1

The Structure and Acquisition of Reading I: Relations between Orthographies and the Structure of Language

Lila R. Gleitman
Paul Rozin

University of Pennsylvania

1 INTRODUCTION

1.1 Learning to Speak and Learning to Read

If meanings could be directly conveyed—in whatever form these are represented in the head—from one human to another, there would be little mystery in language acquisition, and precious little in the acquisition of reading. Problems arise both for listening and for reading because, to convey our ideas, we have encoded them onto a physical medium (a sound wave, a squiggle on paper); to apprehend, we must map heard or seen representations of ideas back onto mental representations. The past two decades of investigation in linguistics and psycholinguistics have enhanced our appreciation of the complexity of the encodings (roughly, semantic, syntactic, and phonological) that mediate between meaning and its physical realization in the sound-stream. In fact, so complex and indirect are the relations between meanings and sounds that the acquisition of language by the child seems an incredible accomplishment, at least in terms of the explanatory concepts available within psychology to account for learning. Yet, whatever the mechanisms, it must be acknowledged that each normal child readily learns the mappings between his thoughts and the sound stream of speech.

Is the case the same for reading? At first, the parallel between learning to speak and understand and learning to write and read does seem close. Written language becomes for many of us second nature, requiring no more effort or will to deploy than hearing or speaking. Even very young successful readers find it difficult to look at billboards without understanding "what they say." This state of affairs superficially seems a natural one–given that the spoken language has been learned–for written language appears to be a simple cipher on spoken language: each letter of the alphabet, roughly, is a visual analogue of some known sound category. In this view, the skill of reading reduces simply to learning a very small set of associations between the sounds of the spoken language and squiggles on a page of print. Yet if this is so, learning to read should be easy, for it involves little that is new to the prospective reader. The rich structures of syntax and semantics, and the relations between patterns here and the sound patterns of language, are already known in large part to the kindergartner. To read, it appears that we need "merely" transfer all of this knowledge of language to a visual modality whose symbols map onto it roughly one to one.

Then considering all that must be learned in order to speak and understand, beginning (as an infant) with no language in advance, acquiring one's native tongue should be a difficult and lengthy task. In contrast, learning to read should be simple. This follows from the parasitic relation between speaking and reading.

But the facts clearly violate these expectations. Significant numbers of individuals fail to learn to read even after years of lessons at school. To be sure, a few children learn to read almost overnight, but they are exceptions, and are sources of wonder to their proud parents. On the contrary, virtually no one fails to talk. Every normal individual comes to speak his native tongue in the ordinary context of events of the first years of life, and without the intervention of teachers and formal drill. The acquisition of speech proceeds successfully across broad ranges in intelligence and across vastly different cultures and child-rearing practices. Speech emerges more or less autonomously in response to a biologically determined maturational schedule that is immune to any but the most radical pathological or environmental interference (for a major statement, see Lenneberg, 1967). In contrast, except in special circumstances (Read, 1971), the acquisition of reading requires a long period of schooling, and varies in success and scope according to intelligence (Thorndike, 1971), motivational and cultural factors (Downing, 1973), and internal differences in the nature of the writing system that is to be acquired (Downing, 1973; Klima, 1972; Leong, 1973).

What account can be given for the fact that what appears to be the more general and complex task (learning to speak and understand) is less difficult and less variable than what appears to be a trivial derivative of this (learning to read and write)? In our view, the difference has only indirectly to do with the visual modality itself. In fact, writing systems whose symbols are words or "meanings" present no special problem for initial acquisition (Rozin, Poritsky, & Sotsky, 1971; Gleitman & Rozin, 1973a). The problem of reading has to do with the *cognitive prerequisites* to understanding alphabetic systems: properties of these orthographies require their users to become aware of and to focus attention on the phonological substratum of speech. Failure to achieve this fundamental insight about the nature of alphabets characterizes the preponderant number of normal individuals who do not achieve literacy (Rozin & Gleitman, this volume).

There seems to be something of a paradox here. Surely hearing and understanding presuppose decomposition and analysis of utterances into their sound properties, else how could we know the difference between such utterances as "The sow bit the cat" and "The cow bit the rat"? This analysis of the sound stream by the ear is carried out below the level of consciousness by an evolutionarily old and highly adapted mechanism for speech and hearing (A. M. Liberman, 1970; Lieberman, 1973). To be sure, reading and writing also appear to require decomposition and analysis: evidently one must analyze a sequence of alphabetic letters by eye "as if" the letters reproduced properties of the sound stream. But why should this be so much harder? Surely it is not the sheer visual aspect of reading (discriminating the visual forms of letters and letter sequences) that creates the difficulty. More likely visual decoding must be carried out at a *higher level of awareness* than acoustic decoding: reading is a comparatively new and arbitrary human ability for which specific biological adaptations do not, so far as we know, exist. One must quite consciously learn to carry out the analysis of visual symbols on a page "in the same way" one naturally analyzes speech signals without insight and without learning. In this derived sense only, the visual modality is implicated in the difficulty of learning to read. The problem with reading is not a visual perception problem; the problem is rather that the eye is not biologically adapted to language.

In sum, learning to read requires a rather explicit and conscious discovery and building from what one already knows implicitly for the sake of speech: the structure of one's language and, particularly, the sound structure of one's language.[1] To the extent that this is so, it is small wonder that the psychology of learning (and the practice of education) provides little guide to how to teach reading. Current theories of learning can scarcely distinguish between acquiring new and arbitrary behaviors–such as memorizing the Pledge of Allegiance and

[1] This position is a controversial one, in a field notable for its controversies (for an earlier statement along similar lines, see Gleitman & Rozin, 1973b; for a more general discussion of "metalinguistic" awareness, see Gleitman, Gleitman, & Shipley, 1972; for a more general discussion of the concept of accessibility, see Rozin, 1975). A number of investigators have similarly proposed that the requirement of explicit awareness of phonology is the fundamental cognitive barrier to the acquisition of alphabetic reading (Mattingly, 1972; Downing, 1973; Cazden, 1973; and I. Y. Liberman, 1970), but many investigators take a radically different view of the reading problem (see particularly Smith, 1971; Goodman, 1969, 1973; and discussion of their position in Rozin & Gleitman, this volume).

using a decimal notation—and becoming aware of knowledge already in the head—as in learning to parse English sentences and following a proof of the Pythagorean Theorem (at least according to Plato).

The present authors do not know any better than anyone else, for the purpose of teaching reading, how to make use of the fact that the child who is learning to read already "knows" that his own language consists of the sentences and words and sounds that he is encountering anew, in a visual analogue, on the printed page. We will argue in this chapter, in terms of facts about the history and conceptual nature of writing systems, and in the following chapter, in terms of facts about novice and fluent readers, that certain levels of language encoding are easier to bring to awareness than others, and thus easier to deploy in the service of the cognitive act of reading. Specifically, we maintain that it is easier to become aware of meaning than of syntax, easier to become aware of words than of sounds, and easier to become aware of syllabic segments than phonemic segments of speech. To the extent that this is so, we claim, the acquisition of reading will be made more coherent for the learner if the task is dissected in terms of successively more abstract encodings of meaning, beginning with relatively concrete visual representations (of words) and progressing by steps toward the phonological representations that underlie alphabetic writing.

In this presentation, we try to describe the constraints on reading acquisition that derive from the nature of alphabetic units and the way these could be perceived; primarily, we describe the categories *morpheme, syllable, phoneme,* and *phone.* After some definitional comments (Section 1), we sketch the history of writing systems in human culture (Section 2). From our point of view, the historical development of orthographies is conceptually orderly, for it reflects a step-by-step increase in the abstractness of the explicit knowledge required of the learner; that is, primitive scripts allow the learner to focus his attention on concrete linguistic categories (morpheme and syllable) while alphabets require him to focus on the abstract phoneme and phone. After outlining the history of writing, we provide in Section 3 a more detailed discussion of the English alphabetic system and the relations of this orthography to the sound patterns of language. In Section 4, we discuss those aspects of speech production and perception mechanisms that correspond to units of writing systems. In a final comment (Section 5), we relate the historical strands of evidence to the descriptive linguistic and perceptual facts, and make a suggestion about teaching reading.

Elsewhere (Rozin & Gleitman, this volume) we take up in detail facts about fluent and novice readers that give further empirical substance to the claims made here on the basis of historical and descriptive facts about scripts. We demonstrate there that the fluent reader makes use of the many levels of linguistic representation available in alphabetic writing, in terms of a rapid and critically integrated sequence of information manipulations. Further, we claim that the prereading child, in the normal course of his language development, becomes increasingly aware of successively more abstract linguistic levels and categories. In this sense, we maintain, the child's metalinguistic history (explicit knowledge of language) mirrors a cultural history seen in the evolution of scripts. We therefore finally propose and report on an approach to teaching reading that recapitulates in many respects the historical development of writing: we begin (as did paleolithic Man) with pictorial representations of words, and proceed through rebuses and syllabic scripts before introducing the highly abstract alphabetic notation of modern English writing.

1.2 On Defining Literacy

What is reading? An old story concerns an illiterate Greek who had an American friend who could decipher the Greek alphabetic notation, but did not understand the language. When mail arrived from Greece, the American would render the words aloud with fine incomprehension, while the Greek listened and understood. Does either of these men read Greek? But an act of reading took place.

This story highlights some difficulties in defining literacy. It points as well to some interlocking questions about what and how to teach children in a reading class; since this second issue is addressed in the accompanying chapter, we leave it aside here. It is hard to find a serious definition of reading that will encompass both protagonists in the apocryphal tale above. But a preliminary step can perhaps be taken by defining literacy in the terms of some test procedure: one who can perform competently in the test will be called a reader, and it is his performance we will subsequently try to describe. What is an appropriate test for literacy?

One possibility is a direct test for certain concepts underlying the script: for example, one might require the subject to read nonsense syllables aloud at a reasonable clip. This test would capture whatever skill the hypothetical American brings to reading Greek, a skill called "decoding" print into sound in the education literature. This test has some evident validity with respect to reading, but it is inadequate to engage some obvious skills of fluent readers of a known language. Clearly, the activities of a skilled reader reading a language he understands are guided, in part, by attention to syntactic and semantic constraints in the text. This accounts for the fact that while readers can recite nonsense text at the rate of about 100 word-like items per minute, their oral reading speed with real English text is more like 250 words per minute (Rozin & Gleitman, this volume). Even if the fluent reader does some "sounding out" or decoding of words, he also makes judicious guesses from context. A nonsense-reading test cannot recruit whatever skills are involved here.

An alternative is the reading-comprehension test. This requires the subject to answer questions about the meanings in a text rather than merely to reconstruct its sound. This plausible procedure has its own flaws. The reading-comprehension test is contaminated by matters external to conceivable reading skills, and rather more so than the nonsense-reading test just rejected. That is, vast differences in comprehension arise that do not involve the printed page in any way. Consider the differences in comprehension among college students listening to the same lecture, or the differences within an individual who is asked to read either *Finnegan's Wake* or *The Little Prince.* The reading comprehension score confounds the ability to read with the ability to understand. The confounding increases as the complexity of the test materials increases. Clearly, such a test confuses verbal intelligence and reading skill. Many so-called standardized tests of reading used in our schools are contaminated in just this way. This is one of the reasons why the correlation between IQ and reading achievement rises substantially between early and late elementary school (Singer, 1974).

What is required in a literacy test is measurement of some confluence of skills, separate in part from understanding complicated meanings but tapping the squiggle-to-meaning mapping, that relates to the difference between a reader and an illiterate, with intelligence equated. This dictates that an appropriate test for literacy consist of (1) meaningful and grammatical material, so that the proficient reader can manifest his various skills, and (2) content we know the reader could understand if he were to hear it rather than read it. It follows that an appropriate test is most properly a comparison of what a person can apprehend from a tape recording with what he can apprehend from a transcript of that recording. (The practical development of such a test is of course complicated immensely when we ask how to define "equal exposure time" to the two conditions, etc.)

At the beginning of reading instruction, the pupil understands spoken materials at some level, while he understands the same materials in print not at all. At the opposite extreme, many educated people seem to acquire information at least as easily from books as from lectures. We will define the individual as fluent when he performs about equally in both tasks. This definition is the sense of literacy we have in mind throughout this and the accompanying chapter. A reasonable goal for education, in reading instruction, is to narrow the original gap between oral comprehension scores and reading comprehension scores. Barring some basement cases, then, the absolute level of performance in the listening-to-looking comparison is of no moment in teaching or testing reading skill in an individual: one may be literate and clever or literate and dull. Of course, for the purposes of cognitive psychologists investigating the components of literacy (as opposed to those who test individual literacy for school purposes), there is much to be learned from measuring absolute performance with nonsense materials, with conceptually complex text, and the like. The point here is merely that the sources of differential performance under these paradigms are too complex to permit either one to stand alone as a definition of "reading" or as the sole evaluative tool in measuring individual literacy. Consequently, we define

reading as the skill of extracting meaning from print to the same degree that one extracts it from the sound stream.

2 THE EVOLUTION OF WRITING

We now turn to the origins of writing in human culture, introducing each script (*orthography*) in the order in which it appeared in human society. This historical overview may appear to be tangential to a description of the cognitively ordered sequence of conceptualizations involved in reading acquisition. After all, why should the history of writing reflect psychological complexity and thus—even more derivatively and less securely—a teaching sequence? It is quite possible to suppose that the various orthographies used in different societies at different times arose quasi-independently of each other (and of principles of cognitive organization) in consequence of haphazard invention or local cultural need. If so, the historical record would provide no guide to issues of learning and using *our* alphabetic script. But a careful reading of the history of writing reveals a conceptually orderly progression. Orthographic convention proceeds, almost uninterruptedly over time, in a single direction: *at every advance, the number of symbols in the script decreases; concurrently, and as a direct consequence, the abstractness of the relations between the written symbols and the meanings increases.*

In retrospect, it is easy to see why the pictures of objects that appeared in primitive writings (see Fig. 1) were so useful: they immediately evoke meaningful interpretation. But it is also clear that since there are very many meanings, there will have to be very many symbols in a script that renders these so directly. Unless all scribes are realistic artists who can render nature veridically, and unless each picture reliably evokes a unique meaning interpretation from all viewers, the learner of a picture script (*pictography*) will have to commit an enormous number of picture-to-meaning pairs to memory. Obviously, if the number of symbols in the script can be reduced, learning will be easier and broader.

But how can the number of written signs be decreased if the writing system must continue to convey everything one wishes to write? The solution is to render language visually in terms of more abstract levels of linguistic representation, which contain fewer units. There are more meanings than words, more words than syllables, and more syllables than speech sounds (phonemes or phones). If, then, a writing system provides only a symbol for each phoneme—describing syllables and words and meanings only as compositions of phonemes—it will increase efficiency for the learner. But there is a cost to decreasing the number of symbols. If the writing system abstracts away from the meanings it conveys, the decipherer will have to recover the meanings from the now encoded form in which they have been rendered. And of course the prospective inventor of a script will have to notice and symbolize these more abstract units. The abstract analytic writing systems were invented later, were reinvented less frequently in different cultures (Gelb, 1952), and pose considerably greater learning problems for many individuals than do the primitive scripts (Rozin *et al.*, 1971; Gleitman & Rozin, 1973a). All of these facts suggest that some guidance for the issue of how to teach reading may come from the natural history of writing.

The observed orderliness in the successive invention of scripts is not surprising or fortuitous because writing did not develop through independent—or evidently very conscious—"invention" (here a syllabary, there a hieroglyph, somewhere else, perhaps, an alphabet). On the contrary, 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. We regard it as very interesting, from the point of view of a possible teaching sequence, that the old conceptualizations did not disappear from later orthographies but were typically retained and embedded within them. For example, English orthography continues to represent words (see Section 3), as did primitive orthographies, but it of course represents abstract sound structure as well. In this sense, the primitive orthographic concepts are proper parts of the alphabetic writing system. That is, modern alphabetic writing is reasonably viewed as consisting of a set of layered concepts, only some of which were deployed systematically in primitive writing.

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. The next chapter (Rozin & Gleitman, this volume) provides facts about the learning and use of alphabets that bolster this case considerably. Here we rest content to claim that the evolutionary context and many internal descriptive facts about the historical scripts are suggestive enough that they ought to be looked at, if the concern is to develop a conceptually orderly curriculum for teaching reading.

As we have implied, it misrepresents the historical facts to suppose that successively more analytic writing systems arose discretely, or that each overwhelmed and supplanted its immediate predecessor. To the contrary, orthographies tend to rigidify under the influence of conservative religious, ethnocentric, and possibly conceptual forces. Writing systems are rarely much changed within the society in which they originally emerge (Gelb, 1952). Advance comes primarily when one society gets hold of the system used in another, and improves it in the course of adapting it to the new language. By such successive borrowings with change, new and more efficient orthographies were adopted over time, without hurt nor leave of psychologists and professional educators, culminating in almost world-wide conquest by the Greek alphabet. After we have surveyed the writing systems that have had their day in the historical laboratory, we will describe how modern English writing fits into this picture.

2.1 Semasiography (the Writing of Concepts or Meanings)

It is not logically required that writing be based on the sounds or words of language. Concepts can sometimes be rendered directly without the mediation of language; a general idea (*sememe*), rather than a sequence of words in a spoken language; a general idea (*sememe*), rather than a sequence of words in a sentence, can be expressed. *Semasiography,* in this sense, is attested from rock paintings and inscriptions dating back at least as far as 20,000 B.C. (Diringer, 1962; Jensen, 1969). If we do not demand that writing consist of permanently scratched or painted signs, we can infer a yet earlier beginning that left no trace. Many preliterate societies employ a kind of "object writing" for communicative purposes; Gelb (1952) cites this instance of rudimentary object writing by the Yoruba, a tribe of Nigeria:

> During an attack of a king of Dahomey upon a city of the Yoruba one of the natives was taken captive and, anxious to inform his wife of his plight, sent her a stone, coal, pepper, corn, and a rag, conveying the following message: the stone indicated 'health', meaning 'as the stone is hard, so my body is hardy, strong'; the coal indicated 'gloom', meaning 'as the coal is black, so are my prospects dark and gloomy'; the pepper indicated 'heat', meaning 'as the pepper is hot, so is my mind heated, burning on account of the gloomy prospect'; the corn indicated 'leanness', meaning 'as the corn is dried up by parching, so my body is dried and become lean through the heat of my affliction and suffering'; and, finally, the rag indicated 'worn out', meaning 'as the rag is, so is my cloth cover worn and torn to a rag [Gollmer, 1885; as quoted by Gelb, 1952, p. 5].*

The objects do not convey particular words of the Yoruba language, and have no direct relation to these. Instead, through arousing semantic associations, the objects have the effect of putting across a meaning or concept. The North American Indian bead writing (*wampum*) apparently served a similar function. Yet it is apparent that this must be a sometime device, subject to rampant ambiguity and vagueness of interpretation.

The most widespread of semasiographic devices is the use of pictorial representations to refer to meanings, attested from upper Paleolithic times and in widely dispersed cultures. Since ideas (not the words of language) are directly expressed in these drawings, it is often possible to reconstruct something of the intent of the writer without knowledge of the language. An example is shown in Fig. 1 (from Jensen, 1969), from the Pasiega cave in Northern Spain. Jensen reconstructs, at the left, cave dwellings, at center a pair of footprints, and to the right, possibly a prohibition sign. We might render the idea, then, as *No Trespassing in Our Caves!* but the signs presumably do not stand for a sequence of concrete words ("no–feet–caves") in the protolanguage of the scribe. A further example is shown in Fig. 2 (adapted from Gelb, 1952), which depicts proverbs of the Ewe tribe of Togo. The first proverb, as the caption states, is literally "The thread

*From I. J. Gelb, *A Study of Writing.* Chicago: University of Chicago Press, 1952.

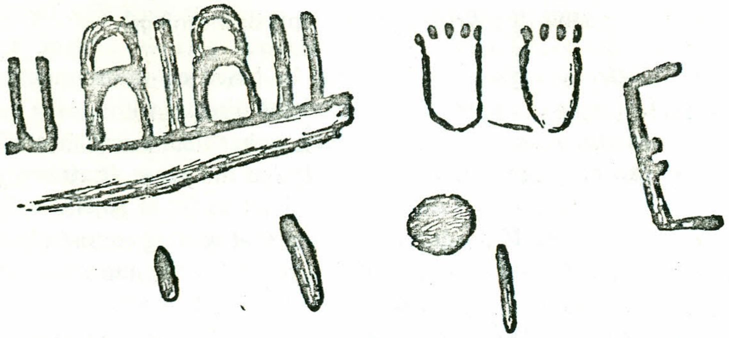


FIG. 1. The first *No Trespassing* sign: semasiographic writing from the Pasiega Cave, Northern Spain. (From Jensen, 1969, p. 40.)

follows the needle," which Gelb interprets as something like our "The apple doesn't fall far from the tree" or a "chip off the old block." The second is "Two opponents cannot last" meaning essentially that one disputant must give way. Notice, particularly in this second instance, that the literal words of the proverb are not rendered: the picture is of two warriors, neither of whom is shown yielding. The picture simply "brings to mind" the previously known proverb, rather than writing its words. Pictures used for these identifying mnemonic purposes may become abstract (*diagrammatic*), as in the circle that represents "the world" in the third panel of Fig. 2. Even so, if a concept as opposed to a word of the language is what is directly expressed, this diagrammatic representation is still properly called a *semasiogram,* as opposed to the *logograms* that we will presently discuss (Gelb, 1952).

So far as is known, semasiography was always severely restricted in use, being employed primarily to record stereotyped, culturally familiar material, such as

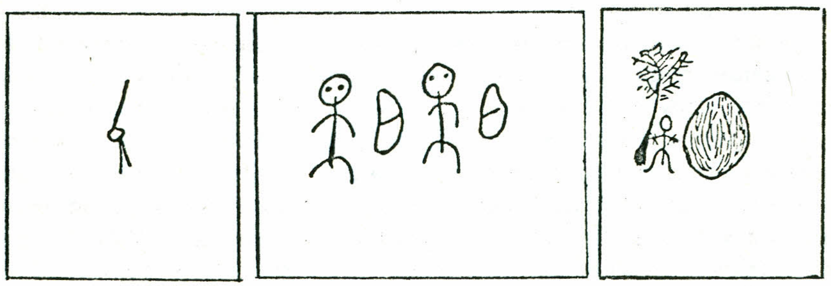


FIG. 2. Semasiographic writing: Proverbs of the Ewe tribe of Togo. The left-hand panel represents the proverb, "The thread follows the needle." The center panel reads, "Two warriors cannot stand" (see text). The right-hand panel reads, "The world is like a baobab tree," that is, so great that it is impossible to embrace it. This is expressed, as Gelb notes, by the picture of a man trying in vain to stretch his arms between the tree and a circle symbolizing the world. (From I. J. Gelb, *A Study of Writing.* Chicago: University of Chicago Press, 1952, p. 49. Copyright © 1952 by The University of Chicago Press.)

ritual celebrations and wise sayings. There is no evidence of an exhaustive system of this type—one that can render all items in the conceptual or linguistic domain. This is not surprising, for the variety of concepts available even in stone age societies must be very large, and then the number of diagrammatic indicators needed for unpicturable items would ultimately be crushing.

The historical decision to abandon semasiographies and to base writing on the words and sounds of the spoken language represents an implicit recognition that within human language infinite expressiveness is derived from finite means. Each subsequent system used a language unit with fewer elements than the last. All observed change is in this direction.

2.2 Logography (the Writing of Words or Morphemes)

The use of written signs to express the words of the spoken language is a device that cropped up independently in a large number of cultures, for example, the Mayan, the Egyptian, and the Chinese. The oldest writings we have evidence for appear on vessels of the First Dynasty of Egypt—about the fourth millenium B. C. The signs appear to represent root forms of substantives and verbs (so-called *morphemes:* roughly, minimal meaning-bearing elements of the language), the "grammatical" terms being ignored. Thus, as an example from English, the item *wanted* consists of two morphemes (*want + ed;* that is, verb + tense marker); in a primitive script, the sentence *I wanted to go* would be represented as *I want go,* the tense-marking and mood-marking elements of the verb being ignored.

The physical shapes of early logograms relate in obvious ways to the embryonic ideas of semasiography; chiefly, the signs derive from four simple bases (Gelb, 1952):

Pictograms: Concrete actions and objects are represented by pictures.

Diagrams: Arbitrary geometric signs are created for unpicturable words.

Ideograms: Semantic associates of pictograms are represented by the same sign; for example, if a picture of the sun originally signified *sun,* then the same sign might be used to express *bright* and *day,* with the results that less easily picturable words can be represented and that excessive proliferation of the set of signs is avoided; however, the signs consequently become ambiguous with respect to the set of words. It is important to realize that the logographic sign is meant to express a single particular word, not the general idea that subsumes *bright, sun,* and *day.* The ancient scribe often went to a good deal of trouble to disambiguate his use of these ideograms, using semantic complements as described below, and also contextual clues when these were available (Gelb, 1952).

Semantic complements: Additional, *unpronounced,* signs are appended to word signs to signify the superordinate semantic category to which the sign (in some particular occurrence) refers, especially where ambiguity arises among ideograms. An example from Gelb (1952) is the cuneiform writing of *Aššur,* the name both of a city and of a deity. A further sign, either 'deity' or 'city', is customarily appended to the *Aššur* sign; thus, $^{\text{deity}}$Aššur, or else $^{\text{city}}$*Aššur*. In either case, the sign is to be read "Aššur," (not "Aššur city" or "the God Aššur") but now the reader also knows the sense which is intended. In many ancient orthographies, the unpronounced complement came to be inserted habitually at each occurrence of the appropriate semantic category, even if no ambiguity arose.

Logography appears to be a useful kind of writing. Indeed it is sufficiently useful that one of the world's major languages, Chinese, is written essentially in logographic form today (see Fig. 3 and discussion in Section 2.3). While there are a great many words, and thus a great many signs to be memorized separately in learning a logography, yet the set of meaningful words is finite and changes its membership quite slowly. At least on a commonsensical view, there are many more concepts than words, for new concepts can generally be expressed by combinations of old words. In this sense, word writing seems to be more efficient than idea writing. However, this argument collapses when the case of new proper names is considered: these are incorporated frequently and freely into language usage, and are unpicturable. Since proper names are a good deal of

父 買 黑 車

哥 哥 說 母 用 白 書

好 哥 哥 不 給 人 紅 車

FIG. 3. Logographic writing: a few sentences in modern Chinese. These three sentences were among those read by second-grade subjects in an experiment by Rozin, Poritsky, and Sotsky (1971). Reading across from left to right, these sentences can be translated as: *top:* father buys (a) black car; *middle:* older-brother says mother uses (the) white book; *bottom:* good older-brother (would) not give (the) man (a) red car. Note the one-to-one mapping of English words to unitary Chinese characters (words in the translation that are not directly represented in the Chinese characters are included in parentheses). The principle of a phonetic complement is illustrated by the sign for *red*, the penultimate character on the bottom line. This character is a compound of two characters: the element on the left signifies *silk*, and the element on the right, which looks like I, is pronounced *ung*. The pronunciation of the word *red* in Mandarin Chinese is *hung*. Thus, *ung* is a phonetic clue for the proper reading of this character.

what the ancients wanted to write (in the commercial early uses of writing, ownership was a major topic), the resources of logography were insufficient. The stage was set for a further advance.

2.3 Phonography (the Writing of Sounds)

A description of orthographic history would be materially simplified if there had been a gradual and uniform development, in each society, from the simple logographic origins. However, the fact is that in some cultures, development proceeds no further (for example, the Mayan and Aztec scripts reveal only a few marginal instances of sound writing). For the case of Chinese, development proceeds to the use of "hints" for pronunciation, as we shall describe. On the other hand, the widespread use of phonological (sound-stream) devices dates from very early records in the case of Egyptian and other Near and Middle Eastern scripts. It is true that the earliest preserved writings are semasiographic and that the latest-invented writing is alphabetic; it is also true that no society moves backward from, say, alphabet to logography. But it is not clear, since the origins of writing are often shrouded, that each culture begins with the most primitive kind of writing and then proceeds unerringly to analytic, sound-stream scripts.

We can clearly observe, however, an orderly evolution of just this sort in the forms of the signs themselves; these become more extensive and highly conventionalized within a culture, with the passage of time. Figure 4 (from Gelb, 1952) shows early and later forms of some cuneiform signs, revealing the loss of pictorial clarity over time. The loss of the pictorial mnemonic, and the need to express unpicturable proper names, no doubt are related to the increasing tendency to represent sound properties of the language in the script. Two main devices developed to supplement the logographies with phonological information:

a. Phonetic complements. A phonetic complement is the sound-stream analogue of the semantic complement we previously discussed. Recall that the semantic complement gave a clue to a semantic range, such as "name of a city" or "name of a kind of cereal." Like the semantic complement, the phonetic conplement is a sign appended to the logogram, and is unpronounced: it gives a clue to the pronunciation of the logogram. For example, if English were written this way, the logogram for *river* would be accompanied by an additional sign interpretable as "sounds like *rover*" or, perhaps, "sounds like *liver*."

The large majority of the 50,000 or so Chinese characters are *phonetic compound signs* formed of a semantic complement (an ideogram, often called a *radical*) indicating the general semantic field to which the intended word belongs; and a phonetic complement (another ideogram, of a word similar in sound to the intended word). The pair, taken together, give—at least theoretically—sufficient information to isolate a particular spoken word of the

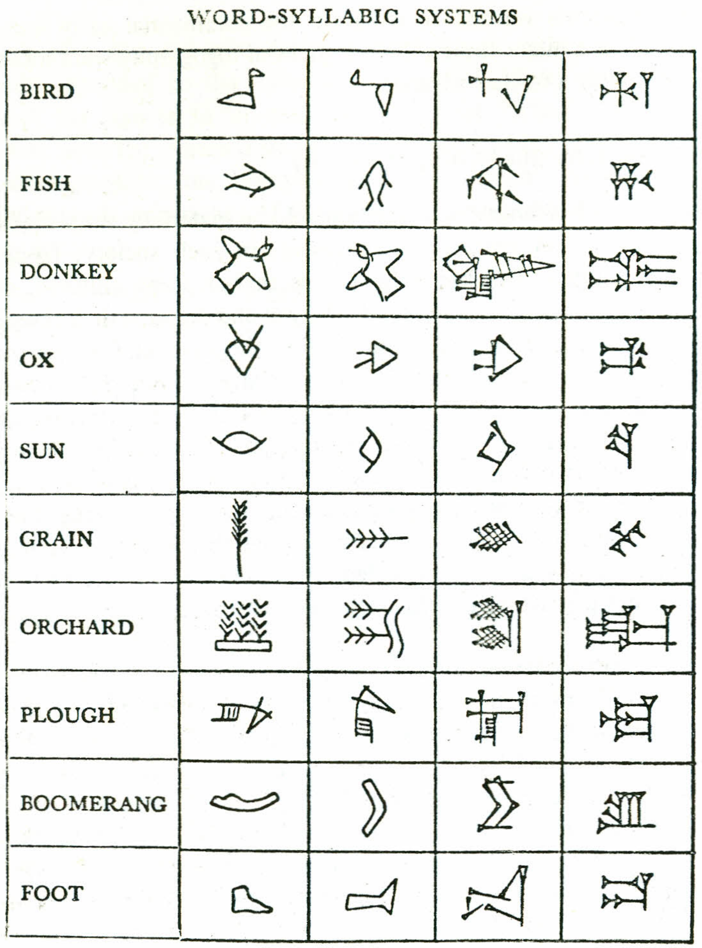
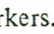


FIG. 4. Pictorial origins of ten cuneiform signs. (From I. J. Gelb, *A Study of Writing:* Chicago: University of Chicago Press, 1952. Copyright © 1952 by The University of Chicago Press.)

language; hence, the whole two-part character is a logogram. Thus, for example, a Chinese style logogram for English *river* would read, literally, "a body of water that sounds like *liver,*" while *creek* would read "a body of water that sounds like *crack.*" On the face of it, it would seem that one would have to decipher rather than read novel items in a script of this kind. Not only are there many spoken words to search, given the superordinate "body of water," but the phonetic clue may be a rhyme, a relative with one vowel difference, etc. (see Fig. 3, bottom line and caption, for an example from modern Chinese). Matters are further complicated by the fact that the semantic and phonological elements are not spatially ordered in the compound sign. Worse, the phonetic and semantic portions may physically merge and become unanalyzable as the sign is con-

ventionalized and simplified over time (Martin, 1972; Chao, 1968). Even more confusing: since there is obviously a choice among items to use as the phonetic clue, the Chinese prefer to use a phonetic relative that also has some semantic relation to the meaning which is expressed by the radical. Thus a decoding act for a previously unlearned item would take on difficulties akin to those of word puzzles in which part of the problem is to decide which segment of the clue is definition, which an anagram on the solution. In practice, it seems that while the semantic and phonological components of a Chinese phonetic compound sign are worth telling the learner as a mnemonic for memorization, one reads Chinese logographically, as though each whole character was an indivisible symbol for a word in the spoken language (Leong, 1973; Chao, 1968).

The major conceptual fact about phonetic complements is that they are not directly "writings of sounds." They do include information about sounds, but merely as unpronounced clues. The Chinese use a further phonographic device, the rebus, mainly to render foreign loan words; but ordinarily they use the logographic characters. The Near and Middle Eastern scripts, however, increasingly incorporated the rebus principle, a concept which is much closer to modern sound writing than is the phonetic complement.

b. The rebus. The rebus concept is the true precursor of sound writing: essentially the notion is that two words that sound the same can be written with the same sign, regardless of meaning differences. Once this idea was conceived, it was theoretically no longer necessary to invent an arbitrary diagrammatic sign for each new unpicturable word; an already available picture sign, related in sound, could represent it. For example, *I* can be spelled  , *deer* can be spelled  , and then *idea* can be spelled  +  , at least for New Yorkers. Now the sound is rendered directly, and the meaning only by mediation of the sound. Notice that while the phonetic complement similarly indicates a rough sound value, it does so as an *additive* to a meaning sign. It does not, like the rebus, insert the sound value *in place of* the meaning value. Two main types of sound writing (phonography) developed from the logographic systems by development and extension of the rebus principle: *syllabary* and *alphabet.*

2.3.1 Syllabaries

Every word in every language can be written as an integral sequence of syllables; roughly, a syllable is a single vowel or diphthong, which may be preceded and/or followed by one or more consonants (see Section 4 for discussion of the syllable unit). Writing notations based on the syllable unit have appeared recurrently, by independent invention, and in different parts of the world; in Africa, North America (among the Cherokee, see Walker, 1969), the Far and Middle East, and in the Aegean. For the case of the Near and Middle Eastern scripts, where the surviving historical record is fairly rich and covers a great time depth, it is clear that the syllable signs evolved immediately from word signs and the rebus idea;

but now a phonographic concept becomes explicit: the syllable signs no longer represent morphemes or words directly, but solely phonological entities.

Suppose, for example, that a number of spoken words include an identically pronounced subpart, such as English "man," "manuscript," "Manhattan," "emancipate." Each of these would require a different logogram, since they are distinct words and since the overlapping portion "man" does not represent the same elemental idea in all these words. The sign for one of them, say,  = *man*, may be designated as the sign uniformly representing that pronunciation, regardless of the meaning of the word in which it is used. For example, "mandate" can be written:

man + *date*

If the script uniformly embodies this concept, the sign  is no longer a logogram, even when used to designate the word *man*. It represents this meaning through mediation of the sound properties. When a script includes a codified set of signs of this phonographic type, and when the unit size of each is the syllable, the system itself is called a syllabary.

Discovery of this syllabic–phonological principle did not necessarily or entirely lead to the dismantling of the logography that preceded it. Table 1 (from Gelb, 1952) compares the number of syllable signs and logograms coexisting in various scripts. Generally it is the foreign adapters of scripts, not the native users, who dispose of some of the residue from past systems. Logographic and syllabic systems may coexist for many centuries (Japanese writing still includes several thousand logograms, beside a full syllabary). The easy coexistence of varying conceptual principles within a script is one more indication of the relative lack of analyticity with which writing systems are invented and apprehended.

Since the number of syllables is for many languages vastly smaller than the number of differently pronounced words, adoption of a syllabic system can effect great economy of signs. However, this efficiency varies for different languages. Insofar as the language is largely monosyllabic (as Classical Chinese: Chao, 1968), the number of syllables approaches the number of words, and a

TABLE 1
Relationship of Word Signs to Syllabic Signs in Word–Syllabic Writings

	Total number of signs	Syllabic signs
Sumerian	about 600	about 100–150
Egyptian	about 700	about 100
Hittite	about 450 +	about 60
Chinese	about 50,000	[62 in fan-ch'ieh].

syllabary yields no advantage over a logography.[2] The efficiency of syllabaries depends also on internal features of the individual syllables in the language. Ideally suited to a syllabary are those languages which allow only a few syllable types, either V (vowel alone) or CV (consonant + vowel), and nothing else. Japanese is such a language, and it can be represented adequately with a vocabulary of about 50 syllabic signs (see Fig. 5). English, at the other extreme, allows certain consonant clusters to a limit of four in syllable-final position (as in *lengths*) and also allows consonants in both syllable-initial and syllable-final position. Thus, among others, there are syllable types CCV, CCVC, CCVCC, CCCV, CCCVCCCC; examples are *spy, spied, spend, stray, strengths.*[3] In English, then, there are several thousand different spoken syllables. Moreover, since consonants occur at both the end and beginning of English syllables, the problem of segmenting polysyllabic words is theoretically and practically difficult, if not arbitrary (is it *fa-ster, fas-ter,* or *fast-er*?). Not surprisingly, syllabaries were invented and adopted within societies whose language had a simple syllable structure. These syllabic forms of writing have sometimes persisted to this day (as in Japanese), for an alphabet will hardly reduce the number of symbols in the script for a language with a highly constrained syllable stock.

A true syllabary, if conceptually consistent, has a separate sign whenever there is either a different consonant within the syllable (e.g., a sign for "pa" and a sign for "ba") or a different vowel (e.g., a sign for "ba" and a sign for "be"). But in practice the ancient syllabaries did not make all these distinctions.

For example, a single sign might be used for two syllables that are identical except for the voicing (vocal cord vibration) of a single consonant. Thus, the sign might represent both the syllable "nōz" and the syllable "nōs." And then  +  might be a fit writing for "nostrum" (the unvoiced S is rendered by voiced Z; the unvoiced T is rendered by voiced D). A more important smudging of distinctions among syllables, for the subsequent development of writing, was the use of a single sign to represent all syllables sharing the same consonant, even

[2] As noted earlier, a syllabic notation is used in Chinese under conditions where it would be most effective: to render polysyllabic loan words. The Japanese similarly use their *Katakana* syllabary to transcribe Western loans. For example, Chao reports that *Katakana* syllables are used for the sequence *chi-ya-i-na-ta-u-n* on the sign of the "Chinatown" nightclub in Kyoto. It is worth noting that modern Chinese, unlike classical Chinese, cannot be characterized as simply monosyllabic in morpheme or word structure (Chao, 1968). It is also worth mentioning, with only the most timid implication of cause and effect, that the Chinese are now moving toward adoption of a phonographic script.

[3] Of course the *th* in *lengths* and in *strengths* transcribes a single phoneme, not a sequence of two phonemes. English writing uses a number of such digraphs, such as *th, sh, ph.* There are various historical reasons for this, discussion of which would carry us too far afield; it is also true that such writings are not descriptively irrelevant to the current phonological system (notice that all the sounds rendered by these digraphs with H share a feature of aspiration), but again we must ignore this detail. It is important here simply to realize that although five-consonant letter sequences appear in English words, these refer maximally to four-consonant spoken sequences.

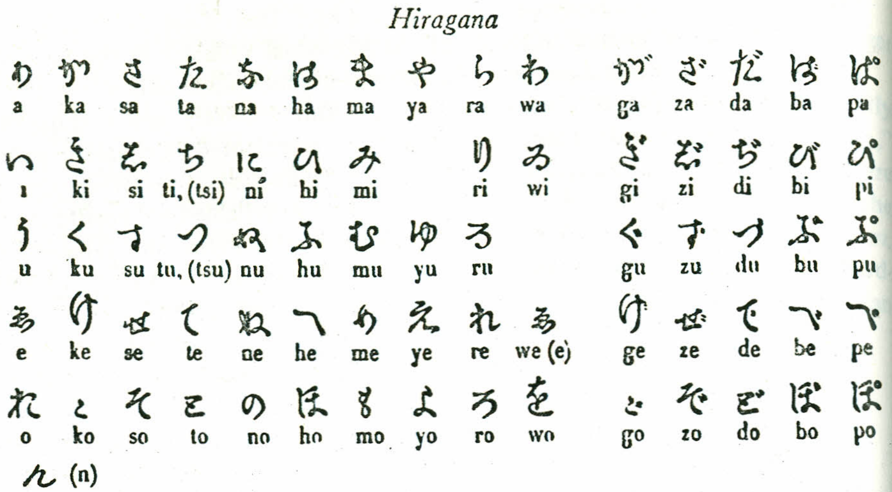


FIG. 5. The Japanese Hiragana syllabary. This syllabic system has a separate symbol for each vowel-alone syllable (the first column) and for each consonant + vowel syllable. Since no other syllabic types appear in the spoken Japanese language, it can be transcribed in terms of this small set of syllabic symbols. But notice that the fact that, e.g., *ka, ki,* and *ku* or *ka, sa,* and *ta* are phonologically related in the spoken language is not acknowledged in this script; from the symbols themselves one can glean only that there are 74 phonological entities, apparently unrelated to each other. (From I. J. Gelb, *A Study of Writing.* Chicago: University of Chicago Press, 1952, p. 161. Copyright © 1952 by The University of Chicago Press.)

though the vowel differed (for example, the Hebrew character *Beth* would represent, indifferently, "be," "bi," "bo," "bu," "ba").

In a way, representations of this sort appear (in hindsight) to *be* the alphabetic principle, lacking only separate signs for the vowels. Certainly the decision to write all consonant-identical syllables with the same sign is a major step toward an alphabet. Notice that the Japanese syllabary (Fig. 5), which did not evolve toward an alphabet, represents consonant-identical syllables differently; that is, five separate signs correspond to "ba," "be," "bi," "bo," "bu" in the Hiragana notation. On the contrary, one Hebrew sign stands for all five. In short, the first consonant of the name of the syllable (e.g., *Beth*) comes to be used as the sound value (e.g., "B") that sign takes in the script. This *acronymic* principle is essentially the same as the one that yields spoken English "radar" from the phrase "*ra*dio *d*etecting *a*nd *r*anging"). But it is important to state that this move toward the idea of phonemes (letter sounds) did not emerge self-consciously, as an analytic notion that was then readily generalized. On the contrary, it seems that the users of early syllabic scripts were unclear about whether the symbols of the script represented syllables or just the consonants of those syllables, and their usage was inconsistent in this respect (Chao, 1968).

The vagueness of the ancient writers concerning the alphabetic notions already manifest in their orthography is more understandable when considered in terms

of the Semitic language which supplied the context for these writing advances. In many Semitic languages, consonants play a primary role in distinguishing the root forms, while vowels often render grammatical functions. As stated earlier, the ancient writers usually did not transcribe the grammatical functions such as tense markers, plural markers, and the like. Then if vowels render grammatical functions in the spoken Semitic language, and grammatical functions are not rendered in writing, vowels will tend not to be written or, perhaps, even noticed. The idea of a vowel within a syllable seems not to have been recognized explicitly, however odd this seems in retrospect. In essence, the Semitic writer tolerated a single sign for words with different vowels (as an example from English, a single writing would be possible for *goose* and *geese*) but not for words with different consonants (for instance, English *grief* and *grieve* would have to be written differently).[4]

Even though the insufficiency of vowel representation in Semitic syllabaries seems acceptable in terms of the language structure, it did yield some ambiguity of interpretation, and did so much more radically when the system was borrowed through cultural diffusion by the Greeks. This problem was relevant to the evolution of the syllabaries into an alphabet.

2.3.2 The Alphabet

To counter the ambiguity of consonant-syllabic representation, the practice of appending vocalic phonetic complements (see Section 2.3) to these signs gradually developed. At first, this added vowel mark was just a clue to the pronunciation of the ambiguous syllable, rather than an explicit writing of vowel letters. In this sense, the writing looks essentially alphabetic; only the "idea" of an alphabet is missing. As an example of what was probably going on conceptually at this stage, consider a hypothetical syllabary which has signs for whole CV syllables–but the sign is the same, regardless of the vowel in that syllable, and signs for vowel-alone syllables (*a, i,* etc.). Then to disambiguate a written syllable, one of the vowel syllables can be appended to it as a clue. Thus suppose we wanted to render the hypothetical polysyllabic spoken word "okoki" in this syllabary; assume we have the syllable signs:

[4]If these spelling notions seem odd or undecipherable, it may help to notice that conventional English spelling frequently, though inconsistently, makes similar compromise with the representation of sound, a problem to which we will return at length (Section 3). Yet there is little problem of decipherment. Consider the word *read* in English, which is pronounced like "reed" (in a present-tense interpretation) and like "red" (in a past-tense interpretation). The writing here is reminiscent of the Semitic. The root meaning of *read* ("apprehend language from a codified visual representation . . .") is represented by the consonants, that is, *r–d*. The grammatical form (tense) is rendered by a particular vowel change in speech, but the vowel alternation is unmarked in the print. Another case is the single writing *house* for the meaning "dwelling," rendered in speech with a final S sound, and for the meaning "to provide a dwelling for," rendered in speech with a final Z sound; again, grammatical function (noun versus verb) is here ignored in the script, with little consequent confusion.

# = any of the syllables *ko, ki, ka, ke, ku*
o = the syllable consisting of the vowel O alone
i = the syllable consisting of the vowel I alone.

Then we could write o – # – #, but this could be read *okoki, okiki, okiko, okuka*, and so on. By using the vowel syllabics in an additional role, we can write unambiguously: o – # – o – # –i or, quite literally, o ("the syllable O") – # –o ("I mean the # with an O-sound") – # – i ("I mean the # with an I-sound"). But if this vocalic phonetic complement is used at every instance, then it can be interpreted as providing a redundant transcription of the vowel. That is, if the syllable *ko* is written #, and the syllable *o* is written *o*, then the *o* in the # = *ko* must be a vowel, O, and the residue, K, is a consonant. Now reinterpreting # as a consonant sign, we derive the alphabetic o#o#i, *okoki*.

Was the "invention" of the alphabet from a transitional consonant-syllabic script really an analytic step of this sort, a conscious insight? Apparently it is better described as a quasi-artistic development and extension. Bolstering this view is a reconstruction of the way the Greeks came to derive symbols for vowel letters from the syllabary they borrowed from the Phoenicians. The "discovery" of the vowel letters (possibly the crowning post-Phoenician step in the history of the alphabet) was really something of an accident. Because of a limitation on the Greek phonological resources, they simply misheard the names of the Phoenician signs for syllables beginning with laryngal consonants (consonants produced by constricting the larynx): failing to perceive the initial consonantal sound, they heard these syllables as vowels alone. (The nature of such systematic "mishearings" between two linguistic communities is discussed in Section 3; see particularly Footnote 9). Since, by the acronymic principle, the initial sound of a syllable was taken as its sound value in the script, the vowels as well as the consonants could now be symbolized directly within CV syllables (H. M. Hoenigswald, personal communication; Jeffery, 1961).

A full alphabet, marking vowel as well as consonant phonemes, developed over a period of about 200 years during the first millenium B.C. in Greece (Kroeber, 1948). It did not arise whole cloth, as a single insight with fully generalizable consequences like, say, the idea of the internal combustion engine. But once the alphabet was designed, it spread across the ancient world by direct borrowing and adaptation with surprising rapidity. It was never reinvented in a separate cultural development. Our own alphabet is a descendant from this same source.

The fuzziness of the inferential line from syllable to alphabet may explain why it was invented only once. Perhaps the obscurity of the alphabetic principle is relevant, as well, to the question of why it is hard for many six-year-olds to acquire English reading. The consonantal segments of syllables are hard to analyze out of the syllabic context in which they (must) always be embedded (for a discussion of the speech-production and perception facts here, see Section 4). Whether individual users of modern alphabets must rediscover and extract the subsyllabic (phonemic) unit in order to achieve fluency is a question we take up elsewhere (Gleitman & Rozin, 1973b; Rozin & Gleitman, this volume).

2.4 Summary

The history of writing, as we have sketched it, involves the use of successively more analytic units of language to correspond to the squiggles on the page (or rock). Change has always been in this direction: from the representation of ideas or meanings, to the representation of words and morphemes, thence to syllables, and then to the yet smaller alphabetic units of vowels and consonants (so far undefined) popularly called phonemes. No revisionists are recorded in history. Nostalgia for bygone scripts and more global units has been absent from the history of writing until the most modern times when professional educators and psychologists—faced with the difficulty of achieving mass literacy—in effect began to resurrect the past. True, no one has come along who wishes literally to impress wedge-shaped marks in clay or paint hieroglyphs on papyrus. Except for a few aberrant cases, revisionists are content to leave The Library of Congress intact. Yet they often argue that the concepts underlying the letter forms can be ignored and that the method for learning and using English writing look back to more halcyon times. There are serious advocates of a return to paleolithic semasiographies, those who suggest that children should—and that adults literally do—learn to read meanings directly off the printed page (Smith, 1971; Goodman, 1969). Another group proposes that children should learn the whole words embodied in Chinese and early Egyptian hieroglyph for, purportedly, fluent adults read whole words (Huey, 1908). We ourselves have suggested a crucial role for the syllable in learning to read English orthography (Gleitman & Rozin, 1973a). In addition, the most diverse theoretical opponents claim to be champions of the alphabet (Bloomfield, 1933; C. Chomsky, 1970).

We believe that the alphabet is a useful invention, and that learning its principles is an essential component of English-language literacy. The alphabet is a highly analytic representation of language that took some two millenia to work out from word–syllabic origins. We cannot give the novice reader this leisure, but we believe that an encapsulation of the evolution of scripts will help the learner achieve the ultimate alphabetic insights. This is because all the conceptions and units of earlier orthographies are proper subparts of the conceptions and units embodied in the alphabet.

In light of this declaration, it is of some interest to ask how modern English writing fits into the history of writing. After all, it does seem at first glance that English and other European orthographies represent a certain backsliding from the alphabetic principle. Certainly that is the view of the persistent reformers of *alfabetic notashun* such as G. B. Shaw (1965) who would return to verities muddied in our script. This position misinterprets the ways alphabets could represent language, a problem to which we now turn.

3 MODERN ENGLISH ORTHOGRAPHY

It is well known that English orthography poses serious difficulties for some would-be learners. Often, it is supposed that the major problem lies in the irregularity of English spelling; that is, in its partially nonalphabetic character. *Psychology* begins with an S sound, but it is spelled with a P. The items *cough, through, though, trough,* and *tough* seem to show only the most haphazard spelling-to-sound correspondences. Although there is some dispute about the extent to which there is a rule-governed basis for much of this irregularity (N. Chomsky, 1970; Francis, 1970), we believe this issue is peripheral to the initial problem of learning to read. The major difficulty of the poor reader is not the inability to read *cough* or similar monstrosities. The poor reader has trouble reading *off* long before the case of *cough* arises so the irregularity of English spelling cannot fully explain why it is hard to learn to read English.

We will show that the major initial difficulty in reading is posed by the regular alphabetic principle itself: the relation between spoken "mat" and the spelling M-A-T (a conspicuously regular case) is highly abstract. The spelling system is related to sound only through the mediation of complex encodings that take place at various stages in the linguistic process:

1. there is a complex relation between the *phoneme* (a cognitive–perceptual category) and the *phone* (a perceptual category), and
2. there is a complex relation between the *phone* and the *acoustic stimulus.*

The preliterate individual must have some access to the nature of these encodings if he is to construct an interpretation of how alphabetic symbols relate to the spoken language. Unless he does this, he will not learn to read.[5]

3.1 From Phoneme to Phone

The alphabetic principle embodies a strong claim about language organization: that perceived speech consists of temporally linear sequences of minimal segments of sound, smaller than a syllable. Thus "pat" consists of three segments; "tap" and "apt" are rearrangements of the same three segments. It will be shown later on (Section 4) that there is no simple discrete analogue of these units of spelling in the acoustic wave; nevertheless, perception seems to yield up units of about this size. Let us call these units *phonemes.* If the phoneme can be defined for the spoken language, then alphabetic notation (in some hypothetical pure form) can be understood as a simple cipher on this unit: a squiggle is assigned to each phoneme. In brief, the perceived sound "puh" would be written P, "duh" would be written D, and so on.

[5] Many excellent discussions of the relations of generative phonology to hypothetical and real writing systems have appeared in recent years; the reader is referred particularly to Klima (1972), Venezky (1970), C. Chomsky (1970), N. Chomsky (1970), and Smith (1971).

Discussed below are attempts to define and enumerate the set of phonemes as perceptual classes that have simple acoustic correlates (Section 3.1.1), as in the formulations of the descriptive linguists (see especially Bloch, 1941; Bloomfield, 1933; Chao, 1934, and Harris, 1951). In discussing the difficulties with this formulation, we follow an argument by N. Chomsky (1964). Of necessity, the version here is much abbreviated and many technical issues are left unstated. Next, we sketch the alternative generative–transformational formulation (N. Chomsky & Halle, 1968), in which the phoneme is defined as an abstract perceptual–cognitive category (Section 3.1.2).

Psychologists and educators approaching this material, which is tangled and highly inferential, may wish to throw in the towel in favor of a rough-and-ready alphabetic unit, on the grounds that Phoenicians or Greeks and seven-year-old readers of English required no great technicalities in inventing the alphabet and learning to read. But without a clear definition of the basic phonemic unit, one must give up hope of accounting for much of the broad range of behavioral language phenomena. A unit approximating the phoneme is the basis of the rhyme of "bat" and "cat" and the alliteration of "bat" and "bag." A unit of this approximate scope is involved in the account of tongue twisters such as "Peter Piper picked a peck of pickled peppers. . ." A unit of about phoneme size disappears from speech at the time in linguistic history that *knight, knee,* and *knickers* came to be pronounced "nite," "nee," and "nikkers." Substitutions of sound in young children's speech ("I wiv at Six, Wantern Wane . . .") are at this same approximate level. Finally, the phoneme is the unit selected by many to represent language visually.

The description of phonemes as relevant to the discussion of reading an alphabet proceeds in two steps:

1. the definition of a phoneme is shown to require information from the morphemic (word-structure) level above and the phonetic (sound-structure) level below; and
2. many regular conventions of everyday English spelling, as well as a large set of apparently irregular spelling conventions, are shown to follow once the phoneme is defined as a cognitive category.

3.1.1 Is the Phoneme a Simple Perceptual Category?

How shall we classify the many sounds that occur in the course of speaking? The reasonable first assumption made by the descriptive linguists was that the sounds of language could and should be described quite independently of meaning, syntax, or any other global level of linguistic representation:

Assumption (1) The *phoneme* is autonomous in the sense that it is a unit of the sound structure of language; information about the relation of sound differences to meaning or syntax differences is for this purpose beside the point.

The next obvious assumption is that the phoneme must be defined as a perceptual class or set of sounds; that is, the linguists agreed to disregard all differences among speech segments that are not discriminated by the hearer, for obviously such differences play no role in the use of language. There is much variation of pronunciation in real speech, both within and across individual speakers, and yet mutual intelligibility is maintained. Clearly many little variations go unnoticed, either because the ear cannot distinguish them or because after suitable training (learning the language) the individual comes to disregard them. The separate speech sounds are termed *phones.* Each collection of phones, all of which are regarded for purposes of language use as "the same" by speakers of that language, is called a *phoneme.*[6] The phonemes are the exhaustive set of *perceived* sound distinctions in a language:

Assumption (2) Two phones are regarded as belonging to two separate phonemes if and only if the speakers of the language in question habitually make and hear a distinction between them in the context of speech.

A "good" alphabet will render phonemes rather than phones, for the differences among phones of the same phoneme are irrelevant to meaning distinctions.

The classical phonemicists developed an apparently straightforward method for isolating the phonemes of the language: find words that differ only by having a single perceived difference in one of the putative segments, *a minimal pair.* For example, the pair *bit/pit* establishes the phonemes B/P under Assumption (2).[7] This assumption and its associated minimal pair test turns out to have considerable power in isolating a perceptual unit very like the one usually symbolized in our alphabet. Some apparently difficult cases turn out to have manageable explanations under this scheme.

[6] The set of phonemes so defined varies somewhat from language to language, and even from dialect to dialect. Speakers of one dialect find it difficult to hear and pronounce distinctions readily made in another. For example, many New Yorkers have a pronunciation of voiced *th* in "these" and "those" that sounds to other English speakers exactly like D ("dese" and "dose"). Yet the New Yorkers in question reliably distinguish between their own pronunciations of "those" and "doze" ("sleep"). That is, some sounds that New Yorkers discriminate among (assign to different phonemes) are variants of a single phoneme (that is, are phones of a single phoneme) in other dialects of the English language.

[7] This is only an informal test, in the practice of the linguists. Whether it can be converted to a psychophysical procedure is not really clear. At any rate, even in its own terms, the test must be refined considerably. Subject reactions tell us only that *pit/bit* differ, but do not tell us exactly where this difference resides. Presumably, a number of further pairs must be compared (*bit/pit; pit/pat; bit/bid*) before we can converge on the initial consonant segment as the source of the difference judgment. For a detailed description of this method, see Harris (1951).

An example of the usefulness of this analysis, often discussed in the linguistic literature, compares the t sound in a word like *grate* (to scrape harshly") with the t sound in *grater* ("one who scrapes harshly"). With a bit of coaxing, speakers of most dialects of American English can be brought to the realization that these two sounds are different. The articulation of "t" in *grate* involves, roughly, placing the tip of the tongue on the ridge behind the teeth, thus momentarily stopping the air column on its way out of the mouth (hence, "t" is called a *stop* consonant).[8] In the articulation of "t" in *grater,* the tongue is allowed to flap against the ridge, and the consequent sound is perceptibly different from the "t" in *grate* (let us symbolize this flapping noise as ϕ). Since American speakers can be cajoled into noticing the t/ϕ distinction, and since they reliably produce these sounds during speech, Assumptions (1) and (2) might seem to require that we recognize t and ϕ as representing two separate phonemes.

This result would be disappointing on two grounds: (a) the speaker (and reader) of English still has the strong feeling that t and ϕ sound very much alike, and (b) the word *grater* seems to be simply the word *grate* with *er* ("one who") added at the end—how much less satisfying to say that the word *grater* is the word *grate* with the t removed, a ϕ inserted for the t, and *er* added at the end. The linguists' account already has enough definitional apparatus to maintain the simpler description; luckily, the minimal-pair test cannot be carried out in the case of t and ϕ. No two words of English differ solely in having t or ϕ in the same place. That is, the environments (of surrounding phones) in which these sounds occur are always different; in the usual linguistic terminology, the relative distributions of t and ϕ are entirely *complementary.* Restating this fact:

Assumption (3) Two phones, even if reliably discriminable by speakers of the language, are assigned to the same phoneme if they are in complementary distribution (i.e., if there is no minimal pair involving these two phones in the language).

A direct parallel to the t/ϕ case exists for the voiced consonant "d." The "d" sound in *grade* ("a level or degree") and the "d" sound in *grader* ("one who puts things in levels or degrees") differ; The "d" of *grade* is a stop, and the "d" of *grader* is a flap (which we will symbolize as ϕ'), but there is no minimal pair of English words differing solely in this respect. Thus d and its flap ϕ' are assigned

[8] The pronunciation of the consonants (under a partial fiction useful at this point in the discussion) is accomplished by making several articulatory gestures in parallel. If the gestures include one that fully stops the air column, the sound is called a *stop* (e.g., P, T, B, G). If the vocal cords are allowed to vibrate rhythmically, the sound is said to be *voiced* (e.g., B, D, G, Z); if this vibrating is delayed, the sound is said to be *unvoiced* (e.g., P, T, K, S). A third dimension is the place of articulation (closure): T and D are *dental* (tip of the tongue behind front teeth), B and P are *bilabial* (the two lips close), and so on. Thus T is an unvoiced dental stop; D is a voiced dental stop; and B is a voiced bilabial stop.

to a single D phoneme. By this quite unexceptional route, linguists arrived at a simple description of the relations between *grate/grater* and grade/grader, under this analysis:

Phonemes	Component phones
T	t, ϕ
D	d, ϕ'

Unfortunately, this analysis is about to collapse. Could it be that the flaps of *grater* and *grader* are physically identical to each other (that $\phi = \phi'$)? In fact, for most dialects of American English, the medial consonants in these two words are indistinguishable, while their initial vowels differ.[9] However, the definitions of the phoneme so far accepted will not allow a statement of the facts in this simple way, for in what sense could the "same sound" refer to two different phonemes? That is, the same sound (the phone $\phi = \phi'$) cannot be assigned both to the phoneme D and the phoneme T, for this would violate Assumption (2). Neither can we claim that the phone $\phi = \phi'$ is an instance of a new and separate phoneme $\emptyset$, for the phone $\phi = \phi'$ is in complementary distribution with both d and t: no spoken English word differs from "graϕer" solely by having d or t where this word has ϕ. Then by Assumption (3), $\phi = \phi'$ must be assigned either to D or T, but by (2), it cannot be assigned to both at once. This complication generated extensive discussion within classical phonemics: how is the choice to be made? Agreeing that there is but one ϕ, let us arbitrarily assign it to D, and

[9] Some readers will have great difficulty agreeing (1) that the flap sounds in *grater/grader* are identical and (2) that the initial vowels in these words differ in duration. Though these are the physical (phonetic) facts, they will turn out not to be the phonemic (cognitive–perceptual) facts, as is the point of this whole discussion. Evidently, one systematically "hears" the sounds in words on the basis of prior language knowledge rather than simply, in terms of sheer physical distinction. We have remarked already on this fact (see Footnote 6). This feature of language organization is massively demonstrated in the systematic misunderstandings among speakers of different languages and of different dialects of the same language. The resulting confusion is the major vehicle for language change in a nonhomogeneous linguistic community (see, for discussion, Bloomfield, 1933; Hoenigswald, 1960). Shaw's comedy, *Pygmalion,* turns on this problem. While Eliza Doolittle believes she is distinguishing between initial H (in, e.g., "Hartford") and initial vowel (in, e.g., "owl")—and perhaps she is—the trained phonetician Henry Higgins, whose dialect is different, in effect hears H when it isn't there in her speech and hears no H when it is until, with great difficulty, she acquires productive use of a new subphonemic (phonetic) distinction. Earlier in our discussion (see Section 2.3.2) we mentioned that the Greeks failed to hear the Phoenician laryngal consonants, and assumed that certain words—here the words for Phoenician laryngal-syllable names—began with a vowel, a lucky misapprehension that led to the vowel-letter concept. So powerful are the conceptual organizing principles here that we have no optimism that all readers of the present chapter will now agree that the medial consonants of *grater/grader* are "the same" and their initial vowels "different." Further discussion is designed in large part to convince the reader that such an analysis applies to his own speech.

see how things turn out:

Phonemes	Component phones
T	t
D	d, ϕ

The next step is to recognize that *grater/grader* form a minimal pair. We already know that they do not differ in their medial consonant (both have ϕ, now claimed to be a phone of D). Since the two nevertheless can reliably be distinguished in pronunciation, a physical difference must lie elsewhere. Fortunate for sanity, it does: The "a" of *grater* is perceptibly shorter in duration than the "a" of *grader.* Symbolizing short a as Ă and long a as Ā, we now have a more complete phonemic inventory:

Phonemes	Component phones
T	t
D	d, ϕ
Ā	ā
Ă	ă

The ā and ă are, under this analysis, assigned to separate phonemes because they distinguish the minimal pair *grater/grader* (and many others, e.g., *traitor/trader, waiting/wading*). English spelling now seems to be incorrect, for it assigns the difference in *grater/grader* to the fourth (T/D consonant) segment, while the present phonemic analysis assigns the difference to the third (Ă/Ā vowel) segment.

Common sense rejects this outcome on the grounds that (a) the assignment of ϕ to D rather than to T—or to T rather than D—is wholly arbitrary and asymmetrical; and (b) the real ("perceived" or, better, "conceived") difference between *grater* and *grader* is in the D/T segment, just as the conventional spelling system claims. But notice that in order to refute the outcome of the analysis so far, one must give up the claim that the phoneme is a physically definable class; the domain of psychophysics cannot handle the desire to assign a single physical event (ϕ) to two different categories. That is, psychophysics provides no excuse for violating Assumption (2).

Before going on to the alternative view of the phoneme as a more complex cognitive category, let us simply show the effect, in the words *grade/grader* and *grate/grater,* of the perceptual definition of the phoneme. As is clear from Table 2, for the case of *grate/grater,* the morphological and meaning relation is obscured by the writing *grăt/grāder*. Notice that a decision to write phonetically (in terms of phones) rather than phonemically (in terms of phonemes) would just as surely obscure the same morphological relations (*grate* would be written *grăt* and *grater* would be written *grăϕer*). This state of affairs was viewed as

TABLE 2
Effect of the Perceptual Definition of the Phoneme

English spelling		"Phonemic" spelling[a]	
Word in isolation	Word with "er" added	Word in isolation	Word with "er" added
grade	grader	grād	grāder
grate	grater	grăt	grăder

[a]Read the symbol "a" in all cases as the vowel in *maid, date,* etc.

acceptable by some linguists despite its apparent descriptive asymmetries (Bloch, 1941). However, it is inadequate on a variety of psychological grounds.

3.1.2 The Phoneme Is a Cognitive–Perceptual Category

The concept of a phoneme must be rehabilitated to account for rhyme, alliteration, tongue twisters, language change, baby talk, and the invention and learnability of an alphabet. The particular ways that these phenomena are manifested in English (or any other language) can be understood only by looking to a set of general *phonological processes,* presumed to be part of each individual's mental apparatus (for discussion see N. Chomsky, 1964). For the examples we have been looking at (*grate/grater, grade/grader*), the particular process that is relevant is a regular increase of vowel length with voicing of the consonant that follows. That is, all vowels are long before a voiced consonant, and all vowels are short before an unvoiced consonant. (Compare the vowel duration in *rib/rip, maze/mace, save/safe, bag/back, wig/wick,* etc.) This "rule" is regularly honored by native American speakers (indeed, failure to observe this constraint in appropriate circumstances will lead speakers to identify accents in foreigners). A description of the sound system that cannot capture this generalization is inadequate. Yet the perceptual definition of the phone fails to capture it: our phonemic writing of *grater* was *grăder* (a counterinstance of the voiced consonant D being preceded by the short vowel Ă).

The generative–transformational description (attributable to N. Chomsky & Halle, 1968, and their co-workers) asserts that the phoneme is not *a perception of a pronunciation* in any direct way, but rather *an abstract internal representation* related to pronunciation only indirectly, through the mediation of a sequence of phonological processes (rules) that take the abstract representations as their inputs. Each morpheme is said to have a *phonological representation* in the mental lexicon (stored dictionary entries) in terms of a sequence of *systematic phonemes* or *lexical representations;* The systematic phonemes in turn are encoded into the *surface phones* through the application of mentally stored phonological rules. Finally, the surface phones are converted into the various articulatory gestures that yield pronunciations.

At the cost of postulating these unobservable internal representations, not

TABLE 3
Successive Representations of Example Words as Phonological Rules Are Applied

Steps from systematic phoneme to phonetic representation	grate	grater	grade	grader[a]
The presumed phonological representation in the mental lexicon	grat	grater	grad	grader
The outcome of postulated phonological Rule 1: lengthen vowels before voiced consonants	grat	grater	grād	grāder
The outcome of postulated phonological Rule 2: convert dental stops preceding unstressed V(C) syllables to flaps	grat	graϕer	grād	grāϕer

[a]Read "a" as the vowel sound in *grate, maid, may,* etc.

directly manifest in pronunciation, the generative grammarians can describe the perceived phonemes (as given by intuition and rendered directly by conventional spelling in our example cases) which are the inputs to the also unobservable phonological rules. They can also describe the perceived phones (perceptibly different sound segments) now described as the output of the phonological rules. The problem with the previous definitions was that these two levels of representation got mixed together, with chaotic results.

For the examples we have been working with, the results of the cognitive approach can be stated informally as shown in Table 3. Notice that the English spelling is very close to the systematic phonemic spelling presumed to appear in the mental lexicon. After application of relevant phonological rules (only two of which are sketched in Table 3), the representation is of the pronounced form. Thus, this kind of analysis allows us to preserve a single representation for a single morpheme in the presumed internal lexicon; to describe the intuition of native speakers about the perceptual–cognitive systematic phonemes; and to describe the surface phonetic (sound) properties of English utterances, even when these phonetic properties are in conflict with phonemic representations.[10]

Now we have shown what was only asserted to begin with: the phoneme can be defined and described only by looking at more global levels of analysis (the

[10]It is important to note that neither this analysis, nor the traditional phonemic analysis that preceded, can be carried out without taking into account even more global facts about the structure of the language. The domain of all the phonological processes described above is the word; these "rules" do not operate across word boundaries. That is, the flap is not substituted if the following V(C) unstressed syllable lies in the next word. For example, the following sentences do not receive the ϕ phone at the relevant point: "I often wri*te* *er*roneous essays," and "I often ri*de* *er*ratic horses."

morpheme and word) as well as more molecular levels (the phone). The phoneme is a perceptual–cognitive entity which cannot be defined autonomously by considering speech sounds alone.

One further note will become important in a later discussion of speech perception and production. Even the very informal remarks above about English phonological rules have been stated by referring to attributes or *features* of the sound segments rather than to whole phonemes and phones. In order to say that "all vowels" behave in such and such a way, or that "all voiced consonants" behave in such and such a way, we obviously conceived each segment as a simultaneous bundle of (binary) features such as *+voiced/–voiced, +vocalic/–vocalic,* and so on. Clearly this is no place to describe the set of all features and their linguistic–articulatory specifications. (See, for description, Jakobson, Fant, & Halle, 1963; N. Chomsky & Halle, 1968.) Two examples will suffice to show that the features cross classify the phonemic and phonetic segments in a number of ways. The segments T and D are *–continuant* while F and V are *+continuant*; that is, the articulation of T and D completely stop, or block, the air stream momentarily so that one cannot hold (continue saying) the consonant, while F and V allow a little air through, so one can continue this articulation (you can say 'ffffff' but not "bbbbbbb"). At the same time, T and F are *–voiced* while D and V are *+voiced.* That is, during D and V, the vocal cords begin vibrating about 25 msec earlier than they do for T and F. On one dimension, then, T is classed with D; on another, it is classed with F. In sum, the classification is

	T	D	F	V
voiced	–	+	–	+
continuant	–	–	+	+

At the systematic phonemic level, the features can be interpreted as a classification; at the phone level, they are the direct inputs to the neural machinery that will trigger gestures of the vocal apparatus. That is, each feature has reference to an aspect of articulation (the features may not be strictly binary at this level). Not surprisingly, then, such matters as the acquisition of the sound system and sound change in language are best expressed in terms of these features, for they are the units closest to the articulation of real speech.[11] The

[11] The question arises whether a script based on features would be more efficient than one based on whole phones or phonemes, since this is apparently the "deeper" phonological unit. In a script of this sort, letters like *b, d, g* would share some visual features (a mark representing voicing; a mark representing stop; etc.). Whether such a notation would aid processing of visual language is problematical. But scripts employing this concept are not entirely unknown. Chao (1968) reports that the Korean *Han-gul* visually analyzes phonemes in terms of subphonemic features. "For example, the symbols for the tense consonant phonemes are made of doublets of the symbols for the corresponding non-tense consonants, such as ㅅ for ordinary *s,* ㅅㅅ for tense *s*. . ." (Chao, 1968, p. 107).

extensive psychological and educational literature on confusability among the sound segments also is stated in featural terms. The notions *phoneme* and *phone,* then, have no real systematic import within the linguistic description. They are cover names (labels) for certain recurrent (classificatory and articulatory) bundles of features in the language in question. Different languages choose differently from the possible bundles of features. At the level of feature bundles, linguistic theory leaves off, bequeathing the speech and hearing apparatus to the psychologists.

3.2 English Orthography Often Represents Aspects of Systematic Phonemics

We have shown that English writing is phonemic rather than phonetic for instances such as *grater/grader.* Restated: the writing system frequently renders just the *unpredictable* (item-specific) facts about the sound of each word; it leaves unstated the *predictable* (rule-determined) facts that follow once the item-specific facts are known. Phonetic facts follow from the application of internalized phonological rules to the phonemic facts. Insofar as the orthography adopts this stance, it is useable only by a knower of the language, for he putatively has stored the phonological rules and will supply their effects automatically, recovering correct pronunciations. Reviewed below is one more well-known example of phonemic representation in English spelling, to show that the instance of *grater/grader* is not an isolated effect in the spelling system. In short, a broad look at the English writing system reveals that many apparently haphazard spelling conventions can be reinterpreted more charitably as transcriptions of systematic phonemes.

One important case is the tendency in English spelling not to mark changes in vowel quality that are totally predictable consequences of the stress on the syllable in which the vowel appears. To understand this phenomenon, consider the pronunciation of the boldface vowels in the following words: *t**e**' le graph; C**a**' na da; c**o**n' tract.* All these vowels sound different. All occur within stressed syllables. When suffixes are added to these words, their stress shifts to the next-right syllable as *t**e** le' graph y; C**a** na' di an; c**o**n trac' tu al,* a common effect for many thousands of English words. Now notice that the vowel sounds in boldface are different from what they were in the unsuffixed forms. While they were all different before, they now all sound alike. All now sound like a brief "uh" (usually symbolized ə).[12]

Is this ə a new phoneme and, consequently, should a proper phonemic alphabet, unlike our conventional one, render these words as *tə le' graph y; Cə na' di an; cən trac' tu al*? For the classical phonemicists the answer was yes, at

[12] The second vowels in *telegraphy* and *Canadian* also change in the suffixed form, for partly similar reasons; but the description of this fact is more complicated (see N. Chomsky & Halle, 1968).

least for the sake of describing the language system presumably known by speakers (Bloch, 1941). Such a solution would describe, at the required level of exactness, the actual sounds that are uttered. But notice that for these three words (and many thousands of English words that have this moving stress), this formulation requires two phonemic writings of each word in the lexicon.

Is it plausible to suppose that the speaker stores all these double entries in his mental lexicon? This seems a cumbersome method for the language user to adopt, especially since a single unifying "rule" will describe all these thousands of alternations: "Pronounce all vowels as ə when they occur in unstressed syllables." The grammarians hypothesize that speakers have acquired the rule instead of the double entries as part of their language knowledge: speakers implicitly know that ə is just the neutralized form of many vowel phonemes. Perhaps the best indication of the correctness of this psychological claim is that the writing system–a natural product, developed by nonlinguists–embodies an identical claim. The ə is assumed in conventional spelling not to be a separate phoneme, that is, there is no symbol for it in the script. The writing system thus implies that, since the speaker–reader has internalized the rule neutralizing unstressed vowels, he will supply the effects of that rule (ə insertion) automatically when his task is to recover pronunciations. Writing ə, on the other hand, would add phonetic detail to written words for which the speaker–reader has no need (as would, in the *grater/grader* example, writing ɸ instead of t or d). Much worse, the surface phonetic writing would wipe out distinctions that the speaker–reader needs in order to recover meanings from the script (such as the information that *grater* is *grate* + *er*, while *grader* is *grade* + *er*). Thus the linguistic description supports the rationality of the writing system. But, much more importantly, facts embodied in the conventional writing system support the psychological validity of the linguists' description.

The assumption that different sounds may map onto a single phoneme (e.g., t and ɸ may both map onto T and ə and e may both map onto E) materially simplifies the formal description of relations between morphemes and sounds; alphabetic writing often embodies the same assumption. Symmetrically, the same sound may map onto more than one phoneme (e.g., ɸ maps onto T and D, and ə maps onto a variety of vowel phonemes); this assumption is also manifest in English writing. In effect, then, the writing system often transcribes a level of organization somewhere between the meaning structure (*morphology*) and the sound structure (*phonology*); it is to this extent a *morphophonemic* writing. It follows that similarities in spelling are partly predictable on the basis of similarities in sound; for example, the two suffixes *est* and ist, which have different meanings, are spelled similarly because they sound similar. It follows as well that identity of sound does not uniformly predict identity of spelling; for example, the suffixes *est* and *ist* sound identical in the words *purest* and *purist*, but they are spelled differently to represent the different morphemes. In fact, we can confidently predict different spellings for the same pronunciation, even for new (invented) words, if these have a recognizable morphemic substructure. For example, suppose that *Mesabrate* is the name given to a new nation, created by redrawing some political boundaries in the Balkans, a not unlikely circumstance. Suppose also that there is a verb *to mesabrate* meaning the newly observed action "to kiss one's elbow." It follows, given English morphology and spelling convention, that a man from Mesabrate is a *Mesabratian* while an act of kissing one's elbow is *mesabration.* In short, spellings in English optimally have both morphological and phonological coherence.

It has been demonstrated that morphophonemic writings have a basis in the cognitive–linguistic organization of fluent speakers. But, obviously, so do phonetic writings. Then would a surface phonetic script also serve the purposes of a reader? Evidently not: to the extent that morphology and surface phonetics are complexly related through a series of intricate recodings, a surface phonetic script would have the effect of obscuring the meanings to be conveyed. Would it really help the reader if, for example, *know* and *no, comintern* and *Common Tern, nitrate* and *night rate, homonym* and *hominem,* or *too, to, two,* and *Thieu* were spelled the same? Indeed this might help the learner, who then has but a single (surface phonetic) principle to acquire. But different spellings for the same sound ultimately act to keep different morphemes apart for the reader, while in the spoken language, intonation and junctural cues, pauses, repetitions, and arm waving are available instead to mitigate the massive ambiguity in the sound stream. Chao (1968) provides a nice example from French displaying the confusion that might result from a phonetic writing. For the expression "If six hundred and six saws saw six hundred and six cypresses..." French orthographic convention yields *Si six cents six scies scient six cents six cyprès...* while a surface French orthography would yield instead *Si si sã si si si si sã si sipre...*

In sum, morphophonemic writing reflects the language organization of its users. Nevertheless, it is unclear whether a script embodying these principles is easy to learn or use. This is an empirical question that cannot be answered solely by pointing to the simplicity of the script with respect to the spoken language. Further, although we have demonstrated that English spelling sometimes reflects deep phonological and morphological facts about the spoken language, we have not shown that it always or nearly always does so. This latter claim has been made by N. Chomsky (1970) who asserts that English spelling convention is near optimal when interpreted as a transcription of systematic phonemics. In our view, it is more straightforward to assume that English writing intermixes a number of levels of linguistic representation, and some arbitrary features.

3.3 English Orthography Represents Many Levels of Language Structure

Although phonologically natural spellings characterize much of English writing, there is evidently not a one-to-one correspondence between letter and systematic phoneme. For example, consider the variety of letters and letter sequences that represent the sound "ay" in *by, buy, bite,* and *bind.* Not all refer to different systematic phonemes. Similarly, the letter A corresponds to a variety of phonemes in *bad, ball, wad,* and *radio.* Procrustean maneuvers would be required to relate these apparent diversities to the phonology of English.

The problem here, in some further detail, is whether we can relate each particular fact of conventional spelling to a morphophonemic fact in the spoken language. The concept of deep phonological representations in the mental lexicon arises as an account of the "best fit" between the sounds (surface phones) and the units of meaning (morphemes): the unobservable deep representations are hypothesized as descriptions that simplify the relationships between these observed levels. If conventional writing neatly transcribes at this intermediate level of organization, it follows that there must be an account in deep phonology of why, for example, we have the differing spellings *lute* and *loot* for two morphemes that sound identical (in most American English dialects). Let us assume, for the sake of the argument, that we do have an analysis of morphemes and their phonemic representations that rationalizes these two spellings for a single sound. But then the same analysis ought to account for the pair *chute* and *shoot,* or it is hopelessly ad hoc with respect to the spelling system. That is, in the proposed analysis, *lute* must be to *loot* as *chute* is to *shoot.* It is hard to conceive of any morphological analysis that would achieve this end (for none of these words has any meaningful relation to any of the others); it is harder yet to suppose that the language user has tacitly internalized such a relationship. The problem is exacerbated when we notice that the pair *route* and *root,* which again sound identical in many dialects, introduce yet another spelling variant, and for no obvious reason of word or sound structure. Worst of all, *mute* and *moot* (which sound quite different) show the same spelling relations as *chute* and *shoot* (which sound the same). Then there are *beaut* and *boot, cute* and *coot,* and so on. The most straightforward analysis, in light of the number of inexplicable variations, acknowledges that English spelling is only a loose transcription of English phonology.

It is important to acknowledge also that a "pure" phonemic or morphophonemic script may not really be feasible because of individual and developmental differences among users. If the deep phonological description most simply connects the morphemes to their pronunciations, then the individual's deep phonological organization depends on the richness and completeness of the etymological relations he notices among his own vocabulary items. That is, if the speaker does not implicitly realize that spoken "bomb" and "bombardier" or "hymn" and "hymnal" are meaningfully related, then the final B of "bomb" and N of "hymn" would not appear in his deep lexical representations; and then conventional spelling would seem to him irrational—as it does to many whom we would call normal adults. N. Chomsky (1970) seems to assume that all adults share enough vocabulary that their organization would tend to be the same, but some counterarguments seem plausible (for further discussion, see Francis, 1970). However this may be, it is certainly clear that the young child's phonological organization is quite different from that of adults (Moskowitz, 1973; Read, 1971). If the child's phonological analyses are different from our own, and if it is the child we must teach to read, then certain of the deep phonological analyses projected for adult competencies would be worse than useless as a basis for initial reading acquisition.

Summarizing, there is evidently no unique level of phonological analysis that accounts simply for all the byways of English spelling, and will also support the acquisition of reading in young children. The writing system is a hybrid, sometimes representing deep phonological and morphological facts, sometimes the facts of a more surface phonetics, and sometimes a fossilized historical process (as in the K of *knight* and *knee*). How can this hybrid system be characterized?

Despite rampant irregularity, redundancy, and downright misrepresentation, English writing is at bottom phonographic: it is an alphabet. Only a tiny fraction of words are represented as true logograms, lacking any phonological information (mainly, these are the numbers and assorted signs of weight and quantity such as %, $, #, and +). Aside from these, each spelled word contains guides to pronunciation; that is, if a word starts with a "buh" sound, it generally begins with the letter B. The historical move from logography to phonographies like our own has a clear basis in efficiency of the script for acquisition, and for the recognition of rare and novel items: there are about 50,000 words in the speaker's vocabulary, as well as a proliferation of proper names. There is no practical way to memorize 50,000 unrelated visual representations of these items. The alphabet provides a phonological mnemonic to reduce the burden on memory. Although we acknowledge that the alphabetic system does not reduce memorization to 26 items—it is not a simple cipher on the systematic phonemes—it probably does reduce memorization to a mere several hundred or at worst a few thousand items (see Rozin & Gleitman, this volume, for a general discussion of the efficiency of phonographic scripts).

Although the script has phonological coherence, it is just as clear that its representations are not solely or simply phonological, else why would *know* and *no, pain* and *pane,* and *in* and *inn* be spelled differently? Each of these writings does give guidance for pronunciation but—at the cost of muddying the clues to sound—each gives meaning information as well. If this meaning information were given systematically for all cases, as it is for the particular classes of examples described in Section 3.2, we would conclude that English writing is strictly morphophonemic, that it represents directly the phonology of mentally stored lexical items. But we have now shown that morphological information is represented inconsistently, even haphazardly, for many words in conventional spelling. This *item-specific* representation of an individual meaning by means of a unique orthographic configuration is the same principle that appears in stone-age scripts: English orthographic representations are partly logographic.

Is this mix of phonographic and logographic information a barrier to English readers? Quite the contrary, we hold that the multileveled representations are efficient for the fluent reader's purposes. Many aspects of the reader's performance suggest that he monitors visual language at many levels in a rapid and critically integrated series of information manipulations (Rozin & Gleitman, this volume). Fluent readers do not proceed solely by "sounding out" or solely by recognizing "whole meanings" of words. A script that represents meanings only (a logography) cannot ultimately serve the purpose of effortless recognition of rare or novel items. And on the other hand, a script that represents sounds only (a phonography) poses problems for the rapid reconstruction of meanings from text (see Fig. 6).[13]

A final comment. We have asserted that English writing usefully intermixes a number of levels of language representation. Surely this system did not get to be the way it is by conscious decisions to represent the sound structure here, the morphological structure somewhere else. Rather, like the ancient scripts, the English writing system comes down to us in its present form through errors by semiliterates, conditional sound changes, printers' preferences, grammarians' strictures, and the like. But then–in light of the traditionalism of the current spelling–is it not likely that the script will become less and less phonologically relevant with the passage of time? Of course this is possible in theory, but it does not seem likely. Rigid as the writing system now is, it does seem to succumb to reform just when the representations stray so far from current phonology as to become useless at that level. The most grotesque feature of current English

[13] The Japanese writing system, from this point of view, seems ideal. We have already mentioned that the Japanese employ a syllabary of about 50 items (see Section 2 and Fig. 5). In fact, they have two syllabaries, which appear intermixed on the page. Mainly grammatical elements (suffixes, and the like) are written in the *Hiragana* syllabary, while borrowings from Western languages are written in the *Katakana* syllabary. Notice, then, that two subcategories of the morphological stock are picked apart visually in terms of two separate sets of visual symbols (see Fig. 6). Further, most meaningful words (nouns, verbs, and the like) are written in a third notation (*Kanji*), in borrowed Chinese characters; this system is purely logographic. Theoretically, every Japanese word can be written in either of the syllabaries. Why are two syllabaries and a logography maintained? Doubtless some of the answers have to do with tradition rather than cognition. But the system is undeniably effective. Our theoretical sympathies with Japanese writing have occasioned incredulity in some quarters (Goodman, 1973; see also Gleitman & Rozin, 1973b) for, whatever one's view of the "best" writing system, it does seem unlikely on the face of it that the best one is three different ones–unless, of course, one assumes that linguistic information processing might be complex! If efficient readers monitor many levels of language representation in rapid succession–even, perhaps, in parallel–to extract meaning from the printed page, then a system that formally makes explicit the appropriate level to be sampled does not seem so strange. About 1800 of the most familiar, frequently recurring substantives appear in *Kanji* (logographic) form; for these words, it may be most efficient to take in the sought-for meaning "direct" without recourse to phonological intervention. But for words that are rare, relational, or external to the morphophonemic system (foreignisms), a more analytic decipherment might be indicated. Our claim is that English writing, by hook and accidental crook, has evolved into an informal representation of the same mixed character as Japanese.

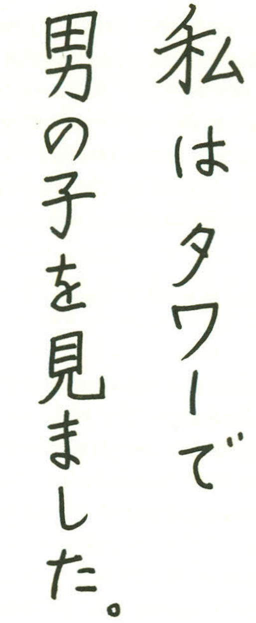


FIG. 6. Japanese orthography; three different scripts are used intermixed on the page in Japanese; the *Hiragana* and *Katakana* syllabaries and the logographic *Kanji.* The writing is read from the upper right corner down. In the illustrative example here, the English sentence, *I saw a boy at the tower,* is shown as it would be written when translated into Japanese. The three different scripts are all used within this sentence, as is usual in Japanese, for each script tends to be used for specific types of items (see the discussion in Footnote 13). The first symbol at the upper right is a Kanji for the first person. The symbol underneath it is a particle indicating the case (in this case, nominative). It is a Hiragana syllable, and is pronounced following the saying of the first person. These two symbols together thus constitute *I.* The following three symbols represent the English word *tower.* Japanese has borrowed this word from English, with essentially equivalent pronunciation. As a foreign loan word, it is written in Katakana. The first symbol (looking like a crossed seven) stands for "tau," and the second for "wa." The third symbol, a vertical line, indicates that the "wa" is lengthened. The final symbol in the first column is the syllable meaning *at,* and this is represented by a Hiragana syllable. The second column begins with the word *boy,* represented by three symbols. The first is a Kanji for *man,* and the third is a Kanji for *child,* with a Hiragana syllable sandwiched in between. The last four symbols represent *saw.* The first is a Kanji for the meaning *see,* and the last three, in Hiragana, represent the pronunciation indicating the past tense (the three syllables "ma-shi-ta"). Notice, then, that substantives and verbs are generally represented in Kanji, grammatical formatives in Hiragana, and loan words from other languages in Katakana.

writing is in the series with final *gh* and *ght.* Just these oddities have been subjected to spelling reform, at least in informal settings. Thus we find, ever more frequently, the spellings *thru, tho, nite,* and so on.

Summarizing, the particular orthographic forms of language that are learned by English readers have, simultaneously, morphological and phonological coherence in some complicated mixture. This mixture is not consistent from word to word and cannot—without unnatural stretching and squeezing—be described satisfactorily on any one unifying language principle or set of categories. The internal inconsistency is a result of the "natural" or unprofessional sources of spelling conventions. As long as the script maintains a loose but perceptible relation to language organization, there is little pressure for reform. On the contrary, the system works by providing meaning and sound information in a manner efficient for the judiciously sampling and constructing reader.

3.4 Problems of Acquiring the Phonemic Aspect of English Orthography

The phonographic aspect of the writing system poses a formidable conceptual problem for the six-year-old. Somehow the child must come to realize that the alphabetic characters and sequences are related to the language he brings to school. But how is a teacher to communicate that the alphabet is related to such conceptual categories as systematic phonemes and morphemes of the language? The learner has no clear idea of the meaning of such metalinguistic terms as *word* or *sentence* or *speech sound.* Much less is he aware of the abstractness of the relationships between speech sounds and phonemes, nor has he any vocabulary for discussing such matters.

In practice, teachers who wish to make explicit for the learner relations between alphabetic writing and phonology usually do so by attempting to relate each discrete letter to the utterance of a discrete surface phone. Thus a teacher may say, "The letter T here stands for the sound at the beginning of *tap.*" A problem must eventually arise because the letter T does not really correspond to the spoken language at the level of one sound to one letter of the alphabet: T corresponds to quite different sounds in *tap, grater, creation, often,* and *think.* Some educators consider the possibility of misunderstanding here to be so serious that they counsel against teaching letter-to-sound correspondences, even as rough and ready first approximations for the learner.[14] Others, ourselves included, believe that the child is well served in initial acquisition stages by

[14] In particular, some educators (Goodman, 1973; Huey, 1908) suggest finessing the alphabetic principle by falling back on reference-making units the child will more readily recognize; that is, one might teach whole words or meanings. But then the huge number of items to be learned eventually imposes an overwhelming burden on the child's memory for arbitrary visual displays, a problem recognized six millenia ago by the Egyptians. It appears, then, that the teacher has no direct apparatus available for presenting the deep phonological units in a way that the child can grasp, and at the same time the whole game is lost if the phonological basis of writing is ignored. Some useful methods for introducing children to deep phonological concepts (and, hence, to the spelling pattern of English) have been suggested by C. Chomsky (1973). In her approach, children discuss morphological (meaningful) relations among items such as *sane* and *sanity* and so come to a realization of why they are spelled so similarly. But these techniques are useable only later in the course of literacy acquisition than the stage on which we focus.

learning some simple phone-to-letter correspondences, even though a later shift to deeper phonological analysis will be necessary if literacy is to be achieved. After all, a major step toward literacy has been taken merely by recognizing that *tip, tap,* and *top* do—and should—begin with the same letter. This is an insight about surface phones and reading; no concepts of deep phonology need be invoked to achieve this first insight. Yet with the limited terminology shared in common by teacher and learner, it is often very difficult to teach even these rudiments of "decoding." The teacher may talk about "the first sound in *tap*" or may give an instance of this sound (that is, "tuh"), but there is no reason to expect the learner to know how to refer these remarks to the appropriate firing patterns in his nervous system. The child does not (at least consciously) conceive speech as a sequence of phone-sized units such as "t" strung along in time on the speech wave like beads on a string. This is the issue to which we now turn: how is the phone "t," or any other, realized as a physical event, as an articulatory gesture and a consequent acoustic wave? The complexity of this problem goes some way toward explaining why it is hard for children to achieve the initial alphabetic insights.

4 FROM PHONE TO ACOUSTIC STIMULUS: THE PROBLEM OF SEGMENTATION

We have seen that the output of a linguistic description (Section 3.1.2) is a set of sequences of phones, described as bundles of features, which are related to the articulatory gestures involved in speech. The symbols of the alphabet refer in part to these phones, in part to phonemes. The concrete link between the alphabetic symbols and the phones is presumably provided by the physically real sound stream of speech. That is, the teacher of reading can hardly refer the learner to generative phonology for an understanding of the nature of alphabetic writing. He can only try to establish contact between the two available physically realized entities: the visible alphabetic symbols and the audible stream of speech. Here arises what we take to be the fundamental conceptual problem in early reading acquisition: the alphabetic signs (and their referents, the set of phones and phonemes) are discrete units, while speech itself does not consist of physically segmentable discrete units, but is a continuous, gradually varying event. Examination of the sound pattern that leads to the perception of a phone sequence shows no trace of segmentation. For example, Fig. 7 demonstrates the continuous complex pattern of acoustic energy emitted in uttering the word "bag." (See also the bottom of Fig. 9 for an idealized computer-generated sound pattern which is heard as "bag.") There is no way a physicist looking at such spectrograms could infer segmental components. The segmentation we perceive comes from within the head. As much for prereaders as for readers, there is surely a representation of language in the head in terms of discrete phones and

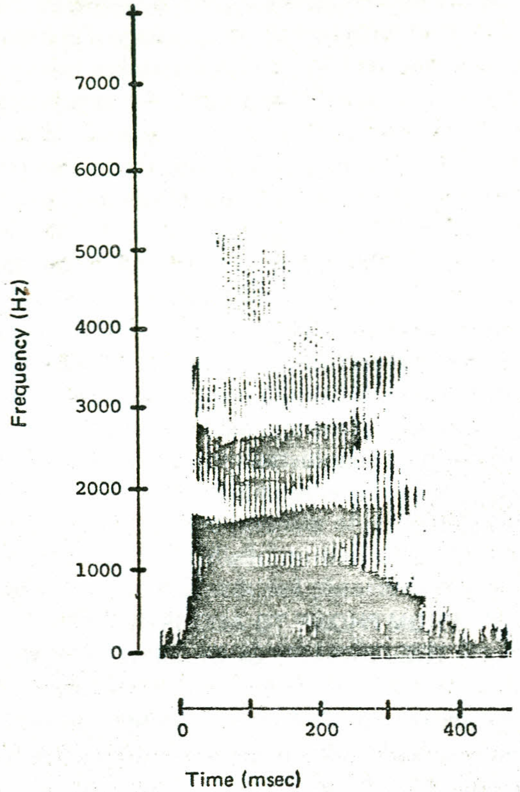


FIG. 7. Sound spectrogram of the spoken word "bag." The darker an area or point, the more acoustic energy is present at that frequency and time. The bands of energy which are relatively consistent over time are called formants. The lower two formants are sufficient to produce the perception of "bag." The computer-generated sound spectrogram of "bag" at the bottom of Fig. 9 represents only the two lowest formants.

phonemes. (Otherwise, as Savin, 1972 has observed, how could a little child "know enough" to get mixed up when saying tongue twisters?) However, since the circuitry of this system is internal, it cannot be physically manipulated for pedagogical purposes, nor is there any guarantee that it will be accessible to introspection. That is, the child is faced with an orthography that refers to a discrete representation of language that cannot be pointed to in speech. There is no clear time at which B ends and A begins in spoken "bag," only in written *bag.*

How is the teacher to refer to an internalized level of the language system at which the continuous stream of speech is segmented and discrete? The problems in giving the child contact with this level of organization are complicated enormously by the fact that the teacher cannot pronounce the three parts of "bag" separately for instructional purposes. The sequence of phones that is the output of the phonological system has been converted into an utterance in the

form of a continuously varying wave. We discuss below the relations between phone perception and production and properties of the continuous speech wave. The discussion follows, for the most part, the positions set forward by the Haskins Laboratory group of investigators (A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; A. M. Liberman, 1970; Cooper, 1972; See also Stevens & House, 1972; Cole & Scott, 1974). The evidence from this very important work, and from related findings in the perception of running speech, points generally to the syllable as the basic unit of speech perception. From this, there is good reason to suppose that the kind of access to molecular levels of linguistic representation required to learn to read is simpler for the syllable unit than for the phone or phoneme unit (see also I. Y. Liberman et al., this volume). Reconstituting the sequential phone from the encoding in which it is embedded in real speech is a difficult cognitive act, not easy to elicit from prereaders.

4.1 Are the Phones Manifest in the Speech Wave?

The simplest hypothesis for understanding what the alphabetic letters refer to (or for understanding how we perceive speech in discrete terms) would be that there exist discrete and invariant acoustic patterns in the speech wave that are directly associated with the perception of the sequential phones. Each letter of the alphabet, approximately, would "stand for" a linguistically significant sound pattern; the alphabet would then bear a simple relation to sequential properties of the speech wave. In fact, such a characterization of the acoustic properties related to phone identification can be made fairly reliably for the case of nondiphthongized vowels. In utterances of such vowels, the energy that is emitted over a period of approximately 150 msec is largely concentrated in restricted frequency regions. The spectrographic display of Fig. 7 shows those characteristic vocalic energy bands (*formants*). The relationship of the frequencies of the first two (lowest) formants is closely related to the perception of various simple vowels. Spectrographic patterns consisting solely of the first two formants are sufficient to synthesize the simple vowels so that they are recognizable to human listeners. At least for these vowels, then, the relation between the phones of linguistic description and the speech wave appears to be straightforward: in A. M. Liberman's (1970) terms, a cipher or one-to-one mapping. The relative simplicity of this mapping can be seen in the synthetically produced spectrograms in Fig. 8 (from A. M. Liberman et al., 1967) where the perceived vowel can easily be referred to the relation between the first two formants.

Invariant properties of the acoustic wave yield information about consonant perception as well. For example, A. M. Liberman, Delattre, and Cooper (1952) showed, using synthetically produced speech, that very high-frequency bursts (about 3240 Hz) of about 15 msec duration were uniformly interpreted as T before various vowels, while low-frequency bursts (about 360 Hz) were interpreted as P before the same vowels. Frequency of an initial burst, then, is an

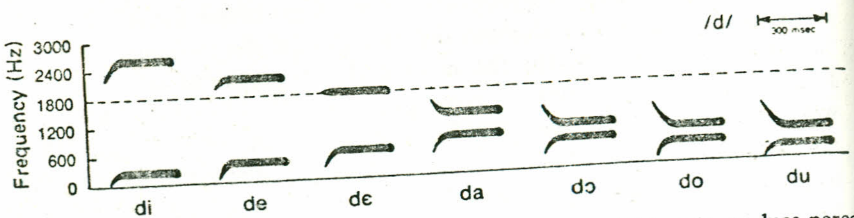


FIG. 8. Synthetic, computer-generated sound spectrograms sufficient to produce perceptions of various D + vowel syllables. The dashed line at 1800 Hz shows the locus for D (A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, p. 436. Copyright © 1967 by the American Psychological Association. Reproduced by permission.)

invariant (context-free) cue to the distinction between P and T (see Cole & Scott, 1974, for a review of the literature on invariant cues to consonant perception).

If, in all instances, invariant cues to the perception of each phone could be found in the acoustic wave, the perception of discrete phones in the continuous sound wave would be easier to understand. However, invariances in the acoustic wave are insufficient to allow unique identification of all consonants. For example, A. M. Liberman et al. (1952) also showed that a single synthetically produced burst of noise at 1800 Hz was understood as P before I and U, but as K before A. Thus the identification of consonants is in part context dependent: the same physical object can correspond to alphabetic writings of both P and K, for these are writings of perceptions, not writings of acoustic patterns.

Figure 8 further demonstrates the nature of this problem. The figure shows spectrographic patterns sufficient for the synthesis of various D + vowel syllables. The steady-state formants occupying all but the first 50 msec of each syllable are sufficient for the synthesis of the various vowels. That is, when we clip off the first 50 msec of these syllables and play back the remainder, subjects reliably report hearing the appropriate vowels. Then, obviously, the rapid frequency modulations (*transitions*) of the first 50 msec are somehow responsible for the perception of the consonant D. Yet inspection shows that these portions of the displays differ radically: for the syllable DI, the second formant is rising rapidly from about 2000 to 2500 Hz; for DE, the second formant is flat at 1800 Hz; for DU, the second formant is falling rapidly from about 1200 to 600 Hz. In short, the patterns of acoustic energy over time differ radically for D before various vowels. Yet in all cases the hearer reports a recognizable D. We have previously seen that the same physical event can represent two different perceived phones (a burst of energy at 1800 Hz is perceived as P before I, but as K before A). We now see that many different physical events can represent one perceived phone (D is realized differently before various vowels). The simple view of "one sound, one perceived phone" is obviously inadequate. On the same grounds, the view of "one sound, one symbol" for the alphabet is also inadequate. The complexity in the relations between these two levels of language

organization is reminiscent of the complexity between the phone and phoneme levels (see Section 3).

A further understanding of the complexity at this level comes from the realization that the consonantal portion of the patterns in Fig. 8 cannot be segmented and recovered at all. While the vocalic portions of these patterns can be identified correctly in isolation, the consonantal patterns, clipped from these patterns and played in isolation, sound to the hearer like a chirping noise. Not only does the listener not hear D; he does not recognize a speech event (A. M. Liberman et al., 1967). Clearly, then, even if a teacher could pronounce the referents of the alphabetic symbols in isolation for the child (which he cannot, as we later show), the child could not hear them. Thus, a practical problem in teaching reading is that a word like "go" cannot be taken apart and pronounced and heard as "g" + "o." One cannot produce (as teacher) or hear (as learner) the elements of the system to be learned. This is the oft-mentioned "puh-ah-tuh" problem: the child is asked to hear, identify, and later *blend* three items that the teacher instances as three syllables. The reduced vowel (ə) is an obligatory part of any isolated utterance of a stop consonant, and yet no such vowel appears in "pat." Teachers' exhortations to pupils to blend by saying "puh-ah-tuh" very fast thus are quite misleading. The child's real task (approximately) is to identify three phones that appear within the teacher's three spoken syllables, and then to pronounce the three-phone monosyllabic outcome (see I. Y. Liberman et al., this volume).

We have shown that a variety of grossly different physical events all represent one phone in different vocalic contexts; symmetrically, the same physical event represents different phones, again depending on the vocalic context. There is no simple mapping between phone and acoustic event; the relation between the two levels of representation is highly complex and context dependent. The various representations of D in Fig. 8 do nevertheless have something in common: a somewhat hypothetical starting point. Notice that the initial transient portion of the second formant for each syllable seems to be aiming backward in time toward a fixed frequency, at about 1800 Hz. [No initial energy is actually emitted *at* 1800 Hz; that is, approximately 15 msec of relative silence precedes the onset of energy emission for all these syllables. This period of silence is, in fact, a context-free cue to the set of stop consonants; if it is removed, the perception is not of a stop consonant (Delattre, A. M. Liberman, & Cooper, 1955).] Thus 1800 Hz is said to be the *locus* for D (A. M. Liberman, Delattre, & Cooper, 1955; A. M. Liberman et al., 1967). What the D syllables seem to have in common, then, is a common starting point that is not physically represented in the speech wave.

The perceived invariance of all the D syllables appears to derive from a common articulation; the articulators are in a fairly uniform position (tip of tongue at alveolar ridge, etc.) as the D syllable begins. Our ear (mind?) has no trouble noting this invariance, although it is well disguised in the sound signal.

The invariance is more simply describable in terms of motor-speech acts than in terms of the acoustic events. These relationships have suggested to some (A. M. Liberman et al., 1967; Stevens & House, 1972), that the process of speech perception is related to the process of speech production, and that speech is recognized in some way by a process of analysis by synthesis: by determining what motor program or command phone matrix could produce the perceived signal. The full implications of such models are too complex for us to deal with here, but they clearly do point to fundamental relationships between alphabetic units and analysis of the wave by referring to its articulatory origins.

4.2 The Production of Speech Sound Is Related to the Letters of the Alphabet

The description of phones as bundles of component features (Section 3.1.2) is a convenient starting place for the description of speech production. If the phone is conceived in the speech production machinery, as in phonological theory, as if it began its journey from the nervous system to the mouth as a set of articulatory features, the facts of speech are easier to understand. We assume here, again following the formulations of the Haskins Laboratory group, that the phonetic features of the linguistic system are represented physiologically as neural commands that ultimately will control speech. We accept the simplifying hypothesis that the abstract linguistic features relate more or less directly to components of pronunciations. The neural commands are still remote from speech. The speech-production machinery converts the simultaneous discrete features into unit phones (Section 4.2.1) and the discrete sequences of phones into a smooth continuous flow (Section 4.2.2). A significant restructuring takes place as the commands to produce phone sequences are carried out simultaneously and become integrated into syllabic units (A. M. Liberman, 1970). One consequence of this restructuring is that it is difficult to achieve the alphabetic insights.

4.2.1 The Simultaneous Production of Features Yields the Unit Phone

There are no remarks in the linguistic account about how the feature bundles are to be realized as coordinated movements by the various articulators (the bone and muscle complexes of the articulatory apparatus–tongue, lips, pharynx, etc.). Because the articulators are essentially independent and can move at different rates, the accomplishment of an overall single phonic act evidently requires cooperative activities of the musculature; for example, to say "m," one vibrates the vocal cords, moves the tongue, lips, and so on, all at one time. We know that the smooth production of all these gestural elements of the phone by integration of the component gestures into larger-unit gestures is in part learned and highly automatized in the adult, since the feature bundles of different languages are partly different and since relearning by adults is difficult. Although all languages choose from the same list of possible features and overlap considerably, the simultaneous bundles differ. (Thus, for example, certain notorious French vowels are difficult for English speakers to utter and are rarely acquired perfectly by adult learners. These vowels may differ from English vowels only in a single feature, indeed in a feature that English speakers can easily pronounce within a different simultaneous bundle. Thus we can approximate the French vowel under such articulatory directions as, "Try to say 'ee' as in 'meet,' while holding the lips in the rounded position you would use for 'u,' as in 'soup.' ") The requirement to produce the component gestures of a phone smoothly and in parallel may involve a restructuring of each underlying gesture. That is, a feature may be realized differently by the musculature depending on its context of simultaneous features. At any rate, it is clear that the simultaneous bundle, or phone, is highly practiced, overlearned, and relatively immune to change by adulthood. Adults rarely acquire perfectly the bundles of a second language—they speak with an accent.

The letter-size unit in alphabetic writing must refer in part to these singulary gesture chunks.[15] To the extent that we can identify and isolate psychologically relevant levels of linguistic representation, it is natural to view "the whole phone" as an operative unit in the articulartory system: it is the habitual, unitary, smoothed pronunciation of a bundle of features. From this point of view, it would seem that *the alphabetic notation makes contact with perception at the level of the neural command to produce the phone chunk, a language-specific compositional unit of pronunciation.* The featural substratum seems to be beneath the level of awareness, inaccessible in most circumstances (Cooper, 1972).

4.2.2 The Shingling of Sequential Phones Yields the Unit Syllable

A significant restructuring of the representation of speech takes place because we articulate "whole syllables" as opposed to discrete phones. The phone-chunks we have just recognized as articulatory organizations of the feature bundles are in turn swallowed up into a larger unit. The transmission of phone sequences up to about the level of syllables takes place in a partly parallel or overlapping fashion, thus further complicating the relation between articulatory gesture and acoustic outcome. Figure 9 (from A. M. Liberman, 1970) is a schematization of the pronunciation of "bag." Information in the sound stream related to the medial vowel (AE) is shown to overlap the transmission both of

[15] Of course we know this is not the whole story. An additional encoding has been recognized between phone and phoneme (Section 3.1). Moreover, even the identification of phonemes is highly dependent on yet more global analyses in terms of morphemes, syntax, and meaning. But part of the perception that there is a correspondence between letters and the "individual sounds" of the language derives, in our view, from access to some level of the program for integrated articulatory movements.

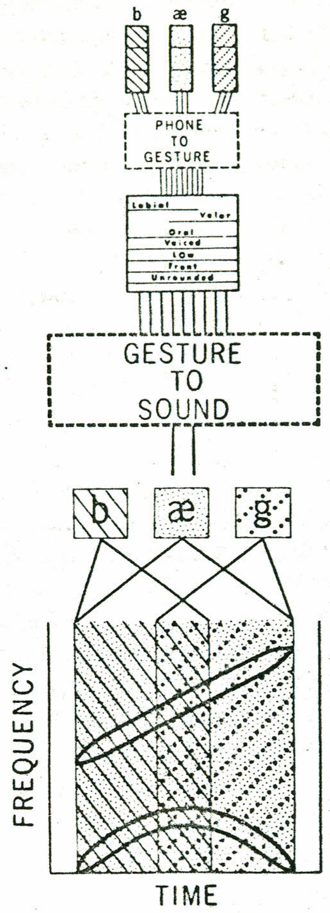


FIG. 9. Schematic model of speech production as formulated by the Haskins group. The figure shows production of the word "bag" in two stages: a conversion from phone to articulatory gesture, and from articulatory gesture to sound. (From A. M. Liberman, 1970, p. 314.)

the initial B and final G consonantal portions of the syllable. This parallel transmission is a consequence of the fact that the articulators respond relatively sluggishly, so that the command for certain muscles to contract arrives while previous commands are still being carried out. Further, the gestures are different depending on the position of the vocal apparatus just prior to and following the position for the gesture in question, so that the transitions between phones (the rapid frequency modulations described earlier; see Fig. 8) contain much of the useful information in the acoustic wave. The articulatory gestures and, hence, the sound pattern, will be influenced by what the vocal apparatus *was* and *will be* doing during the saying of the whole syllable. If this is so, we can understand why the pronunciation of the individual phones cannot be invariant: the pronunciation of each is dependent on the context of phones in which it occurs.

The net effect of the conversion of sequential phone commands to articulatory movements and sounds is that in the sound stream the phones are *shingled*, much in the sense of overlapping roof shingles (A. M. Liberman et al., 1967). Then the transition (frequency glide) between adjacent consonant and vowel gives information about both surrounding phones at once. It follows that there is no discrete sequential analogue of the phone in the wave-form.

A direct consequence of these facts is that it is impossible to utter most context-dependent phones in isolation, for they are realized in articulation shingled across the whole syllable. For the case of the stop consonants, there is silence at the time the articulators are in the required position (at the locus, as described in Section 4.1). Energy is emitted after the onset of the vocalic segment. Thus a teacher cannot utter D (or B, G, K, P, T) in isolation for the learner. He must add an artifactual vowel sound. For other sounds, a more satisfactory degree of isolation can be attained, although in normal speech there is a gradual, continuous shift from one phone to another.

Given the Haskins group's powerful demonstrations and explanation of the context dependence of realizations of phones, the relations between the presumed high-level neural commands (corresponding to the discrete features and phones) and articulatory movements must be complex. That is, as Stevens (1972) and Stevens and House (1972) have pointed out, the high-level command must program an articulatory goal or target (e.g., "move the tongue to a midpalatal position") but the actual articulatory movement must depend heavily on where the tongue was prior to this point. The actual command to achieve the midpalatal target might result in forward or backward tongue movement, depending on where the tongue was when the initial command was received. The same phone, differently approximated in the musculature, is invariant with respect to the initial command: invariant phone commands leading to various movements *toward* an articulatory target are heard as equivalent. On the contrary, the articulatory endpoints are not neatly related to the underlying phone command because of context dependence and shingling owing to sluggishness in the motor system.

Summarizing, the actual production of speech is not in terms of discrete phone chunks produced in a temporal row. A longer unit, approximately the length of a syllable, is preplanned so as to produce in a smoothed flow the set of gestures called for by each phone command in the sequence. The result shows simultaneous interactions among all parts of the syllable. The smoothed whole-syllable gestures, like the whole-phone gestures above them in the hierarchy, must be partially learned. (For example, Spanish words never begin with ST consonant clusters. Adult Spanish learners of English can often be heard to produce a vocalic context for the initial S of this cluster, thus relegating it to a separate syllable, one that occurs in Spanish. They do this because, presumably, the new ST syllable is hard to produce; for them, "stake" is more easily pronounced "es-take," making use of whole-syllable vocal gestures learned in childhood.) The discrete phone stratum above the whole syllable may be largely inaccessible, this level, like the featural components of the phones, residing only in the motor and perceptual systems (Cooper, 1972).

We have argued that the component features of a phone become a habitual unitary act and that the sequence of phones in a syllable become inextricably bound together. One may ask: how far can this game go? After all, one must learn to organize whole polysyllabic words, whole phrases, even whole lectures. Higher level preplanning of units at least as large as the sentence must be assumed, for otherwise it would be difficult to understand how prosodic features (stress and intonation) are inserted at the right times and places (see Cole & Scott, 1974, for a discussion of "envelope features" of the sound stream). The segmentation issue arises for these molar units too: pauses do not regularly appear at syllable, word, or phrase boundaries in real speech. (For some speakers we know, pauses cannot be shown to occur at a lower than whole-lecture boundary.) Thus the concept of chunking cannot by itself explain the fact that notions like syllable, word, and sentence seen to be relatively accessible to consciousness, while the notions of phone and phoneme are relatively inaccessible. Nevertheless, units more molecular than the syllable have been shown to share one property that complicates the process of learning to read: they are in varying degrees, in relation to their context dependence, difficult for the speaker to pronounce in isolation or for the listener to isolate from the sound stream.

4.3 The Syllable Is the Basic Unit of Speech Perception and Production

The fact that transitional cues are required for unique identification of certain consonants in certain environments is the central argument (A. M. Liberman, 1970) for the claim that the syllable represents the lower bound as a unit of speech perception (Cole & Scott, 1974, make a rather different argument for the same conclusion). In general, syllables are the smallest coherent units of speech: they tend to be physically undissectable, they are the smallest separately pronounceable units of speech, and they may be produced in preplanned units.

There is additional evidence for the primacy of the syllable in the speech processing of adults. The syllable maintains a salient position in perception, even for practiced adult readers of English phonemic writing. Adults can identify syllabic targets (e.g., BA) in running speech more rapidly than initial phonemes (e.g., B) of that same target (Savin & Bever, 1970; Warren, 1971). "Sounding out" by fluent readers seems to be organized around syllables when the material to be read is novel. Adults characteristically break an unfamiliar word into pronounceable syllables, e.g., *malagasy* would be broken down into *ma-la-ga-sy* on route to a pronunciation, not into *muh-a-luh-a-* and so on, despite years of experience with phonemic writing systems.

These findings about perception and characteristic organization in reading dovetail with the general Haskins findings. We have noted earlier (Section 2) that syllabic systems were the first phonographies to appear historically and that, unlike alphabets, they recurred by independent invention. We show elsewhere (Rozin & Gleitman, this volume) that the syllabic unit is accessible earlier in individual development than the phonemic unit and that syllabary scripts are easier for young children to acquire than phonemic scripts. We believe these effects derive from the fact that the syllable is the smallest independently perceivable unit of speech: "before" is physically quite close to a quick rendition of "be" + "fore." On the contrary, there is no physical basis for building "be" from "buh" + "e."

If the syllable represents the lower bound as a unit of real-time speech perception, what is the upper limit? It was mentioned earlier that speech production will require, at least for appropriate representation of prosodic features, preplanning of units as large as the sentence. Cooper (1972) points out that limits on the sophistication of auditory analysis and also on short-term memory must set a stringent upper bound on the perceptual processing unit; he concludes that minimal units of about syllable, word or short phrase length are about right because of these constraints on a possible auditory process.

It is important to notice that, even if the syllable has primacy as a unit in real-time speech processing, as we have asserted, it is a relatively superficial unit in terms of its relations to the underlying linguistic organization of word and syntactic form. Relations to this underlying system at the syllabic level have been obscured in the considerable recoding that takes place for the purpose of speaking and hearing. Most natural language mechanisms, such as language acquisition and language change, are conveniently describable only in terms of the underlying phonological units of phoneme and phone. Reading, however, is not a natural part of language organization, and it appears only under special circumstances of environment and careful training. Fluency does not emerge in response to quite automatic language-acquiring predispositions in humans. Our view is that, for such highly derived skills as learning to read an alphabet, the encoded speech signal must be unpeeled through quite conscious attention to its obscured underlying origins in tacit language organization. The syllabic unit may have special usefulness in this process of becoming aware because it is the smallest physically realized unit in real speech.

5 SUMMARY

The evolution of writing is orderly on the following analysis: each succeeding script required fewer different visual symbols to transcribe the spoken language than did the script before, thus reducing the burden on long-term memory for visual items required of learners and users; but in consequence, each new script represented language in terms of units more remote from meaning than the units of the preceding script, thus posing new problems for learning and decipherment. The alphabet (phonemic writing) was the latest script to be invented.

The nature of the phonological concepts embodied in modern English writing throws some light on why alphabetic scripts are hard to invent or learn: the relations between the cognitive–perceptual alphabetic units (phonemes and morphophonemes) and the perceived sounds of speech (syllables and phones) are very abstract, mediated through a set of covert rules whose operations are essentially closed to consciousness. That is, the meanings (morphemes) are encoded onto phonological representations in terms of a complex, context-dependent conversion.

An even more fundamental barrier to reading acquisition is the problem of segmenting the speech wave in discrete terms: the reader of an alphabet must relate the continuously varying acoustic wave of speech to a writing system that represents this in terms of a linear array of discrete symbols. Some of the difficulty in acquiring alphabets is therefore understandable through an analysis of speech production and perception. The perceived speech sounds (phones) are encoded onto syllables in the sound wave in terms of a complex, context-dependent conversion. The encoding of the phone onto the syllable in speech may account for why syllabic scripts were so much easier to invent than phonemic scripts, and why they are so much easier to learn.

The child's natural history of explicit language knowledge proceeds in a sequence similar to the evolution of writing. The young child first becomes explicitly aware of meaning units, and only later becomes aware of the syntactic and phonological substrata of language. Thus it is easy for the young child to learn the principles of a script that tracks meanings directly and hard for him to acquire a script that tracks the sound system. These parallels suggest an approach to teaching reading which we will expand upon in the next chapter (Rozin & Gleitman, this volume). It might be useful for the child to be introduced to visual language as a logography. Thereafter we suggest that the syllabic unit, which maintains its shape and sequential integrity in speech perception and production, may be useful for introducing the learner of an alphabet to the general class of phonographic scripts. In this approach, the abstract phonemic (alphabetic) concepts would be introduced to the learner relatively late. Summarizing, we propose that an initial reading curriculum that essentially recapitulates the historical evolution of writing will mirror the metalinguistic development of the child.

ACKNOWLEDGMENTS

The preparation of this paper was supported by NIH Grant No. 23505. We had the advantage of critical readings of earlier drafts of this manuscript by R. Gelman, E. Gallistel, and E. Newport. We thank H. Hoenigswald who made substantive contributions to our description of the history of writing and other sections of this paper. We also thank Ms. Mutsumi Ishida for invaluable assistance in preparing Fig. 6. And we particularly wish to thank Henry Gleitman, whose mediation between the print and the meaning of this and the accompanying chapter was crucial at all stages.

REFERENCES

Bloch, B. Phonemic overlapping. In M. Joos (Ed.), *Readings in linguistics: The development of descriptive linguistics in America since 1925* (2nd ed.). New York: American Council of Learned Societies, 1958. (Originally published in *American Speech*, 1941, *16*, 278–284.)

Bloomfield, L. *Language.* New York: Henry Holt, 1933.

Cazden, C. *Play with language and metalinguistic awareness: one dimension of language experience.* Paper presented at The Second Lucy Sprague Mitchell Memorial Conference, "Dimensions of Language Experience," Bank Street College of Education, May 19, 1973.

Chao, Y-R. The non-uniqueness of phonemic solutions of phonetic systems. In M. Joos (Ed.), *Readings in linguistics: The development of descriptive linguistics in America since 1925* (2nd ed.). New York: American Council of Learned Societies, 1958. (Originally published in *Bulletin of the Institute of History and Philology, Academia Sinica*, 1934, *4*, 363–397.)

Chao, Y-R. *Language and symbolic systems.* London: Cambridge University Press, 1968.

Chomsky, C. Reading, writing, and phonology. *Harvard Educational Review*, 1970, *40*, 287–309.

Chomsky, C. *Beginning reading through invented spelling.* Selected papers from New England Kindergarten Conference. Boston, November 30, 1973.

Chomsky, N. *Current issues in linguistic theory.* The Hague: Mouton, 1964.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading.* New York: Basic Books, 1970.

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

Cole, R. A., & Scott, B. Toward a theory of speech perception. *Psychological Review*, 1974, *81*, 348–374.

Cooper, F. S. How is language conveyed by speech? In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass.: MIT Press, 1972.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, *27*, 769–773.

Diringer, D. *Writing.* New York: Frederick A. Praeger, 1962.

Downing, J. (Ed.). *Comparative reading: Cross-national studies of behavior and processes in reading and writing.* New York: Macmillan, 1973.

Francis, W. N. Linguistics and reading: A commentary on chapters 1 to 3. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading.* New York: Basic Books, 1970.

Gelb, I. J. *A study of writing: the foundations of grammatology.* Chicago: University of Chicago Press, 1952.

Gleitman, L. R., Gleitman, H., & Shipley, E. The emergence of the child as grammarian. *Cognition*, 1972, *1*, 137–164.

Gleitman, L. R., & Rozin, P. Teaching reading by use of a syllabary. *Reading Research Quarterly*, 1973, *8*, 447–483. (a)

Gleitman, L. R., & Rozin, P. Phoenician go home? (A response to Goodman). *Reading Research Quarterly*, 1973, *8*, 494–501. (b)

Gollmer, C. A. On African symbolic messages. *The Journal of the (Royal) Anthropological Institute of Great Britain and Ireland*, 1885, *14*, 169–181 as cited by I. G. Gelb, *A study of writing; the foundations of grammatology.* Chicago: University of Chicago Press, 1952, p. 5.

Goodman, K. Reading: A psycholinguistic guessing game. In K. S. Goodman & J. Fleming (Eds.), *Selected papers from the IRA Pre-Convention Institute*, Boston, April 1968. Newark, Del.: International Reading Association, 1969.

Goodman, K. The 13th easy way to make learning to read difficult: A reaction to Gleitman and Rozin. *Reading Research Quarterly*, 1973, *8*, 484–493,

Harris, Z. S. *Methods in structural linguistics.* Chicago: University of Chicago Press, 1951.
Hoenigswald, H. M. *Language change and linguistic reconstruction.* Chicago: University of Chicago Press, 1960.
Huey, E. B. *The psychology and pedagogy of reading* (1908). Cambridge, Mass.: MIT Press, 1968.
Jakobson, R., Fant, C. G. M., & Halle, M. *Preliminaries to speech analysis.* Cambridge, Mass.: MIT Press, 1963.
Jeffery, L. H. *The local scripts of archaic Greece.* Oxford: Clarendon Press, 1961.
Jensen, H. *Sign, symbol and script* (3rd rev. ed.). New York: G. P. Putnam's Sons, 1969. (First American edition translated by George Allen & Unwin, Ltd.)
Klima, E. S. How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass.: MIT Press, 1972.
Kroeber, A. L. *Anthropology: Race, language, culture, psychology, prehistory* (new ed., rev.). New York: Harcourt, Brace, 1948.
Lenneberg, E. H. *Biological foundations of language.* New York: Wiley, 1967.
Leong, C. K. Hong Kong. In J. Downing (Ed.), *Comparative reading: Cross-national studies of behavior and processes in reading and writing.* New York: Macmillan, 1973.
Liberman, A. M. The grammars of speech and language. *Cognitive Psychology,* 1970, *1,* 301–323.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review,* 1967, *74,* 431–461.
Liberman, A. M., Delattre, P. C., & Cooper, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *American Journal of Psychology,* 1952, *65,* 497–516.
Liberman, A. M., Delattre, P. C., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America,* 1955, *27,* 769–773.
Liberman, I. Y. Segmentation of the spoken word and reading acquisition. *Bulletin of The Orton Society,* 1970, *23,* 65–77.
Lieberman, P. On the evolution of language: A unified view. *Cognition,* 1973, *2,* 59–94.
Martin, S. E. Nonalphabetic writing systems: Some observations. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass.: MIT Press, 1972.
Mattingly, I. G. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass.: MIT Press, 1972.
Moskowitz, B. A. On the status of vowel shift in English. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press, 1973.
Read, C. Pre-school children's knowledge of English phonology. *Harvard Educational Review,* 1971, *41,* 1–34.
Rozin, P. The evolution of intelligence and access to the cognitive unconscious. In J. Sprague & A. N. Epstein (Eds.), *Progress in psychobiology and physiological psychology,* Vol. 6. New York: Academic Press, 1975.
Rozin, P., Poritsky, S., & Sotsky, R. American children with reading problems can easily learn to read English represented by Chinese characters. *Science,* 1971, *171,* 1264–1267.
Savin, H. B. What the child knows about speech when he starts to read. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass.: MIT Press, 1972.
Savin, H. B., & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior,* 1970, *9,* 295–302.
Shaw, G. B. *Selection of his wit and wisdom.* (Compiled by C. T. Harnsberger.) New York: Follett, 1965.

Singer, H. IQ is and is not related to reading. In S. Wanat (Ed.), *Intelligence and reading.* Newark, Del.: International Reading Association, 1974.
Smith, F. *Understanding reading.* New York: Holt, Rinehart and Winston, 1971.
Stevens, K. N. Segments, features, and analysis by synthesis. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, Mass.: MIT Press, 1972.
Stevens, K. N., & House, A. S. Speech perception. In J. Tobias (Ed.), *Foundations of modern auditory theory,* Vol. 2. New York: Academic Press, 1972.
Thorndike, W. E. Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology,* 1971, *8,* 323–332. (Also in *Reading Research Quarterly,* 1970–1971, *6,* 425–434.)
Venezky, R. L. Regularity in reading and spelling. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading.* New York: Basic Books, 1970.
Walker, W. Notes on native writing systems and the design of native literacy programs. *Anthropological Linguistics,* 1969, *11,* 148–166.
Warren, R. M. Identification times for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics,* 1971, *9,* 345–349.