DID ORTHOGRAPHIES EVOLVE?*

Ignatius G. Mattingly†

Abstract. According to Gelb (1963), writing has "evolved" from picture writing to logography to syllabic writing to alphabetic writing. It is argued here that this widely accepted theory of orthographic evolution does not really fit the historical facts very well, and that the variety of orthographies is better explained on linguistic grounds. Orthographies have to be productive, and they can manage this only by providing devices for transcribing the possible words in the lexicon. The very limited number of different ways in which this is accomplished in different orthographies is accounted for by the structural peculiarities of the languages that the orthographies transcribe.

It is generally believed by linguists, psychologists, psycholinguists and educators that writing has "evolved." First there was picture writing, then came logographies, then syllabaries, and finally, the alphabet. At each of these stages of development, writing became more efficient, because a smaller inventory of signs was required to do the job. The alphabet is the culmination of this evolutionary process, and its nearly universal triumph over less efficient orthographies has been well deserved.

The evolutionary view of writing probably originated during the nineteenth century, when most of the decipherments that led to our present knowledge of ancient writing systems took place, and theories of cultural evolution, inspired by the theory of biological evolution, were in vogue. The evolutionary view can be found in one form or another in many of the standard accounts of the history of writing. Thus Jensen (1970):

In the broader history of writing we can see then certain evolutionary tendencies emerging. Above all it is governed by the law of least resistance, according to which every change must in the normal way run from the more difficult to the more easy, from the more complicated to the more simple; we find, furthermore, in keeping with the general development of civilization, an increasing abstraction, a certain assimilation of the form to the self-increasing intellectuality of the content. (p. 22)

(Cf. also Pedersen, 1962, chap. VI.) And the evolutionary view has been elaborated into a theory by Gelb, whose A Study of Writing (1963) most of us

*To appear in RASE Topics in Learning and Learning Disorders.
†Also University of Connecticut

Acknowledgment. Support from NICHD Grant HD-01994 is gratefully acknowledged. Arthur Abramson, Alvin Liberman, Isabelle Liberman, and Harriet Magen read early drafts of this paper and offered valuable criticisms and useful examples.
Mattingly: Did Orthographies Evolve?

who are interested in the psychology of reading turn to for enlightenment about the natural history of writing.

Gelt says that "writing had its origin in simple pictures" (p. 190), advanced to "semasiography" (that is, picture writing), and then to "phonography," which comprehends word-syllabic, syllabic, and alphabetic writing (p. 191). The development of writing is said to be "unidirectional" (p. 200):

What this principle means in the history of writing is that in reaching its ultimate development writing, whatever its forerunners may be, must pass through the stages of logography, syllabography, and alphabetography in this, and no other, order. Therefore, no writing can start with a syllabic or alphabetic stage unless it is borrowed, directly or indirectly, from a system which has gone through all the previous stages. A system of writing can naturally stop at one stage without developing farther. Thus a number of writings stopped at the logographic or syllabic stage. (p. 201)

Thus, just as biological evolution explains the variety of natural species, orthographic evolution is said to explain the variety of orthographic species.

What I wish to do here is to reconsider the theory of orthographic evolution. I will argue that the evolution of writing has been more apparent than real, and that the variety of orthographic species is better understood from a standpoint more linguistic than Gelt adopts. The alphabet, I will suggest, is not necessarily the best way to write all languages. For the evidence that leads to these conclusions, I rely mainly on the remarkable erudition of Gelt himself.

Is this a matter of more than marginal concern for the psychology of reading and spelling? I suggest that it may be, for the evolutionary view is echoed by psychologists concerned with the reading process (Crowder, 1982, p. 148; Henderson, 1982, p. 7), and the supposed evolution of writing is sometimes taken to reflect psychological facts and even to suggest teaching strategies. Citing Gelt (1963), Gleitman and Rozin (1977) say:

...each orthography arose as a gradual refinement and generalization of resources already implicitly available in its predecessors, as though the early scripts formed the necessary conceptual building blocks required for further development.... On these grounds, one can build a plausibility case (though only that) for organizing reading instruction in terms of a similar accumulation of conceptions: perhaps ontogeny recapitulates cultural evolution. (p. 8)

Let us begin with the claim that logography evolved from picture writing. There are seven ancient traditions of logographic writing: the Mesopotamian, Proto-Elamite, Proto-Indic, Sino-Japanese, Egyptian, Cretan, and Hittite. Decipherment has not progressed very far in the cases of Proto-Elamite, Proto-Indic, and the early Cretan writing, but in the case of the other logographic traditions there is evidence that the signs were at first iconic (Gelt, 1963, chap. III), only later becoming arbitrary and non-iconic. The obvious explanation for this development is that while iconic signs were suitable for monumental inscriptions, hieratic, commercial, and literary uses required signs that could be rapidly written rather than slowly drawn. There
was thus an evolution from iconic to non-iconic writing. But regardless of their graphic form, the signs were from the beginning logograms: they stood for words (or more correctly, morphemes), not, as is sometimes said, "concepts" or "meanings." An iconic sign designated a particular word by suggesting some aspect of its meaning, but the meaning of the logographic text did not depend on these pictorial hints, but on the selection and ordering of the words, just as it does in spoken and written language in general. Non-iconic signs, arbitrarily associated with words, served the purpose equally well.

Picture writing, on the other hand, is non-linguistic. The term is a convenient cover label for a fascinating miscellany of assorted artifacts from preliterate societies: rock-drawings warning of danger nearby, pictorial "letters," narratives and proverbs, tribal and commercial identification symbols, calendar systems, and so on (Gelb, 1963, chap. II).

In what sense can logography be said to have evolved from picture writing? The claim would have some substance if it could be shown that the signs of some logography were borrowed from or paralleled those of a particular tradition of picture writing, but there appears to be no example of this sort in any of the logographic traditions. The Mesopotamian Sumerians used both cylinder seals and logographic writing on commercial identification tags, but there is no relationship between the seals and the writing (Gelb, 1963, p. 65). If cultural evolution means anything, it must imply some kind of structural development: thus the computer can reasonably be said to have evolved from the loom. But linguistic writing merely took over the communicative functions of picture writing, as the internal combustion engine took over the locomotive functions of the horse; it did not, in any interesting sense, evolve from picture writing.

The second part of Gelb's theory is that syllabaries evolved from logographies. This claim implies that within a particular orthographic tradition, there is a period of strictly logographic writing, then, perhaps, a transitional period, and then a period of strictly syllabic writing. But what we actually find, in the Mesopotamian, Hittite and Sino-Japanese traditions (Egyptian will be discussed shortly) is just the transitional period.

The writing in these traditions is what Gelb aptly calls "word-syllabic" writing, in which logograms and syllabary signs supplement each other. Thus, in Sumerian and in Japanese writing, the syllable signs are used regularly to write inflectional morphemes and can also be used to write base morphemes. Alternatively, a base morpheme can be written with a logogram, and in this case, a supplementary syllable sign is sometimes used to indicate the phonological form of the morpheme. In Chinese writing, some of the characters are simple logograms, but most of them consist of two component signs: the "radical," one of 214 signs that serve as semantic classifiers, and the "phonetic complement," a sign that in isolation has a phonological value similar or identical to that of the compound character. The compound character for /kuə/, blind, for instance, is composed of the simple signs for /kuə/, drum and /muə/, eye (Jensen, 1970, p. 170). Since the phonetic complements have logographic values of their own, and there are in general quite a few phonetic complements for a particular syllable (10 for /liə/ for example; Wieger, 1927), it might seem a bit eccentric to regard Chinese writing as systematically syllabic, rather than simply as a case of massive phonetic transfer. But the fact that a common error in the writing of Chinese is the use of an
incorrect but phonologically accurate phonetic complement (H.-B. Lin, personal communication) attests to the psychological reality of the syllabary system.

In all these word-syllabic orthographies, the syllable signs clearly derive from logograms. Thus the syllable sign for /gal/ in Sumerian derives from the logogram for /gall/, great (Gelb, 1963, pp. 110-111); one of the phonetic complements for /ku̯/ in Chinese, as we have seen, derives from the logogram for /ku̯/, drum; and the Japanese kana for /mo/ derives from the character for /mo/, hair, borrowed from Chinese /maoo̯/, hair (Jensen, 1970, p. 201). But is derivation necessarily to be equated with evolution? Gelb himself makes it quite clear that there is no period in any of these traditions during which the writing was strictly logographic; syllable signs occur in the earliest specimens (Gelb, 1963, pp. 67, 83, 85). Nor did any of these traditions lead eventually to a strict syllabary, though some of the later Mesopotamian systems came fairly close (p. 165).

The Cretan tradition is perhaps the one case that supports the claim. Whether there was a strictly logographic stage cannot be determined until the early Minoan scripts are deciphered, but the strictly syllabic Cypriote orthography appears to have developed from the earlier word-syllabic stage represented by Cretan Linear B (Gelb, 1963, p. 154).

Finally, Gelb's theory claims that alphabetic writing evolves from syllabic writing. But this part of the theory depends crucially on Gelb's particular interpretation of the structure of the Egyptian and West Semitic orthographies, and on his presumption that the latter derive from the former.

In the Afro-Asiatic family of languages, to which both Egyptian and Semitic belong, the base morphemes are, in general, simply consonantal patterns, for example, Egyptian n-f-r, lute; p-r, house; and Semitic k-t-b, to write; n-l-k, to rule. In actual words, vowels are morphologically inserted and, together with prefixes and suffixes, distinguish the various forms derived from the base. Thus the base k-t-b yields in Hebrew [ka'tav], he wrote; [jix'tov], he will write; [jik'atev], he will be inscribed; [mix'tav], letter; [ktu'ba] marriage, and many other forms.

Egyptian writing is a mixture, often redundant, of logograms and signs for consonants and for sequences of two consonants. These consonantal and biconsonantal signs are derived from the logograms by phonetization. Thus the sign for d-t, snake, is used for the consonant /d/, and the sign for /w-r/, swallow, is used for the consonantal sequence /w-r/ in writing /w-r-d/, to be weary (Jensen, 1970, p. 60). There are no obviously syllabic signs. Vowels are not ordinarily indicated, but in special cases, such as foreign proper names, the signs for the consonants /u/, /j/, /w/ are used for vowels /a/, /i/, /u/, respectively. This assignment of consonantal signs to vowels is not arbitrary. /j/ is homorganic with /i/ and /w/ with /u/. While /u/ is not homorganic with /a/, it is nevertheless phonologically reasonable to transcribe the low back vowel with the sign for the glottal stop, the lowest and most back consonant. As with Sumerian and Chinese, there appears to be no historical period during which the writing is strictly logographic; the consonantal signs are there from the first (Gelb, 1963, p. 74).

Ancient Semitic writing consists simply of signs that ordinarily stand for single consonants: thus Hebrew [ka'tav] is written ktb, and [mix'tav], mktb. But as with Egyptian, consonantal signs are used, when necessary, to
indicate vowels: the signs for /æ/, /j/, /w/, aleph, yod and waw, could indicate /a/; /i/ or /e/; and /u/ or /o/, respectively. This device was used not only for proper names: [da'wid], David, being written dwid, but also to avoid ambiguity in other words, [jix'tov] being written jktwo to distinguish it from [jik'atev], written jktb.

Pace Gleitman and Rozin, it was surely not the case that the West Semites didn't "notice" the vowels in their language (1977, p. 19): when it was important to write the vowels, they wrote them. On the contrary, what is especially significant about the Afro-Asiatic languages is that their morphological structure must have fostered awareness of segmental structure to a far greater degree than in the case of Indo-European languages. As I have argued elsewhere, such "linguistic awareness" is not automatic and is essential for alphabetic reading and writing (Liberman, Liberman, Mattingly, & Shankweiler, 1980; Mattingly, 1972).

The reason that both Egyptian and Semitic could be written without consistent indication of vowels is that, in general, the vowels carried only inflectional information. Since word-order is relatively fixed, this information is for the most part redundant. On the other hand, in Greek and in Indo-European languages generally, the base morphemes include vowels. Thus, when the Phoenician alphabet was adapted to Greek, it became a plene alphabet: vowels as well as consonants were regularly transcribed, aleph, yod, and waw being used for /a/, /i/, and /u/ as before, and three other Phoenician consonantal signs, he, /h/, heth, /h/ and ayin, /œ/, for /e/, /œ/ and /o/, respectively.

To maintain his theory of orthographic evolution, Gelb has to argue, since there are no preceding West Semitic logographies or syllabaries, that the West Semitic scripts derive from the Egyptian. And since he denies the direct development of an alphabet from a logography, he has to argue that the Egyptian consonantal and biconsonantal signs are really syllabic.

In asserting the derivation of the West Semitic script from the Egyptian, Gelb very properly rejects the farfetched attempts of other scholars to demonstrate similarities in the forms of the signs of the two scripts. His argument relies on the similarity of "inner structure" (p. 146), that is, the use of a limited set of signs to express consonants but not (ordinarily) vowels. But this argument loses what force it might have in view of the fact that it is the same peculiarity in morphological structure that made it possible for both languages to be written in this way. Gelb might have adduced a further similarity of inner structure: when vowels did have to be written, the signs for the same three consonants, /æ/, /j/, and /w/, were used to write the same three vowels, /a/, /i/, and /u/. But the similarity of Egyptian and West Semitic phonological inventories explains this. Since both had the consonants /æ/, /j/, /w/ phonologically related to the vowels /a/, /i/, /u/, respectively, the signs for these consonants were the obvious choices to write the corresponding vowels. Though the possibility cannot be ruled out, there is no need, in the absence of other evidence, to conclude that West Semitic script is derived from Egyptian script. The linguistic similarity of the Egyptian and Semitic languages is quite sufficient to account for the similarity of the two scripts.
As for the Egyptian consonantal signs, Gelb's proposal is that each of them represents a set of syllables or disyllables with the same consonants but varying (or zero) vowels. Thus the biconsonantal sign that other scholars transliterate as mn or m-n is transliterated by Gelb $m^x n^x$, $m^x n(x)$, $m(x) n^x$, $m^x n(x)$; $x$ standing for whatever vowel is required in context (1963, pp. 77-78). From the reader's point of view, this might seem a distinction without a difference, but for Gelb it is crucial:

The Egyptian phonetic, non-semantic writing cannot be consonantal, because the development from a logographic to a consonantal writing, as generally accepted by Egyptologists, is unknown and unthinkable in the history of writing, and because the only development known and attested in dozens of various systems is that from a logographic to a syllabic writing. (pp. 78-79; original in italics)

But, obviously, this argument is entirely circular; only the theory itself justifies the syllabic interpretation. One might have supposed that the West Semitic scripts, at least, could be allowed to be alphabetic without damage to the theory, but to concede this would obviously undermine the claim of inner structural similarity between them and the Egyptian script. Thus the West Semitic script must be syllabic, too, waw, for example, being transliterated wa, wi, wu (Gelb, 1963, p. 148), and the development of alphabetic writing must await the Greeks.

This claim is not only uncorroborated; it also makes it much more difficult to account for the emergence of the Greek plane alphabet. If the Phoenician orthography was syllabic, there is no particular reason why the Greeks, any more than other Indo-Europeans, should have become aware of the segmental character of their language when they borrowed this orthography. We should expect to find them using, at least at first, a patched-up syllabary like that of the Persians. But if it is recognized that the West Semites, thanks to the peculiar morphology of their language, had already arrived at the alphabetic principle, then the development of the Greek alphabet from the Phoenician alphabet can be seen to be simply a matter of adding two more vowel signs and using them consistently.

If we do not accept the claim for the development of West Semitic writing from Egyptian writing, and for the syllabic nature of at least the latter, then Gelb's theory is in trouble, for it would seem that, insofar as derivation can be equated with evolution, an alphabet can evolve from a logography without an intervening syllabic stage, as in the case of Egyptian; and may even, perhaps, emerge without any precursors, as in the case of West Semitic; but that no alphabets have developed from syllabic or word-syllabic systems, for apart from the Ugaritic cuneiform alphabet, of unknown origin (Gelb, 1963, p. 129), all other alphabets are derived directly or indirectly, from the West Semitic consonantal alphabets.

The theory of orthographic evolution cannot be correct, for logography cannot be shown to have evolved from picture writing in any meaningful sense; syllabaries do not generally develop from logographies; and alphabets do not develop from syllabaries. What we find instead are either logosyllabic traditions: Mesopotamian, Hittite, Cretan, and Sino-Japanese; or alphabetic traditions: Egyptian and West Semitic. We can, if we choose, regard as evolutionary the development of non-iconic logograms from iconic ones, or the development of the Greek plane alphabet from the Phoenician consonantal alphabet, but these are not the sorts of evolution the theory calls for.
Mattingly: Did Orthographies Evolve?

But without the theory, how can we account for the variety of orthographies? Let us consider this question from a rather different point of view. The orthography of a language must be productive; that is, it must enable the user to write any of the infinite number of possible utterances of the language. Because there are many levels at which an utterance is mentally represented in production and perception, there are, in principle, many possible forms that a productive orthography might take. For example, any utterance of a particular language (in fact, any utterance of any language) can be written in a general system of phonetic transcription. If such a transcription were used as an orthography for all languages, any literate person could read aloud in any language. Or one could imagine an orthography that would be based on the acoustic properties of utterances (cf. the "visible speech" of Potter, Kopp, & Green, 1947, and the stylized spectrographic patterns used for speech synthesis by rule at Haskins Laboratories by Liberman, Ingemann, Lisker, Delattre, & Cooper, 1959); such an orthography would include just the information on which the listener to spoken language relies. Or one could imagine an orthography based on the semantic representations of utterances (cf. Kazz & Fodor, 1963), if indeed such representations really exist (Fodor, Fodor, & Garrett, 1975); after all, it is the meaning, not the linguistic structure, that the writer really wants to convey to the reader. But it is obvious that none of these alternatives would do for a practical orthography, though it is not easy to say exactly why (see Mattingly, 1984, for some speculations).

There is in fact a very severe limitation on orthographic variety. In practical orthographies, only one basic principle has ever been used, that of transcribing utterances of a language as sequences of lexical items, that is, words. I would argue that all known orthographies are in this sense lexical, varying only in the specific ways in which they happen to transcribe the words. The lexical character of logographies seems obvious, but it might be objected that alphabetic systems are essentially transcribing the phonemes of utterances, and only incidentally the words. With a well-behaved orthography, like that of Serbo-Croatian, only the spaces between the words indicate its specifically lexical character. The point becomes clearer in the case of an eccentric orthography, like that of English, in which there is usually more than one way to write a particular sound. Thus English [ay], phonologically /l/, can be written -igh-, -y, -y(-)e, 1(-)e, -uy. But despite this variability, there is but one way of writing each of the words sight, try, lye, dyne, lie, lime, buy.

A lexical orthography can only be productive if it incorporates a system for transcribing all the words in the languages. There is, however, no principle that can specify just the actual words of a language, and provide the basis for such a system. Thus /8ay/8, v., to gather truffles on Wednesday could perfectly well be an English word; its absence from the lexicon is accidental. Nor, since the membership of the lexicon, though finite in theory, is indefinite in practice, would it be satisfactory simply to list all the words and provide an arbitrary sign for each. Any word that was inadvertently omitted, or entered the language after the list was compiled, would be unwriteable. And the writer who could not remember the sign for a word that was on the list would be driven to paraphrase. Thus there can be no strict logographies, for a strict orthography would not be productive; and accordingly no such stage is actually found in Sumerian, Egyptian, Hittite, or Chinese.
Mattingly: Did Orthographies Evolve?

There is, however, a way to specify all possible words in a language. The phonetics and phonotactics of a language determine the set of phonological forms that qualify for membership in its lexicon. Thus, while /kayf/ could be a word in English, and /kæt/ really is one, /λø/ and /stwøyɡ/ could not be. By exploiting the phonological structure of the language, that is, by some form of phonetization, an orthography insures that any possible word can be transcribed. This does not mean that a writer will always know the standard way to write a particular word, or that the reader will always know what word is transcribed by a particular orthographic form. It does not preclude a particular word's being standardly transcribed in some exceptional or arbitrary way, e.g., one. What it does mean is that if /kayf/ should enter the English language, there will be at least one, in fact several, ways to write it; that the writer who cannot recall the standard spelling of cat can at least write kat, and that the reader confronted with a word unfamiliar in its written form will have a basis for guessing what the word is.

Although lexical items have syntactic and semantic as well as phonological properties, only the last allow the specification of the set of possible words of a language. Syntactic properties are not sufficient to specify different words uniquely, and a principled characterization of word meaning has thus far eluded the efforts of linguistic semanticists (Fodor, 1977, chap. 5). As we have seen, however, semantic properties can nonetheless play a useful auxiliary role in orthographies.

Every orthography, then, achieves productivity by incorporating a system for transcribing phonologically the possible words of the language. Since the only relevant phonological units are syllables and phonemes, there are really only two ways to do this: the syllabic way and the alphabetic way, and we have seen that all orthographies make use either of the one or the other. But why must there be even two ways? Why are not all orthographies plene alphabets? The answer is that, to a large extent, the morphological and phonological structure of a language defines the orthographic options. There are some languages for which a plene alphabet would be cumbersome and redundant, and others for which there is no really satisfactory method of phonetization. Moreover, the alphabetic option becomes an obvious one only under rather special linguistic circumstances.

A Semitic language, unless it has borrowed heavily from a non-Semitic language, has no need of a plene alphabet. Since lexical items are consonantal patterns, the vowels carrying only inflectional information, an extremely parsimonious system of phonetization is possible, as the West Semitic orthographies demonstrate. Under similar linguistic circumstances, Egyptian writing was able to achieve productivity in much the same way. The extensive and often redundant use of logograms does not alter the fact that the uniconsonantal and biconsonantal signs are the true basis of this orthography.

Because of their restricted syllable structure, Sumerian, Chinese and Japanese are less orthographically amenable. Japanese has only 74 phonotactically possible syllables (or more exactly, moras). Chinese has about 1200 possible syllables, but by no means all of them are actually used. Sumerian appears to have been similarly restricted. Restricted syllable structure surely promotes awareness of syllables, and in these cases a syllabary might seem to be the obvious phonetization device. But the morphological consequence of restricted syllable structure unfortunately, is pervasive homophony, exacerbated when, as in Chinese and Sumerian, the base morphemes are mostly

258
monosyllabic. For example, there are 38 different Chinese words with the phonological form /ni/ (Wieger, 1927). Under these circumstances, a strict syllabary is hardly practical, for it would give rise to pervasive homography, far less tolerable in writing, because of the lack of prosodic information to help specify syntactic structure, than pervasive homophony in speech. For these languages, a word-syllabary system, in which the ambiguity of syllable signs is reduced with the help of logograms, is a reasonable, if not highly efficient solution. Alphabetic writing would be no improvement. To replace the syllabic signs in Chinese and Japanese writing by alphabetic ones would do nothing to reduce homography, and to use only an alphabet to write these languages, convenient though it might be for printers, would be disastrous for readers.

For many other languages, a plene alphabet is the most efficient system of phonetization. But the alphabetic principle is not an obvious one. It did not occur to the Hittites, who used a word-syllabary system even though they did not have a homophony problem and could have used an alphabet. It occurred to the Egyptians and the West Semites only because the morphology of their peculiar character of the languages made them aware of phonological segments. It is certainly owing entirely to the West Semitic example that alphabetic writing is now so widespread.

It would, however, be pressing the point too far to say that variations in linguistic structure account for all orthographic variety. Non-linguistic factors assuredly play a role. The Akkadians, for example, spoke a Semitic language and would certainly have been well advised to use a consonantal alphabet. But being impressed by the culture of the Sumerians, they adopted the Sumerian orthography and made writing unnecessarily complicated for themselves and their Mesopotamian successors (Jensen, 1970, p. 94). Greek speakers on the island of Crete used a word-syllabary system, Linear B, no doubt influenced by the example set by the speakers of the unknown Minoan language written in Linear A (Gelb, 1963, p. 91 ff.). The bewildering complexities of the Japanese kanji, borrowed from the Chinese, have a similar historical explanation (Martin, 1972).

To summarize, Gelb’s widely accepted theory of orthographic evolution must be rejected. Orthography has no relationship to picture-language, and there is no sequential development from logography to syllabary to alphabet. The forms that orthographies have taken are constrained by the requirement that they must be productive, and must transcribe lexical items. The limited variety of orthographies can be explained largely on linguistic grounds.

References

Mattingly: Did Orthographies Evolve?


Footnote

1Subscripts denote phonological tones.