

# Records, Writing, and Decipherment

I. J. Gelb

Written records together with material remains derived from excavations form the main bases for our understanding of past civilizations and their underlying language systems. There is no systematic treatment of written records, and little attention has been paid to the interrelationship between ancient writing and language. Full systems of writing express language at two levels—morphological and phonetic—which give rise to three basic writing system types—logo-syllabic, syllabic, and alphabetic. Four categories of decipherment—based on our relative knowledge of the writing system and the language—are discussed: known writing/known language; unknown writing/known language; known writing/unknown language; and unknown writing/unknown language. From a cryptanalytic point of view there are two general decipherment methods: (1) utilization of external information to determine probable contents (e.g., bilingual texts), and (2) internal information from an analysis of the text itself (structure and typology). The assumption of the underlying language is critical for deciphering procedures and provides the test of successful decipherment.

## 1. *Written Sources*

Under written remains, sources, or records we understand texts, inscriptions, manuscripts, books, etc., which represent the written output of an individual, a nation, or any larger human configuration.

Knowledge about these written records either has been passed on traditionally from generation to generation, or it has been obtained through field explorations and excavations in modern times. Similarly, the understanding of these records either has been preserved through the ages, as in the case of Hebrew or Chinese; or it has been, or will have to be, recovered through decipherment in our times, as in the case of Egyptian hieroglyphic or Mesopotamian cuneiform. Under the general term “decipherment” we include the recovery of the writing systems and languages, the use of which was brought to a stop, and consequently lost, between the time of their use in the past and their recovery in the present.

293 *Gelb : Records, Writing, and Decipherment*

Written records together with material remains derived from excavations, such as buildings and tools, form the two main bases for our understanding of past civilizations. Evaluation of the written records also leads to the reconstruction of their underlying systems of writing.

It is very difficult, if not impossible, to separate the concept of written records from that of literature. Authors of works on literatures of ancient peoples often include under literature the whole body of written remains, as distinguished from material remains, such as buildings, tools, weapons, dress, etc. Literature to us means mainly belles-lettres; that is, works created under some kind of artistic, inspirational, or imaginative impulse. What may be called literature in one period or area may not be literature in another. For that reason, I am avoiding the term "literature" altogether in the following discussion.

The use of writing in written records can be either primary or secondary. Its primary use is found in records whose sole purpose is to convey the written message. Its secondary use is found in records which have another purpose in addition to that of communication—as on inscribed sculptures, reliefs, vases, gems, seals, coins, or measures.

The frequency and variety of records available from different areas and periods depend on many factors, the most important of which are the availability and cost of writing materials, their perishability, and the widely varying degrees of the application of writing.

The recovery of written records of dead civilizations is tied up with the progress of field explorations and excavations. Before the systematic explorations and excavations in Asia and Africa began early in the nineteenth century, our knowledge of the ancient Near Eastern civilizations was limited to what could be gathered from the Bible and Classical sources. This was very little indeed, certainly for the great civilizations which surrounded Palestine, if not for Palestine proper. It can be said without any exaggeration that almost our entire knowledge of ancient civilizations, such as Egypt and Mesopotamia, is derived from the efforts of the explorers and excavators in the field. Certain areas of the ancient Near East have been explored more thoroughly than others. Mesopotamia and Egypt are well explored, but hundreds of sites still remain to be excavated. Relatively little

known are eastern Anatolia, Syria, and Iran. No Median sites have as yet been discovered, and as a result our knowledge about the Medes does not go much beyond what we have known for years from Greek sources. The best explored and excavated area in the Near East is Biblical Palestine; contrary to the expectations, the attestation of written records from Palestine is very meager.

Progress of discoveries in the ancient Near East often follows the proverbial "excavator's luck." It is at two Assyrian sites, Nineveh and Dur-Sharrukin, that the excavators discovered the great archives which paved the way for our understanding of the Mesopotamian languages; not at Babylon, which yielded great palaces but scanty written attestation. And it is Knossos and Pylos, not Troy and Mycenae, that provided us with written records which formed the basis for the recovery of the older languages of the Aegean area. For years Ras Shamra was one of the hundreds of sites in Syria unknown to and untouched by archeology. In 1928 a chance discovery of a stone inscription at the site by a native plowman led to great excavations in subsequent years which have unfolded for us the most varied collection of written records ever made in the ancient Near East: Ugaritic, Amorite, Akkadian, Sumerian, Hurrian, Hittite, Egyptian, and Aegean.

The extent to which the frequency and variety of written records is connected with the degree of literacy in a given area cannot be discussed due to lack of data. My feeling is that the connections are negligible.

Another point which cannot be profitably discussed here because of lack of data is the relation of written to oral tradition. It is quite plausible that certain genres—such as epics, legends, historical narratives, popular songs, and proverbs—have traditionally been preserved from generation to generation by word of mouth rather than in writing. In certain areas and periods, the taboo against the use of writing may explain the lack of attestation of certain genres of written records.

I know of no systematic treatment of written records, their form, material, function, etc. For preliminary thoughts on the topic, cf. Gelb 1973.

## 2. *Epigraphy and Paleography*

The investigation of writing from the textual point of view has been traditionally the prime domain of the epigrapher and paleographer. While epigraphy is concerned mainly with inscriptions written in characters which are incised or scratched with a sharp tool on hard material (such as stone or metal), paleography deals mainly with manuscripts written in characters which are drawn or painted with pen, pencil, or brush on soft material (such as leather, papyrus, or paper). Since epigraphy means "writing on something" and paleography means "old writing," it is clear that the distinction made above between epigraphy and paleography cannot be justified on etymological grounds. This distinction has grown artificially over the years, as one scholar or another began to apply one or the other term to his own branch of study of written sources. Owing to the close interrelations between epigraphy and paleography, some scholars refuse to admit any distinction between the two, and prefer to use one of the two terms to cover both disciplines.

The main characteristics of epigraphy and paleography as listed above may be applied, with some leeway, to the ancient Near East (Mesopotamia, Egypt, Anatolia, etc.), the Classical world, China, India, the Islamic world, and, in general, to the Western writings from the Middle Ages on. But there are many difficulties and exceptions.

Paleography and epigraphy are involved in the study of written sources from two points of view: the purely formal aspect and the hermeneutical aspect.

The study of the purely formal aspect—possible without any understanding of the contents or without an extended study of the contents—is concerned, for example, with the kind, form, and size of the materials, techniques of writing, and with the form, order, and direction of writing. Hermeneutics—possible only with study of the contents—is concerned, for example, with the dating and localizing of written sources, their authorship, linguistic interpretation, and content evaluation.

A general scientific discipline of epigraphy and paleography does not exist. There are no studies which treat of the subject from a general, theoretical point of view, encompassing all the written sources wherever they may be found. The narrow fields which are

represented are, for example, West Semitic epigraphy, Arabic paleography, Greek and/or Latin epigraphy and/or paleography, or Chinese epigraphy and/or paleography. In all cases, these narrow fields of study form subdivisions of wider, but still linguistically or geographically defined areas of study, such as Semitic or Arabic philology, Classical philology, Assyriology, Sinology, etc. (Cf. also Gelb 1974).

### 3. *Language and Writing*

Very little work has been done in the field of relations of writing to language. The philologists have been concerned mainly with the historical evolution of writing and have paid little attention to the interrelations between writing and language. The linguists have been more concerned with the spoken language than with the written language. When interested in written languages, they have often limited their study to living written languages, neglecting the rich sources of information which can be culled from ancient written languages, and with it of pre-alphabetic systems. The question of the relationship of writing to language has been pursued in recent years mainly by scholars with a background in linguistics. Because of their interest in modern languages and writings, this implies generally relations between the alphabet and language. Due to the preference for a synchronic-descriptive approach rather than diachronic-historical, linguists generally have stressed the independent character of writing and have studied it as an independent system, rather than as a system ultimately based on and related to the underlying language. Both approaches seem justified to me. Scholars have as much right to point out the close interrelations between writing and language as they have to study writing as a relatively closed system, without being involved in matters of relationships between writing and language and the degree of their interdependence.

While the connections between language and writing are very close, there has never been a one-to-one correspondence between the elements of language and the signs of writing. The "fit" (a term used by Gleason 1955, p. 302; Voegelin and Voegelin 1961, pp. 85 f.) between language and writing is generally stronger in the earlier stages of a certain system of writing and weaker in its later stages. This is due to the fact that a writing system when first introduced

generally reproduces rather faithfully the underlying phonemic structure. In the course of time writing, more conservative than language, generally does not keep up with the continuous changes of language and, as time progresses, diverges more and more from its linguistic counterpart. A good example is the old Latin writing, with its relatively good fit between graphemes and phonemes, as compared with the present-day French or English writing, with their tremendous divergencies between graphemes and phonemes. In some cases, recent spelling reforms have helped to remedy the existing discrepancies between writing and language. The best fit between phonemes and graphemes has been achieved in the Korean writing in the sixteenth century and in the Finnish and Czech writings of modern times.

Families of writings are not related to families of languages. Note, for example, that English and Finnish are written in the Latin writing, but they belong to two different families of languages, and that the cuneiform writing was used in antiquity by peoples speaking many different languages.

The temporal primacy of language over writing has been taken for granted by most scholars, especially the American linguists. It has been contested by some European scholars who claim that writing is as old as oral language and gesture language. The fact is that full writing, expressing linguistic elements, originated only about five thousand years ago in Mesopotamia and Egypt and full writing is therefore much younger than language. Only if we include the semasiographic stage under writing can the assumption of equal temporal hierarchy of writing and language be admitted. As noted elsewhere, however, the semasiographic stage should not be treated as full writing, but as a forerunner of writing (Cf. also Gelb 1974).

#### 4. *Philology and Linguistics*

There is a good deal of confusion regarding the aims and methods of the fields of philology and linguistics. Philology, involved mainly in the study of the linguistic sources of a people or a group of peoples, forms the basic means for the comprehension of their respective cultures. It deals less with oral sources than with written sources, mainly literature (whatever its exact meaning). Philology deals with

the formal aspects of writing under the topic of epigraphy and paleography. Linguistics is concerned with the study of linguistic systems as reconstructed mainly from oral sources. Pursued less than the study of "oral language," the study of the "written language"—that is, of the language as it is used in written sources—is also a matter of linguistics. Linguistics deals with the structural aspects of writing under the heading of graphemics (Cf. Gelb 1974).

#### 5. *Semiotics*

Men communicate with each other by means of various systems of signs, of which the most universal are oral language or speech, a system of auditory communication; and gesture language and writing, two systems of visual communication. For the general science of signs several terms have been proposed, of which the term "semiotic" (used by Morris 1946, p. 3 f.) or "semiotics" (as here preferred) are the most appropriate. The term "semantics," which deals with the meaning of linguistic elements, should be carefully distinguished from the much broader term "semiotics." (For these and other terms, see Read 1948, pp. 78–97; Sebeok, et al., 1964, pp. 5 f. and 275 f.; and Gelb 1974.)

#### 6. *Grammatology*

The general field of study which deals with writing in the broadest sense has been called "grammatology" by Gelb (1952, p. 23), following partially the term "grammatography," which was used in 1861 in a title of a book on writing by Ballhorn published in England. The original German edition, from which the English translation was made, does not use the term. Equally appropriate as "grammatology" are the terms "grammatonomy" used by Boodberg (1957, p. 113) and "graphonomy" used by Hockett (1958, p. 539). Not acceptable are "graphology" of Halliday (1961, p. 244) because this term already has another, well-established meaning; and "graphemics," because this term is too narrow (Cf. also Gelb 1968 and Gelb 1974).

#### 7. *Graphemics*

The field of graphemics deals with full writing or phonography, as represented in systems of writing in which written signs generally have set correspondences in elements of language. Graphemics is concerned

mainly with the graphemes of a system—such signs as phonemo-grams, syllabograms, and morphograms—which find their equivalents in phonemes, syllables, and morphemes of language. Thus the field of graphemics deals with writing after it became a secondary transfer of the language, a vehicle by which elements of the spoken language were expressed in a more or less exact form by means of visual signs used conventionally. This took place for the first time about five thousand years ago in the Sumerian and Egyptian writings.

Instead of “graphemics,” other scholars use the terms “graphics” (Francis 1958) or “graphic linguistics” (Crossland 1956). All three terms are frequently misused by scholars who limit the terms to the study of alphabetic writings, overlooking or paying scant attention to all other types of writing, such as the logo-syllabic and syllabic systems (Cf. also Gelb 1968 and Gelb 1974).

#### 8. *Auxiliary Disciplines*

There are a number of scholarly disciplines which have some relation to the study of writing. The most important among them are: art, religion, cultural anthropology, child psychology, human pathology (amnesic aphasia), and animal ethology (biosemiotics) (Cf. the extensive treatment in Gelb 1974).

#### 9. *Types of Writing Systems*

The accompanying chart showing the typology of writing is given to enable readers to understand the terminology and classification of writing as used in the section devoted to methods of decipherment.

Scholars interested in descriptive and historical presentations of the various systems of writing should consult Gelb 1963, Jensen 1969, Diringier 1968, Février 1959, Cohen 1958, Istrin 1965, Friedrich 1966, or Barthel 1972.

#### 10. *Grammatology and Decipherment*

A basic prerequisite to a successful decipherment consists of a thorough acquaintance with the field of grammatology, specifically the structure and typology of writing. This should enable scholars to answer primary questions as to whether the texts to be deciphered represent real writing or no writing, original writing or forgeries, and should be helpful in indicating to scholars the type to which the

# THE TYPOLOGY OF WRITING

NO WRITING: Pictures

FORERUNNERS OF WRITING: Semasiography

1. Descriptive-Representational Devices
2. Identifying-Mnemonic Devices

FULL WRITING: Phonography

## 1. Logo-Syllabic

Sumerian	Egyptian	Proto-Elamite	Proto-Indic	Cretan	Hittite	Chinese
+Akkadian					hieroglyphics	+ Chinese
+Hittite						derived
cuneiform						systems

## 2. Syllabic

Type A (with vowel indication)	Type B (without vowel indication)	Type A (with vowel indication)	Type A (with vowel indication)
Cuneiform	West Semitic	Aegean	Japanese
Syllabaries:	Syllabaries:	Syllabaries:	Syllabary
Elamite	Proto-	Linear A	
Hurrian	Sinaitic	Linear B	
Urartian	Proto-	Phaistos disk	
Hattic,	Palestinian	Proto-Byblian	
Luwian,	Phoenician	Cypro-Minoan	
Palaic,	Hebrew	Cypriote	
etc.	Aramaic		
	etc.		

*Mixed Syllabic B + A*

Ugaritic  
Persian  
Cuneiform

*Mixed Syllabic A + Alphabetic*

Iberian

## 3. Alphabetic

Greek  
Aramaic (vocalized)  
Hebrew (vocalized)  
Latin  
Indic  
Korean  
etc.

particular writing belongs, e.g., alphabetic, syllabic, or word-syllabic.

It may sound rather preposterous to ask a scholar to take a stand on the question of whether the texts he is working on, do or do not represent writing, but even such well-known systems as the cuneiform and the hieroglyphic Egyptian were for a time considered by some scholars as merely ornamental or symbolic. It was not until the nineteenth century that scholars were able to provide evidence that the two systems represent real writings.

There are written remains, especially those appearing in bilingual versions, that have so many of the obvious characteristics of writing that scholars may be forgiven if they fail to provide evidence that the texts in question represent writing; but this is not always the case, and all too often scholars have failed to provide such evidence when it was required. I have in mind primarily the question of the Rongorongo boards of Easter Island which are generally taken to be composed in a real writing system, but which I have characterized as nothing else but a series of pictorial representations concocted for magical purposes.

The question as to whether certain written remains are original or modern forgeries can often be answered with the help of grammatology. Modern forgeries are spurious fabrications usually made by or for dealers in antiquities for the purpose of financial profit; some well-known fakes were made for the purpose of national glorification, while other hoaxes were perpetrated by scholars, I assume for amusement.

Among the best-known forgeries of modern times are the "Glozel finds" in France, whose authenticity was defended tenaciously by the renowned French archaeologist S. Reinach. On the other hand, the Phaistos Disk in Crete and the Stone of King Mesha of Moab, once thought spurious by many scholars, are now almost universally considered to be authentic.

Knowledge of grammatology appears most useful to scholars in the definition of the type of writing on which they are working. From a practical point of view, this is also the most common application of grammatology to the field of decipherment. It is relatively easy to count the number of graphemes (i.e., of signs with distinctive features) in a given writing. The number thus reached might be about

sixty graphemes. The teachings of grammatology tell us that a writing consisting of about sixty graphemes should represent a syllabary. Similarly, if the number of the counted graphemes reaches several hundred, it is safe to assume in the light of grammatology that the underlying writing represents a logo-syllabic system.

### 11. *Methods of Decipherment*

In contrast to the extensive literature in the field of cryptology, both scholarly and popular, there is a lamentable dearth on the subject of the decipherment of extinct writings and languages. More often than not, the decipherers themselves have failed to preface their work with any remarks on their methodology. The history of decipherment is treated in Friedrich 1966a, a book which also appeared in English and Russian versions. The latter contains extensive additions by I. M. Diakonoff.

More or less popular accounts of decipherment can be found in Doblhofer 1957 (also in English and French editions), Cleator 1959, and Gordon 1968. A bulky volume on cryptanalysis, written by an ex-journalist, David Kahn 1967, has a chapter on extinct writings and languages which is uneven in quality. Good discussions of the decipherment of individual writings are to be found in standard manuals on writing by Jensen (1969), Diringer (1968), Février (1959) Cohen (1958), Istrin (1965), and Friedrich (1966).

I know of only two relatively brief studies dedicated to the methodology of decipherment: Modrzejewski 1930 and Aalto 1945. One can also consult the section "Methodisches zur Entschlüsselung verschollener Schriften und Sprachen" in Friedrich (1966a, pp. 123-8). Some thoughts on the cultural basis of decipherment are expressed in Voegelin and Voegelin 1963. My own ideas on the methodology of decipherment were offered in Gelb 1973. Rewritten completely with the collaboration of Robert M. Whiting, this article was presented as the opening paper at the three-day International Symposium on Undeciphered Languages organized in celebration of the 150th anniversary of the Royal Asiatic Society in London.

There are many stories connected with the decipherment of ancient writings and the recovery of forgotten languages but these stories need not be retold here. Furthermore, these stories usually deal only

with the discovery of the key, that brief moment of insight when some datum is arrived at, which when inserted causes the rest of the puzzle to fall into place. But what we are interested in here is the tremendous amount of work—routine but necessary—which precedes that moment and makes it possible, and the even more tremendous amount of work which follows that moment and results in the recovery of the language.

There is a clear connection between the decipherment of extinct writings and languages used for the purpose of normal communication and cryptanalysis, which deals with the decipherment of writings used for the purpose of secret communication. For this reason, it is frequently convenient to use the terminology of cryptology when referring to the problems encountered by the decipherer of ancient writings. However, one important difference should be stressed. Cryptography or secret writing attempts to lay obstacles in the path of the non-intended reader which will hopefully make the interpretation of the message impossible, while the writing systems used by the ancients were meant for direct communication with the reader. This is important because of a linguistic feature known as redundancy which the designers of cryptographic systems try to eliminate to the greatest extent possible.

In spoken language, redundancy allows us to understand what has been said even if some of the sounds are not heard or are badly distorted, because of the fact that certain sequences of sounds are more frequent than others while some sequences do not occur at all. Thus if several phonemes of an utterance are missed, our knowledge of the sound patterns of the language allows us to eliminate very quickly the sound sequences which are not allowed and to reconstruct the one which is. If more than one sequence is allowable, context usually permits us to make the final decision.

This same feature of redundancy is also present in writing systems: certain combinations of signs are more frequent than others while some do not occur at all. Most of the methods of decipherment make use of this fact, and the degree of difficulty or ease with which a writing system can be deciphered depends in large measure on the extent to which this feature can be recognized and exploited.

Full systems of writing express language on two levels, morphological (logography) or phonetic (syllabary, alphabet). This gives

rise to three basic types of full writing systems: logo-syllabic, syllabic, and alphabetic. The latter two types are phonetic while the first combines morphological elements (logograms) and phonetic elements (syllabograms). Even so, the syllabic and alphabetic systems usually contain morphemic elements to a greater or lesser degree. In cryptography, coding at the phonetic level is called cipher, while replacement of morphemes is known as code. There are two basic types of cipher: substitution and transposition.

Redundancy is present at the morphemic level, but since the number of morphemes in a language is so much greater than the number of phonemes, the effect of redundant morphemes is not as great as that of phonetic redundancy. For this reason, a code is generally harder to break than a cipher, and a writing system based mainly on logograms will be more difficult to decipher than one based on phonograms, given the same degree of availability of materials and of knowledge of the underlying language.

The preceding statement implies that the degree to which the underlying language of a writing system is known affects the decipherment of the writing system. This can be stated more strongly: provided sufficient text is available, a phonetic system of writing can and ultimately must be deciphered if the underlying language is known. It should be noted, however, that the converse is not true. Interpretation of a language is a matter of linguistic analysis which determines the morphological and syntactic rules governing it. These rules are much more extensive and complicated than the phonological ones; hence it is much more difficult to deduce the former from the latter than vice versa.

In view of this, we should really speak of decipherment in connection with writing systems and speak of the recovery or interpretation of languages. In popular usage, however, the term decipherment is also used in connection with languages as one speaks of "the decipherment of Etruscan."

Using the term decipherment in its wider sense we can classify decipherments into several types based on the extent of our knowledge of the two elements involved: the writing system and the language. There are four possible situations, only three of which present problems to the decipherer. The four categories are:

- Type O: known writing and known language
- Type I: unknown writing and known language
- Type II: known writing and unknown language
- Type III: unknown writing and unknown language

It must be pointed out here that “known” and “unknown” in this application are not absolutes but shade into each other in a manner which cannot be expressed quantitatively. This is especially true of languages where we can speak of a well known one such as Latin, a less known one such as Sumerian or Hurrian, or a virtually unknown one such as Etruscan. Furthermore, a language may be unknown itself but be more or less closely related to a known language or group of languages; a fact which, once established, moves it closer to the category of a known language. Keeping this in mind, we shall return to our four types of decipherments.

Type O, a known language written in a known writing system, is generally considered trivial and requires little discussion. However, despite the fact that this type offers no difficulty in decipherment, examples of it can provide us with valuable information especially if the writing system is other than the one normally used for the language.

I am thinking here (1) of the inscriptions written in the Phoenician (Punic) language but using the Greek or Latin alphabet, which furnish useful information for the vocalization of that language which is not expressed by its normal writing system; and (2) of the limited amount of Sumerian and Akkadian material written in the Greek alphabet which not only increases our phonological knowledge of these languages, but also serves as a convenient check on the validity of the decipherment of cuneiform.

Type O corresponds to what is known in cryptography as plain-text; that is, an uncoded message.

Decipherments of Type I, an unknown writing system used for a known language, vary in degree of difficulty depending on the nature of the writing system. We have already seen that there are three basic types of writing systems: alphabetic, syllabic, and logo-syllabic. Deciphering these writings makes extensive use of the techniques of cryptanalysis.

If the writing system is alphabetic, the problem resembles that of a simple substitution cipher and, provided sufficient text is available, is very easy to solve utilizing the redundancy features of the known language.

If the writing is syllabic, the problem is slightly more complex because of the larger number of graphemes and their more complex phonetic structure. However, syllabaries have their own redundancy features and such a system will eventually fall to analysis. It should be noted that the amount of text required to guarantee a unique solution of such a system is considerably larger than for an alphabetic system. It is for this reason that the Phaistos syllabary, which probably hides a known language, remains undeciphered. There is just not enough text available to provide an unambiguous solution.

If the writing is logo-syllabic, the problem can become quite complex. It amounts to code mixed with polygraphic encipherment. It has already been noted that code is considerably harder to break than cipher. For this reason the proper cryptanalytical procedure would be to attack the cipher (syllabograms) first and then deduce the code (logograms) from the knowledge of the morpholexical structure of the language. On the practical level, other factors often intervene which simplify this process, as we shall see when we discuss methodology.

The history of Type I decipherments bears out this general picture. Thus of the writings in this category, the most easily deciphered have been the Phoenician and Ugaritic writings, which consist of only 22 to 30 signs. (Although syllabic in nature, from the point of view of cryptology these systems behave like alphabets.) More difficult problems were posed by the Old Persian, Cypriote, and Linear B systems—all of the class of syllabaries—with the number of signs varying from approximately 40 to 80. Much more difficult were the decipherment of logo-syllabic systems such as Egyptian hieroglyphic, Akkadian cuneiform, and Hittite hieroglyphic—involving hundreds of signs.

As we have already noted, Type II decipherments are not, strictly speaking, decipherments, but linguistic analyses. The category of an unknown language corresponds in cryptography to code, which involves the substitution of an unknown linguistic element for a known

one. Such systems are extremely difficult to break, and a language which is truly unknown is virtually impossible to reconstruct using cryptanalytic methods because of the tremendous amount of text required. This is not to deny the possibility of reconstructing an unknown language, but only to indicate the very low probability of recovering a significant part of it by internal means alone. In every case of a Type II decipherment, external sources of information have played a large role. Even if a language cannot easily be reconstructed by internal means, it is still possible to construct a formal (descriptive) grammar of such a language by making a catalog of short repetitive elements and classifying them as affixes or function words (as opposed to content words), even though their meaning or use may be unknown. This describes our present state of knowledge about the Etruscan language, for example.

Type III decipherments, involving an unknown writing system and an unknown language, are clearly the most difficult of all. If such a case occurs in cultural isolation where no outside information can be brought to bear, it can be considered undecipherable. External sources can sometimes provide clues which can reduce it to a Type I or Type II problem. If the language is truly unknown, reduction to Type II is not a significant improvement in understandability. As we have seen, our knowledge of the phonetic shape of Etruscan does little to reduce its obscurity.

Type III situations have been compared by some authors to the cryptographic category of enciphered code, but this is not true. Cryptographic codes have no phonological shape since the code words are usually series of unrelated symbols. An unknown language is code whether it can be read or not. The purpose of enciphering code is to further reduce redundancy by making frequently used code words appear differently at different places in the text. In Type III these repetitive morphemes are clearly visible and can be collected to form a preliminary grammar of the language in exactly the same manner as described for Type II.

Having outlined the types of decipherment which may be encountered and having assessed their relative difficulty, let us consider some of the methods involved in decipherment.

The approaches to the recovery of extinct and unknown writings and languages have in the past been almost uniformly characterized by haphazard touch-and-go procedures. With very few laudable exceptions, the would-be decipherers have approached their task without any idea of cryptanalytic techniques. In the light of this almost total lack of systematic methodology, it is astonishing to note how frequently the tenacious efforts of scholars have led to a successful decipherment.

From a cryptanalytic point of view, we can distinguish two broad areas of methodology. These deal with the utilization of external information, or what can be determined about the probable contents of the cryptogram from outside sources; and internal information, or what can be learned from an analysis of the cryptogram itself.

Every cryptogram has a certain a priori probability of containing a given message. If the cryptanalyst can guess what the message is, the solution of the cryptogram will be much easier. Even having only a general idea about what the message might be will facilitate decipherment.

In considering the application of this concept to ancient writings, it is interesting to note that frequently the key to a decipherment has been provided by a source external to the writing under study. In most cases this has been one or more proper names known traditionally, such as the Persian royal names found in Herodotus, or provided by a bilingual inscription, such as the Rosetta Stone—to name just two of the best known examples.

Therefore, before doing any work on the decipherment of a specific writing or language, a would-be decipherer must become acquainted with the historical-geographical background of the area from which it comes. One should remember Champollion, who spent years in familiarizing himself with the history, geography, religion, and languages of Egypt as preserved in the Classical sources or by tradition, before he dared even to suggest the reading of a single sign of the Egyptian writing.

It is generally considered that a bilingual text is a type of external information which will immediately produce a unique solution of an ancient writing or language. The reasons why this is not true will become apparent if we remember the cryptological analogies which we have established. If a cryptanalyst has both the plain and

enciphered text of a message, his problem is solved since ciphering is a linear and reversible transformation of one writing system into another. Furthermore, this information allows him to determine the key, that is, the rules under which the transformation was made. But we have already established that the only class of decipherment which corresponds to cryptographic cipher is Type I, a known language written in an unknown writing, and that this corresponds to simple substitution cipher. It can easily be shown that for substitution cipher, the choice of the cipher alphabet does not affect the general appearance of the ciphertext, or in other words, all substitution encipherments of a given message are equivalent and all that is significant is the pattern of letter repetition. But the patterns of letter repetition are generated by the language, or, more specifically, by the redundancy features of the language. The immediate conclusion is that, provided there is sufficient text for these redundancy features to assert themselves, the only key necessary for the solution of a substitution cryptogram is the language in which it is written!

The reason that having a bilingual text is not equivalent to a cryptanalyst having the plain and enciphered text of a message is the simple fact that the two parts of the bilingual *are not written in the same language*. Hence we are not dealing with an enciphered version of the same text, but rather with an encoded version of it. While encipherment is a linear reversible process, translation is not—as anyone who has given any thought to the problem of machine translation quickly realizes. Translation of a passage of moderate length into another language and then back into the original by another translator would almost never result in exactly the same text as the original. If a cryptanalyst has the plaintext and encoded version of a message, he cannot guarantee a unique solution for each code group; and even if by chance he could, knowledge of a hundred or so code groups would hardly make a dent in a code consisting of ten thousand groups.

Of what value, then, are bilingual texts to the decipherer of ancient writings and languages? Let us consider this in the light of the possible types of decipherments. Since the known language of the bilingual represents a coded version of the unknown one—for Type II and III decipherments, which represent code—this means that we now have the same text encoded in two different systems one of which is known. This allows us to determine some, if not all, of the unknown code

groups. As we have seen, unless this number is quite large no significant increase in our knowledge is gained. Thus for Type II decipherments our understanding of Sumerian has been advanced quite far by the tremendous amount of Akkadian-Sumerian bilingual material available, while the relatively short Etruscan-Punic bilingual recently discovered at Pyrgi in Italy has added little to our knowledge of Etruscan.

For Type III decipherments the situation is similar. But there is an important extra. Apparent Type III situations frequently hide a less difficult decipherment, usually Type I, and a bilingual sometimes allows us to break a code group which makes this apparent. This is exactly what happened during the decipherment of the Cypriote syllabary using a Cypriote-Phoenician bilingual. The language turned out to be Greek, but because the redundancy features were altered by the writing conventions of the syllabic script from those expected of classical Greek, it was unrecognizable as such when enciphered.

Another feature of bilingual inscriptions which requires discussion brings us back to our original reason for rejecting the bilingual as an immediate solution, namely that it represents code, not cipher. Cryptographic codes always have groups for a syllabary and an alphabet so that words not included in the code can be spelled out. The most frequent need for these groups is to express proper names. Proper names are usually not translated into another language, but are simply transcribed. Thus if a bilingual text has a proper name in one version and the exact position of this name can be located in the other, one can consider the two occurrences as one phonological sequence expressed in two writing systems, or cipher. Since the plaintext is known, the phonetic values of the signs used to encipher it are established. If enough such occurrences could be found, a Type III situation could be reduced to Type II, and if during the process enough information about the language were found to indicate that the language was known, it would quickly reduce to Type O.

Finally, a bilingual can be useful as a check on a decipherment made by other means. It should be noted here, however, that not all bilingual inscriptions are well suited to this use, since the two sections might be either very reliable, verbatim translations, or one might be simply a loose paraphrase of the other.

Although other sources of external information may come to bear on the decipherment, it is best to shift at this point to internal analysis and see what preliminary steps should be taken.

Equally as important as a thorough study of the area from which the writing to be deciphered comes from is a sound acquaintance with the field of grammatology, specifically the structure and typology of writing (see above, Section 10).

From a consideration of the writing system as a whole, we move to graphotactics or the evaluation of such graphic characteristics as the position, sequence, arrangement, and direction of signs, and word division. The sequence of signs in a sign group may be orderly (as in the classical Latin writing) or disorderly (as in the earliest Sumerian writing). The sequence of signs may be from left to right, from right to left, or both, or from top to bottom, or from top to bottom in the sequence of individual signs but from right to left or left to right in the sequence of columns or rows of signs. Word division may not be indicated at all in the writing, or it may be indicated by special marks in the form of dots or strokes, or by a space.

More information about the probable content of the inscription can be deduced from repetitive schematic arrangements in the text. Such a repetitive sequence may be called a "routine." Routines offer exceedingly valuable information of a quasi-bilingual nature and are usually easy to detect without any elaborate statistical analysis. For example, if the last line of a text includes a fairly low number, it is a reasonable assumption that it is a date routine, and one should expect the word "year" and perhaps a royal name. Similarly if there are numbers throughout the text and the last line includes a higher number with a word before or after it, it is almost certain that the word is "total" or "sum."

Still another source of information about the probable context of an inscription can be utilized if the script is logo-syllabic. This source is the interpretation of logograms.

It was pointed out earlier that an unknown logo-syllabic system could be considered as a mixture of enciphered text and code. While the comparison is typologically valid, no cryptographer would ever use such a system. Because even a small amount of context can be extremely valuable in solving a cryptogram, there is an absolute prohibition in cryptography against plaintext appearing in the body

of an encrypted message. For the same reason, no cryptographer would risk having code groups recognized among the enciphered text and would insert the code groups at the plaintext level and then encipher both. But ancient writing systems did not strive for secrecy and the logograms are in full view of the decipherer.

If a logo-syllabic writing appears in a Type II situation (known writing, unknown language), the interpretation of all logograms may be known, a situation which generates numerous context clues. The key to cuneiform Hittite was actually discovered in this manner.

If a logo-syllabic system is unknown but largely pictographic, it may be possible to deduce the interpretation of some logograms from their pictorial representations. This is more often true of logograms denoting nouns than of those denoting verbs. But there are pitfalls in this approach because in a logo-syllabic system, syllabograms have normally developed from logograms by means of phonetic transfer, and one must be certain that he is not trying to interpret a syllabic sign as a logogram.

There are several other ways of distinguishing and interpreting logograms. Logograms are at times distinguished from syllabograms by a special mark. In a sign group composed of several signs, it is frequently possible to assume that the first or sometimes the first two signs represent the logogram, and the rest of the signs denote the syllabic indicators. Finally, the use of semantic indicators or determinatives preceding or following a logogram helps in distinguishing logograms and at the same time—because they are logograms themselves—helps in ascertaining the sphere of meaning to which they belong.

The limitations involved in the understanding of logograms without being able to read them form no great obstacle to the understanding of the texts. There are well-known and fully developed writing systems, such as Sumerian and cuneiform Hittite, which abound in logograms with clear meaning but unknown reading. If we can interpret all the logograms of a logo-syllabic system, the residue is obviously a syllabic system with plaintext clues scattered through it. Such a system would a priori be easier to solve than a normal syllabic one.

Having exhausted all sources of information about the probable content of the text, the next step is a systematic application of

statistics to determine the redundancy characteristics of the writing or the language.

Statistical analysis involves making frequency lists of the individual signs, of sequences of signs, and of signs which appear more frequently than others in initial or final position in a word. These statistics about the frequency with which certain signs appear in combinations or in certain positions are in fact the redundancy features of the language which generated them, perhaps modified by the limitations and orthographic conventions of the writing system employed. Thus, for example, an orthography which does not express double consonants will distort the redundancy features of a language that does. This is a problem which has to be resolved by outside information, since the distortion of the redundancy features will make the enciphered language appear to be different, or, more likely, unknown. However, if some phonetic values can be established the language will soon be recognized, even with the reduced redundancy, because redundancy is merely a measure of how much of a transmitted message can be omitted without impairing its intelligibility.

Another useful outcome of statistics, especially for syllabic systems of the Aegean type (only single vowels and consonant plus vowel signs), is the possibility of constructing a grid system which limits the phonetic values which may be assigned to a given sign by grouping the signs which interchange under certain circumstances together. By this method it is possible to group together signs which should have the same consonant but a different vowel, or alternatively, the same vowel but a different consonant. These groupings may be arranged to form a grid with the consonant as one coordinate, and the vowel as the other. All this may be accomplished without being able to read a single sign. The usefulness of the grid stems from the fact that once a few phonetic values are determined, the rest fall into place almost automatically.

Discussion of statistics leads to the application of computers to decipherment. The compilation of the statistical analyses I have described can be a very laborious process, especially if the amount of textual material available is large. In accomplishing this work, computers can be utilized most profitably. I estimate that the use of a computer in my work on Amorite has saved me several man-years of tedious repetitive labor.

While the computer is very useful for collecting, sorting, and counting, analysis should, at least at present, be left to the human mind. The computer has the same problems with ancient languages as with modern ones, and the same limitations which presently make computer translations unreliable (except for very narrow purposes) hinder their usefulness in analyzing ancient languages.

The first attempt to decipher an ancient writing with the help of computers was the so-called "Siberian" decipherment of the Maya writing. While the results are disappointing, the method is sound and may hold promise for the future. The first step was the collection of both linguistic and graphic data from available sources. Next the redundancy features of both the linguistic and graphic data were determined by computer. Finally, the output of the linguistic and graphic data was correlated in order to reach certain conclusions as to the reading or meaning of the Maya signs on the basis of comparable frequencies and distribution within the Maya linguistic material.

It is my belief that the failure of the decipherment is not due primarily to flaws of methodology, but to the fact that Mayan glyphs do not represent a full phonetic writing system. The decipherment is apparently of Type I (unknown writing, known language), and we have already shown that for this category—provided sufficient text is available—a phonetic writing can, and ultimately must, be deciphered if the underlying language is known. Since there is plenty of Mayan text available and the language is known, the fact that the writing has defied even the most sophisticated attempts at decipherment leads me to the conclusion that it is not a phonetic writing system. Therefore the lack of success by the computer does not diminish its potential for this type of decipherment.

I have left until now discussion of the assumption of an underlying language because it is involved in the test of a decipherment. But I have tried to stress, both in my discussion of the types of decipherments and elsewhere, that the extent to which the underlying language is known, or can be recognized as known, governs almost completely the difficulty of the decipherment. Hence the assumption of an underlying language is one of the most basic premises of the decipherment. If the assumption of the underlying language is wrong,

then the decipherment is wrong. Note, however, that the decipherment may progress up to a point even if the assumption of the language is wrong. A good example is Linear B where a grid structure for many of the signs had been worked out, but which remained undeciphered as long as it was assumed that its underlying language was Etruscan.

A preliminary assumption of an underlying language is the result of logical deductions about the linguistic situation of the area from which the writing or the language to be deciphered comes. The assumption is more or less self-evident in the great majority of cases. Thus it was logical to assume that ancient Egyptian would be a language ancestral to Coptic, spoken in Egypt until quite recently; just as it was plausible to start with the presupposition that the language of the Ugaritic texts would be a Semitic one, closely related to the other Semitic languages known from the general area, such as Phoenician, Aramaic, and Hebrew. The probabilities as to the underlying language had to be weighed in varying degrees in the case of other decipherments, such as hieroglyphic Hittite, Linear B, Cypriote, and Iberian.

The assumption of an underlying language may be plausible or even probable, but no decipherment is possible if the assumed language has no parallels in any known group of languages. This may be the case for Cretan hieroglyphic, which may be written for Minoan, a native Mediterranean language possibly unrelated to any other known language.

As a test of a decipherment, we should insist on the translation of a full text, not simply excerpts. It is frequently possible to provide a persuasive interpretation for a small portion of the text, such as a phrase or even a sentence, but this cannot be a decipherment if the rest of the text is gibberish.

The translation of the text must be consonant with the preliminary expectations about the contents of the text. It is reasonable to assume that Proto-Elamite texts, full of numbers and measures, would represent simple administrative accounts dealing with the day-by-day routine of running a household, such as the listing of incoming and outgoing commodities. A decipherment that would read magic conjurations into these texts would be highly suspect.

But perhaps the best test of a decipherment is repeatability. That is, it should be possible to decipher another text, preferably one the

original decipherer has not seen, using the decipherer's methods. Perhaps the most spectacular example of this was conclusive proof of the correctness of the decipherment of Linear B provided by the discovery of the famous "tripod" text, which not only deciphered into good Greek words, but also had pictographic representations which showed exactly the objects which those words represented.

## REFERENCES

- Aalto, Pennti. 1945. Notes on methods of decipherment of unknown writings and languages. *Studia Orientalia* 11/4.
- Barthel, Gustav. 1972. Konnte Adam schreiben? *Weltgeschichte der Schrift*. Köln.
- Boodberg, Peter. 1957. The Chinese script: An essay on nomenclature (the First Hecaton). Studies presented to Yuen Ren Chao, pp. 113–20. Taipei, Taiwan.
- Cleator, P. E. 1959. *Lost languages*. London.
- Cohen, Marcel. 1958. *La grande invention de l'écriture et son évolution*. Three volumes: Texte, Documentation et Index, Planches. Paris.
- Crossland, R. A. 1956. Graphic linguistics and its terminology. *Mechanical Translation* 3/1; republished in *Proceedings of the University of Durham Philosophical Society* I B (1957).
- Diringer, David. 1968. *The alphabet: A key to the history of mankind*. London and New York, 1948; 2nd edition, 1949; 3rd edition, 1968.
- Dobhofer, Ernest. 1957. *Voices in stone: The decipherment of ancient scripts and writings*. London and New York. [Translation of Dobhofer's *Zeichen und Wunder* (Wien, Berlin, Stuttgart, 1957).]
- Février, James G. 1959. *Histoire de l'écriture*. Paris, 1948; nouvelle édition 1959.
- Francis, W. Nelson. 1958. *The structure of American English*. New York.
- Friedrich, J. 1966a. *Entzifferung verschollener Schriften und Sprachen*. Berlin, 1954; 2nd edition, 1966. [Also American and Russian editions.]
- . 1966b. *Geschichte der Schrift, unter besonderer Berücksichtigung ihrer geistigen Entwicklung*. Heidelberg.
- Gelb, I. J. 1963. *A study of writing: The foundations of grammatology*. London and Chicago, 1952; revised edition, 1963. [Also editions in German, French, Italian, etc.]
- . 1968. Grammatology and graphemics. Papers from the Fourth Regional Meeting, Chicago Linguistic Society, ed. by Bill J. Darden, Charles-James N. Bailey, and Alice Davidson, pp. 194–201. Chicago.
- . 1973. Written records and decipherment. *Current trends in linguistics* 11, ed. by Thomas A. Sebeok, pp. 253–84.
- . 1974. Article: Writing. *Encyclopaedia Britannica*.
- Gleason, H. A., Jr. 1955. *An introduction to descriptive linguistics*. New York; 2nd edition, 1961.
- Gordon, Cyrus, H. 1968. *Forgotten scripts: How they were deciphered and their impact on contemporary culture*. New York.
- Halliday, M. A. K. 1961. Categories of the theory of grammar. *Word* 17.241–92.

- Hockett, Charles F. 1958. *A course in modern linguistics*. New York.
- Kahn, David, 1967. *The codebreakers: The story of secret writing*. New York.
- Modrzejewski, Annelise. 1930. *Zum Problem der Schrift: Ein Beitrag zur Theorie der Entzifferung*. Dissertation; Breslau.
- Morris, Charles. 1946. *Signs, language, and behaviour*. New York.
- Read, Allen W. 1948. An account of the word "semantic." *Word* 4.73-97.
- Sebeok, T. A., et al. 1964. *Approaches to semantics: Cultural anthropology, education, linguistics, psychiatry, psychology*. The Hague.
- Voegelin, C. F., and F. M. Voegelin. 1961. Typological classification of systems with included, excluded, and self-sufficient alphabets. *Anthropological Linguistics* 3.55-96
- . 1963. Patterns of discovery in the decipherment of different types of alphabets. *American Anthropologist* 65.1231-53.

This article has been adapted from a paper read at the symposium—Language and Texts: the Nature of Linguistic Evidence—organized by the Center for Co-ordination of Ancient and Modern Studies at The University of Michigan in Ann Arbor in March 1974.