Orthography, Phonology, Morphology, and Meaning
R. Frost and L. Katz (Editors)

 840

CHAPTER 4

The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis

Leonard Katz
Department of Psychology
University of Connecticut

Ram Frost
Department of Psychology
Hebrew University

It has been said that most languages get the orthography they deserve and there is a kernel of truth in that statement. There is generally an underlying rationale of efficiency in matching a language's characteristic phonology and morphology to a written form. Although the final product may turn out to be more suitable for some languages than for others, there are certain basic principles of the fit that can be observed. The attempt to make an efficient match between the written form, on the one hand, and morphology and phonology, on the other, typically determines whether the orthography chosen is a syllabary, a syllabary-cum-logography, or an alphabet. Further, within the group of alphabetic orthographies itself, there are varying degrees of dependence on the strict alphabetic principle: the range of correspondence between grapheme and phoneme varies both in consistency and completeness. The degree of this dependence is to some extent a function of a language's characteristic phonology and morphology, just as was the choice of the kind of orthography itself. We discuss here what this varying dependence on the alphabetic principle may mean for the mental processes involved in reading and writing.

Diversity in writing systems

Although writing systems are, in general terms, systems for communication, what they actually communicate is the spoken language—as opposed to communicating nonverbal ideas and meanings. DeFrancis (1989) reinforced this point with his analysis of so-called pictographic languages, writing systems whose elements are pictures and symbols that do not stand for words. DeFrancis argued that true pictographic systems are not, in principle, effective and showed that existing examples of pictographic systems had been designed only as novelties or playful communication systems. In practice, they were never used to communicate without substantial ancillary aid from spoken language. The example of pictographic writing emphasizes the poverty of written communication that is not based on language. Therefore, because writing systems are systems for representing the spoken language, it is reasonable to suggest that an understanding of the psychological processing involved in using a writing system must include an understanding of the processing of the spoken language. Although the former will not necessarily parallel the latter, it will be constrained by it. A major constraint arises from the spoken language's morphemes which are the smallest units that carry meaning. It is these morphemes of speech that will be the focus of communication, both spoken and written. Word stems are all morphemes and so are their derivational and inflectional affixes; these units of the spoken language must be easily recoverable from the written language.

A large number and variety of writing systems have flourished, evolved and developed, and in many cases, died, over the centuries. Each of the known systems can be categorized as either logographic-phonetic, syllabic, or alphabetic (DeFrancis, 1989). These distinctions are made on the basis of how a script (a set of symbols) relates to the structure of its language. This relationship between a script and its language is what is described by the term orthography (Scheerer, 1986). The kind of script system and its orthography are typically not wholly the result of accident. It is not accidental that the Chinese languages, for example, have a logographic-phonetic system. In the Chinese orthography, the typical character has two parts to it: a logographic and a phonetic component, the former providing a visually distinctive cue to the semantics and the latter giving the reader a partial guide to the pronunciation. Together, the two components make a combination that specifies completely a unique spoken morpheme. Words may be mono- or polymorphemic.

Chinese morphemes are mainly monosyllabic and, because the variety of possible syllables is limited, there is a high degree of homophony in the language. Such a language is best served by an orthography that distinguishes between the different meanings of morphemes that sound alike. Instead, if the orthography had represented the spoken form alone (e.g., only the phonetic component in the printed word), the reader would not be able to determine the intended meaning of each homophone except, possibly, from the word or sentential context (many polymorphemic words are unique compounds of homophonous morphemes)—but not without the additional cognitive cost needed to resolve the ambiguity. The homophony problem is more of a problem for a reader than for a listener because the listener has more nonverbal contextual information available to assist in determining word meanings.

But a pure logography would not suffice either, for Chinese. If the orthography had no phonetic component, a reader would have to remember, without a phonetic cue, a pronunciation for each of several thousand logographs. This would have effectively limited the number of printed characters that a reader could remember and name to an unacceptably small number. Instead, in modern Mandarin, it is necessary to remember only a small number of phonetic components together with a smaller number of semantic signs. DeFrancis gives the number of these as 895 and 214, respectively (DeFrancis, 1989). A phonetic and a semantic component are paired to produce a character; the effective set consists of 4300 characters, the approximate number considered necessary for full literacy.

In contrast to Chinese, spoken Japanese is polysyllabic and is composed of regular syllable-like components, called moras. Because the number of syllables is small (fewer than 113), it is feasible to represent them by means of a syllabary. The Japanese orthography called kana was such a system, adapted from Chinese characters. However, because there is a good deal of homophony in Japanese, the use of a syllabary alone would not have been without problems and a logography also came into use. That logography is still in use today and is routinely mixed with the use of the kana, the syllabaries being used primarily for morphological affixes and grammatical function words, foreign loan-words, and words not already covered by the Chinese.

Indo-European languages have less homophony and more polysyllabic morphemes than Chinese and Japanese. In addition, the structure of the Indo-European syllable itself is generally more complex than Chinese or Japanese, containing a larger number of phonologically permissible clusters. English is said to have at least 8000 syllables in its phonology, compared to fewer than 1300 for Chinese (DeFrancis 1989). Eight-thousand is far too large a number for an effective syllabary. For English, an alphabet, representing phonemes, is more efficient for learning to read and write. Similarly, the Semitic languages (which include Arabic and Hebrew) would be less suitably represented by a syllabary than by an alphabet. Like Indo-European languages, they too have complex syllable structures. Historically, a consonantal alphabet that developed for West Semitic was the alphabet from which we trace the evolution of modern alphabets.

One of the characteristics of Semitic languages that may have led to the invention of the alphabet is the Semitic triconsonantal root: Semitic words that are related by derivation or inflection have a common core (usually three consonants). Although the vowels in each of the different relatives may be quite different and although there may be additional consonants in a prefix or suffix, there remains an invariant series of three consonants, the root. This core has a strong linguistic salience—it represents a morphological communality of meaning among its family members (see Frost & Bentin, this volume). One can speculate that several early attempts were made at developing a writing system but only an alphabetic system could have captured this communality of morphology efficiently. Syllabic representations would not be optimal: in Hebrew, morpheme boundaries fail to coincide with syllable boundaries (a condition that is true of Indo-European languages as well). Therefore, syllabic representations would not be appropriate for representing morphological units.

The causes of diversity in alphabetic orthographies

Even among the various alphabetic writing systems themselves there are major differences in the degree to which they mirror the phonemic structure of their respective spoken languages. Again, the reason for the differences is largely accounted for by the particular phonological and morphological characteristics of each language. For example, standard written Hebrew is an orthography in which all diacritics (or *points*) are omitted. These diacritics represent nearly all of the vowels and are also used to disambiguate some of the consonants. Nevertheless, writing without diacritics is usually sufficient to indicate the exact intended (i.e., spoken) word if it is supported by a phrasal or sentential context. Thus, although the printed root may be insufficient to allow an unequivocal identification when presented in isolation, when it is presented in a normal context—even a printed one—the combined sources of information are enough for word identification.

In strong contrast to the Hebrew orthography is the Serbo-Croatian. Serbo-Croatian is a major language of the Balkan Peninsula. Its present alphabet was introduced in the early nineteenth century following the principle, "Spell a word like it sounds and speak it the way it is spelled." Each letter represents only one phoneme and each phoneme is represented by only one letter. Moreover, no phoneme in the spoken word is ever excluded in the spelled word. The relation between letters and phonemes is isomorphic and exhaustive. To this day, the Serbo-Croatian spelling system follows the phonemic structure of spoken words. So regular is the relation between speech and writing that the writings of people with different regional pronunciations will show different spellings, mirroring the minor differences in their spoken language. This simple writing system works well for Serbo-Croatian because morphemic variations in the language due to inflection and derivation do not often produce alterations in the phonemic structure of word stems; word stems largely remain intact.

A different state of affairs exists in English. English is somewhere between Hebrew and Serbo-Croatian in the directness with which its phonology is represented in its spelling; there is a large amount of regular phonologic change among words in the same derivational family. Examples of this are the contrasts between the derivational relatives HEAL and HEALTH, between STEAL and STEALTH, etc. Chomsky and Halle (1968) argue that English spelling represents a morphophonemic invariance common to these word pairs, an abstract phonological communality that is below their surface difference in pronunciation. In reading English aloud, the reader must either remember the pronunciation of such a word as a whole or remember the appropriate context-dependent rule for pronunciation. An alternative writing system might have spelled English in the same way that Serbo-Croatian is spelled: with an isomorphic relation between letter and phoneme. However, that method would rob printed English of the advantageous common spelling for words with common morphology. The words HEAL and HEALTH, for example, might then be spelled HEEL and HELTH, disguising their common meaning. The printed form HEEL would also suffer from a double meaning as a consequence of its being homophonic with the word meaning, "part of a foot." English spelling represents a compromise between the attempt to maintain a consistent letter-phoneme relation and the attempt to represent morphological communality among words even at the cost of inconsistency in the letter-phoneme relation.

Thus, alphabetic writing systems reflect the spoken forms of their respective languages with different degrees of consistency and completeness between letter and phoneme. Some of the differences in writing systems have purely political, cultural, or economic causes. But many differences have been motivated by two factors that are purely linguistic. The first has to do with how complex the spoken language is in the relation between phonology and morphology—only a phonologically complex language can have a deep alphabetic orthography. For example, Serbo-Croatian is not phonologically complex. All morphologically related words have a common phonologically invariant core. Two words that are morphologically related will share a common word stem that will necessarily sound the same in both words. Both instances of that common stem will, of course, be spelled the same. Thus, when evaluated by the characteristic of phonological complexity, a language that is not complex can be written (and generally *will* be written) in a shallow orthography, an orthography that tracks the phonology. Secondly, if a language is one that is phonologically complex then the orthography has the option of representing either morphological invariance (a deep orthography) or following grapheme-phoneme invariance (a shallow orthography). As we suggested above, English qualifies as quite complex, phonologically. In principle, it could have been written either as a shallow or a deep orthography. The advantage to English in choosing a deep orthography is in the consistent spelling of morphemic invariances. However, that choice having been made, there are then different pronunciations of the same spelling on occasion (e.g., HEAL-HEALTH) and, inadvertently, identical pronunciations for different spellings (e.g., PEEL-DEAL).

A different situation exists in Hebrew. Hebrew's phonology is complex; morphemes may undergo considerable sound change under either inflectional or derivational change. On the other hand, because of the pervasiveness of the triconsonantal root in Hebrew, a great deal of morphological constancy exists. Therefore, there was an historical choice, so to speak, for the evolution of the Hebrew orthography: It could have opted for either morphemic or phonemic invariance but, unlike Serbo-Croatian, it could not have contained both in a single orthography because of its phonological complexity. Hebrew initially evolved as an orthography in which the morphology was preserved at the expense of phonological completeness. Vowels were omitted thereby emphasizing the morphologically based consonantal invariance in a given family of word roots. Vowel points were added to the script at a later stage in the orthography's development only because the language was no longer being spoken as a primary language and it was feared that its pronunciation would become corrupted unless vowels were included in the script. Nowadays, the orthography used by adults is the unpointed one, which is graphemically incomplete and somewhat inconsistent to the reader because it omits nearly all of the vowels and makes some of the consonants ambiguous.

In summary, all alphabetic orthographies can be classified according to the transparency of their letter-to-phoneme correspondence, a factor that has been referred to as orthographic depth (Liberman, Liberman, Mattingly, & Shankweiler, 1980). An orthography in which the letters are isomorphic to phonemes in the spoken word (completely and consistently), is orthographically shallow. An orthography in which the letter-phoneme relation is substantially equivocal is said to be deep (e.g., some letters have more than one sound and some phonemes can be written in more than one way or are not represented in the orthography). Shallow orthographies are characteristic of languages in which morphemic relatives have consistent pronunciations.

Differences among alphabetic orthographies in processing printed words: The orthographic depth hypothesis

Our discussion to this point has made the standard argument that there are differences among alphabetic orthographies in orthographic depth and that these differences are a result of differences in their languages' phonology and morphology. In this section, we propose that the differences in orthographic depth lead to processing differences for naming and lexical decision. This proposal is referred to as the orthographic depth hypothesis (ODH). It states that shallow orthographies are more easily able to support a word recognition process that involves the language's phonology. In contrast, deep orthographies encourage a reader to process printed words by referring to their morphology via the printed word's visual-orthographic structure.

We would like to make two points, each independent of the other. The first states that, because shallow orthographies are optimized for assembling phonology from a word's component letters, phonology is more easily available to the reader prelexically than is the case for a deep orthography. The second states that the easier it is to obtain prelexical phonology, the more likely it will be used for both pronunciation and lexical access. Both statements together suggest that the use of assembled phonology should be more prevalent when reading a shallow orthography than when reading a deep orthography. Because shallow orthographies have relatively simple, consistent, and complete connections between letter and phoneme, it is easier for readers to recover more of a printed word's phonology prelexically by assembling it from letter-phoneme correspondences.

Suggested by the above is our assumption that there will always be at least some dependence on phonological coding for the process of reading in any orthography. That is, the processing of (at least) some words will include assembled phonology (at least in part). This assumption can be easily motivated for alphabetic orthographies. The assembling of phonology has a certain precedence in a reader's experience; instruction in reading typically means instruction in decoding, i.e., learning how to use letter-phoneme correspondences. It is well established that beginning readers find it easier to learn to read in shallow orthographies, where those correspondences are most consistent (see, for example, Cossu, Shankweiler, Liberman, Katz, & Tola, 1988). Even in learning to read Hebrew, instruction is typically given in the shallow pointed orthography instead of the deep unpointed one (the transition to the unpointed form beginning in the third grade). In any orthography, after learning to read by using assembled phonology routines, skilled readers may continue its use to the extent that the cost of doing so is low. This will be particularly true when the orthography is shallow. However, given the experimental evidence, some of which we discuss later, it seems certain that assembled phonology is not used *exclusively*. More likely, a mix of both assembled phonology and visual-orthographic codings are nearly always involved, even in shallow orthographies.

Two versions of the orthographic depth hypothesis

Two versions of the orthographic depth hypothesis (ODH) exist in the current literature. What can be called the *strong ODH* states that phonological representations derived from assembled phonology alone are sufficient for naming and lexical decision in shallow orthographies. Thus, according to the strong ODH, rapid naming in shallow orthographies is a result of only this prelexical analytic process and does not involve pronunciation obtained from memory, i.e., the lexicon. However, we submit that the strong form of the ODH is patently untenable when applied to the orthographies that have typically been used in research on word perception. It is insufficient to account for pronunciation even in a shallow orthography like Serbo-Croatian. This is so because Serbo-Croatian does not represent syllable stress and, even though stress is often predictable, it is not always predictable. Because the final syllable is never stressed, stress is completely predictable for two-syllable words but for words of more than two, it is not. However, one- and two-syllable words make up a large part of normal running text so much or most of the words a reader encounters can be pronounced by means of a prelexical subword analysis. But, of course, many words will be greater than two syllables in length and these can be pronounced correctly only by reference to lexically stored information. In addition, there are some exceptions to the rule that a letter must represent only one phoneme; some final consonant voicing changes occur in speech that are not mirrored in the conventional spelling (these changes are predictable, however). Thus, Serbo-Croatian, although it should be considered an essentially shallow orthography, is not the perfect paradigm of a shallow orthography. We should not expect a strong ODH to make sense for such an orthography.

We support the *weak ODH*. In this version, the phonology needed for the pronunciation of printed words comes not only from prelexical letter-phonology correspondences but also from stored lexical phonology, that is to say, from memory. The latter is the result of a visual-orthographic addressing of lexicon, i.e., a search process that matches the spelling of a whole word or morpheme with its stored phonology. The degree to which a prelexical process is active in naming is a function of an orthography's depth; prelexical analytic processes will be more functional (less costly) in shallow orthographies. However, whether or not these prelexical processes actually dominate orthographic processing for any particular orthography is a question of the demands the two kinds of processes make on the reader's processing resources, a question we discuss further below. We proposed (and supported) this weak form of the ODH in Katz and Feldman (1983) and Frost, Katz, and Bentin (1987); further details are given later in this chapter.

With regard to word recognition (as in lexical decision), some of our colleagues have argued that Serbo-Croatian necessarily involves prelexical (i.e., assembled) phonology (Feldman & Turvey, 1983; Lukatela & Turvey, 1990a). Others have made a similar claim for the obligatory involvement of prelexical phonology in English (Van Orden, Pennington & Stone, 1990; Perfetti, Bell, & Delaney, 1988). However, these researchers have not argued for the *exclusive* involvement of assembled phonology. Logically, assembled prelexical phonological information, without syllable stress information, is sufficient to identify the great majority of words in the English lexicon. However, irregularly spelled words, foreign borrowings, etc., would pose problems for an exclusively phonological mechanism, and, therefore, such a view seems less plausible. Finally, note that we are confining this discussion to the problems of naming and lexical decision; it is an entirely different question to ask whether phonological representations are necessary for *postlexical* processes like syntactic parsing and text comprehension.

Evidence on the questions of phonological recoding and the weak ODH

We discuss next the evidence for the hypotheses that the lexicon is addressed by assembled phonology, presumably in combination with visual-orthographic addressing, and that the specific mix of the two types of codes depends on orthographic depth. We show why single-language studies, in general, are not suitable for testing the weak ODH and mention the few exceptions. Experiments that directly compare orthographies with each other provide the most direct evidence. We will argue that these cross-language comparisons are absolutely critical to an investigation of orthographic depth effects.

It is important to realize that, in the controversy over whether visual-orthographic recoding or phonological recoding is used in word perception, there is little direct evidence of visual-orthographic effects. Instead, the burden of proof is placed on assembled phonology; if no effect of phonology is found, then visual-orthographic coding is said to win, by default. The assumption is not unreasonable because a visual-orthographic representation is obviously available in principle and seems to be the only alternative to assembled phonology. However, it should be kept in mind that, because of this, the experimental evidence hinges on the sensitivity of the experimental task and its dependent measures to phonology. If they fail to indicate the presence of phonology, it may not be because phonology is not operative.

In fact, several experiments have demonstrated that phonological recoding effects can be found even in deep orthographies. In Hebrew, Frost has shown that, if available, the full phonology given by the pointed orthography is preferred for naming; this seems to suggest that even if word recognition does not normally proceed via assembled phonology in Hebrew, the recognition process is prepared to default to its use (Frost, forthcoming). In English, Perfetti and his associates and Van Orden and his associates have presented strong evidence for phonological recoding in lexical access (Perfetti, Bell, & Delaney, 1988, Perfetti, this volume; Van Orden, Pennington, & Stone 1990; Van Orden, this volume).

However, it is one thing to find the active presence of phonological recoding, another to determine the conditions under which phonological recoding occurs, and yet another to determine the degree to which naming or lexical access is dependent on it. Is assembled phonology obligatory or is it rarely used; if neither of these extremes, is it the more preferred or the less preferred code? In this vein, Seidenberg (Seidenberg, 1985; Seidenberg, this volume) has argued that word frequency is the primary factor that determines whether or not assembled phonology is used to access the lexicon. His argument is that in any orthography, whether deep or shallow, frequently seen words will become familiar visual-orthographic patterns and, therefore, rapid visual access will occur before the (presumably) slower phonological code can be assembled from the print. Low frequency words, being less familiar, produce visual-orthographic representations that are less functional; lexical activation builds up more slowly. This gives time for phonological recoding to contact the lexicon first. However, we cannot presently answer questions that are concerned with the relative importance of word frequency to orthographic depth or concerned with the relative dominance of the two kinds of representation, visual-orthographic and phonological. But we can meaningfully address the question of whether or not the relative amount of assembled phonological coding decreases with increasing orthographic depth: the orthographic depth hypothesis.

Comparisons across orthographies

Cross-language experimentation, in which different languages are directly compared, are the critical methodology for studying the orthographic depth hypothesis. Single-language experiments are adequate for testing only the strong form of the ODH, in which shallow orthographies are said to never use lexically stored information for naming—but, as we showed, this is a claim that can be rejected on logical grounds. Single-language experiments are not without interest, however; they can be useful in indicating how easy it is to find effects of phonological coding. This may suggest—but only suggest weakly—what the dominant representation is for an orthography. If it is easy to find effects of phonological coding in Serbo-Croatian and difficult to find those effects in Hebrew, using more or less similar experimental techniques, we may suspect that phonological coding is the dominant (preferred) code in Serbo-Croatian but not in Hebrew. However, such experiments can not rule out the additional use of the alternate type of representation for either orthography. In this vein, we know that it is difficult to find effects of phonological coding in English using standard lexical decision paradigms but, nevertheless, phonological effects can be found using Perfetti's backward masking paradigm, which is apparently more sensitive (Perfetti, 1988). Finally, however, an accurate answer to the question of which type of representation is dominant for a particular orthography can not be given by the current experimental paradigms; the results from these experiments may only reflect the adequacy of the paradigms in capturing the true word recognition process.

The weak form of the ODH proposes that (1) both orthographic information and prelexically assembled phonological information are used for lexical access and (2) the degree to which one kind of information predominates is a function of the structural relationship between orthography and the lexical entry. The Serbo-Croatian orthography, with its simple and consistent letter-phoneme relationships, makes it easy for the reader to learn and maintain the use of assembled phonology. This assembled phonology must address, presumably, the same abstract phonology addressed by a listener's spoken language lexicon (although it may be only a subset of the full spoken phonology, because of the absence of stress information, at the least). In contrast, the Hebrew orthography, because it lacks most of the vowels and has many ambiguous consonants, is incapable of providing enough assembled phonology that will consistently identify a unique word in the phonological lexicon (only the consonants can be assembled); therefore, there are fewer benefits in generating phonological information by assembling it from grapheme-phoneme correspondences. These are lessons that the developing reader can learn tacitly, lessons that may lead, eventually, to different dominant modes of printed word processing for Serbo-Croatian and Hebrew readers. For languages that are in between these two extremes, the relative balance of assembled phonology to orthographic representation should reflect the relative efficacy of the two kinds of information in that orthography; some letters or letter-sequences may be simple and consistent and the assembled phonology derived from these may be used along with orthographic information. It is not possible to make a more precise statement without an understanding of the details of the lexical recognition process and the processing resources that are required. For example, for a visual-orthographically coded word to be recognized, a mental representation of that word must have been created previously and stored in the reader's memory. We do not know how to compare the resources needed to create a new orthographic representation with the resources needed to generate assembled phonology; which is more demanding? Neither can we automatically assume that it is easier to access lexicon via a visual-orthographic representation.

Additional complications arise when we try to be more specific about the phonological nature of the information in the lexicon itself; what, exactly, is the information that is represented in lexicon that is, presumably, addressed by assembled phonology? Alphabets mainly represent phonemes but are words in the spoken lexicon to be represented as phoneme sequences? If so, why do syllabic orthographies work at all, since the printed units of syllabaries map onto syllables, not phonemes? In fact, there are several different theoretical descriptions that have been proposed by speech researchers for the lexical representation of a word. Perhaps, the spoken lexicon contains multiple phonological descriptions of a single word, e.g., phonetic, phonemic, syllabic, gestural, etc. The phonology produced by reading may be a subset of the information constituting the spoken lexicon or it may even be different in kind (although related). We do not propose to discuss this problem in detail here, but only wish to point out that there is a companion to the question of how phonological information is used in reading, namely, the question of the nature of the phonological information that is used in spoken word recognition.

Evidence supporting the orthographic depth hypothesis

Some early evidence that there is a relationship between orthographic depth and lexical access was obtained by Katz and Feldman (1981). They compared lexical decision times in Serbo-Croatian and English for printed stimuli that were divided with a slash character. The stimuli were divided either at a syllable boundary (e.g., WA/TER) or one character to either the left or right of the boundary (e.g., W/ATER or WAT/ER). If word recognition involves recoding of the stimulus to a phonological form, and if that phonology includes the syllable as a unit, then division at a syllable boundary—which preserves the syllable units—should be less disruptive than division off the boundary. Pseudowords were similarly divided. Lexical decisions to words and pseudowords that were irregularly divided were slower than lexical decisions to their regularly divided counterparts and the disruptive effect of irregular division was stronger for Serbo-Croatian. The data, then, were consistent with a model of word recognition that assumes the operation of at least some phonological recoding of print prior to lexical access and, further, assumes that phonological recoding is a more consistent determiner of access in shallower orthographies (e.g., Serbo-Croatian) than deeper ones (e.g., English).

In a second study, Katz and Feldman (1983) made a direct test of a second prediction of the orthographic depth hypothesis: that pronouncing a word (naming) depends more on assembled phonology in Serbo-Croatian than in English. A lexical decision experiment and a naming experiment were run in both languages; stimulus words had the same (or similar) meanings in both languages. The subjects were native speakers who were tested in their native countries (Serbia or the United States), in order to avoid subjects who were fluent in both languages. This was done because the experience of a bilingual speaker might have affected his or her strategy for reading.

Each test stimulus (the target), whether it was a word or a nonword, was always preceded by the brief (600 ms) presentation of a real word. On half of those trials when the target was a word, this predecessor was semantically related to the target (e.g., MUSIC-JAZZ); on the other half, it was unrelated. Words also preceded the nonword targets. It is well established, that preceding a stimulus with a semantically related word will facilitate and speed a lexical decision response to the target. Thus, it was expected that reaction time to the target JAZZ would be faster for those subjects who saw it preceded by the word MUSIC than for those subjects who saw it preceded by the word GLASS, which has no strong semantic relation to the target. The critical fact for this kind of experimental technique is that the facilitating effect of MUSIC can only occur by activating the semantic link between it and JAZZ and this linkage necessarily must be within the lexicon. Thus, to the extent that there is facilitation in the subject's recognition of JAZZ, it indicates that recognition is being assisted by activity within the lexicon. Such lexical activity may facilitate whether the lexicon is addressed by orthography or by phonology.

For naming, however, the prediction is different. Naming can be accomplished largely without accessing the lexicon by means of subword letter-to-phonology recoding. Of course, in neither English or Serbo-Croatian can the process be entirely without reference to lexical memory, because the stress of polysyllabic words is not specified in the orthography. Nevertheless, the process of naming can, in principle, be carried out substantially without reference to the lexicon. Thus, if naming in Serbo-Croatian is more dependent on phonological recoding (and less dependent on lexical look-up) than English, naming in Serbo-Croatian ought not to be affected by the semantic priming manipulation, which is necessarily lexical in its locus of operation. Results supported this suggestion: target words that were preceded by semantically related words (e.g., MUSIC-JAZZ) were pronounced faster than target words that were preceded by unrelated words (e.g., GLASS-JAZZ) in the case of English but not in the case of Serbo-Croatian. In contrast, there were equivalent strong effects of semantic priming for lexical decisions, in both languages.

A three-way comparison of Hebrew, Serbo-Croatian, and English increased the range of orthographic depth examined in a single study (Frost, Katz, & Bentin 1987). Necessary to the success of any cross-language experiment is a set of stimulus words that have equivalent critical characteristics in all the languages under consideration: for example, equivalent frequencies of occurrence. Subjectively estimated frequencies of occurrence were obtained and equivalent sets of stimulus words were used in all three test languages. The important comparison, over the three orthographies, was the relation between naming and lexical decision reaction times for words versus nonwords. The importance of this comparison was based on the rationale that in shallow orthographies the lexicon plays only a minor role in the naming process compared to its role in the lexical decision process. The opposite assumption, i.e., that even in shallow orthographies, a skilled reader always employs the orthographic route to accessing the lexicon, predicts that readers in all three orthographies should perform similarly. On the other hand, if the orthographic depth hypothesis is correct, the greatest difference between naming and lexical decision reaction times should be in Serbo-Croatian, which has the shallowest orthography while Hebrew should show the greatest similarity. Results were in line with the orthographic depth hypothesis; naming times were considerably faster than lexical decision times in Serbo-Croatian but, in Hebrew, lexical decision and naming looked quite similar. In Hebrew, it took as long to name a word as to recognize it: a suggestion that naming was accomplished postlexically. In addition, in Serbo-Croatian, the faster responding for naming versus lexical decision was even greater for pseudowords than for words. In these comparisons, English was intermediate. Thus, the results support the hypothesis that the shallower the orthography, the greater the amount of phonological recoding that is carried out in naming. Subsequent experiments in this study, which maximized the potential for lexical processing by semantically priming target words and by varying the relative number of pseudowords further supported this interpretation.

In all the experiments we have discussed to this point, the experimental paradigms used have been naming and lexical decision tasks. These tasks have a disadvantage as methodologies because the phonological variation that is used to affect the subject's response (e.g., the consistency of the grapheme - phoneme relation) is obtained through manipulating the orthography (e.g., different alphabets) and not by manipulating the putative phonology directly. The experimenter never observes any phonologic recoding; its presence is only inferred. Thus, one can not be certain that the differences that are observed are true effects of phonological recoding or, instead, are only the result of orthographic effects which happen to be correlated with phonology. Frost and Katz (1989) addressed this issue by introducing a paradigm in which subjects had to compare a spoken word and a printed word. This paradigm requires subjects to perceive and use phonology in their task processing. Subjects were required to simultaneously read and listen to two words presented by computer and judge whether or not they were the same (i.e., represented the same lexical item). In order to make the comparison, the subject had to mentally place both spoken and printed stimuli into a common representation. This could have been done, in principle, in several ways, although only two possibilities seemed reasonable. The spoken word could have been imagined as a spelled (printed) word or subjects could have generated the phonology of the printed word. The evidence indicated that subjects chose the latter. This was not surprising: subjects have had far more practice reading than spelling. After converting the printed stimulus to a phonological representation, both phonological representations could then have been compared in order to determine if they matched. Over a list of 144 or more trials, subjects made the judgment "Same" or "Different" about each pair of printed and spoken words on each trial. There were three conditions: clear speech and clear print, degraded speech (noise added) and clear print, and clear speech and degraded print (visual noise added). Serbo-Croatian and English native speakers were tested on comparable materials. The effects of degrading were marked; when either the print or the spoken word was degraded, performance declined sharply. However, the difference in latency between the slower responses to print or speech that had been degraded compared to clear print or speech was four times greater in the orthographically deep English than the shallower Serbo-Croatian; degradation had a much stronger deleterious effect in English.

An interactive activation network model can be extended easily to account for these results. The model contains parallel orthographic and phonologic systems that are each multilevel with lateral connections between the two systems at every level. In particular, the sets of graphemic and phonemic nodes are connected together in a manner that reflects the correspondences in a particular orthography: mainly isomorphic connections in a shallow orthography and more complex connections in a deep orthography. The simple isomorphic connections in a shallow orthography should enable subjects to use the printed graphemes to activate their corresponding (unambiguous) phonological nodes, supplementing weaker activation generated more directly by degraded speech. This higher, aggregated, activation should reach threshold fast compared to a network representing a deep orthography with its weaker grapheme-phoneme connections.

Evidence against the orthographic depth hypothesis

We mentioned above the study by Seidenberg (1985) who studied word naming in Chinese and English. In English, he found no difference between regularly spelled words and exception words as long as their word frequency was high, suggesting that phonologic representations play no role in naming frequent words. Differences were found, however, for low frequency words, exception words having the longer latencies. In Cantonese, an analogous pattern was found: There was no significant latency difference between phonograms (compound characters that contain both a phonetic component and a logographic signific) and non-phonograms (characters that contain no phonetic)—as long as they were high frequency. For low frequency items, phonograms were named faster. However, what seemed to drive naming latency most strongly in both languages was word frequency. The results suggest that the effect of frequency, an effect that was similar in both orthographies, may be of overriding importance in determining which kind of lexical access code is successful; differences between orthographies that can affect the coding of phonology may be irrelevant when frequent words are compared.

Besner and Hildebrandt (1987) capitalized on the fact that Japanese is written in three script systems. One of the scripts is a logography that is derived from the Chinese but one in which the characters are pronounced differently. Two of these scripts are essentially syllabic orthographies, katakana and hiragana. Historically, their graphemes evolved from separate sources but they both address the same phonology, similar in this regard to the dual Cyrillic and Roman alphabets of Serbo-Croatian. Unlike Cyrillic and Roman, however, the Japanese scripts are rarely used to write the same words; instead, they "specialize," being used for mutually exclusive vocabularies. In Japanese, then, the pronunciations of those words that are logographic must be recalled lexically. However, those words that are normally printed in katakana and those words that are normally printed in hiragana can, in principle, be pronounced via grapheme-syllable correspondences. In a simple but direct experiment that compared only the katakana and hiragana scripts, subjects named words that were printed either in their normal script or in the other script. When printed in the normal script, naming times were 47 to 65 milliseconds faster than when printed in their atypical script. Besner and Hildebrandt interpreted the results to mean that subjects were not using grapheme-syllable correspondences in order to pronounce the normally printed stimuli because changing the visual-orthographic form had been detrimental; if subjects had been assembling the phonology for naming, they would not have been slowed by the change in grapheme-syllable script system. Thus, there is reason to suspect that Japanese readers always adopt a visual-orthographic mode for naming no matter the depth of the script system they are reading: the deep logography or the shallow syllabaries. In his chapter in this volume, Besner offers further details.

Additional evidence against the orthographic depth hypothesis was reported by Baluch and Besner (1991). They studied naming in Persian, an orthography which offers a comparison between words that omit the vowels, like all words in Hebrew (opaque words) and other words that are spelled with a relatively full phonological specification, like Serbo-Croatian (transparent words). The difference lies in the representation of the vowels; opaque words have one or more of certain specific vowels that can be written as diacritics but, instead, are typically omitted from the spelling (as in Hebrew) while transparent words contain only those vowels that are never omitted. The authors found that semantic priming equally facilitated both transparent words and opaque words; the weak orthographic depth hypothesis would predict less facilitation for the transparent words, which can be pronounced largely via assembled phonology, needing lexical information only for syllable stress. Differences between opaque and transparent words did appear when pseudowords were included in the list of words to be pronounced; then, only the opaque words were facilitated. The inclusion of pseudowords presumably biased the subject toward the use of addressed phonology as the default because the pseudowords would have had no addressed phonology and grapheme-phoneme correspondence rules were therefore an effective alternative. Apparently, when the recognition process is biased in this way, transparent words are no longer processed via visual-orthographic coding and they are no longer facilitated by semantic priming. Their results suggest that in normal reading, where there are no pseudowords, subjects may always use the direct lexical route, without the use of assembled phonology. Baluch and Besner note that in all the studies in which subjects used phonological recoding, pseudowords had been included in the stimulus list of words, perhaps biasing subjects toward an atypical processing strategy (e.g., (Katz et al., 1983; Frost et al., 1987; etc.).

A similar point was made by Tabossi and Laghi (1992). In a clever series of experiments, they showed that semantic priming effects in naming, which are indicative of lexical processing, disappeared or were attenuated when pseudowords were introduced. The implication is, then, that visual-orthographic coding was the preferred strategy for their subjects. The authors interpret their results to suggest that assembled phonology is produced only under artificial conditions such as when pseudowords are present. Because their experiments used a shallow orthography (Italian), the authors suggest that all orthographies, shallow and deep, use the same mechanism for processing print, i.e., the visual-orthographic route.

However, the alternative explanation, the ODH, is not directly addressed by their study (with one exception, discussed shortly). As we suggest above, no standard experiment on a single script system will be able to test the claim of the ODH that the amount of lexical involvement is greater in shallow than in deep orthographies. The ODH is a statement about relationships *among* orthographies; it does not categorically disallow the use of either assembled phonology or visual-orthographic processing for any orthography (except for the strong form of the hypothesis, which is clearly unacceptable on rational grounds). Thus, it is still not known if some phonological processing is occurring in Italian but is not being observed because of special characteristics of the experiment itself (e.g., the task, stimuli, etc.). As in all situations in which the experimenter is pressing the null hypothesis ("Can we show that *no* phonological processing is occurring?"), the better the set of alternative models for comparison, the more convincing the outcome. In the case of the ODH, the more convincing argument against it would be to show that manipulations that affect semantic facilitation cause identical effects (do not cause differential effects) between Italian and some orthography that is deeper than Italian. Tabossi and Laghi (1992) do, in fact, make this test; when they do, they find evidence that is consistent with the ODH. Semantically primed words were named faster than controls in English but not in Italian, suggesting that naming involved the lexicon more in the deeper English than in Italian. However, both word lists contained pseudowords which would tend to increase the amount of phonological processing for both.

A similar demonstration was made in the Frost, Katz, and Bentin (1987) study. In Experiment 3, the authors explicitly examined the same hypothesis that put forward by Baluch and Besner (1991). In this experiment the ratio of words to pseudowords in the stimulus list was manipulated and its effect on naming was measured in Hebrew, English, and Serbo-Croatian. The results showed marked differences in the pseudoword ratio effect in the three different orthographies. Whereas in Serbo-Croatian the inclusion of pseudowords had almost no effect on naming latencies (consistent with the notion that assembled phonology is the preferred strategy for that orthography), much larger effects were found in English and Hebrew. The point raised by Baluch and Besner is indeed important; pseudowords can, in fact, affect naming strategies. However, this issue has no direct relevance to the ODH. The ODH suggests that the relative effect of pseudoword inclusion should be different in deep and in shallow orthographies. The results of both Frost et al. (1987), and Tabossi and Laghi (1992) are compatible with this notion.

Sebastián-Gallés (1991) presented evidence that was said to be inconsistent with theories that propose different mechanisms for shallow and deep orthographies. Spanish subjects pronounced pseudowords that had been derived from real words; each pseudoword was orthographically similar to its counterpart, except for one or two letters. In some pseudowords, the correct pronunciation of a critical letter (*c* or *g*) changed from its pronunciation in the real word according to grapheme-phoneme correspondence rules, because of a change in the following vowel. In other pseudowords, there was no such change in the vowel (and, therefore, no change in the pronunciation from the real word model). Subjects pronounced about 26% of the change pseudowords contrary to the correspondence rules while only about 10% of the no-change pseudowords were pronounced in that way. Sebastián-Gallés interpreted this result to mean that subjects were using a lexical strategy for pronouncing pseudowords. But this interpretation is warranted only if the theory being tested allows *no* lexical involvement at all in naming: The author was attacking the strong form of the ODH. A closer look at the evidence suggests that the data are, in fact, consistent with the weak ODH. Seventy-four percent of the change pseudowords were pronounced in accordance with spelling-to-sound correspondence rules and only 26% were pronounced "lexically." Thus, a mix of the two processes may have been at work. In a second experiment, comparing lexical decision and naming times, Sebastián-Gallés found a moderate correlation (.455) for latencies between the two tasks. When the latency on each task was correlated with word frequency (presumed to be an index of lexical involvement), the correlation was of greater magnitude for lexical decision (-.497) than for naming (-.298). This result is consistent with a continuity of lexical involvement, naming being under weaker lexical control than lexical decision. A final experiment in this series showed semantic priming for naming under conditions where it is usually not found in shallow orthographies, viz., when pseudowords are included in the list of stimuli. Such results do suggest more lexical involvement in naming for a shallow orthography than other research has suggested. Nevertheless, in the author's conclusion, Sebastián-Gallés interprets the sum of the results to mean that "... lexical access from print in Spanish involves the use of orthographic information during at least some of the processing time," a statement that is consistent with the weak ODH.

Concluding remarks

Our working hypothesis has been that all alphabetic orthographies make some use of assembled phonology for word recognition. The proposal that a mixture of prelexical and visual-orthographic information is used for word recognition is consistent with the weak form of the ODH. The approach we suggest is in line with those models that contain dual phonological and orthographic representations, such as the dual route models (see Paap, this volume) and network models (see Seidenberg, this volume). The question of just how prevalent the use of phonology is, relative to the use of visual-orthographic coding, for any given orthography is an open one and is not addressed by the ODH itself. It could even be the case that the predominant lexical access code for frequent words in the shallow Serbo-Croatian is actually visual-orthographic or, on the other hand, that the predominant code in Hebrew is based on the partial phonological information that can be assembled from the unpointed letters (although either possibility seems unlikely from the present evidence). The ODH does not specify what degree of orthographic depth determines predominance for, say, visual-orthographic coding. Of course, orthographic depth will not be the only determiner of dominance; the reader's experience should play a major role as well.

Which of the codes is dominant might be determined by a mechanism like the following. Assume that the processing system is capable of using either code: a dual-code model. Suppose that a phonological representation is the *default* code for any given word but processing of a word via its phonological code can be replaced by processing via its visual-orthographic representation when the word has been experienced by the reader a sufficient number of times: a word frequency criterion. The premise that phonology is the default code is based on the fact that it is typically the code of instruction and the beginning reader receives much practice in its use. (One piece of evidence supporting the notion that the default code is phonological is that even adult readers of Hebrew prefer to use phonological information if it is available; Frost, forthcoming). The criterion word frequency that is required in order to replace processing by assembled phonology with processing by visual-orthographic representations should be a function of the costs involved—in part, the cost for assembling the phonological representation and using it to access the lexicon. A higher replacement criterion will obtain in a shallow orthography where assembled phonology is easy to generate than in a deeper orthography.

What other factors affect the cost? Besides the ease of assembling phonology, a second factor is the ease with which phonology can be used in the lexical search process itself. Likewise, the ease of generating a visual-orthographic representation that is suitable for lexical search, the ease with which *that* information can address the lexicon, and, of course, the cost involved in establishing the visual-orthographically coded lexical representation in the first place, all need to be evaluated in establishing the criterion. The tradeoff between the generation of information that can be used for lexical access and the access process itself is important. Visual-orthographic codes may be costly for both the beginning and skilled readers to generate, particularly if a word is morphologically complex (it may require decomposition, in that case). However, the search process based on a visual-orthographic representation may be rapid for the skilled reader once he or she has a well-established visual-orthographic representation in lexical memory. Phonological codes may be difficult to generate but, once obtained, may be a fairly natural (i.e., low cost) way of addressing lexicon; after all, our primary lexicon, the speech lexicon, is based on phonology. Although the phonological lexicon may have taken years for a child to develop, it is thereafter available free to the reading process. However, it is obvious that we know very little about this tradeoff in terms of processing costs. The answer to the question about which of the two representations, phonological or visual-orthographic, is dominant depends on knowing more than we presently do about the perceptual and cognitive resources involved in word recognition. This question is discussed in some detail by Seidenberg, (this volume). Nevertheless, we repeat that even in the absence of a fuller understanding of how the word recognition process draws on these resources, it is still a plausible (and testable) hypothesis that word recognition in shallow orthographies will depend more on phonological representations simply because such information is available at less cost in those orthographies. Much of the evidence we have presented here is consistent with that hypothesis.

There has been substantial progress over the past decade in understanding the mechanisms behind naming and recognizing printed words. Importantly, research has involved an increasing variety of languages and writing systems, forcing theory to be general enough to encompass this wider scope. The vitality of this research is great and even shows signs of increasing in its pace. We hope that the main import of this article will be to clarify some issues in this area on the differential effects of writing systems on word perception. If there is presently significant (although perhaps not universal) agreement among researchers that visual-orthographic and assembled phonological representations may both play roles in word perception, then the next phase of research activity must include ways of assessing the conditions under which they are active, the relative contributions of each, and the mechanisms of their action. This requires a switch from research designs that address qualitative questions (e.g., "Is the lexicon accessed phonologically or not?") to designs that address the relative balance of phonological and visual-orthographic coding.

The best candidate for a heuristic framework for this research may be network modeling which offers a natural way of simulating the relationships between orthography and phonology, orthography and morphology, and phonology and morphology (at one level) and between these coded representations and the lexicon (at another). Likewise, differences between and within orthographies concerning the consistency, regularity, and frequency of these relationships can be implemented as initial constraints on such networks. Seidenberg (this volume) suggests that network architectures offer ways of modeling how the various general cognitive resources involved are adapted to the processing of printed words: how the system assembles itself under the constraints of language, orthography, and memory. Implicit in this characterization is the additional possibility of modeling the historical evolution of a writing system. While the resources of memory, perceptual discrimination, and the like, may be constant, languages and their orthographies have not been immutable and their histories of change are well known in many cases. The orthographic depth hypothesis itself is a statement about a part of this larger issue of the fit between writing systems and human capabilities. The hypothesis embodies the assumption of covariant learning (Van Orden et al., 1991). That is, the structure and operation of the network should reflect the contingencies among phonology, morphology, and orthography that exist for printed words and, therefore, the contingencies that will be experienced by a reader. Each orthography, shallow or deep, defines its own pattern of contingencies. Additional progress in this area should come from requiring our ideas about the differences among orthographies to be made precise enough to be modeled.

Acknowledgment
Work on this paper was supported by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. The authors wish to thank I. G. Mattingly for comments on an earlier draft.

References

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 644-652.

Besner, D., & Hilderbrandt, N. (1987). Orthographic and phonological codes in the oral reading of Japanese kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 335-343.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9,* 1-16.

DeFrancis, J. (1989). *Visible Speech: The diverse oneness of writing systems.* Honolulu: University of Hawaii Press.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croation is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 288-298.

Frost, R. (forthcoming). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17,* 302-310.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104-115.

Katz, L., & Feldman, L. B. (1981). Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In A. M. Lesgold &. C. A. Perfetti (Eds.), *Interactive processes in reading.* Hillsdale, NJ: Erlbaum.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 157-166.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. L (1980). Orthography and the beginning reader. In J. F. Kavanagh &. R. L. Vanezky (Eds.), *Orthography, reading and dyslexia* (pp. 137-153). Baltimore: University Park Press.

Lukatela, G., & Turvey, M. T. (1990a). Phonemic similarity effects and prelexical phonology. *European Journal of Cognitive Psychology, 18*, 128-152.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Memory and Language, 27*, 59-70.

Scheerer, E. (1986). Orthography and lexical access. In G. Augst (Ed.), *New trends in graphemics and orthography* (pp. 262-286). Berlin: De Gruyter.

Sebastián-Gallés, N. (1991). Reading by analogy in a shallow orthography. *Journal of Experimental Psychology: Human Perception and Performance, 17(2)*, 471-477.

Seidenberg, M. S. (1985). The time-course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M., Waters, G., Barnes, M. A., & Tannenhaus, M. K. (1984). When does irregular spelling and pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.

Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition, 20*, 315-328.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review 97*, 488-522.