R, Deutsch A, Plaut D. The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. Language and Cognitive Processes. 2005; 29:169–206.

Velan H, Frost R. Cambridge University Vs. Hebrew University: The impact of letter transposition on reading English and Hebrew. Psychonomic Bulletin & Review. 2007; 14:913–918.

Velan H, Frost R. Letter-transposition effects are not universal: The impact of transposing letters in Hebrew. Journal of Memory and Language. 2009; 61:285–302. [PubMed: 20161017]

Velan H, Frost R. Words with and without internal structure: what determines the nature of orthographic processing. Cognition. 2011; 118:141–156. [PubMed: 21163472]

Wang, WS-Y. Language structure and optimal orthography. In: Tzeng, OJL.; Singer, H., editors. Perception of print: Reading research in experimental psychology. Erlbaum; Hillsdale, NJ: 1981. p. 223-236.

Wang, F.; Tsai, Y.; Wang, WS-Y. Chinese literacy. In: Olson, D.; Torrance, N., editors. Cambridge Handbook on Literacy. Cambridge University Press; 2009. p. 386-417.

Whitney C. How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. Psychonomic Bulletin & Review. 2001; 8:221–243. [PubMed: 11495111]

Whitney C. Supporting the serial in the SERIOL model. Language and Cognitive Processes. 2008; 23:824–865.

Whitney C, Cornelissen P. Letter-position encoding in Dyslexia. Journal of Research in Reading. 2005; 28:274–301.

Whitney C, Cornelissen P. SERIOL reading. Language and Cognitive Processes. 2008; 23:143–164.

Wydell TN, Kondo T. Phonological deficit and the reliance on orthographic approximation for reading: a follow-up study on an English-Japanese bilingual with monolingual dyslexia. Journal of Research in reading. 2003; 26:33–48.

Yamada J, Banks A. Evidence for and characteristics of dyslexia among Japanese children. Annals of Dyslexia. 1994; 44:105–119.

Ziegler JC, Goswami UC. Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. Psychological Bulletin. 2005; 131:3–29. [PubMed: 15631549]

Ziegler JC, Bertrand D, Tóth D, Csépe V, Reis A, Faísca L, et al. Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. Psychological Science. 2010; 21(4):551–559. [PubMed: 20424101]

Zorzi M, Houghton G, Butterworth B. Two routes or one in reading aloud? A connectionist dual-process model. Journal of Experimental Psychology: Human Perception and Performance. 1998; 24:1131–1161.