English orthography: its graphical structure and its relation to sound

RICHARD L. VENEZKY *University of Wisconsin*

PRESENTS AND organizes sets of orthographic patterns, based upon an analysis of the spellings and pronunciations of the 20,000 most common English words. Two basic sets of patterns are discussed. The first pertains to the internal structure of the orthography: the classes of letters (graphemes) and the allowable sequences of these letters (graphotactics). The second set contains those patterns which relate spelling to sound. Each functional orthographic unit is classed as a relational unit or a marker, depending upon whether it relates to sound itself or marks the functioning of some other unit. Relational units are mapped first onto an intermediate (morphophonemic) level by one set of rules, and then into sound by another set of rules, thus allowing a clear separation of rules based primarily upon orthographic considerations from those based primarily upon morphemic and phonemic considerations.

L'orthographe anglaise: structure graphique et rapport avec le son

CET ARTICLE présente et organise des séries de «patterns» orthographiques basés sur une analyse de l'orthographe et de la prononciation des 20,000 mots les plus usités de la langue anglaise. Une discussion suit sur deux séries fondamentales de «patterns». La première série se rapporte à la structure interne de l'orthographe: classes de lettres (graphèmes) et séquences permissibles en Anglais de ces lettres (graphotactique). La seconde série comprend les «patterns» qui rattachent l'orthographe au son. Les unités fonctionnelles sont divisées en deux classes: unités de rapport et signaux, selon qu'elles sont en rapport avec le son lui-même ou qu'elles signalent le fonctionnement d'une autre unité. Les unités de rapport sont schematisées d'abord à un niveau intermédiaire (morphophonémique) selon un groupe de règles, et ensuite au niveau de son selon un autre groupe de règles; ceci permet une nette séparation des règles basées principalement sur des con-

75

sidérations d'ordre orthographique et de celles basées principale-
ment sur des considérations d'ordres morphémique et phonémique.

Ortografía del Inglés: su estructura gráfica
y su relación con el sonido

PRESENTA Y organiza una serie de patrones ortográficos, basados
en el análisis de la ortografía y la pronunciación de 20,000 de las
palabras inglesas más comunes. Se discuten dos series de patrones
básicos. La primera se refiere a la estructura interna de la orto-
grafía: las clases de letras (grafemas) y el orden de sucesión
permisible de estas letras (grafotácticas). La segunda serie com-
prende aquellos patrones que relacionan la ortografía con el so-
nido. Cada unidad ortográfica funcional se clasifica como una
unidad relacional o un marcador, dependiendo de si se relaciona
al sonido mismo o si marca el funcionamiento de alguna otra
unidad. Las unidades relacionales son organizadas primero en
un nivel intermedio (morfofonémico) mediante un conjunto de
reglas, y despues en sonidos mediante otro, permitiendo así hacer
una separacíon clara de las reglas que se basan principalmente
en consideraciones ortográficas de aquellas basadas principal-
mente en consideraciones morfémicas y fonémicas.

English orthography contains two basic sets of patterns. The first is the internal structure of the orthography: the classes of letters (graphemes) and the allowable sequences of these classes (graphotactics). The second, and the more complex, is the set of patterns which relate spelling to sound. In the first set are patterns based solely upon graphical considerations and which an illiterate must acquire in learning to read. The second set includes not only patterns which are based upon the idiosyncrasies of the orthography, but also patterns which result directly from English phonological habits. The illiterate speaker of English already has mastered the latter patterns in learning to read, therefore, he does not have to learn them *de novo,* but rather must learn to relate them to orthographic stimuli. The present study describes these patterns, not just in terms of regular spelling-to-sound rules, but in terms of the more general phonemic and morphemic elements which characterize the present system of orthography. Furthermore, suggestions are made for incorporating these patterns into the teaching of reading.

For centuries philologists have approached the study of English orthography with the purblind attitude that writing serves only to mirror speech, and that deviations from a perfect letter-sound relationship are irregularities.

Whatever may have been the relationship between writing and sound when the first Old English writings were inscribed in Latin script and whatever may have been the reason for the subsequent development of this system, be they due to random choice or to an all-pervading National Orthographic Character, the simple fact is that the present orthography is not merely a letter-to-sound system riddled with imperfections, but, instead, a more complex and more regular relationship wherein phoneme and morpheme share leading roles. The present study attempts to show these different levels of patterning in the current orthography.

Survey of the literature

Today neither a comprehensive study of Modern English spelling-to-sound correspondences nor a detailed history of English spelling exists in print. This kind of information must be gathered from a multitude of diverse sources and pieced together much as ancient vases are reconstructed from unassorted potsherds. The earliest writings on English orthography are based upon an alphabetic principle derived from the fourth and sixth century Roman grammarians. Each

letter of the alphabet has, besides its name (nomen) and appearance (figura), a power (potestas) or sound. A description of the orthography involves simply a classification of the letters according to their powers.[1] Thus, orthography from the time of Alfred to the present day has been delimited by the letters and their powers. So ingrained has this principle become that some contemporary linguists have attempted, by substituting *grapheme* for *letter,* to sanctify it with the countenance of linguistic science without examining how unsound it is.[2]

Some critical exceptions to this view are found in the writings of Francis (1958), Hockett (1958), and Vachek (1959). While they all have recognized that the orthography is more complex than an irregular letter-to-sound system, only Francis attempted, with some success, to analyze the relationship between spelling and sound and to enumerate the non-phonemic elements entering into this relationship. Most other linguists have paid little attention to the orthography. For example, Bloomfield (1933) held that writing was not a part of language, but simply an imperfect image of speech. Even though he wrote at length on the teaching of reading, he maintained that English orthography was simply a grossly irregular alphabetic system (1933, pp. 500–501).

Spelling reformers, although having contributed an enormous quantity of literature on English orthography, seldom analyzed the object of their scorn beyond the more common examples of scribal pedantry. Their arguments were, and are still, based upon the *a priori* assumption that alphabets should be perfectly phonetic. It is no surprise, therefore, that most spelling reformers concerned themselves with direct letter-to-sound relationships and ignored all other facets of the writing system.[3]

Grammarians, like spelling reformers, also viewed the orthography as a mirror for speech. The earliest grammarians were primarily concerned with correct pronunciation and directed their attention towards relating spelling directly to sound.[4] As spelling was regularized, they turned more and more toward establishing spelling rules, while still retaining a direct spelling-to-sound standpoint. From the time of the earliest English grammars, however, a small number of grammarians discerned non-phonetic features in the orthography,

1. This trichotomy is discussed by Einar Haugen (1950, pp. 41–42).
2. The so-called grapheme-phoneme parallel is discussed in Venezky (1965).

3. See in particular R. E. Zachrisson (1931).
4. Summaries of statements on orthography from Aristotle to Murray can be found in Brown (1859).

although few carried out comprehensive analyses of such features. Most prominent among these are Alexander Hume (1647) and James Douglas (1740). The most recent description of direct letter-sound relationships is Alex Wijk's *Rules of pronunciation for the English language* (1966). While carefully written and informative, this work differs little in its basic approach from similar works by Douglas and Craigie (1927).

Source materials

The research reported here began at Cornell University in 1961 as part of an inter-disciplinary study of the reading process (Levin, 1963). After an initial study of the spelling-to-sound correspondences in monosyllables, a computer program was written to derive and tabulate spelling-to-sound correspondence in the 20,000 most common English words (Venezky, 1963).[1] This program was used to obtain, for a corpus of the 20,000 most common words in English, the following information:

1] A complete tabulation of the spelling-to-sound correspondences found in the corpus, based upon the position of consonant and vowel clusters within the printed words. For any continuous string of vowels or consonants found in a printed word, the tabulations include all of the pronunciations found for that string, along with the totals and percentages for each pronunciation in each word-position (initial, medial, and final).

2] Complete word lists for each correspondence found in 1 above. For example, if the cluster *gh* with the pronunciation /g/ occurred in 1 above, then the word list would contain all of the words in the corpus in which *gh* was pronounced /g/, arranged into separate, alphabetized lists for the three word positions.

3] The same as 1 and 2 above for the 5,000 most common words in the corpus and the graphic monosyllables.

4] A dictionary of the corpus in which spellings were reversed and then alphabetized. This list was used extensively for studying suffixes and other word endings.[2]

The Cornell data collection phase was followed by several years of analysis and research at Stanford University in collaboration with the late Professor Ruth H. Weir. The primary goal in this phase was to construct a theoretical framework for deriving sound from spelling and to search for the most general patterns in the orthography and the most plausible linkages for fitting these relationships into the total language structure. Most of the present report is concerned with the results of this work. The third phase, not described here, is the testing phase, now under way at the University of Wisconsin and elsewhere. From the results of phase two, experiments are being designed to explore how literates translate from spelling to sound and how children acquire this translating ability. Theoretical patterns from phase two are being used as guides for exploring reading habits. That is, given that a certain general pattern is constructed to account for a spelling-sound relationship, the question that remains to be answered is: "Does the literate actually use this pattern?"

The synchronic material presented is based upon the spelling-to-sound correspondences from the tabulations mentioned above, although words outside of the 20,000 word corpus are cited occasionally to illustrate interesting or unusual patterns.[1] Proper nouns, contractions, hyphenated words, and variant pronunciations of the same spellings are not included in this corpus. Most pronunciations are derived from *A pronouncing dictionary of American English* (Kenyon & Knott, 1941). Broad IPA transcriptions are employed for the pronunciation of Modern English.[2] Except where non-phonemic contrasts are discussed, these symbols are enclosed in slant lines, e.g., /e/. The same symbols are used between double slant lines, e.g., //e//, to represent morphophonemic forms. Graphemic units are given in italics, e.g., *e*.

Schwa is used both for the stressed vowel as in *come* and the unstressed neutral vowel, as in the first syllable of *away*. To indicate the correspondences between graphemic, morphophonemic, and phonemic units, the abbreviations *x*—//y//, //x//—//y//, and //x//—/y/ are written. The dash (—) indicates that the unit on the left corresponds to the unit on the right. This is always to be interpreted as a one-way correspondence from left to right; correspondences in the

1. In the *Thorndike-Century senior dictionary* (Thorndike, 1941), the most common 20,000 words according to the Thorndike frequency count are identified. Many low-frequency words from that list, especially proper nouns, were omitted by the present writer, and a number of words not included in the original Thorndike list were included.
2. See Kurath (1964, Ch. 1) for a discussion of the transcription system used here.

other direction, that is, sound-to-spelling correspondences, are beyond
the scope of this paper.

Preliminaries to orthographic analysis: graphemic features

Spelling units

Even from the direct letter-to-sound standpoint, the graphemic
system is more complex than is revealed in the notion that there are
twenty-six letters or graphemes which, through careful manipulation,
can be mapped into the phonemes of English. From the enumeration
of the twenty-six graphemes to the point where correspondences to
phonemes can be considered, a number of complexities must be un-
tangled. One is the designation of the spelling units themselves. Obvi-
ously, there are more than twenty-six functional units: *th*, *ch*, and *oo*,
for example, are as basic to the current orthography, as *a*, *b*, and *t*.
But are *tch*, *ck*, and *dg* primitive units, on a level with *a* and *th*, or are
they in some sense compound units whose correspondences to sound
can be predicted from their immediate constituents?

To map from spelling into sound, regardless of the inter-
mediate levels which are introduced, graphemic words must be seg-
mented into their basic graphemic units. This requires a systematic
procedure for handling letters like the final *e* in *rove* and the *b* in *debt*.
Is, for example, the *e* in *rove* connected to *o*, forming the discontinuous
unit *o . . . e*, or is it part of the unit *-ve*, or is it a unit by itself? And,
similarly, how is the *b* in *debt* to be handled? As part of the unit *eb*, or
of *bt*, or as a separate unit? The solution to this problem should not
only be consistent with the way similar graphemes are handled, but
also general enough to handle new cases which may arise. The designa-
tion of spelling units in *Webster's new collegiate dictionary* (Bethel,
1956), for example, fails to meet both of these aims. The editors give
no general rules for handling silent letters, but instead, haphazardly
and quite inconsistently, classify individual cases as they arise in
traversing an alphabetical list of spelling units. For example, *gh* "in
aghast, ghostly and *ghost* is a useless spelling for 'hard' *g*. . . ." (1956,
p. xii). However, *rh*, as in *rhetoric* and *rhesus*, is not mentioned as a
separate unit. *ng* is a single unit in words like *long*, corresponding to
/ŋ/, but *mb*, on the other hand, is two units, the second being
silent.

These problems cannot be settled satisfactorily by simply labeling all unpronounced letters as silent. Consider the so-called silent *b*'s in *subtle* and *bomb*. One could say as *Webster's* does, that the *b*'s in these two words are silent, and let the matter rest. But, by doing so, an important difference that exists in these two cases is neglected. The form *subtle* occurs only with the *b* corresponding to zero, but *bomb*- in *bombard* and *bombardier* has non-silent *b*. It is incorrect, therefore, to say that the second *b* in *bomb*- is silent; the more exact statement is that it is silent before word juncture and before certain suffixes (cf. *bombing, bombs, bombed*). This is one of the forms of orthographic patterning that almost all traditional treatments of spelling overlook.

Another inherent feature in the orthography is the distinction between functionally simple and functionally compound consonant units. One of the most general, although not entirely regular, spelling-to-sound rules is that the vowel spellings *a, e, i, o, u* are mapped into one form before a single consonant unit which is followed by a vowel spelling, and into another form in all other environments. In the vocabulary of the direct letter-to-sound school, these forms are the *long* and *short* pronunciations of the vowels, as shown in the examples below (*free* and *checked* in the vocabulary used here):

Table 1   Long and short pronunciations of vowels

	Free (long) pronunciation	Checked (short) pronunciation
a	/e/ anal ache	/æ/ annals ratchet
e	/i/ fetal ether	/ɛ/ fettle hedge
i y	/aɪ/ hypo writhe	/ɪ/ hippo whittle
o	/o/ phonograph kosher	/a/ sonnet noxious
u	/(j)u/ super fuchia	/ə/ supper luxury

To apply this rule, simple and compound consonant units must be differentiated and this task cannot be done by counting the

number of letters involved. *ch,* for example, is composed of two letters, yet it functions as a simple unit as in *fuchia. x,* on the other hand, contains only one letter, yet it functions as a compound unit, as in *luxury* and *noxious.* What must be formulated is a consistent criterion for classing consonant units as simple or compound. While the classifications of *x, ch, th, ph,* and *rh* may be intuitively obvious, those of *ck, dg,* and *tch* are not. What is important is that, first, the rule mentioned above—and, as will be shown soon, almost all spelling-to-sound rules —be based not upon letters or graphemes as such, but rather upon functional spelling units; and, second, that functionally simple and functionally compound units be distinguished.

Graphemic alternations

A feature of the graphemic system which has arisen partially from scribal necessity and partially from pedantry is the alternation of various letters according to their graphemic environments. In such cases, two different letters which correspond to the same sound occur in complementary (or near-complementary) distribution. For example, the functionally simple vowel spellings *i* and *y* alternate, *y* occurring generally in final position and *i* in all other positions. In addition, regular rules control the alternation of final *y* to *i* before certain suffixes. This alternation holds not only for the simple vowel spellings *i* and *y,* but also for the compound spellings in which these two letters occur as the second elements, e.g., *ai/ay, ei/ey.* In the compound units, the *y* spellings generally appear before other vowel spellings and in morpheme final position, and the *i* spellings appear in all other positions. For example, *bait:bay, receive:grey, boisterous:boy.*

u and *w* when each corresponds to /w/ also occur in complementary distribution. As a simple consonant spelling, *u* occurs in a limited number of environments—after *q, g, s,* and a few others—in all other cases *w* occurs. As second units in compound vowel spellings, *u* and *w* alternate similarly to *i* and *y.* Thus, *w* occurs as the second element before other vowel spellings and in morpheme final position, while *u* occurs in all other positions, e.g., *auction:awe:draw; feud: ewer:flew; ounce:coward:vow.*

Several other alternation patterns should also be considered in an exhaustive analysis of the graphemic system. These patterns are listed in Table 2.

In addition to these alternations, a number of graphemic substitutions, introduced, for the most part, between the times of Chaucer

Table 2   Additional grapheme alternation patterns

1. *ous/os*	Word final *ous* becomes *os* before the suffix *ity*, e.g., *curious:curiosity*.
2. *i/e*	*i* in the suffix *ity* becomes *e* when the suffix is preceded by *i*, e.g., *society, variety, sobriety*.
3. *er/re*	With the addition of certain suffixes, word final *er* becomes *re* and then the *e* is dropped. Thus, *center:central, theater:theatrical*. In addition, *er* and *re* spellings alternate in word final position. *re* occurs after *c* and *g*, and *er* occurs in all other positions.
4. Consonant gemination	Gemination of a final consonant occurs before certain suffixes, e.g., *run:running, hop:hopped*.
5. *e/o*	Final *e* alternates with zero under certain types of suffixation. Thus, *dive:diving*.

and Shakespeare, must be treated separately. One of these is the substitution of *t* for *c* in suffixes like *tion* and *tial*, e.g., *nation, essential* (cf. ME *nacion, essenciall*). Early Modern English scribes effected in these substitutions one of the few true spelling reforms in English orthographic history. Their notion of reform, however, was to restore the appearance of the Latin root in favor of a more phonemic spelling. Another graphemic substitution was the Middle English replacement of *u* with *o* in the vicinity of *m, n, u* (*v*). This substitution accounts for many of the so-called irregularities like *some, love,* and *ton.*[1] All of these alternations and substitutions are parts of the current orthography and must be considered in a description of orthographic patterns.

Types of graphemic units

In this article, *grapheme* refers to one of the alphabetic characters *a* through *z* and the term *graphemic level* is used as a general reference for the spelling level. Spelling units are not related directly to sound, but to an intermediate (morphophonemic) level first, and then to sound. This indirect approach allows a clear separation of rules based upon orthographic considerations from those based upon morphological and phonological ones. Rules employed in mapping from graphemic to morphophonemic forms are those which are theoretically unique to the reading process. All other rules exist apart from the orthography and are, in general, a part of the language habits of all speakers of English, literate or illiterate. The model into which this

---

1. "In ME texts of a more recent date (Chaucer, etc.) we find *o* used still more extensively for /u/, namely in the neighborhood of any of the letters *m, n,* and *u* (*v, w*). The reason is that the strokes of these letters are identical, and that a multiplication of these strokes, especially at a time when no dot or stroke was written over i, rendered the reading extremely ambiguous and difficult . . . this accounts for the present spellings of won, wonder, worry, woman . . . above, love . . . and many others" (Jespersen, 1909–1949, vol. 3, p. 482).

mapping procedure is incorporated is briefly outlined later in this paper. Language-dependent units on the graphemic level which are significant for the prediction of sound are called *functional units* and are divided into two classes: *relational units* and *markers*.

A *relational unit* is a string of one or more graphemes which has a morphophonemic correspondent which cannot be predicted from the behavior of the unit's smaller graphemic components.

A *marker* is a string of one or more graphemes whose primary function is to indicate the correspondences of relational units or to preserve a graphotactical or morphological pattern. It has no sound correspondence.

Relational units   The division of graphemes into functional units depends partially upon the environments in which they occur. Thus, *gn* in *cognac* and *poignant* is a single relational unit which corresponds to the morphophonemic cluster //nj//. However, *gn* in *sign* and *malign* is not a relational unit, but rather a combination of two relational units which separately correspond to the morphophonemes //g// and //n//. Morphophonemic alternation rules map //g// into either //ø// or into //g//, depending upon allomorphic considerations (cf. *sign:signal, malign:malignant*). The final selection of relational units is based upon function and composition. Any string of graphemes that corresponds to a non-zero morphophoneme is a potential relational unit. However, only those strings whose morphophonemic correspondences cannot be predicted by general rules based upon smaller units contained in the string are classed as relational units. *ch* in *chair*, for example, is a relational unit since the morphophoneme //č// cannot be predicted from general rules based upon *c* and *h* separately. Geminate consonant clusters, however, are not single relational units since their morphophonemic forms can be obtained from rules based upon their separate constituents. (The leveling of clusters like //ff// to //f// can be accounted for by a general phonotactical rule, as long as morpheme boundaries are marked.)

Consonant relational units, furthermore, are classed as functionally simple or functionally compound [1]—a distinction needed for an accurate statement of a general correspondence rule. In the sequence *primary vowel + consonant + final e*, *vowel* is generally mapped into its free alternate if *consonant* is a functionally simple unit (or this type of unit plus *l* or *r*), and into its checked alternate if

---

1. The major functionally compound units are *tch, dg, ck, wh,* and *x*. The remaining consonant units (including *sh, th, ch,* etc.) are functionally simple.

consonant is a functionally compound unit, or a cluster.[1] Thus, vowel units are classed as primary (*a, e, i, y, o, u*) or secondary (all others).

Table 3   Primary vowel pronunciations before simple and compound consonant units

Free alternate	Checked alternate
b*a*ke	*a*xe
*a*che	b*a*dge
conc*e*de	*e*dge
l*i*chen	h*o*dge
cl*o*the	k*i*tchen
cr*u*de	l*u*xury

Relational units are classed as consonants or vowels depending upon the class of the morphophonemes into which they are mapped (glides are classed as consonants). Some *relational units* are classed as both consonant and vowel, e.g., *u* in *language* (consonant) and *during* (vowel). Within these classes major and minor patterns are distinguished on the basis of frequency of occurrence. Thus, *ch* is classed as a major consonant unit, but *kh* (khaki) is classed as a minor unit. While the major-minor classification may appear arbitrary, it distinguishes frequently occurring, productive patterns from infrequent patterns which generally occur only in a small number of borrowings.

Table 4   Major and minor relational units

Major relational units								
Consonants					Vowels			
Simple				Compound	Primary	Secondary		
b	gh	n	s	w[a]	ck[a]	a	ai/ay	ie	ue
c	h	p	sh	y[a]	dg	e	au/aw	oa	ui
ch	j	ph	t	z	tch	i	ea	oe	
d	k	q	th		wh	o	ee	oi/oy	
f	l	r	u[a]		x	u	ei/ey	oo	
g	m	rh	v			y	eu/ew	ou/ow	

Minor relational units		
Consonants		Vowels
Simple	Compound	Secondary
kh	gn	ae
sch		eau
		eo
		uy

The consonant and vowel relational units are enumerated below. Units followed by superscript ^{a} are discussed in the following paragraphs.

u is a consonant unit when it corresponds to //w// as in *quack, language,* and *assuage.* It may also be a vowel unit, or part of a vowel unit (*ou*), or a marker (*guest, plague*).

w is a consonant unit when it corresponds to //w//, e.g., *warm, beware.* It also appears as part of a vowel unit (*ow, aw*) but never as a vowel unit itself.

y is a consonant unit when it corresponds to //j//, e.g., *yes, beyond.* It also appears as a vowel unit and as part of a vowel unit, e.g., *cycle, boy.*

ck is a consonant unit in words like *rack* and *tack.* In instances like *picnicking,* however, the *k* is a marker. That *ck* in *picnicking* is identical to the relational unit *ck* is immaterial, since the base form *picnic* ends in *c,* not *ck.*[1]

Table 5 Examples of minor units

kh	khaki, khan	ae	aesthetic, algae
sch	schist, schwa,	eau	bureau, plateau
	seneschal	eo	jeopardy, leopard
gn	cognac, poignant	uy	buy, guy

Markers Examples of markers are the final *e* in *mate* and *peace,* the *u* in *guest,* and the *k* in *trafficking,* all of which indicate the pronunciation of a preceding grapheme. *ue* in *plague, catalogue,* etc., is a sequence of two markers. The *u* after *g* marks the correspondence *g*—//g// rather than *g*—//ǰ//. Since *u* does not appear in final position in English words (except after a vowel in a few recent borrowings), a final *e* is added, as in *continue* and *blue* (cf. the alternations ou/ow, au/aw). Only graphemes mapped into zero can be classed as markers (this is a necessary, but not sufficient condition). However, graphemes with non-zero morphophonemic correspondences, though properly classed as relational units, can also perform marking functions. For example, the *i* in *city,* besides corresponding to //ɪ//, marks the correspondence *c*—//s//. A geminate consonant cluster also performs a marking function since it regularly indicates the correspondence of the preceding vowel.

The strongest evidence for a separate class of markers in English orthography is found in orthographic alternation patterns. For

1. An alternate suggesiton is to assume that *c* is replaced by *ck* in such alternations.

example, final *e* as a marker for the pronunciation of a preceding *c* or *g* is dropped before a suffix which begins with a letter that will perform the same function as *e*. Therefore, *notice* drops the final *e* before *ing* (*noticing*) since *i* also marks the correspondence *c*—//s//, but retains the *e* before *able* since *noticeable* would have *c*—//k//. Similarly, the *e* added to an otherwise terminal *u* is dropped before any suffix since the only function of the *e* is to avoid having word-final *u*, e.g., *argue, arguing.*[1]

Preliminaries to orthographic analysis: spelling-to-sound correspondences

Types of correspondences

After pursuing the graphemic labyrinth through its intra-graphemic complexities, the next task is to analyze the relationships of these units to sound. The first object here is to show that even if the direct spelling-to-sound view is assumed, more types of relationships must be considered than the simple regular-irregular classes that bisect the traditional approach to this subject. Furthermore, it is shown that the concepts of regular and irregular are far more complex than is generally assumed, and, indeed, require quite sophisticated notions for adequate definition. For the present, however, *regular* and *irregular* will be used in a loose sense, meaning high frequency and low frequency without careful enumeration of what objects are to be counted to arrive at such statistics.

Regular spelling-to-sound correspondences can be classed first as either *invariant* or *variant. f,* for example, is invariant since it corresponds regularly to /f/. In fact, this correspondence is so regular that only one exception, *of,* occurs among the 20,000 most common English words. Several other consonant spelling units like *ck, m, v,* and *z* are also invariant or nearly so. The vowel spellings are rarely invariant, though not *irregular* in most cases.

Variant correspondences are those correspondences that are still regular, but that relate the same spelling to two or more pronunciations depending upon regular graphemic, phonological, or grammatical features. The letter *c,* as an example, corresponds to /s/ when it occurs before *e, i, y* plus a consonant or juncture; in most other positions, it corresponds to /k/. The spelling *k* corresponds to zero in initial position before *n,* e.g., *knee, know, knife;* in all other positions, *k* cor-

---

1. For a summary of markers in English orthography, see Venezky (in press).

responds to /k/. This is graphemic conditioning from the letter-sound standpoint. (The silent initial *k* is explained more adequately by phonotactical rules. The cluster /kn/ does not occur within a single morpheme in English; where such prohibited consonant clusters would otherwise occur in morpheme-initial position, the first consonant is dropped, as in *knee, gnat, ptarmigan, pneumonia, psychology.*)

Position alone may determine the correspondence of a spelling unit. For example, initial *gh* always corresponds to /g/: *ghost, gherkin, ghoul* (but never to /f/ as assumed in the spelling reform creation *ghoti*), but medial and final *gh* have pronunciations besides /g/, as is too often pointed out in spelling reform tracts. Stress may also be a conditioning factor for regular, variant correspondences. The most prominent role that stress plays in spelling-to-sound correspondences is in the pronunciation of unstressed vowels. While the reduction of unstressed vowels to schwa is not entirely regular, it can still be predicted in many cases. The patterns, however, are highly complex and are beyond the scope of this paper. A more interesting example of stress conditioning occurs in the correspondences for intervocalic *x*, which generally corresponds either to /ks/ or /gz/, depending upon the position of the main word stress. If the main stress is on the vowel preceding *x*, the pronunciation is /ks/ as in *axiom* and *exercise*. Otherwise, the pronunciation is /gz/ (cf. *examine, exist*). While this rule is similar to Verner's Law for the voicing of the Germanic voiceless spirants, it is not a case of pure phonological conditioning. Words like *accede* and *accept* have the identical phonetic environments for /gz/, yet have /ks/.

Another type of correspondence in which stress is important is the palatalization of /sy, zy, ty, dy/ to /š, ž, č, ǰ/. This form of palatalization generally occurs when /sy, zy, ty, dy/ are followed by an unstressed vowel, as in *social, treasure, bastion,* and *cordial.* The retention or deletion of medial /h/ in most cases also depends upon the position of the main word stress. Compare *prohibit:prohibition; vehicular:vehicle.* In each pair, the first member, which has the stress on the vowel following *h,* has a fully pronounced /h/, while the second member, with an unstressed vowel after *h,* has no /h/. This rule also holds for *vehement, shepherd, philharmonic, annihilate, rehabilitate,* and *nihilism,* all of which generally have no /h/. Some forms like these may have /h/ occasionally preserved by over-correct pronunciations.

Irregular spelling-to-sound correspondences also show important differences. *Arcing* and *cello,* for example, both have irregular cor-

respondences for *c*, yet there is an important distinction between these two irregularities. *Arc*, from which *arcing* is derived, has the correct correspondence for *c*. When suffixes beginning with *e, i, y* are added to words ending in *c*, a *k* is normally inserted after the *c*, as in *picnicking* (cf. *picnic*) and *trafficked* (cf. *traffic*). The irregularity in *arcing*, therefore, is in the irregular formation of the derivative. *Cello*, on the other hand, contains an aberrant correspondence for *c*, paralleled only by a few other Italian borrowings.[1] In the examination of the influence of morphemic features upon spelling-to-sound correspondences which follows, even more patterning appears, only now the patterning tends to show regularities where the direct letter-sound approach shows only irregularities.

Morphemic features

Morpheme boundaries   Morpheme boundaries must be known to predict certain spelling-to-sound correspondences. The spelling *ph*, for example, regularly corresponds to /f/ as in *phase, sphere*, and *morpheme*. In *shepherd*, however, *ph* clearly does not correspond to /f/, but to /p/. One way to explain this is to say that *shepherd* is an exception to the more general rule of *ph*-/f/. For consistency, then, the same analysis must be repeated with *uphill* and *topheavy*, which is somewhat specious. A more satisfactory procedure is to say that *ph* corresponds to /f/ when it lies within a single graphemic allomorph and that across morpheme boundaries *ph* is treated as the separate letters *p* and *h*. Therefore, one factor that should be considered in the spelling-to-sound relationship is morpheme boundaries. That this factor is not unique to *ph* can be seen from the following examples.

1]   Within graphemic allomorphs geminate consonant clusters (as in *letter, add*, and *canned*) are pronounced as single consonants. Across morpheme boundaries, however, both graphemic consonants may correspond to separate phonemes, as in *midday* and *finally*.

2]   All of the digraph and trigraph spellings are subject to the same morpheme boundary problem as *ph*, e.g., *hothead, changeable*.

3]   The spelling *n*, before spellings in the same morpheme which correspond to /g/ or /k/ corresponds to /ŋ/, as in *congress, finger, anchor*. Across morpheme boundaries this generally does not hold, e.g., *ingrain, ingenious, ingratiate*.

1.   *Concerto* and proper nouns like *Cellini* are the most common examples.

4] Many word final clusters contain silent letters, e.g., *gm, gn, mb* (e.g., *paradigm, sign,* and *bomb*). Before certain morpheme boundaries, the silent letter remains silent, as in *paradigms, signer,* and *bombing.* As long as the morpheme boundary is recognized, the correct pronunciation can be predicted. If the morpheme boundary is not recognized, then the three forms above would be thrown together with *stigma, ignite,* and *bamboo.*

In some cases the discrimination of a morphemic spelling from an identical, non-morphemic spelling is necessary for the prediction of sound from spelling. Consider the following two word lists.

Table 6 Morphemic and non-morphemic *s* spellings

A	B
boys	melodious
judges	stylus
cats	apropos
man's	careless

The pronunciation of final *s* in any word in column A can be predicted by the following rules (these rules must be applied in the order shown here):

1] /ɪz/ after /s, z, č, ǰ, š, ž/.
2] /z/ after any other voiced sound.
3] /s/ in all other cases.

These rules, however, apply only to *s* when it is one of the following morphemes:

1] regular noun plural
2] third person singular, present indicative marker for the verb
3] singular or plural possessive marker
4] any of the contractions like John's (from John is)

The past tense marker (*e*)*d* functions similarly, but is not entirely regular. In all of these cases, nevertheless, the direct spelling-to-sound approach fails unless it is based upon morpheme identity—and, if so, the approach is no longer a direct spelling-to-sound approach.

Another area in which the direct correspondence approach

fails to recognize inherent patterning is in the treatment of the final clusters *gm* and *gn*, mentioned above. Consider the forms *autumn: autumnal, damn:damnation, paradigm:paradigmatic, sign:signify*. It is not sufficient to state that *gn* and *gm* in final position correspond to /n/ and /m/, while in medial position to /-gn/ and /-gm-/. Such rules fail in cases like *autumns, designing,* and *signer*. There is no way to avoid reference to morphemes in this case, unless one simply enumerates the words for each pronunciation. A regular pattern is present in these forms, the most important aspect of which is the preservation of morpheme identity. The alternations of /g/ and zero in these examples, along with the alternations of the vowels preceding *g* are predictable. The direct spelling-to-sound approach once again breaks down when morpheme identity becomes important.

Form class   The school-book approach to orthography, as exemplified by *Webster's new collegiate dictionary* (Bethel, 1956), recognizes *ng* as a spelling for /ŋ/ when this phoneme is not followed by /g/.

> The digraph *ng*, as in *sing, singing*, represents the voiced tongue-back velar nasal continuant, corresponding to the voiced tongue-back stop *g*, and the voiceless tongue-back stop *k* (1956, p. xiv).

In contrast, the parallel cluster *mb* is analyzed phonotactically: "*b* is usually silent after *m* in the same syllable, as in *bomb, climb, thumb,* etc." (1956, p. xi). That *Webster's* treats identical phenomena in contrastive ways is only one of the problems here. Another problem is that an adequate description of the pronunciations of *ng* and *mb* must be based upon both morphemic and phonotactical relations. The pronunciation of any form ending in *nger* or *ngest* cannot be predicted unless the morphemic identities of *er* and *est* are known. If these are the comparative and superlative markers, then *ng* is pronounced /ŋg/ as in *stronger;* in most other cases, the /ŋg/ cluster is leveled to /ŋ/, just as it is in word final position. Morphemic identity is also important for predicting the pronunciation of word final *ate*. In adjectives and nouns, this ending is generally pronounced /-ɪt/, e.g., *duplicate, frigate, syndicate,* while in verbs, /-et/, *deflate, duplicate, integrate*.

   A final example of where form class identity is necessary for correct pronunciation is in initial *th*. Functors beginning with this cluster have the voiced inter-dental spirant /ð/: *the, then, this, those,* while contentives have the voiceless spirant /θ/: *thesis, thin, thumb*.[1]

---

1. Functors and contentives are defined by Hockett (1958, pp. 264 ff.).

Phonotactical influences

Consonant clusters   A knowledge of phoneme arrangements which are not allowed in English words is a necessary prerequisite for analyzing many spelling-to-sound correspondences. Sequences like /bp/ and /pb/ do not occur within English words—where they would occur, as in *subpoena* and *clapboard,* the speaker drops one sound or the other (with /pb/ and /bp/, the first sound is always omitted). While the spellings do not change, the pronunciations do. Yet, to label the pronunciation of *b* in *subpoena* as irregular, just as one does for the *b* in *debt,* is to ignore a pattern of English phonology. The elision of sounds in consonant clusters can be predicted, not only across morpheme boundaries, but also in initial and final positions, as in *knee, gnat, bomb,* and *sing.* In all of these cases, the correct pronunciation can be derived by first mapping all spelling units into some prephonemic level and then applying the rules for leveling non-English clusters to obtain the phonemic forms. Thus, *knee, gnat, bomb, sing,* become first //kni//, //gnæt//, //bamb//, //sɪŋg//, and then the non-allowed clusters are leveled, giving /ni/, /næt/, /bam/, /siŋ/.

Palatalization   To predict consonant cluster leveling is not the only reason for observing the arrangements of phonemes in English words. The palatalization of /sy, zy, ty, dy/ to /š, ž, č, ǰ/ and the deletion of /j/ from the cluster /ju/ also depend upon this knowledge. In addition, many spelling-to-sound patterns which can be described only clumsily in direct spelling-to-sound terms are more adequately described in phonological terms. A preceding /w/, for example, tends to change /æ/ into /a/ when this vowel is not followed by a velar consonant or /f/, e.g., *swamp, assuage, quadrant, swan, quality, quantum: wag, quack, twang, wax.* To describe this process in direct spelling-to-sound terms is difficult. The various spellings which correspond to /w/ and to /k/, /f/, /g/, and /ŋ/ must be enumerated and, even if this is done, the phonological nature of the /æ/ − /a/ shift is not revealed.

Descriptive model for relating spelling to sound

Any system of rules chosen to relate spelling-to-sound must be not only accurate and as simple as possible, but also it must allow a differentiation of the various patterns in the system. To present the *x* patterns, which depend upon a graphemic distinction and stress

placement as parallel to the /w/ pattern discussed previously is, as an example, an unsatisfactory account of the current orthography. An adequate description of spelling-to-sound correspondences is not something that could or should be implemented by a machine or directly applied to the teaching of reading, but, rather, is a complete analysis of all that the orthography holds and advances understanding of spelling-to-sound relationships.

To achieve this goal, a model has been constructed for describing spelling-to-sound relationships. In this model, graphemic words are divided into their graphemic allomorphs and, then, these allomorphs are related to intermediate (morphophonemic) units by an ordered set of rules. Other rules then relate the morphophonemic units to phonemic forms. All rules which are based upon non-graphemic features are applied in an ordered sequence on the morphophonemic level, yielding various sub-levels of intermediate forms for each word. The final morphophonemic form is then mapped automatically onto the phonemic level. While the intermediate level is not strictly a morphophonemic level, it is labeled as such hereafter. Its primary function is to separate graphemically dependent rules from grammatically and phonologically dependent ones.

As examples of how this model organizes spelling-to-sound rules, the processes for predicting the pronunciation of *social* and *signing* are shown below.

> *social* would be mapped into //sosɪæl// by the grapheme-to-morphophoneme rules for the separate units *s, o, c, i, a, l*. On the first morphophonemic level, the main word stress would be placed on the first syllable, resulting in //sósɪæl//. Then, through vowel reduction, //ɪæl// would become //jəl// and the resulting //sj// would be palatalized to //š//. The form //sóšəl// would then be mapped onto the phonemic level, giving /sóšəl/.

> *signing* would first be broken into *sign* and *ing* and then each of these graphemic allomorphs would be mapped onto the morphophonemic level, yielding //sɪgn// and //ɪng//. Upon combination of the two forms and the application of stress and certain phonotactical rules, the form //sígnɪŋg// would result. By the rules for leveling consonant clusters, final //ŋg// would become //ŋ// and //gn// would become //n//

with compensatory alternation of //ɪ// to //aɪ//. These operations yield //sámɪŋ// which is automatically mapped into /sáɪnɪŋ/.[1]

Vowel correspondences

Following is a brief sketch of spelling-to-sound correspondences for vowel spellings. A complete summary of consonant correspondences can be found in Venezky (1965).

Patterns of primary vowels

The vowel spellings *a, e, i/y, o, u,* called primary vowel spellings in this paper, carry the major burden of vowel representation in the current orthography.[2] They occur in all positions and have a vast complexity of morphophonemic correspondences and alternations which reflect an even more complex history. When viewed from the direct spelling-to-sound standpoint, the patterns for these units reveal no regularity. *o* corresponds to seventeen different sounds, *a* to ten, *e* to nine, and the combined group to forty-eight. When the morphemic structure and consonant environments of the words in which these units appear are analyzed, however, a single major pattern emerges, from which regular sub-patterns can be derived. Exceptions still remain, but the underlying pattern is so dominant that the exceptions, which were once the rule, become mere oddities, begging for historical justification. In the discussion which follows, the major pattern for the stressed vowels is introduced in a general form and then refined through the introduction of its regular sub-patterns, alternations, and exceptions.[3]

Major pattern   Each of the primary vowel units corresponds regularly to two different morphophonemes, a checked one and a free one,

1. A more detailed discussion of this model is given in Weir (1964) and Weir and Venezky (1965).
2. Henceforth *i* will stand for both *i* and *y*.
3. Word stress patterns play a significant role in the relationship of spelling to sound, especially in the correspondences of the primary vowel spellings. With a few exceptions, only stressed vowels are treated in this paper. Three

levels of stress, all introduced on the morphophonemic level, are assumed: primary, secondary, and tertiary (unstressed). While some linguists have claimed without qualification that English word stress is "predictable," no extensive analysis of this topic has ever been published. Two recent publications (Waldo, 1964; Wijk, 1966), however, indicate that work is being done in this area.

according to the morphemic structure of the word in which it occurs and the consonant and vowel units which follow it. These correspondences are shown in Table 7.

Table 7   Major patterns for primary vowels

Spelling	Free alternate	Checked alternate
a	//e//	//æ//
	s*a*ne	s*a*nity
	m*a*te	m*a*t
	r*a*tion	r*a*ttle
e	//i//	//ɛ//
	athl*e*te	athl*e*tic
	m*e*te	m*e*t
	p*e*nal	p*e*nnant
i	//ai//	//ɪ//
	r*i*se	r*i*sen
	mal*i*gn	mal*i*gnant
	s*i*te	s*i*t
o	//o//	//a//
	c*o*ne	c*o*nic
	r*o*be	r*o*b
	p*o*sy	p*o*ssible
u	//ju//[a]	//ə//
	ind*u*ce	ind*u*ction
	r*u*de	r*u*dder
	l*u*cre	l*u*xury

[a] The retention or elision of //j// before //u// is handled as a morphophonemic process.

In monomorphemic words, a primary spelling unit corresponds to its free alternate when it is followed by   1] a functionally simple consonant unit which, in turn, is followed by another vowel unit (including final *e*) or   2] a functionally simple consonant unit, followed by *l* or *r*, and then another vowel unit (including final *e*).[1] It corresponds to its checked alternate in the remaining cases, i.e., when followed by   1] a functionally compound consonant unit, e.g., *x, dg;* 2] a cluster of consonant units, e.g., *-nn, -lth;* or   3] a word-final con-

---

1. The difference between monomorphemic and polymorphemic words, a difference unfortunately neglected in the teaching of reading, is too complex to be discussed adequately here. For the prediction of sound from spelling in a large number of words, however, the distinction is crucial. What are involved, primarily are the morphophonemic alternations which occur with suffixation, as in /ərbén/ /ərbǽnɪtɪ/, /kon/:/kánɪk/. The best linguistic material on English suffixation is found in Newman (1948, pp. 24–36). For a pedagogical approach to this subject, see Thorndike (1941).

sonant unit or units.[1] Examples of these correspondences are shown in Table 8. The column numbers correspond to the numbered qualifications in the sentences above.

Table 8 Examples of primary vowel correspondences for selected environments

Spelling	Free alternate		Checked alternate		
	1	2	1	2	3
a	canine	ladle	badge	saddle	sat
e	median	zebra	exit	antenna	ebb
i	pilot	microbe	chicken	epistle	hitch
o	vogue	noble	pocket	cognate	sod
u	dubious	lucre	luxury	supper	rug

The correspondences for these vowel spellings in polymorphemic words depends not only upon the graphemic environment, but also in many cases upon the morphemic structure of the word. Correspondences for some polymorphemic words have been discussed previously (Weir & Venezky, 1965, pp. 42–44).

Sub-patterns The two most important sub-patterns which can be derived from the major pattern are the final *e* pattern and the geminate consonant pattern.[2] Examples of these are shown in Table 9; complete analyses are given in the two following sections.

Table 9 Examples of final *e* and geminate consonant patterns

Spelling	Final e pattern	Geminate consonant pattern
a	rate—rat	anal—annals
e	mete—met	Peter—petter
i	site—sit	diner—dinner
o	pope—pop	coma—comma
u	cute—cut	super—supper

1. The non-gemination of *v* and *th* has led to a large number of exceptions to the major pattern for the correspondences of the primary vowel spellings. For example, *cover, bevel, level, river, brother, mother, other* have vowel spellings corresponding to checked alternates in environments which indicate free alternates. To indicate the checked alternate, *v* and *th* would have to be geminated, but the graphotactical patterns of English exclude the doubling of these units. In *prison* a slightly different problem exists. While *ss*, which is needed to make the correspondence *i*—//ɪ//, is allowed, it generally corresponds to //s// in medial position, as in *blossom, gossip,* and *lasso,* so it could not be employed where *s* corresponds to //z//. The present use of *ss* is derived from old French orthography where "intervocalic *ss* served to distinguish voiceless s from voiced s (= z)" (Ewert, 1933, p. 113).

2. Sub-patterns for vowels before *r* and *l* and for vowels after *w* are discussed in Venezky (1965, pp. 164–69).

Final *e* pattern   While the final *e* pattern applies primarily to monosyllabic words, it also holds for many polysyllabic words, even when the vowel before the final *e* is unstressed, as in *microbe, decade, schedule, volume, placate*. Besides the patterns mentioned above (*vowel + consonant + e, vowel + consonant + le, vowel + consonant + re*), the environment *vowel + ste* is also part of the final *e* pattern, as in *baste, chaste*. Examples of regular correspondences are shown below, followed by an exhaustive list of exceptions for stressed vowels.

Table 10   Examples of the final *e* pattern

a—//e//	*i—//ai//*	*u—//ju//*	*e—//i//*	*o—//o//*
bake	cycle	cube	accede	cove
decade	domicile	duke	impede	erode
fable	five	mule	obsolete	globe
gage	prize	produce	scheme	joke
haste	profile	resume	serene	mediocre
shake	sublime	secure	theme	smoke

Irregular correspondences for the final *e* pattern are listed below.
  1]   *a* corresponds to //æ// in *bade, forbade, have, morale*. *a* corresponds to //a// in *are, barrage, camouflage, corsage, facade, garage, massage, mirage, sabotage*. Note: *a* in the ending *-ate* corresponds to //e// in verbs, but alternates to //ɪ// in nouns and adjectives. Cf. *duplicate* (vb): *duplicate* (adj., noun). Stress is also important in this alternation.
  2]   *e* corresponds to //ɛ// in *allege, ere, there, treble, where.*[1]
         e corresponds to //ɪ// in *renege*
         e corresponds to //ə// in *were,*
         e corresponds to //e// in *fete*.
  3]   *i* corresponds to //i// in the following words:
         bastile   machine   ravine   valise
         caprice   marine   regime
         castile   police   routine
         elite   prestige   sardine
                                 tangerine
         i corresponds to //ɪ// in *give* and *live*. Note. The spelling *i* is highly irregular in the ending *-ine* when unstressed. Cf. *canine, asinine: examine, famine*.

1. In the following, final *e* is not a marker, but a relational unit: (1) corresponding to //ɪ//: *adobe, coyote, epitome, extempore, facsimile, finale, nike,* *hyperbole, sesame, simile, ukulele, recipe*, (2) corresponding to //e//: *cafe, protege*.

4]   *o* corresponds to //ə// in *above, come, done, dove, glove, love, none, shove, some.*
o corresponds to //u// in *lose, move, prove, whose.*
o corresponds to //ɔ// in *gone.*

Vowel spellings before geminate consonants   The primary spellings *a, e, i, o, u* occur frequently before geminate consonants, the digraph (secondary) spellings, rarely. (*Braille, chauffeur,* and *trousseau* are the only examples from the corpus used for this study.) Before geminate consonant clusters, primary vowel spellings correspond to their checked alternates with the following exceptions:

1]   *a* in *mamma* corresponds to //a// and *a* in *marshmallow* corresponds to //ɛ//.
2]   *o* in *across, albatross,* and *boss* corresponds to //ɔ//.
3]   *o* in *boggy, gross,* and before final *ll* corresponds to //o//.
4]   *u* in *butte* corresponds to //ju// and *u* in *pudding* and *pussy* corresponds to //ʊ//.

Table 11   Examples of regular correspondences before geminate consonants

a—//æ//	*e*—//ɛ//	*i*—//ɪ//	*o*—//a//	*u*—//ə//
abbess	appellate	artillery	accommodate	button
accent	beggar	idyllic	collar	funnel
apple	bellow	blizzard	college	funny
babbittry	cellar	bacillus	comma	hullabaloo
cabbage	cheddar	cribbage	commerce	hummock
callow	dilemma	issue	dollar	mummy
fallacy	ebb	shrill	hobby	puddle
flabby	fellow	symmetry	hollow	pummel
grammar	kennel	vanilla	lobby	rubber
happen	lesson	village	sonnet	shutter
mammal	message	willow	toboggan	supper
rattle	tennis	wriggle	toggle	tunnel

Alternations based upon primary vowel spellings   The major pattern for the primary vowel spellings in stressed positions depends upon two basic features: environment, which was discussed in the previous sections; and morphemic structure, which is discussed briefly in this and the following sections. Morphemic structure forms the basis for describing the morphophonemic alternations based upon the primary vowel spellings. For example, the word *sanity,* if considered solely on the basis of the rules given in the preceding section, would be an exception to the major pattern since *a* before a simple consonant unit followed by a vowel corresponds to its checked, rather than its free,

alternate. If viewed, however, in relation to the sequences, *sane:sanity, humane:humanity,* and *urbane:urbanity,* another regular feature can be seen. If one starts with the forms *sane, humane,* and *urbane,* then regular rules can be written for changing the free alternate //e// to the checked alternate //æ// when the suffix *-ity:* //-ıtı// is added.

   This rule also holds for the spellings *e, i,* and *o,* as can be seen from the following examples:

Table 12   Alternations based upon
e, i, and *o*

e	extreme	extremity
	obscene	obscenity
	serene	serenity
	supreme	supremity
i	asinine	asininity
	devine	devinity
	malign	malignity
	senile	senility
o	frivolous	frivolity
	mediocre	mediocrity
	precocious	precocity
	verbose	verbosity

   (Morphophonemic alternations based upon stressed *u* are rare in Modern English, the most common being those which occur in *assume:assumption, conduce:conduction, presume:presumption, reduce:reduction.* Even with loss of stress *u:* //ju// tends not to change to //ə//. Thus, *compute:computation, execute:execution, usurer:usurious, utilize:utility.*)

   While a complete survey of vowel morphophonemics is beyond the scope of this paper, some of the more common alternations are presented below.

1]   *-ic:*   free   —   checked

  a   angel   angelic
        state   static

  e   athlete   athletic
        hygiene   hygienic
        meter   metric

  i   cycle   cyclic
        mime   mimic
        paralyze   paralytic

	o	c*o*ne	c*o*nic
		neur*o*sis	neur*o*tic
		ph*o*ne	ph*o*nic
2]	*-ion*	free —	checked
	e	conc*e*de	conc*e*ssion
		conv*e*ne	conv*e*ntion
		descr*e*te	descr*e*tion
	i	coll*i*de	coll*i*sion
		dec*i*de	dec*i*sion
		prov*i*de	prov*i*sion
	u	red*u*ce	red*u*ction

Patterns of secondary vowels

The secondary vowel spellings differ from the primary vowel spellings in several important ways. First, they occur less frequently and have a more limited distribution. None appears commonly before geminant consonant clusters; some like *ai, au, ei,* and *eu* rarely occur in word-final position; others, like *ie* and *oa* rarely occur in word initial position. Second, while each primary vowel spelling has two basic correspondences, according to the graphemic environment and the morphemic composition of the word in which it occurs, each secondary vowel spelling generally has a single major correspondent. Third, the morphophonemic correspondences based upon the secondary vowel spellings tend not to alternate in quality with reduction in stress. Compare, for example, the first vowels in *neutral:neutrality, cause:causation* with those in *melody:melodious, potent:impotent.*

Historically, the secondary spellings also differ from the primary ones. Primary vowel spellings are found in the earliest English records; their correspondences can be traced through a complicated chain of sound changes from Old English to the present time. Most secondary vowel spellings, on the other hand, were introduced during the late Middle English period and, consequently, have been involved in considerably fewer sound changes. A synchronic and diachronic summary of the secondary vowel spellings is given by Venezky (1965, pp. 171–83).

Conclusions and suggestions for further study

Learning to read one's native language differs radically from learning to read a foreign language. For learning to read a foreign lan-

guage, the individual most frequently does not have any knowledge of the language he is to read and, in the majority of the cases, he desires to translate directly from writing to meaning. In learning to read one's native language, however, the individual brings a reasonably adequate set of language habits, as evidenced by his ability to speak. Learning to read in this situation requires primarily the translation from written symbols to sound, a procedure which is the basis of the reading process and is probably the only language skill unique to reading. (Comprehension, for example, while a necessary criterion for reading, is a function of both speech and writing.) The primary concern of this paper is the teaching of this translation process; references hereafter to the teaching of reading, unless qualified otherwise, refer to spelling-to-sound translation only.

A good reader in the sense of this paper is one who can not only pronounce all of the words which he has been taught to read, but can also pronounce a high percentage of new words which he encounters. Certainly a person who could do the former task—that of pronouncing words he had seen before—but has displayed no ability to pronounce new words, would be classed as a deficient reader. Most literates form some spelling-to-sound generalizations regardless of the methods which they encountered in their initial reading instruction. What these generalizations are, how they develop, and how they differ from one literate to another is at present unknown.

The patterns summarized here represent an ideal system for translating from spelling to sound, formed by assigning equal weights to each of the words used in the original corpus. It is inconceivable that any human could without special effort arrive at the same rules. Type-token relationships probably are highly influential in the formation of such patterns. The pronunciations of initial *ch,* for example, are irregular according to the present results since no consistent cues can be used for selecting among /š/, /k/, and /č/. The average literate, nevertheless, would most probably pronounce initial *ch* in a strange word as /č/ simply because most of the frequently occurring words with initial *ch* are pronounced with /č/.

Which of the major patterns is learned by literates is not known—in fact, extremely little is known about the extent of generalization in this area. Furthermore, there has been so little discussion of this process that no clearly defined criteria for generalization have been formulated. In the *ch* pattern cited above, for example, what would good generalization be? To pronounce all new initial *ch* words with /č/? To pronounce x per cent with /č/, y per cent with /š/, and z

per cent with /k/, where x, y, and z correlate to the frequency of known initial *ch* words with the different pronunciations? Or what? If the teaching of reading is to be placed on a more substantial foundation than it now occupies, the nature and goals of reading generalization must be explored.

Once the reading habits of literates are known and attainable goals for reading established, some of the sacred cows of reading instruction could be re-examined. The first of these to undergo inspection should be the notion of sequencing materials from the simple to the complex, the heart of the Bloomfield approach and of almost all of the new "linguistic" approaches to teaching reading. While such an approach may, with sufficient experimentation, prove to be more efficient than any other possible approach, there is little linguistic or psychological support for it at present.

For example, understanding one of the most important spelling patterns, that of the correspondences for the primary vowel spellings, requires differentiation of both graphemic environments and responses. The letter *a*, as an example, has two primary pronunciations in stressed position, /æ/ and /e/. The checked alternative, /æ/, occurs when *a* is followed by a final consonant or by a series of consonants, as in *rat* and *annals*. In addition, it occurs when *a* is followed by a single consonant plus one of several possible suffixes, like *-ity* (e.g., *sanity*). The free pronunciation, /e/, occurs when *a* is in the other graphemic environments, like *rate*, *anal*, and *sane*. What must be acquired for the proper pronunciation of *a* is the ability to differentiate the environments and suffixes; final consonant vs. consonant plus final *e* (*rat: rate*), double medial consonant vs. single medial consonant (*annals: anal*), and the base form vs. particular suffixed forms (*sane:sanity*).

The Bloomfieldian sequencing begins with the /æ/ pronunciation for *a*, introducing the /e/ pronunciation at a later time with no special emphasis on the relation between /æ/ and /e/ when derived from *a*. An alternative to this approach is to present both pronunciations at once, working with such pairs as *rat:rate, mat:mate, fat:fate, hat:hate*, and *man:mane*. Both the associations of *a* to /æ/ and *a* to /e/ and the discrimination of the graphemic environments would be emphasized. Whether or not a child first learning to read can handle this task probably depends upon the pedagogy employed. The potential generalization derived from the differentiation approach, however, certainly is greater than that from the simple-sequence method.

Another factor which might relate to the teaching of reading is the differentiation between spelling-sound patterns based primarily

upon the orthography and those based primarily upon phonological habits. Both initial *c* and medial *n* can have two different pronunciations, as exemplified by *cent, cot; anchor, vanity,* but there is an important difference between the rules which describe the pronunciations for each. The choice between /s/ and /k/ for *c* is primarily dependent upon the orthography; /s/ when *c* is followed by the front vowel spellings *e, i,* or *y,* and /k/ otherwise. The choice between /n/ and /ŋ/ for *n,* however, is primarily phonological, in that /ŋ/ occurs only when a velar stop follows.[1] /n/ generally does not occur before /g/ or /k/; where it would, it is backed to /ŋ/.[2] Native speakers of English perform this change without conscious effort; they do not need to be taught to do so when learning to read, although they must be taught the two different pronunciations for *c* and when to employ each. It is evident, therefore, that more new information must be taught for reading initial *c* than for reading medial *n.* The types of reading mistakes one would expect are quite different, also. The substitution of /k/ for /s/ (and vice versa) is fairly common in initial reading for *c,* but the substitution of /n/ for /ŋ/ before /g/ or /k/ is extremely rare, simply because the phonology excludes the /ng/ and /nk/ sequence.

Whether or not this difference can be incorporated into the teaching of reading has to be determined by experimentation. The most important factor to realize is that the *n* pronunciations, like many other patterns in English orthography, depend primarily upon habits which the beginning reader already has, that is, the phonological habits of English. Learning to read is to a great extent learning to relate orthographic forms to already existing phonological forms. The more that reading pedagogy can take advantage of this fact, the more successful the teaching of reading will be.

1. Stress and juncture also affect this pattern, but are ignored for the present.

2. In contrast, both /s/ and /k/ occur before front vowels, e.g., *kit:city, cat:sat.*

REFERENCES

BETHEL, J. R. (Ed.) *Webster's new collegiate dictionary.* (2nd ed.) Springfield, Mass.: Merriam, 1956.

BLOOMFIELD, L. *Language.* New York: H. Holt, 1933.

BOLINGER, D. Visual morphemes. *Language,* 1946, 22, 333–40.

BRADLEY, H. On the relations between spoken and written language with special references to English. *British Academy,* 1914, 6, 211–32.

BROWN, G. *The grammar of English grammars.* (4th ed.) New York: Samuel S. & W. Wood, 1859.

CRAIGIE, W. *English spelling, its rules and reasons.* New York: F. S. Crofts, 1927.

DOUGLAS, J. *James Douglas on English pronunciation, c. 1740.* B. Holmberg (Ed.), Lund Studies in English, 26. Lund: Gleerup, 1956.

EWERT, A. *The French language.* London: Faber & Faber, 1933.

FRANCIS, W. N. *The structure of American English.* New York: Ronald Press, 1958.

GIBSON, ELEANOR J., *et al.* An analysis of critical features of letters, tested by a confusion matrix. In H. Levin (Ed.), *A basic research program on reading.* Ithaca: Cornell University, 1963.

HALL, R. A., JR. *Sound and spelling in English.* Philadelphia: Chilton Books, 1961.

HALL, R. A., JR. Graphemics and linguistics. *Proceedings of the American Ethnological Society,* 1962, 62, 39–53.

HAUGEN, E. First grammatical treatise: the earliest Germanic phonology. *Language monograph no. 25,* Linguistic Society of America, 1950.

HOCKETT, C. *A course in modern linguistics.* New York: Macmillan, 1958.

HUME, A. Of the orthographie and congruitie of the Briton tongue. In H. B. Wheatley (Ed.), *EETS publication no. 5.* London: Early English Text Society, 1865.

JESPERSEN, O. H. *A modern English grammar on historical principles.* Heidelberg: C. Winter, 1909–49. 7 vols.

KENYON, J. S., & KNOTT, T. A. *A pronouncing dictionary of American English.* Springfield, Mass.: Merriam, 1951.

KURATH, H. *A phonology and prosody of Modern English.* Ann Arbor: University of Michigan, 1964.

LEVIN, H. (Ed.) *A basic research program on reading.* (Final report, Cooperative research project No. 639) Ithaca: Cornell University, 1963.

LEVIN, H. (Ed.) *Project literacy reports.* Ithaca: Cornell University, 1964, 1965.

MC INTOSH, A. The analysis of written Middle English. *Transactions of the Philological Society,* 1956, 26–55.

MC LAUGHLIN, J. *A graphemic-phonemic study of a Middle English manuscript.* The Hague: Mouton & Co., 1963.

NEWMAN, S. English suffixation: a descriptive approach. *Word,* 1948, 4, 24–36.

PULGRAM, E. Phoneme and grapheme: a parallel. *Word,* 1951, 7, 15–20.

SAUSSURE, F. DE. *Cours de linguistique generale.* Paris: Payot, 1916. Trans. by W. Baskin, *Course in general linguistics.* New York: Philosophical Library, 1959.

SOFFIETTI, J. Why children fail to read: a linguistic analysis. *Harvard Educational Review,* 1955, 25, 62–84.

STETSON, R. The phoneme and the grapheme. *Melanges de linguistique et de philologie offerts à Jacq. van Ginneken,* Paris, 1937. Pp. 353–56.

THORNDIKE, E. *The teaching of English suffixes.* New York: Teachers College, Columbia University, 1941.

THORNDIKE, E. *Thorndike-century senior dictionary.* New York: D. Appleton-Century, 1941.

ULDALL, H. Speech and writing. *Acta Linguistica,* 1944, 4, 11–16.

VACHEK, J. Two chapters on written English. In *Brno studies in English,* Vol. 1. Praha: Státní Pedagogické Nakladatelstvi, 1959. Pp. 7–34.

VACHEK, J. On the interplay of external and internal factors in the development of language. *Lingua,* 1962, 11, 433–48.

VENEZKY, R. A computer program for deriving spelling to sound correlations. In H. Levin (Ed.), *A basic research program on reading.* Ithaca: Cornell University, 1963.

VENEZKY, R. A study of English spelling-to-sound correspondences on historical principles. Unpublished doctoral dissertation, Stanford University, 1965.

VENEZKY, R., & WEIR, RUTH H. *A study of selected spelling-to-sound correspondence patterns.* (Final report, Cooperative Research Project No. 3090) Stanford: Stanford University, 1966.

VENEZKY, R. The basis of English orthography. *Acta Linguistica.* In press.

WALDO, G. The significance of accentuation in English words. *Proceedings of the Ninth International Congress of Linguists,* 1965, 9, 204–10.

WEBSTER, N. *An American dictionary of the English language.* (Rev. ed.) Springfield, Mass.: Merriam, 1852.

WEIR, RUTH H. *Formulation of grapheme-phoneme correspondence rules to aid in the teaching of reading.* (Final report, Cooperative Research Project No. S-039) Stanford: Stanford University, 1964.

WEIR, RUTH H. & VENEZKY, R. L. *Rules to aid in the teaching of reading.* (Final report, Cooperative Research Project No. 2584) Stanford: Stanford University, 1965.

WIJK, A. *Rules of pronunciation for the English language.* London: Oxford University Press, 1966.

ZACHRISSON, R. *Anglic, a new agreed simplified English spelling.* Uppsala: Anglic Fund, 1930.

ZACHRISSON, R. Four hundred years of English spelling reform. *Studia Neophilologica,* 1931, 4, 1–69.