8 Orthographic structure and spelling–sound regularity in reading English words

Richard L. Venezky and Dominic W. Massaro

Abstract

This chapter is a logical/empirical exploration of lexical and sublexical information that is potentially utilized in reading printed words. It assumes a single recognition process for words wherein information from different sources is integrated over time to yield a recognition decision, based on the relative degrees of support for the various word alternatives. Two of the sources of information that are of particular interest here are orthographic structure and spelling-to-sound correspondences. Among the approaches to quantifying orthographic structure, statistical redundancy and rule-governed measures have been the most often utilized for empirical studies over the past 30 years. These, along with spelling-to-sound mapping, are discussed and their limitations noted. Finally, three classes of metrics for scaling letter strings based on their spelling–sound correspondences are proposed: membership, exertion and fluency. For fluency, different orders of scalings are derived, using a mean fluency score for each word. Then an exertion measure and one of the fluency measures are compared to other lexical and sublexical metrics as predictors for reaction times in pronunciation and lexical decision tasks. The fluency measure accounted for a significant percentage of the variance in these studies, as did word frequency and orthographic structure. These results support the assumption that multiple sources of information contribute to printed word recognition in reading and pronunciation.

LANGUAGE PERCEPTION AND PRODUCTION
ISBN 0 12 [illegible]

Framework

The pursuit of mechanisms for word recognition in reading has occupied a central role in experimental psychology since Wundt opened the doors of his Leipzig laboratory in the 1880s. Word shape, determining letters and sound were the focal points of the earliest speculations and experiments, but spelling constraints (i.e., orthographic structure) were also probed in at least one experiment before 1910 (Dearborn, 1906; see Venezky, 1984, for a review). Since the revival of cognitive psychology in the 1950s, however, the primary focus of word recognition studies has been first on statistically defined and now linguistically defined units. In the past two decades, research on sublexical units has shifted back and forth, often like desert sand, across single-letter positional frequency, bigrams, trigrams, structural regularity, spelling–sound regularity, analogically defined phonological regularity, and even regularity defined by an edition of the Oxford English Dictionary (Parkin, 1982, 1984). Several major attempts at synthesizing this area have appeared in the last five years (Gough, 1984; Henderson, 1982), along with several notable collections of papers that touch on the subject (Kavanagh and Venezky, 1980; Lesgold and Perfetti, 1981; Tzeng and Singer, 1981).

In none of these works, however, have the sublexical properties of printed English words been carefully examined. In this chapter we focus narrowly on two different forms of sublexical information in printed words: orthographic structure and spelling-to-sound regularity. The former refers to information at the level of the letters themselves while the latter refers to information obtained by mapping letters into sounds. The motivation for this work is in part natural curiosity about orthography, but it also derives in part from the disagreements, misconceptions and occasional confusions that we often see in the stimuli selected for word recognition experiments. Our intention, however, is not to make folly or to criticize, but to delineate issues in stimulus selection that have not been attended to sufficiently in the past.

Before addressing this problem, we propose a framework for conceptualizing word recognition in reading. Such a framework, we believe, focuses the field on significant issues and inoculates us to some degree from pseudo-issues and false dichotomies. A literate person faced with a written word is captured by it and seems to have no choice but to read it. Our phenomenal experience attests to this fact, as does experimental demonstrations of the Stroop (1935) effect and its variants (Jonides, Naveh-Benjamin and Palmer, 1985). Adult readers are clearly experts rather than novices in the reading domain of pattern recognition, in the same sense that experts are differentiated from novices in chess or radiography (Chase and Simon, 1973; Lesgold, 1984). Although we do not accept a binary distinction between automatic and voluntary skills, the former more adequately represents word recognition for the adult reader (van der Heijden, Hagenaar and Bloem, 1985). Thus, we envision word recognition as a highly automatic, fast and efficient skill, but one analogous to other domains of pattern recognition such as speech and shape perception.

Following Massaro (1979; Chap 6), we assume that a single recognition process exists for printed words whereby information from various sources (e.g., features of the graphic image, letters, letter combinations and linguistic and non-linguistic context) are integrated over time to yield (usually) a recognition decision. These multiple sources of information are evaluated, and then integrated to determine what degree of support is given to alternative word candidates. The different sources are evaluated relatively independently of one another without crosstalk, and each has its own time course for becoming available. Integration involves conjoining the various sources and has the consequence that the least ambiguous sources available will have the most impact on the outcome. A decision is based on the relative degrees of support among the various word alternatives.

Consider a word being recognized in the context of a sentence. Sentential constraints might facilitate recognition (e.g., decrease naming time) even though the nature of the letter and word processing did not change. A highly constraining context would contribute much more than a less constraining context, again without modifying letter and word processing. That is, featural analysis and letter resolution would occur independently of context, but information derived from context and information derived from the word would be conjoined to give faster recognition relative to the case of either context or word information being presented alone (see Massaro, Chap 6, for a more complete presentation of this model). Depending upon the relative qualities of these inputs and the reader's prior experience, any one will be more or less important to a specific recognition task.

There is an emerging consensus among some researchers that for English highly familiar words are accessed from visual information of the letters with relatively small contributions from morphemic structure, spelling-to-sound correspondences and context. For less familiar words, poorer readers, or reduced stimulus quality, these other types of information appear to play significant roles (Becker and Killion, 1977; Gernsbacher, 1984; McConkie and Zola, 1981; Perfetti, Goldman and Hogaboam, 1979; Rosson, 1985; Waters, Seidenberg and Bruck, 1984; West and Stanovich, 1978). In terms of the present framework, the different results merely reflect differences in the relative ambiguity of the various sources of information. The contribution of sublexical sources is inversely related to the contribution (information value) of the lexical source. Within the metaphor of horse race models, all horses are always in the race, but the contribution of any one horse is a function of the speed of other horses. Lest the reader be misled by the metaphor, however, the outcome of the race reflects the contribution of several horses, not just the winner.

Orthographic structure

Statistical redundancy I — ordered approximations

Ordered approximations to English were the first measures of orthographic structure utilized in modern (i.e., post-behaviorism) reading studies. Using algorithms proposed by Shannon (1948, 1951), nth-order approximations to English were generated and used in free recall tasks (e.g., Gilmore and Egeth, 1976; Miller, Bruner and Postman, 1954). As recently as nine years ago ordered approximations were advocated as psychologically relevant descriptions of orthographic structure (Mewhort and Beal, 1977). Since ordered approximations are left-context and not right-context sensitive, however, the probability of recall of any letter in a letter string would (in this theory) be unaffected by succeeding letter context. Thus, the predicted probability of correctly reporting the *a* in a highly constraining context like *state* would be the same as for less constraining contexts like *stack* (cf. *stick*, *stuck*). In addition, the possibility that real words, pseudo-homophones, and structurally illegal strings could be generated with higher ordered approximations has led to a general abandoning of this definition of orthographic structure. There is simply too much variability *within a given order* of approximation to qualify it as a meaningful index of orthographic structure.

Statistical redundancy II — letter counts

A second approach to statistical measures of English orthography was developed from frequency counts of letters and letter sequences. Underwood and Schulz (1960), for example, tabulated bigrams and trigrams from a sample of the Thorndike and Lorge (1944) word list. These position-free token counts were used in word recognition studies, with mixed results, by Anisfeld (1964), Gibson, Shurcliff and Yonas (1970), and others. A serious drawback to these counts, their lack of position sensitivity, was overcome by Mayzner and Tresselt (1965) and Mayzner, Tresselt and Wolin (1965a, 1965b), who derived position-sensitive single-letter, bigram and trigram counts for words three to seven letters in length. (Position sensitive single letter and bigram counts from a larger corpus have been published in Massaro *et al.*, 1980 and Solso, 1979.)

The Mayzner *et al.* tables have been used by Mason (1975, 1978) to explore single-letter positional frequency effects in good and poor readers. Using a target search task, Mason (1975) found that good readers (6th grade level) were significantly more sensitive to single letter positional frequency than poor readers of the same grade level. Although Mason (1975) confounded rule-governed orthographic structure with positional frequency, a small but significant positional frequency effect remains when structure is removed (Massaro, Venezky and Taylor, 1979).

Both single-letter and bigram frequencies have continued to be used in studies of word recognition (e.g., Bouwhuis, 1979; Krueger, 1979; Massaro, Jastrzembski and Lucas, 1981; McClelland and Johnston, 1977; Rice and Robinson, 1975.) However, Gernsbacher (1984) observed that the effect of bigram frequency was often confounded with word familiarity, as measured by subjective familiarity ratings of subjects. She demonstrated that when familiarity is controlled in a lexical-decision task, the bigram effects found in several earlier studies disappear. This result, plus the high intercorrelations among pronounceability, rule-governed orthographic structure, single-letter frequency, bigram frequency and word frequency make it difficult to choose among these measures.

Rule-governed measures

Rule-governed regularity describes the predictable structure of the orthography in terms of phonological and scribal constraints. However, a critical feature of this description is that the predictable structure of letter occurrences can be utilized without any mediation of the spoken language. In contrast to the spelling-to-sound descriptions in the next section, utilization of orthographic structure in perceptual recognition is assumed to occur independently of how the printed pattern is translated to sound. Although much of the structure of a written alphabetic language necessarily follows its phonology, access to the phonology is not necessary for utilization of the orthographic structure. Direct evidence for this assumption comes from the utilization of orthographic structure by deaf readers (Gibson, Shurcliff and Yonas, 1970; Hanson, 1986). In addition, Singer (1980) found that a phonological rendering of words presented in an artificial alphabet was not necessary for learning and utilizing orthographic structure. In the sections which follow, both the phonological and the scribal constraints of English spellings are outlined.

Phonological constraints

The primary constraints on English orthography derive from the allowable sequences of sounds in English words. Thus, /sk/ is an allowable initial consonant cluster and therefore spellings for it exist (e.g., *sk*, *sc*, *sch*); on the other hand, /tl/ and /dl/ do not occur and never did occur initially in English words, and therefore no spellings for these sequences occur in English words (but they do in Tlingit and several other languages in which initial /tl/ and /dl/ occur). Complicating this relationship between sound and spelling, however, are several factors, including scribal pedantry, which leave us with spellings for (1) sound sequences which once occurred, but have since been dropped from the language (e.g., *wr-*, *kn-*, *gn*), or for (2) sound sequences that were presumed to have occurred in the language from which a word was borrowed (e.g., for *psychology*, this assumption is supportable; for *ptarmigan* it is clearly based on mistaken etymology; and for *debt* (and various others), it is only partially true, in that /b/ and /t/ occurred in the Latin ancestor of *debt* (*debitum*), but not as a final cluster). A further complication is the deviation in serial order of several spellings from the order of sounds which they represent. Chief among these is *wh*, a spelling for which some English speakers have retained /hw/, and *-le*, as in *bottle*, etc., which spells (in deliberate speech) /əl/ (*wh-* and *-le* are scribal reversals of earlier *hw-* and *-el*).

But even with these deviations from expected practice, most English sounds are represented in a moderately rational fashion in the orthography (Venezky, 1970). However, since English possesses more than one potential spelling for almost every sound, a second set of constraints, here called scribal constraints, enter into rule-governed orthographic structure.

Scribal constraints

These constraints have resulted in English spelling from 1300 years of sound change, lexical borrowings, scribal tampering and occasional attempts at standardization. While no complete codification exists, Venezky (1970) summarizes from a linguistic standpoint the major and minor scribal patterns of English. These patterns are divided into four groups: (1) gemination; (2) distribution of u and w, i and y; (3) vowels before geminates and pseudo-geminates; and (4) the distribution of c, j, k, and v, and q. In group one are the rules that regulate the doubling of graphemes in different word positions and the single rule that outlaws tripling of graphemes.[1] English does not allow a, h, i, j, k, q, u, w, x, and y to double (geminate), and restricts the doubling of *v* to a small group of words (e.g., navvy, savvy). For doubled *k*, *g* (pronounced /j/) and *ch* English scribes of the seventeenth century adopted *ck*, *dg*, and *tch*, respectively. These *pseudo-geminates* obey the same distribution rules as true geminates. A few exceptions to the gemination rules exist, such as *aardvark*, *skiing*, *trekked*, but these are rare. In general, doubled consonants (and the pseudo-geminates) can not occur in initial position, and those freely admitted in final position are restricted to ff, ss, and ll,[2] plus the pseudo-geminates (but *dg* must always be followed by *e*). Based on these rules, pseudo-words like *taaf*, *ooming*, *bribb*, *ssoch*, *ckip*, and *swodg* would be pronounceable, but scribally irregular.

In the second group of scribal rules are those that control the distribution of *u*/*w* and *i*/*y* as second elements of digraph vowels, and which restrict *u* and *i* to initial and medial positions. These two graphemes alternate as second elements of digraph vowels: *u* occurs before consonants while *w* is used before vowels and in word final position (e.g., *sound*, *tower*, *cow*).[3] Less frequent are a small group of French borrowings: *caribou*, *tableau*, etc. Similarly, with ai/ay, ei/ey, oi/oy and ui/uy, the *i* form occurs before consonants while the *y* form occurs before vowels and in word final position. (The words that admit *i* in final position are Latin plurals like *radii*, which is also irregular for having a doubled *i*; Italian plurals like *ravioli* and *macaroni*; and a handful of others of various origins: *magi*, *yogi*, *rabbi*, *ski*, *khaki*, etc.) The *u*/*w* distribution rule has a relatively large number of exceptions where the *w* form occurs before a consonant (e.g., *bowl*, *town*, *owl*, *pewter*, *powder*). Exceptions to the *i*/*y* distribution are rarer (e.g., *oyster*). By these rules, pseudo words like *snou*, *tawp* and *nayh* are scribally irregular.

The third group of rules outlaws digraph vowels before doubled consonants and before the pseudo-geminates *ck*, *dg* and *tch*.[4] Thus, pseudo-words like *soick*, *treedge* and *teaffer* are also illegal. In the last group are the distribution rules for *c* (not allowed in final position except in the ending *-ic*), *j* (allowed only in initial position), *k* (generally restricted initially and medially to occurring before e, i, and y, and finally only after digraph vowels), *q* (must be followed by *u*) and *v* (not allowed in final position). Exceptions to these rules are rare, consisting mainly of a handful of French borrowings that end in *c* (*cognac*, *zodiac*, *lilac*), plus a few others (e.g., *zinc*, *talc*, *bloc*, *arc*).

Pseudowords that violate these rules include forms like *blec*, *pojer*, *koff*, *qraft* and *brev*. Massaro *et al.* (1980) showed through subjective ratings and two alternative forced choice tasks that subjects utilize both rule-governed information and statistical information in discriminating among pseudowords. Hart (1980) obtained familiarity ratings (closeness to English words) for grapheme strings that incorporated or violated 12 different types of scribal regularity, holding phonological regularity constant. The lowest ratings (least word-like) were obtained for *dg*, *tch* and *ck* in initial position (e.g., *tcheam*) and doubled consonants in initial position (e.g., *nonish*). In a related lexical decision task, Hart (1980) found no pseudo-homophone effect for pronounceable but scribally irregular pseudo words (e.g., *chode*), a result replicated by Venezky (1981).[5] Finally, Massaro and Hestand (1983) showed that primary school children improve grade by grade from first through third grade in their ability to utilize rule-governed information.

Spelling–Sound regularity

The second form of sublexical information of interest here is spelling–sound regularity, which is based on the consistency with which graphemes can be mapped into phonemes. E. Gibson and colleagues attempted to define a dichotomous pronounceability (i.e., regularity) measure which they then employed in a full recall task (see Gibson and Levin, 1975, for a review). Besides confounding recognition with memory for unfamiliar letter strings (Baddeley, 1964), the Gibson studies also confounded structural regularity with pronounceability, as was found by Gibson, Shurcliff and Yonas (1970) using deaf subjects who also showed a pronounceability effect. This possibility had been demonstrated earlier with primary school children by Thomas (1968), who showed that pronounceable and unpronounceable CCVs (e.g., dri, rdi) were equally inferior to pronounceable CVCs (e.g., dir) in an oral recall (spelling) task. A number of researchers have attempted to show that lexical access in reading English is mediated by phonology generated by rule from graphemic input (e.g., Baron and Strawson, 1976; Coltheart *et al.*, 1977; McCusker, Hillinger, and Bias, 1981; Rosson, 1985; Rubenstein, Lewis, and Rubenstein, 1971; Stanovich and Bauer, 1978). Henderson (1982) discusses in depth the complex and often contradictory empirical evidence for phonological mediation and the theories proposed to account for these data (see also Humphreys and Evett, 1985; and Venezky, 1981).

Since Henderson's (1982) text, a number of studies have, through improved stimulus control, shown that phonological effects are confined to low familiarity words (e.g., Rosson, 1985; Waters and Seidenberg, 1985; Waters, Seidenberg and Bruck, 1984). In a recent study, Seidenberg (1985) showed that in a naming task, not only was the spelling–sound effect restricted to the low frequency words, but was also absent in the fastest one third of the subjects as defined by mean naming latency. Phonological effects have been obtained primarily with pseudowords (e.g., Rubenstein, Lewis, and Rubenstein, 1971); those from real words tend to be relatively small (e.g., Gough, 1984; Stanovich and Bauer, 1978). This latter observation applies whether spelling–sound rules, lexical analogies (Glushko, 1979), or sound-to-spelling tabulations (Rosson, 1985) have been utilized. Nevertheless, phonological effects have been found consistently in lexical access for Serbo-Croatian readers (e.g., Feldman and Turvey, 1980) and in categorization tasks for Japanese kana (e.g., Kimura, 1984) and Chinese (see Hung and Tzeng, 1981 for a review).

Issues in rule derivation

Most studies that use a rule-governed approach to spelling–sound regularity draw on linguistic descriptions of English orthography for rules. However, several issues remain unresolved in the derivation of spelling–sound correspondences. The first relates to the context to consider in rule derivation. For example, if the *d* in *Indian* is viewed only in the context of other *d* pronunciations, we would assign it to the *regular* category. But in relation to the following vowel context and word stress pattern, *d* in *Indian* is irregular because it does not palatalize to /ʃ/ as in *cordial* and *educate* (cf. ʃ in *issue*, ʒ in *azure* and tʃ in *bastion*). In the next section we propose a series of levels or orders for describing increasingly more sophisticated spelling–sound rules such as these.

A second issue relates to the corpus of words from which rules are derived. Venezky (1970) is based, with some editing, upon a corpus from the 1930s that includes many inflectional and derivational forms, plus proper names. A more recent corpus, or one restricted to base forms might alter some of the rule statistics. Then, the notion of *rule* itself must be questioned. Venezky (1970) bases rules on recognizable word features, including stress and morpheme boundaries, but excluding etymology. The latter feature, which figures prominently in Chomsky and Halle (1968), raises issues of validity and reliability. While some words contain foreign spellings (e.g., *pk*), many carry no obvious clues to their origins. Furthermore, origin itself is often insufficient for predicting spelling–sound correspondences. For example, both *chief* and *chef* are derived from the same French word, yet *chef*, which was borrowed relatively recently, exhibits a modern French pronunciation of *ch*, while *chief*, which was borrowed before French initial /tʃ/ changed to /ʃ/, retains the earlier French form. Similar dichotomies can be found in *ch* words from Greek (e.g., *machine*, *machination*), many of which were borrowed via Latin or French.

Exactly what makes a rule is beyond the coverage of this chapter. We will, nevertheless, offer one example of the complexity of this issue. *Bush* and *push* both contain what most people would classify as irregular pronunciations of *u* (cf. *flush*, *rush*, *crush*, *dull*, *hull*). However, after a non-nasal bilabial, *u* corresponds regularly to /ʊ/ in some following consonant environments (e.g., *full, pull, bull, bushel, butcher, fulsome*). Does this make the *u* in *bush* and *push* regular? Semi-regular?

Given a set of rules from whatever source, an issue remains in how to use the rules to scale regularity. While most experimenters have treated regularity as a binary variable, at least one (Rosson, 1985) has tried to create a continuous regularity scale by extracting the log of the weakest (i.e., least common) correspondence in a word. This measure, unfortunately, confounds spelling–sound regularity and grapheme frequency of occurrence. By this measure, a highly regular but relatively infrequent correspondence (e.g., *qu* — /kw/ would be scaled lower than a common but irregular correspondence (e.g., ea — /e/). As the next section shows, these two approaches do not exhaust the possible scaling techniques for spelling–sound regularity.

Lexical analogies

One of the more serious drawbacks to rule-based systems is their clumsiness in accounting for rule change. Developmental studies of letter–sound learning show, for example, that children change from an invariant /k/ pronunciation for *c* in initial position, to a /k/–/s/ differentiation in initial position, to a full differentiation across all word positions (Venezky, 1974; Venezky and Johnson, 1973). If we assume that at each stage only the currently active rules are accessible, then the construction of new rules requires either deductive learning, which is unlikely, or re-exposure to the corpus of *c* words, with attention now to the graphemes immediately following *c*. A more attractive alternative utilizes active access and reorganization of words stored in a mental lexicon.

A step in this direction has been proposed by Glushko (1979), who suggested that both real word and pseudoword pronunciations can be accounted for by reference to the set of real words activated by a stimulus's ending. For real words, lexical access is made on a holistic basis. However, at the same time that a word's pronunciation is activated, so are the pronunciations of all similar words (i.e., neighbors). If any of these have ending pronunciations that differ from the ending pronunciation of the stimulus, response delay occurs. For pseudowords, a similar process is hypothesized, with uniform pronunciation neighbors producing faster and more accurate pronunciations than neighbors with variant pronunciations.

As a first step, the model is appealing, but it leaves many issues unresolved. For example, by the strict criteria Glushko used for selecting neighbors (same length, same spelling after initial grapheme), many pseudowords can be constructed that have no real word neighbors (e.g., *zehe, fibe, tufe, soge*). The Glushko model cannot account for the pronunciations that subjects give to these, nor can it account for initial grapheme pronunciations or pronunciations of most multi-syllabic words.

Bauer and Stanovich (1980) attempted to utilize the Glushko model, but altered the neighborhood criteria to include variable lengths and chaining via pronunciation to alternate spellings. Thus, *size* might activate (via its pronunciation) *sight* which would activate (also via pronunciation) *height*, which activates (by spelling) *weight*, which activates *fate* and *great*, and so on through an unmanageably large corpus. It is not clear how a subject would arrive at a pronunciation via this model. Because of these extensions to the neighborhood criteria, Bauer and Stanovich (1980) is not an accurate replication of Glushko (1979). For example, in their second experiment, five of the nine words which they label 'regular, inconsistent' are 'regular, consistent' by Glushko (1979).[1]

Henderson (1982) has proposed several alterations to Glushko's criteria, including the marking of spelling units and vowel–consonant status, and activation by regions of visually related forms, including those in initial position. These suggestions move the activation model towards the rule-based model, while retaining the ability to account for change. But little empirical evidence has been gathered to test these ideas. In the next section we propose a sequence of increasingly more sophisticated linguistic metrics for scaling spelling–sound regularity, taking into account different scaling techniques and different approaches to defining spelling–sound rules. Our only claim for these is that they offer a systematic approach to exploring the psychological relevance of spelling–sound translation systems.

Spelling–sound metrics

For the present we consider the generation of a pronunciation from the spelling of a visually presented stimulus. Furthermore, we consider only those generation procedures that operate on spelling–sound correspondences. That other approaches to arriving at a pronunciation have been suggested, or that spelling–sound models have been subjected to various criticisms is not at issue here in that our primary concern is to explicate those stimulus properties that need to be considered in testing spelling–sound models. Following the ideas presented at the beginning of this chapter on processing of information from different sources, we are not assigning any particular level of importance to spelling–sound information over other sources such as orthographic structure. The goal of the current exploration is to derive a metric for scaling letter strings such that assumptions about pronunciation via spelling–sound translation can be tested.

Three classes of metrics are considered: *membership*, *exertion* and *fluency*, corresponding to dichotomous, interval and continuous scaling.

Membership

This class of metrics assigns letter strings to regular and irregular classes upon the basis of some canonization of spelling–sound rules. Besides the criticisms mentioned above of spelling–sound rule systems, this class also suffers from low resolving power; it can not make differential predictions within either the regular or irregular class. For example, the mapping of *i* to silence in *business*, which is a unique irregularity, must be treated the same as the more common irregularity of *i* in *triple*.

Exertion

Exertion metrics attempt to scale spelling–sound regularity along an integer scale, utilizing the number of different pronunciations for each spelling unit in a word as a basis for scaling. Consider, for example, the spelling *heir*. Initial *h* has a highly common pronunciation (/h/), and a rarer one as in American English *herb* (ø). *ei* has six 'common' pronunciations, as exemplified by the words *feint*, *either* (/i/), *height*, *albeit*, *counterfeit* and *heir*, of which the first five occur more often (by type counts) than the *ei* pronunciation in *heir*. For *r*, the pronunciation in *hair* represents the most common mapping. For scaling, we can assign a count of zero to each correspondence that is the most common for the spelling involved, a one if it is the second most common, and so on. These separate counts can then be combined as a sum for the entire word. By starting the counts at zero, some degree of independence from the number of spelling units in a word is achieved (for *heir*, the process would yield a score of six.) This is labeled an *exertion* metric because it corresponds (roughly) to the number of different pronunciations that would need to be generated before the correct one is produced, assuming that the order of testing proceeds from more common to less common correspondences.

Fluency

A fluency score utilizes the actual frequencies of occurrence for correspondences to create a continuous scale. One approach to this metric has been tried by Rosson (1985), who utilized the lowest frequency correspondence, based on sound to spelling type counts. Besides confounding spelling unit frequency with spelling–sound frequency, as [illegible]sed earlier, this measure assumes a parallel assignment of cor[illegible]uch that the

slowest single correspondence controls reaction time. While there is insufficient evidence to reject this hypothesis, we prefer at this early stage in the development of metrics to propose alternative approaches for comparison to each other.

Therefore, as one alternative to a weakest rule fluency metric, we propose a mean fluency metric. The *fluency score* assigned to a particular spelling–sound correspondence is based on the occurrences of that correspondence relative to the sum of all correspondences for that same grapheme. That is, for a complete set of correspondences $f_1, f_2, \ldots, f_n$ of a spelling, the fluency score of any f_i is $(f_i/\sum_{1} f_i) \times 100$, where multiplication by 100 transforms the distribution from a 0–1 scale to a 0–100 scale. This approach ensures that consistent but infrequent correspondences like *z* → /z/ will have higher fluency scores than inconsistent but frequently occurring correspondences like *ea* → /e/. For the *c* pronunciations shown in Table 1, the fluency scores are: /k/ = 74.22; /s/ = 21.98; /ʃ/ = 3.67; and /č/ = 0.03. The total fluency score assigned to a word is the mean of its grapheme fluency scores.

Table 1. Correspondences for *c*.

Pronunciation		Word position		
	total	initial	medial	final
/k/	2431	1336	718	377
/s/	719	150	569	0
/ʃ/	120	0	120	0
/č/	1	1	0	0

Rule systems

Fluency scores, like membership and exertion scores, derive from a base set of spelling–sound rules. But the notion of *rule* in this domain is not well-formed, thus allowing different rule sets, depending upon the features which are admitted for rule formation. For the computation of mean fluency scores we define four levels or types of spelling–sound rules, here called *orders*. These are also applicable, *mutatis mutandis*, to membership and exertion scalings, although because of space limitations we apply them here only to mean fluency scores. A zero order spelling–sound scaling is based upon single grapheme counts without regard for position of letters within words or any higher order contextual or linguistic considerations. Notice that this definition leads to treating digraphs and trigraphs as sequences of single graphemes rather than as separate units. Thus, the fluency score for *chin* would be derived from the separate scores for *c*, *h*, *i* and *n* at the zero order. (In this case, we treat the correspondences of *ch* as *c* → /k/ and *h* → /ϕ/.) From both a psychological and a linguistic standpoint, the zero-order metric is uninteresting, and will not be pursued further.

A first-order scaling takes into account functional units,[7] so that *chin* is now treated as a sequence of three units: *ch*, *i*, *n*. Position of graphemes within words, however, is still ignored. This means, for example, that the fluency score for final *e* in words like *flame* and *take* is based on all possible correspondences for *e* rather than just those for *e* in final position. For a second-order measure, we take into account grapheme positions within the word; namely, initial, final and medial, where medial is everything left after the first (initial) and last (final) units are removed. For the *c* correspondences given in Table 1, we now have fluency scores for each word position as shown in Table 2. For final *e*, the fluency score for *e* → /ϕ/ as in *flame* is now 100 rather than 15 for first order *e* → /ϕ/.

Table 2. Second-order fluency scores for *c* pronunciations.

Pronunciation	Word position		
	initial	medial	final
/k/	89.84	51.03	100.00
/s/	10.09	40.44	0.00
/ʃ/	0.00	8.53	0.00
/tʃ/	0.07	0.00	0.00

Third order scalings admit the more general linguistic rules; namely, the long–short vowel rule (including the final *e* pattern);[8] the *c* rule; palatalization of /s/ /z/, /t/, /d/; the phonotactics of consonant clusters, and *r*-influence.[10] For the pronunciations of *c*, rules at this level render almost all of them totally regular, the exceptions being words like *arcing*, *facade* and *cello*. At this level, the correspondences for *c* are divided into three classes as represented by the following word groups: (1) *coal*, *picnic*, *secret*; (2) *ceiling*, *ocelot*, *society*; (3) *ocean*, *social*. Fluency scores are then computed separately for correspondences within each of these groups. The fourth and highest level of scaling takes into account all of the minor spelling sound rules that have potential psychological reality. This includes, for example, a rule for initial *th* based on form class (functors versus content words), but excludes rules based on etymology. Shown in Table 3 are fluency measures for three words: *chin*, *shape* and *cent*, at five levels or orders of scaling.

Table 3. Fluency measures.

Word	Order of scaling				
	0	1	2	3	4
chin	47.6	83.1	83.6	85.8	85.8
shape	33.4	65.5	81.0	99.7	99.7
cent	57.0	57.8	62.4	77.0	99.3

For *chin*, the step from zero to first order scalings yields a large increase in the fluency score because at the zero order *ch* is treated as a sequence of two units with *c* mapped into /k/ and *h* into silence. But higher-order rules have only a marginal effect thereafter because neither position nor context have a major influence over the mappings involved. For *shape* large increases occur in metrics 1, 2 and 3 because of (1) the change from treating *sh* as two separate units at the zero order to treating it as one at order one; (2) the change at order two to considering position in the correspondences for *e*; and (3) the addition of the long–short vowel rule at order three. The fourth-order scaling, on the other hand, produces no further increase in the score for *shape*. *Cent* differs from *chin* and *shape* in that the zero and first order scalings yield identical fluency scores while the second-, third-, and fourth-order scalings produce increasingly higher scores. The largest increases derive from the long–short rule (third order) and the *c* rule (fourth order).

Testing psychological reality

One test of the psychological reality of the various metrics proposed here could be based on English words that discriminate clearly among the different predictions made. Another possibility is to evaluate how well these metrics predict performance on individual words in pronunciation and lexical decision tasks. Following this latter suggestion, we evaluated these new metrics against a recent series of experiments by Waters and Seidenberg (1985). The authors systematically varied word frequency, spelling-to-sound correspondences, and orthographic regularity in six pronunciation and lexical-decision tasks. Although a relatively small number of words was used in each experiment and each task, RT differences between certain stimulus sets were significant and revealed that a systematic analysis was warranted. To overcome the limitations with correlations based on small sample sizes and to simplify the analysis, the results were modified to allow three grand analyses. First, for each word in a given task in a given study, the dependent measure for that word was its RT minus the mean RT across all words in that task and study. This was done for every word in every task in every study except for words that did not occur in the Kučera and Francis (1967) word list. In addition, only words between four and seven letters in length) were included because of certain constraints of assigning measures of orthographic structure.[11] This gave a total of 505 cases for the analysis. Since some words were used in several tasks, there were not 505 unique words although there were 505 unique dependent measures.

Six independent variables were included in the analysis. Log word frequency was based on log to the base 10 of the Kučera and Francis (1967) corpus. The orthographic measures were average position sensitive log bigram and trigram counts based on the Massaro *et al.* (1980) sublexical counts of the Kučera and Francis word list. To provide a measure independent of word length, only initial, medial and final positions were included. The counts were based on all words between four and seven letters in length. Thus, each test word, regardless of word length, could be assigned a bigram count from a single table and analogously for the trigram count. The exertion measure and the second-order fluency measure were included as measures of spelling-to-sound correspondences. Finally, number of neighbors was computed for each word by counting the number of words that differed by a single letter from each test word (again based on Kučera and Francis, 1967).

Partial correlations are given in Table 4, reflecting the correlation of each independent variable, with the dependent variable, with the variance of all other variables pulled out. As can be seen in the table, both orthographic structure and spelling to sound correspondences account for some of the variance. Word frequency also is an important variable although number of neighbors is not. Given the findings of Waters and Seidenberg (1985) that differences were found for only words of low word frequency, the analysis was repeated for high and low frequency words analysed separately. These partial correlations are also given in Table 4. Replicating the conclusions based on group comparisons, the item analysis shows no significant effect of spelling-to-sound correspondences for high-frequency words. However, the effect of orthographic regularity is marginally significant (p≤0.10) for these same words.

Table 4. Partial correlations between RTs and six predictor variables for Waters and Seidenberg (1985).

Predictor	All words	Low frequency	High frequency
log word freq	−0.293*	−0.418*	0.058
fluency	−0.137*	−0.219*	0.073
log bigram	−0.146*	−0.143*	−0.097
neighbours	−0.028	−0.020	0.024
exertion	0.036	−0.027	−0.139*
log trigram	0.082	0.052*	0.069

Number of cases is 505, 280 and 225 for all words, low frequency words, and high frequency words, respectively.

* Significant (p< 0.05) and in the appropriate direction.

Conclusions

The new metrics proposed here for spelling sound regularity reflect some of the uncertainty that remains to be resolved in visual word recognition. Testing of orthographic structure and spelling–sound regularity represents a maturing process that has occurred over the past decade. By contrasting simultaneously different metrics for orthographic structure and spelling–sound regularity, we have the potential of describing the multiple sources of information that are functional in reading printed words.

Notes

[1] Following Venezky (1970), we use *grapheme* to refer to the alphabetic units a . . . z, corresponding to Henderson's (in press) *sense 1*. For compatibility with a highly diverse literature, however, we will continue to use the term *letter* (and its hyphenates) when it is the label commonly employed in the work under discussion.

[2] Exceptions exist for b(ebb), d(add, odd), g(egg), and n(inn).

[3] Two common exceptions to this latter rule are *you* and *thou*.

[4] Almost all exceptions to this rule are French borrowings like *chauffeur*.

[5] cf. Scheerer, Chap 12.

[6] Base, duel, lucid, shown, throng.

[7] Functional units for English are (a) those graphemes or grapheme sequences whose phonological correspondences can not be predicted from the correspondences of their separate components (e.g., ch, tch, a, b, ea) or (b) graphemes whose primary function is to indicate the pronunciation of another functional unit or preserve a graphotactical or morphological pattern (e.g., *e* in *rage* and *purse*, *u* and *e* in *catalogue*). The often employed 'grapheme-phoneme rule' is actually inconsistent, in that mappings are often not between a grapheme and phoneme but between a functional unit and one or more phonemes (e.g., ch-/č/).

[8] A single letter vowel spelling is short (checked) before a compound consonant (e.g., x, ck) or a consonant cluster; otherwise it is long (free).

[9] c is soft (/s/) before e, i, or y; otherwise it is hard (/k/). (Palatalization of /s/ to /š/ as in *ocean* is covered by the next rule.)

[10] For these three rules, see Venezky, 1970.

[11] With these constraints, only 6 of 224 words were dropped.