

Internalization of English Orthographic Patterns

Robert H. Secrist

This paper examines (1) the patterns of regularity underlying the largely superficial chaos of English orthography, (2) the extent—or the lack—of internalization of both these patterns and the graphic representations of specific phonemes in different environments, (3) the reactions of literate native speakers as to the relative “rightness” or naturalness of the different phoneme-grapheme correspondences in these situations, and (4) the reactions of these informants to recognition tests involving alternative solutions to certain special problems involved in English spelling reform—such as lexicographic separation of agnates, polymorphic representation of inflectional affixes, and graphic distinction of homophones.

In analyzing the history of proposals for the reformation of English orthography, from the twelfth century to the present, one becomes increasingly aware of an emerging consensus that radical respelling with new graphemes in rigidly phonemic transcription is at best impracticable, and that a more moderate proposal, based on comprehensive orthographic revision within the context of the traditional alphabet and spelling patterns, is the only realistic approach to this problem.

It becomes quite apparent, however, that, while some would-be spelling reformers who share this view have made some effort to determine the extent of the patterns and regularities already existing in English orthography (Hall 1961, Wijk 1959, Zachrisson 1932, et al.), none has even attempted to determine to what extent those who already write the language are aware of these regularities—nor, when such regularity is lacking, which of the various alternative spellings is most acceptable to the already literate. Neither has there been, to my knowledge, any research that might lead to solutions to the various non-phonetic problems that have troubled these otherwise rational and realistic spelling reformers:

309 Secrist : *Internalization of Orthographic Patterns*

Visible Language, X 4 (Autumn 1976), pp. 309-322.

© 1976 *Visible Language*, Box 1972 CMA, Cleveland, Ohio USA 44106.

Author's address: Dept. of English, Youngstown State University, Youngstown, OH 44555.

namely, the extent to which spelling should be strictly phonemic (rather than partially morphological and/or semantic), whether homophones need be distinguished graphically, whether agnate relationships (between forms with the same base) must be considered or may be ignored, and whether the spelling of inflectional affixes should be invariable or morphophonemically conditioned.

In view of the multiplicity of patterns in the choice and deployment of graphemes, and in the approaches to, and the treatment of agnates, homophones, and affixation among any group of proposals for spelling reform, it seemed obvious to me that research was needed to determine—in an objective and psycholinguistically sound manner—which are most valid, or at least defensible, considering both linguistic principles and psychological factors.

In this paper I shall report and analyze the findings of such research, conducted over a three-year period, into the nature of traditional English orthography (referred to henceforth as TO), the degree to which both its regularities and its vagaries have been internalized by literate native speakers, and the most vexing non-phonetic problems involved in its comprehensive but realistic reform.

Procedures

The basic procedure followed in the research on internalization of orthographic patterns was to dictate imaginary English words to some 270 literate native English speakers, high-school and community-college students enrolled in French or in English-composition courses. (Since I wanted a random sampling, with all levels of intelligence and proficiency in English—given basic literacy—represented, entire classes were involved.) The subjects were asked to write these words as they would most likely be spelled if they were genuine English vocabulary—that is, in what form they would first be sought out in a dictionary.¹

In most cases, ten words were invented to test each phoneme in each environment in question. In some cases, however, when dealing routinely with sounds which are overwhelmingly regular in conventional spelling and the responses to which could reasonably be expected to conform, only five words were used. Since separate, individual reactions and responses were desired, not

simply unthinking repetitions of the first—and since the words had been painstakingly devised to include virtually every possible sort of environment with the position-category being tested (e.g., after voiced consonants, before atonic syllables, and intervocalic)—I ordinarily tested several dissimilar sounds at once, in alternation, in order to minimize the effects of mutual contamination. In several cases, where the results were especially remarkable, unexpected, or inconclusive, the tests were repeated with one or more different groups of subjects—and the results compared and averaged.

In this way data were gathered as to the most “natural,” automatic spellings of all English phonemes in every kind of environment. These data were then compared and correlated with others of a similar nature from various sources—such as literary “eye” dialect, advertising, popular journalism, newly-coined words, and analogical misspellings of unfamiliar words by children and other semi-literates.

Since literacy in an orthographic system involves reading as well as writing, further research was undertaken, on the basis of the data compiled and analyzed by the above methods, to test whether currently irregularly spelled words could easily be recognized when respelled according to the most regular existing pattern and/or the most frequently indicated “felt” response—both in isolation and, if necessary, in sentence contexts. In this way I could determine whether those orthographic patterns already internalized and felt appropriate for unfamiliar words would also be acceptable for real words presently outside the system in an unfamiliar form—i.e., whether the “felt” responses were merely spelling preferences based on familiarity with unpatterned phenomena, or true indications of the “rightness” or naturalness of the phoneme-grapheme correspondence in question, within the larger context of systematic patterns.

For the sounds noted in Tables II and III the testing provided a more or less routine confirmation of the results of the regularity tests. For those in Tables I and IV, however, this research was more necessary—since in neither group did regular responses obtain, though in Table I there is at least some regularity in the conventional spelling (TO) to provide a base or starting point

Table I. Traditional Orthographic Regularities Not Well Internalized

<i>Phoneme-Position</i>	<i>TO</i>	<i>% Responses</i>	<i>Well-recognized alternative(s) *</i>
1. /a/ before final /m/	al	44	o
2. /o/ between /w/ & /r/	a	37	o
3. /æu/ in closed syllable before /n/	aw	62	au
4. /ou/ in closed syllable before /n/	ow	54	ou
5. /æ/ before /n/ or /l/	ai	57	a . . . e
6. /ʌ/ before velar stop	oo	65	uu
7. /k/ after initial atonic vowel	cc	6	c, k
8. /k/ before non-front vowel, /æ/, or diphthong (except /ie/)	c	23	k
9. /k/ before consonant	c	52	k
10. /f/ final after short vowel	ff	58	f
11. /s/ final after short vowel	ss	67	(none)
12. /s/ final after /or/ or /ur/	se	39	ce
13. /s/ final after /en/	se	48	ce, s
14. /j/ final after short vowel	dge	61	(none)
15. /j/ final after long vowel	g(e)	65	(none)
16. /j/ post-tonic before /u/ or /ʌ/	d	36	
17. /ʒ/ between long vowel or /i/ & atonic syll.	s	35	zi, zh
18. /ʃh/ medial after /n/ (exc. /en/) or /l/	si	20	
19. /ʃh/ intervocalic before /ʌ/ or /ʊ/	ssi	57	
20. /ʃh/ between long vowel & /un/	ti	43	
21. /ʃh/ between long vowel & /ul/ or /us/	ci	53	
22. /ʧ/ before atonic /un/	ti	65	
23. /ʧ/ before atonic /ʌ/ or /ur/	t	63	

*For 16-23, see discussion of Results of the Tests of Recognition.

Table II. Strong Preferences Despite Traditional Irregularities

<i>Phoneme-Position</i>	<i>Preference</i>	<i>Per cent</i>
1. /u/ in tonic syllable (except before /r/)	u	97
2. /u/ in atonic syllable	e	85
3. /ʌ/ in closed syllable	u	71
4. /ʌ/ in closed syllable	oo	64
5. /ʌ/ before velar stop	oo	74
6. /z/ final after short vowel	z	85
7. /s/ initial before /i/, /e/, or /ur/	s	80
8. /s/ between tonic short vowel & /un/ or /ul/	ss	78
9. /z/ intervocalic	z	78

Table III. Strong Preferences Contrary to Traditional Regularities

<i>Phoneme-Position</i>	<i>TO</i>	<i>Preference</i>	<i>Per cent</i>
1. /ie/ in closed syllable before /t/	igh	i . . . e	88
2. /v/ final after tonic short vowel	ve	v	93
3. /j/ before /i/, /e/, /æ/, /ie/ or /ur/	g	j	76 *
4. /æu/ in closed syllable before /s/	o	au	79

*This is the average of the five, which were tested separately, with results ranging from 67% to 82%

Table IV. Unpredictable Spelling Variations Tested for Recognizability

<i>Phoneme-Position</i>	<i>Well-recognized alternative(s) *</i>
1. /æ/ before final /r/	a . . . e, eh
2. /æ/ in closed syllable	ee
3. /æu/ in closed syllable before /l/, /t/ or /k/	au
4. /oe/ final	o, ow
5. /w/ final (with invariable pronunciation)	oo
6. /ue/ final (with invariable pronunciation)	iew
7. /w/ } final (with variable pronunciation)	u
/ue/ }	yu
8. /w/ in closed syllable	oo
9. /w/ in non-final open syllable	u, oo
10. /ue/ in non-final open syllable	u . . . e
11. /s/ final after long vowel or diphthong	c(e), s(e)
12. **/k/ between short vowel or consonant & final /s/	
13. **/ Sh/ intervocalic after short vowel	ti, ssi
14. **/ Sh/ medial after /en/	si, ti
15. /u/ tonic before non-intervocalic /r/	u, e, i
16. /u/ tonic before intervocalic /r/	u

*In order of preference.

**See discussion of Results of the Tests of the Reality of Certain "Problems."

point. For the sound-environments in Table IV there is no such readily discernible base or starting point; hence they required more extensive investigation into the preferences of literate informants for different phoneme-grapheme correspondences.

Ordinary English words with the sounds and environments to be tested—plus others about whose recognizability there was little doubt—were placed in lists of fifty, spelled according to the conventional pattern or the most frequently indicated “felt” response, whichever was more regular or reasonable. Every sound sequence in question was represented by at least three examples, with at least five examples for each particularly unpredictable situation. The words chosen were fairly common ones which were otherwise regular in their spelling, thus focusing attention on the single orthographic change being tested—without interference from other, extraneous factors. Obviously, it was necessary to avoid here words whose respelling would have resulted in homographic association with words already spelled that way, though with different pronunciations—e.g., /word/ (*ward*) was not respelled *word*. Testing was done on an oral and individual basis—orally to enable me to note any hesitation or uncertainty on the part of the informants, individually to prevent them from being influenced by others’ responses.

After collation and analysis of the data from these responses in isolation, those cases in which the responses had been significantly inconsistent, hesitant, or dubious were retested in sentence contexts. Here too, contexts were selected to offer minimal interference with the testing of the word in question, i.e., the other words in the sentence were regularly—or at least rationally—spelled, or required no more than minor alterations. Furthermore, the entire sentence was respelled regularly, so that undue attention would not be focused by the informant on the particular word being tested—the informant, of course, being completely unaware of the identity of this word.

Results of the Tests of Internalization

In these ways was the extent of the internalization of English orthographic patterns determined. Table I shows those which were found to be regular in TO, but whose regularity has not been well internalized. By "regularity" is meant, of course, not a one-to-one correspondence between sound and symbol, phoneme and grapheme. "Regularity" implies—and requires—simply that the spelling of a particular phoneme in a particular environment be predictable with a high degree of accuracy. There are, of course, "exceptions" to almost any orthographic "rule" or principle in English. Yet, except in the cases noted in Tables I-IV, a remarkable degree of consistency and regularity is both apparent and recognized.

In this study regularities in TO were not considered "well internalized" unless they appeared in at least 75% of the responses—a watershed percentage determined by an examination of the results, which showed a remarkable tendency to cluster above 75% and below 50%; there were very few in this middle range—"fairly well internalized"—and are thus noted. Analysis also revealed that most non-traditional spellings appearing alongside the "well internalized" patterns were either very rare (e.g., *oe* for /*œ*/), or clearly errors in comprehension (e.g., *v* for /*f*/) or due to carelessness (e.g., *g* for *q*).

With this in mind, it can readily be seen (from Table I) that only six of the normal, regular spellings of vowels in all positions have failed to be well internalized (#1-6)—and that only the first two were actually rejected by the informants, presumably by analogy with the more general patterns involving these sounds in less restricted environments (cf. *calm* with *comma* and *combat*, *war* with *for* and *horse*). The other four (#3-6) have actually been fairly well internalized, but not to a degree consonant with their conventional frequency. Of these, *ai* for /*æ*/ is very common before /*n*/ and /*l*/ (#5) but anomalous in other closed syllables—as is *ow* for /*ou*/ before final /*n*/ or /*l*/ (#4), which spelling is otherwise restricted to open syllables, while the closed-syllable form of this diphthong is *ou*—even before /*n*/ when another consonant follows (cf. *town* with *count* and *bounce*). A very similar pattern obtains with *aw* for /*æu*/ (#3) (cf. *lawn* with *launch* and *flaunt*). Thus it could be

said that, even in these cases, the larger patterns have been internalized—if not the deviations from them. Even in the anomalous spellings there is a pattern of sorts, since nasals and liquids have clearly influenced the spelling of historical diphthongs. In all of these cases, however, a more common or less restricted pattern appears to be interfering with the full internalization of the aberrant pattern; in other words, analogical leveling seems to be operating on the level of orthographic internalization.

This also seems to be the case with many of the consonantal spelling patterns which have failed to be well internalized—especially the intervocalic *cc* for /k/ after atonic vowels (#7), and the gemination of final *f* and *s* after short vowels (#10 & 11)—both of which violate the more general pattern of consonant gemination, which normally occurs only intervocalically (or in weak preterits) after tonic short vowels. Likewise the relatively frequent use of *ce* for /s/ finally after /en/, /or/, and /ur/ (#12 & 13) instead of the traditional *se* apparently reflects the internalization of the normal spelling of this sound after /r/ or /n/ preceded by any other vowel. This normal pattern has, in fact, even affected the orthography in modern times, several words in *ens* having given way to *ence*, especially in British usage (*defence*, *licence*, etc.) American orthographic practice, under the influence of Noah Webster's linguistic nationalism, apparently reflects an attempt to resist, or even reverse, the trend.

There are, however, other factors that may be playing a part in this lack of full internalization—such as the inability to distinguish, in unfamiliar words, between historical affricates or palatal fricatives and the results of recent assibilation and/or palatalization of alveolar consonants in reciprocal assimilation with historical /y/ (#16-23), especially in medial position before common suffixes like *-ure* and *-ion*, particularly the former (cf. *teacher* with *feature*, *bashin'* with *passion*). A different element is apparently operative with /k/ before consonants, non-front vowels, and diphthongs (#7 & 8), where the internalization of the *c* of TO has probably been impeded—or at least eroded—by the prevalence of so many commercial names with *k* in these positions (e.g., *Kool-Aid*, *Go-Kart*, *Kleenex*). Finally there is the problem of those consonants that occur homophonically both in monomorphemic final

phonemic clusters and in bimorphemic inflected forms. These are not indicated in Table I, since they could not meaningfully be tested for internalization—only for recognition (see discussion of Results of Reality of Certain “Problems”).

In contrast to the regular but uninternalized spellings noted in Table I were those which are quite unpredictable in TO, but for which the informants showed remarkably strong preferences (Table II).

There are even a few cases (noted in Table III) for which TO provides fairly regular patterns, but for which the respondents displayed not only a lack of internalization of those patterns but a distinct—even overwhelming—preference for another pattern. In each case of both types, a more common spelling of the sound in other positions seems to have been generalized at the expense of the minor, deviant—albeit traditional—pattern. Here too, analogical leveling seems to operate on the semiconscious level of orthographic internalization.

There are, of course, other phonemes whose unpredictability in certain positions in TO was reflected in the lack of clear-cut patterns among the responses (Table IV).

Results of the Tests of Recognition

Tables I-IV also record the results of further research that I conducted to determine whether the spellings preferred by the informants—or suggested by the application of the principles of analogy, structural symmetry or graphic economy—could be easily recognizable in words currently spelled in other ways, that is, if reading comprehension would be impeded by the indicated respelling.

It should be noted in Table I that some successful experimentation was done with non-traditional digraphs: with *uu* for /*u*/ (#6), when neither *oo* nor *u* was recognized consistently before velar stops; and with *zi* and *zh* for /*ʒ*/ (#15), because neither *si* nor *s* of TO has been well internalized, and both have much more common pronunciations—/*ʒ*/ never having had its own graphic representation. Although *zh* is commonly used in dictionaries for this sound, *zi* proved to more easily recognizable—especially in isolation. While the former is structurally symmetrical with the

TO spelling of /ʃh/ in positions not involving assimilation, this is irrelevant for /ʒ/ since it does not occur in those positions.

In Table IV the preference of *are* for final /ær/ (#1) is probably due to the normal lengthening of vowels—and the consequent neutralization of their usual quantity contrast in this environment, with the subsequent application of the VCe pattern normal for long vowels in closed syllables. The use of *ee* for /e:/ in closed syllables (#2) does, of course, violate the structural symmetry of this more general pattern, but the symmetrical form occasioned considerable hesitation and confusion. This is only one of many cases where data, principles, and/or patterns conflict.

Results of the Tests of the Reality of Certain "Problems"

Finally I will more or less briefly summarize the results of my research into the validity of certain conservative principles often espoused by would-be spelling reformers reluctant to eliminate the few advantages TO possesses: graphic distinction of homophones, graphic differentiation of monomorphemic simple words and bimorphemic inflected forms with homophonic final consonant clusters, and graphic resemblances between heterophonous agnates.

The first "problem" is easily disposed of, since in my tests of reading comprehension, 96% of homophones spelled identically in sentence contexts (i.e., new homonyms) were unhesitatingly recognized, and 91% of them were easily written in TO—the balance revealing problems only in spelling ability, not recognition or discrimination.

In the case of inflectional affixes, the problem is also homonymic, since a respelling of plural nouns, third-singular-present verbs, and weak preterits on the same phonemic basis as uninflected forms would result in a great deal of new homonymy. These fall into four categories: (a) weak preterits and past participles after voiceless consonants (cf. *mist/missed, pact/packed, rapt/rapped*), (b) weak preterits and past participles after long vowels, diphthongs, or voiced consonants (cf. *road/rowed, suede/swayed; side/sighed, proud/proved; build/billed, find/fined*), (c) plural nouns and third-singular-present verbs after /k/ (cf. *tax/tacks, minx/minks, coax/cokes*), and (d) plural nouns and third-singular-present verbs after long vowels, diphthongs, or /æu/ (cf. *freeze/frees, graze/grays; prize/pries,*

browse/brows; clause/claws, pause/paws). However, I found that by spelling base and affix separately, neither appeared to lose its identity, and monomorphemic words could be clearly distinguished from their bimorphemic homophones. My research indicated, though, that preterits in *d* were much more easily recognized than those in *t*—presumably both because *-(e)d* is normal in TO and because /d/ is automatically assimilated to /t/ anyway after voiceless consonants. Another possibility, after tonic short vowels, is the gemination of the final base consonant—but the informants tended to reject this method of avoiding homonymization.

Finally the matter of preserving the graphic resemblances between derivatives and their base forms (and each other)—despite (1) the variations in vowel quantity and/or quality accompanying stress changes (e.g., cf. *compete, competitor* and *competition*), and (2) the palatalization involved in affixation with historical /y/ following alveolar consonants—proved to be a “problem” of considerable complexity, but very small magnitude. The former difficulty could be obviated by employing the same grapheme for both long and short vowels, which is already normal and generally internalized in TO. And in the latter situation, there are only six suffixes involved (*-ian, -ion, -ious, -ial, -ual, & -ure*), and their spellings have been fairly well internalized, while phonemic respellings of them evoked considerable difficulty in recognition. But the degree of difficulty varied according to the degree of resemblance obtaining between base and derivative—and thus presumably the degree of internalization of the relationship. Those with a free base (e.g., *spirit/spiritual, expose/exposure, music/musician*) were almost all (97%) recognized as being related in TO, but only 52% when phonemically respelled. Those whose base is bound but retains its vowel sound (e.g., *construct/structure, motor/motion, visible/vision*) have rates of recognition of 82% and 26%, respectively. When the base vowel changes (e.g., *face/superficial, native/natural, price/appreciate*), these rates fall to 22% and 9%, respectively.

Analysis, Summary, and Commentary

One of the most significant results of these investigations was a verification of Robert Hall's contention that there has been a tendency to overemphasize the capriciousness of English orthography—that, in fact, there are “only a few instances in which an English phoneme does not have an independent regular representation of its own.” (Hall, 1961, p.23). Equally important, there are many combinations and patterns of phonemes that also have “regular representations” in TO—even though they are sometimes at variance with (or at least outside) conventional phonemic analysis. A good example of the disparity—one might even say conflict—between graphemic and phonemic patterns is the distribution of *k* and *c* for initial, prevocalic /k/. While *k* normally occurs only before front vowels and *c* before non-front vowels, this situation is reversed before /*ie*/ and /*æ*/—thanks to the Great Vowel Shift, which altered the sounds after their spellings had been frozen.

A far more complex example of this disparity is the relationship between vowel quantity, syllabic type, and graphic shape. For instance, short (or lax) vowels normally appear only in closed syllables and are represented by a single grapheme. However, while /*æu*/ and /*ʌ*/ are phonemically short vowels, their distribution and their usual spellings differ strikingly from those of the other five—primarily in that they normally require two graphemes each in TO. But while /*ʌ*/ behaves like a true short vowel by appearing solely in closed syllables, /*æu*/ occurs also, and with spelling differentiation, in open syllables, like long vowels.

Similarly, the distinction normally made by linguists (except in Trager-Smith analysis) between “long vowels” and diphthongs is only partially valid graphically—since both sound types share many characteristics and privileges of occurrence in English, such as appearing freely in final position and being usually represented there by digraphs. However, /*oi*/ and /*ou*/ differ markedly from other diphthongs—and “long vowels”—by requiring at least two graphemes at all times, whereas the other six normally use a single grapheme in non-final open syllables and a digraph in closed syllables.

This digraph often involves another important—and often underestimated—feature of TO: so-called “silent *e*”—which is, in reality, usually the second element of a discontinuous digraph; hence it is no more “silent” than the *y* of *ay* or the *h* of *th*. In fact, far from being “silent,” this letter—by affecting the quality of at least seven vowels and six consonants (sometimes two at once: cf. *breath/breathe*, *stag/stage*, *hears/hearse*)—is an extremely active and even useful feature of English orthography. Besides being essential to the pronunciation of many words (e.g., *face*, *shone*, *hinge*), it helps compensate for TO’s deficiency in vowel graphemes (thus distinguishing many graphemic minimal pairs—like *pin/pine*, *her/here*, *cut/cute*, *nap/nape*, *cod/code*, etc.) and its failure to distinguish voiced from voiceless interdental fricatives (cf. *mouth/mouthe*, *bath/bathe*, *teeth/teethe*). (This letter is, of course, truly “silent” in a few words [*have*, *score*, *definite*, *bronze*, *giraffe*, etc.], but these are a distinctly minority group.)

It is facts like these that have rendered this research valid—or even possible. For, as the accompanying data illustrate, not only do regular spellings exist for many—if not most—English phonemes and phoneme sequences, but many are also recognized as such, i.e., internalized, by literate native speakers. Only those not so internalized appear in Table I. In other cases, where the spelling is not predictable, it can be seen (in Table II) that certain representations are, nevertheless, considered by these speakers to be more “normal,” i.e., more “regular,” than others—even if in conflict with TO (Table III). In only a few cases is there real confusion or unpredictability on both counts (Table IV). Similarly, there is only a relatively small group of “hard-core” irregularly spelled words in English. Most bizarre spellings are simply rare—sometimes unique—exceptions to the basic, regular orthographic patterns. Unfortunately, their appearance in words of very high frequency has tended to exaggerate the significance of the problems they pose.

This is, of course, only a brief sampling of some of the results of my research into the internalization of English orthographic patterns. It does, however, give some indication of both the regularities involved in TO and the problems involved in any serious attempt to change it. It was not intended as a brief in favor of

spelling reform per se—much less a blueprint for any specific proposal in that direction. It does, however, provide some hard, objective data in an area which heretofore has been prey to subjectivity, conjecture, and caprice. In so doing, I have attempted to throw a little light into certain dark corners that no one else, to my knowledge, has ever probed in any systematic way.

1. Examples of the tests referred to here and subsequently may be found in my doctoral dissertation, entitled *Investigations Toward a New Approach to the Reformation of English Spelling* (q.v. under "References").

Some of the material in this paper was presented at the Twenty-First Annual Conference on Linguistics of the International Linguistic Association in New York City, March 1976.

REFERENCES

- Hall, Robert A., Jr. *Sound and spelling in English*. Philadelphia: Chilton Co., 1961.
Secrist, Robert H. *Investigations toward a new approach to the reformation of English spelling* (Doctoral Dissertation, New York University, 1965). *Dissertation Abstracts International*, 1966, 27, 762A. (University Microfilms No. 66-5754)
Wijk, Axel. *Regularized English*. Stockholm: Almqvist & Wiksell, 1959.
Zachrisson, R. E. *Anglic, an international language—with a survey of English spelling reform*. Uppsala: Almqvist & Wiksell, 1932.