STUDIA ORIENTALIA
EDIDIT SOCIETAS ORIENTALIS FENNICA
XI. 4

NOTES ON METHODS OF DECIPHERMENT OF UNKNOWN WRITINGS AND LANGUAGES

BY
PENTTI AALTO

HELSINGFORSIAE 1945
SOCIETAS ORIENTALIS FENNICA

HELSINKI 1945
Printed by Suomalaisen Kirjallisuuden Seuran Kirjapainon Oy.

Notes on Methods of Decipherment of Unknown Writings and Languages

BY

PENTTI AALTO

Most of our information about the history and culture of the ancient peoples in the Near- and Middle-East comes from the original written monuments of these peoples, and was almost unknown a hundred years ago. The acquirement of all the knowledge we have to-day has been made possible only by solving the problems of the ancient scriptures, hieroglyphs, cuneiforms etc., which during thousands of years had challenged all the efforts made to decipher and interpret them and their language. The names of the genial men who have performed this miracle are rather well known, but that can not be said about their achievements. Have these, in fact, been miracles or, if not, by what methods, by what kind of logical thinking were these results made possible?

The methods of decipherment of unknown scriptures and of interpreting strange languages, as linguistic tasks, have not been systematically treated, while the deciphering of secret writings, proper cyphers, has been developed to a true science as to both thepreparing and the breaking of them: namely, cryptology.[1]

[1] P. Jensen in *ZDMG* 48 (1894) p. 260 summarizes these methods under the headline »Kritik der an und für sich möglichen und bisher befolgten Entzifferungsmethoden». According to him, the decrypment can be made: »1) Mit Hülfe von Bilinguen, deren einer Theil in dieser Schrift abgefasst ist, der andere entsprechende in einer uns bekannten Sprache in für uns lesbarer Schrift. 2) Mit Hülfe von Texten, die nachgewiesenermassen in derselben oder

Here we have a ready, complete system of definitions, rules and methods, which for the most part are very well applicable to the study and solution of all types of unknown writings and to the interpretation of strange languages. Problems of this kind can always be identified with the deciphering of some general type of ciphergrams. It is known that Sumerian, earlier *e. g.* by Halévy, and for a time also by Fr. Delitzsch, was regarded as a mere secret script used by Babylonian priests. And recently, H. GRIMME in *Die Buchstabendubletten des Sinaialphabets*, *Westfäl. Studien* 1928 pp. 302—312, regarded this alphabet as some secret or private one. (Cited by H. JENSEN, *Die Schrift*, Hamburg / Glückstadt, 1936, p. 181. See also H. GRIMME, *Die neuen Sinaischrift-Denkmäler*, *ZDMG* N. F. XII p. 177—.) Also the so-called Szekli-alphabet — nearly related to Old Turkish[1] script, both deciphered by V. THOMSEN — was by some philologists considered to be an artificial secret writing. (cfr. PEDERSEN, *Språkvetenskapen*, Stockholm 1924, pp. 176—180.) Also the methods of decyphering and interpreting texts of this kind are applicable to solving of ciphergrams and vice versa.[2]

einer möglichst nahe damit verwandten Sprache in für uns lesbarer anderer Schrift geschrieben sind. 3) Mit Hülfe von Texten, die in uns bekannter Sprache in derselben, für uns von vorne herein nicht lesbaren oder in einer nachgewiesenermassen damit verwandten Schrift abgefasst sind. 4) Lediglich durch Studium der Inschriften allein, wobei die Aussicht auf Erfolg um so grösser sein wird, je mehr analoge Inschriften dieser Gattung Vergleiche untereinander ermöglichen.»

Various methodical details have also been treated by some authors, *e. g.* by H. Bauer in his *Zur Entzifferung der neuentdeckten Sinaischrift*, Halle 1918, p. 20; by J. Sundwall in *Altkretische Urkundenstudien*, *Acta Acad. Ab.* 2. 1937; by V. Thomsen in his *Une inscription de la trouvaille d'or de Nagy-Szent-Miklós*, *Danske Videnskabs-Selsk. Medd.* I, i, 1917; by Bauer in *ZDMG* N. F. IX, 1930 p. 251—, etc.

[1] A. C. Emre in his somewhat phantastic study *Sur l'origine de l'alphabet vieux-turc* (Istanbul 1938) contends that the Old Turkish alphabet is nearly related to the Sumerian linear pictograms.

[2] The decipherer of the cuneiforms, G. F. Grotefend tells himself: »Im Juli, als mein Freund Fiorillo, Sekretär der Königlichen Bibliothek, beim Spazierengehen mit mir darüber verhandelte, ob der Inhalt von Schriften festgestellt

Thus, at the beginning of the Great War in 1914, Admiral Sir Reginald Hall, the chief of the British Naval Intelligence, descended upon the British Museum and drafted the savants deciphering hieroglyphs to break codes, which they »did so well that it seemed like sorcery». (*The Living Age*, New York, March 1940 p. 46.).

Therefore, in studying the methods which have been and which are to be used in deciphering all types of writings and languages previously unknown, we are entitled to use the definitions, terms and the methodical disposition of problems used by cryptology. The usual types of cyphers are based on two fundamental methods: transposition and substitution.[1]

Of these only the methods based on substitution are the object of the following study. Three different kinds of problems can be distinguished: 1) the script only, 2) the language only, or 3) both simultaneously — are unknown to us.[2]

I. The Script Is Unknown

If only the scripture is to be deciphered, the language being either known or easily deduced with the aid of related languages we can regard the problem as identical with the simple substitution

---

werden könne, deren alphabet und Sprache gänzlich unbekannt seien, behauptete ich, da ich schon von früher her gewohnt war, Sätze der heimischen Sprache, die mit unbekannten Zeichen geschrieben waren, zu deuten, dass das sicherlich möglich sei» (Cited by Messerschmidt, *Die Entzifferung der Keilschrift*, *A. O.* 5, 2, Leipzig 1903, p. 8.); cf. also the study of H. Bauer cited above.

[1] In the former, the elements of plain language are transposed in some new order unintelligible for outsiders, in the latter, the textual elements are substituted by others according to a system previously agreed upon.

[2] In the following, little or no attention can be paid as to whether the results of decipherment serving as examples are in every detail correct or not; the literature referred to is also regrettably incomplete, because of the fact that many important works have unfortunately not been available at the present moment due to the evacuation of libraries during the recent war.

ciphers in cryptography. In breaking ciphers of this type the analysis has the following points to start from:

1. Statistics: In starting with the analysis of any ciphered text a complete statistic investigation is to be made. It must give us the necessary information about repetitions of all textual elements, such as characters, bigrams, trigrams, words, and sequences of words etc., and thus reveals to us the type of substitution used.

E. g., the statistics about the Old Turkish Orkhon inscriptions revealed that the scripture consisting of 38 letters must be syllabic, or syllabic and alphabetic (V. THOMSEN, Bull. del'Académie Royale de Danemark, 1893, p. 285—.). In the trilingual Achaemenidae inscriptions the Persian text comprised 39 signs, and was thus similar to the one mentioned above, the 111 Elamite signs being thus syllabic, while the Babylonian was proved to have signs for entire words, comprising in total about 500 different characters.[1]

Anther important preliminary fact to be learned from statistics is the direction and order in which the lines and columns are to be read. This is very clearly illustrated by THOMSEN *l. c.*, to whom the parallels in the text gave an easy solution of the problem: the columns were to be read from above as in Mongolian and Chinese and from right to left, since the beginning of col. 40 in mon. I was found in the end of II 40, the end of I 40 is verbatim the beginning of II 39, which is continued by I 39 ending with the beginning of II 38.

It is from statistics that we learn the repetitions of signs and groups of signs in the text which are the most important condition for the solution of any ciphered text, and the avoidance of which is the capital question in composing any safe system for enciphering. They also make it possible to separate the single words, if the scrip-

[1] Very interesting are the statistic facts as to the script on the so-called Disk of Phaistos: there are 242 characters in all on the two sides. On side A occur 35 different signs and on B these 35 plus 10 new signs, total 45 different. Ipsen *IF* 47, 1— came to the conclusion that the alphabet must be syllabic consisting of over 50 but not much over 60 different characters. My friend L. Laitinen, M. Sc., has calculated as limits minimum 49 maximum 63 signs, almost the same as Ipsen.

ture is without any spaces or separating marks between the words. But even a practical help for the decipherment itself can be derived from these statistics: Thomsen first established the vowels, and supposed therefore that in a sequence XYX, X must be a consonant if Y is a vowel and *vice versa*. Having enough material to prove his hypothesis he easily discovered three characters which seemed to be vowels. The relation between certain consonants and certain vowels was proved to be constant and in this way he deduced two consonants and gave also the explanation for the occurrence of so many characters for one and the same consonant as depending on the nature of the surrounding vowels, which was also a test for the existence of a vocal harmony in the language in question. Some other groups seemed to be mere groups of consonants.

In cryptology a ciphergram enciphered by simple substitution can very easily be broken with the aid of statistics if only the language is known. When the respective statistics about the ciphergram and about the normal frequencies of various letters and polygrams in the language are put side by side, they show — often at first glance — which cipher-element corresponds to a certain en-clair-element.

A most remarkable and successful use of this method was made by H. Bauer in deciphering the texts found at Ras Shamra. He learned from statistics the relative frequencies of the 30 letters of this alphabet. The words were separated by a slant and had in general a length of 3—4, rarely 5, more often 2 or even 1 letter, being very well suitable for a Semitic language, as he had supposed. He then tried to find with the aid of the separation mark the signs corresponding to the elements used in flexion in order to compare them with those of Western Semitic languages. The most probable were:

1. $\alpha\ \beta\ \gamma\ \delta\ \varepsilon\ \zeta\ \eta\ \vartheta$ as probable prefixes
2. $\alpha\ \beta\ \gamma\ \delta\ \varepsilon\ \ \ \eta$ » » suffixes
3. $\ \ \ \ \ \ \delta\ \varepsilon\ \zeta\ \ \vartheta$ » monosyllabic words.

In the Western Semitic languages again are:

1.	'	*y*	*m*	*n*	*t*	(even	*b*	*h*	*w*	*k*	*l*)	the most usual prefixes
2.		(*y*)	*m*	*n*	*t*			*h*	(*w*)	*k*	»	» » suffixes
3.			*m*				*b*		*w*	*k*	*l*	» » » monosyllabic words.

All these 3 classes of characters had 2 common signs and the three classes of plain language elements 3 common letters. For δ and ε the only possible equivalents are *w* and *m* alternatively, $\alpha, \beta\ \gamma$, and η being prefixes and suffixes as well, but not words. Among the elements of plain language the most frequent are (*k*), *n*, *t*, thus $\alpha = n$ or *t*, $\beta = t$ or *n*. Consequently the 4 most frequent signs could have 2 possible values each:

$w = \delta$ or ε
$m = \varepsilon \quad \delta$
$n = \alpha \quad \beta$
$t = \beta \quad \alpha$

A Danish scholar, Fr. Münter, proved that the Old Persian inscriptions are written by Achaemenidae in the Avesta-language. Examining then the texts statistically he defined the nature of the script and confirmed with the aid of parallels that the wording of the three texts was almost identical — though in different languages. He regarded the most frequent signs as vowels, compared them statistically with those of the Avesta, but could correctly identify only one (Messerschmidt *o. c.*).

The statistical possibilities being exhausted, additional help for the decipherment must be found. As such, cryptographic litterature knows:

2. Bilingual texts (in which the second language naturally must be a known one) called »compromises» in cryptographic terminology, the utilization of which demands the utmost carefulness. In many cases the parallel text can be paraphrased, as was the Chinese version of the Orkhon inscriptions. If the version is a verbatim one, the decipherment is most easy.

3. Routine, which in the cryptographic sense means all schematism and mannerism in style and expressions. It is also regarded as one of weakest points in a ciphered message, but therefore also one of the best from which to start the analysis. If there are available some texts in a well-known language, such as addresses in letters, dedications, or epitaphs, *etc.*, in which similar or the same expressions are repeated, we must look for repetitions at the corresponding places in the encyphered text, to put in them the routine and to test the probable values of signs, gained hereby, on other groups. Virolleaud had noted on some axes, found at Ras Shamra, a group of six characters, which also occurred in the table Nr 18 preceded by the character ϑ (see above). He deduced that the group must be name of a person and the table a letter addressed to this person. The sign ϑ was thus a preposition corresponding to the Akkadian *ana* = Fr. à, or, the language being regarded as a Western Semitic one, $\vartheta = l$.

4. Probable words, given us by a bilingual text and routine or by other circumstances as well, are perhaps the starting point most favoured by the cryptologists. For example, in the Ras-Shamra texts the letter l, and two proposals for m being known Mr. Bauer looked for the word *mlk* 'King' and, having found it, verified at once the values $\varepsilon = m$ and $\delta = w$. Then he tried to find another probable group *bn* 'son', in the beginning with much difficulty, since in the tables the common routine phrase »A son of B» seems not to occur at all. In the table N:r 10, however, the group $\zeta\beta$ occurred fifteen times before varying groups, and since there was already the supposition $\beta = n$ and the possibility $\zeta = b$, as it occurred once as a prefix, once as a monosyllabic word, never as suffix, Bauer concluded $\zeta = b$ and $\alpha = t$. He then made out the word *ba'al* and in this way at last the values of all the characteres of this alphabet.[1]

[1] The decipherment was completed in a record time. Bauer himself relates that

on April 22 nd he received the texts published in Syria X.
» » 27 th the texts were almost entirely deciphered.
» » 28 th he informed Mr. Dussaud.
» May 23 rd the latter informed the French Academy.

The probable word can also be found with the aid of typology. In the Orkhon inscriptions as well as in those of Yenisei a group ending in the sign supposed to be = *i* was repeated at beginning of greater sections. Thomsen concluded this group must be *täŋri* since there are very rare words ending in *-i* in the Turkish languages. This sign occurred also in a group which, he thought, must be the name[1] of the Prince, in Chinese *K'iuehti-k'in* (but the Turkish form is *Kültegin*, not *Kök-tigin*, as he firstly supposed). As soon as these two groups were solved, he could identify a very frequent group with '*Türk*'. The values of all the remaining signs were then solved by substituting the known values in the text and by conjecturing those lacking.

On probable words, found with the aid of routine, was also based the decipherment of Old Persian inscriptions. Münter had already identified an often repeated group of seven signs as being 'the King' and when it was repeated two times successively as the routine expression, 'the King of the Kings', this group occurred always unchanged, but the next preceding one varied and had consequently to be a proper name. Grotefend then »translated» the inscription following the known Sassanian routine:

»NN the King, the Great (?), the King of Kings, X:s, the Kings, Son, of the House of Achaemenes.»

His translation »X:s, the Kings, Son» was based upon the fact that this group X occurs at the beginning of another inscription and accordingly must be a name. In the inscription in question it occurred increased by one sign, thus in another case, after the sequence »the King of Kings». The names of the Kings he then sought after among those of the Achaemenidae and identified X with Darius and NN with Xerxes, and the father of X, who was not entitled King, with Hystaspes. Substituting these names in corresponding groups in such forms as they occur in the Avesta and the Old Testament, and then the corresponding values of signs gained hereby in the group

[1] Most of the proper names of the Chinese version do not occur in the Turkish text.

'King' he came to a solution *khsheh-*; the Avesta dictionary gave him King = *khshe-io*, and consequently his solution was correct.

As to Rawlinson, he deciphered the Persian inscriptions, independently of Grotefend, in the same way with the aid of exactly the same proper names.

II. The Language Is Unknown

While deciphering an unknown language written with a known alphabet, the meaning of words as well as the grammatical construction is to be cleared up. Such a language can be regarded as almost identical with what in cryptology is called a code, a simple word-by-word respectively expression-by-expression substitution. As to the deciphering of codes, cryptologic literature is very poor in instructions and even the most competent authors — *e. g.*, Colonel Givierge — give few detailed precepts. This seems, therefore, to be an item regarding which the cryptologists have much to learn from the philologists and their methods, as well successful as erroneous. The most important of the methods successfully utilized are:

1. Statistics, which should teach us about the frequencies of all occurring words and sequences of words. Strange to remark, such statistics have not yet been made on Etruscan since Eva Fiesel in her *Etruskisch* (*Gesch. der Indog. Spr. wiss.* II, 5[4] p. 33) states of the standstill in Etruscology »bevor der Abschluss des CIE mit einem Gesamtindex des etruskischen Sprachmaterials vorliegt, wird kaum eine Erneuerung zu erwarten sein.» Such an index is the first a cryptologist needs when beginning to decipher a code. But, contrary to a code, a language usually has a grammatical morphology which also is to be studied and solved. In this work an index »a tergo» can be most useful for clearing up the suffixes, the classes of words and the flexion (cfr. Fiesel *o. c.* p. 40).

The famous French cryptographer Colonel Givierge has divided the words of a language in two categories: »les mots pleins» and »les mots vides». The 'full words' (*e. g.* nouns and verbs) have an independent sense contrary to the 'blank words', which express

only the different relations between the formers. Thus — in statistics on such languages as the Sumerian, Hittite etc., the determinatives prevail, while in the Khattish also the prefixes are very frequent.

Complete statistics, which in general are greatly dependent upon the character of the texts, can already give us the meaning of some frequent words, prefixes and suffixes, etc., but they are also needed while breaking a 'code' by other analytic means.

2. Bilingual texts are the most valuable and desirable aid in this work, and even a paraphrased one can be very helpful. Imbert in *MSL* 8 p. 449 ff. derives beside proper names also some other words from the Graeco-Lycian bilingual epitaphs, *e. g.*:

tideimi — — — *se tideimi se tideime*,

υἱὸς — — — *καὶ υἱῶι καὶ τοῖς ἐγγόνοις*

From these words he also could compose a paradigm for the Lycian noun. Even for the Etruscan some words are explained by bilingual texts such as

hinθial teriasals

εἴδωλον - - Τειρεσίαο

3. Routine gives also a very good starting point here. Imbert in *MSL* 9 p. 211 noted that the Greek epitaphs in the Lycian area are always entirely schematic: »le thème est d'abord un avis que le monument est la tombe construite par l'ordre d'un tel pour son usage personnel et celui des membres de sa famille; ensuite une interdiction conçue en termes précis d'apporter là d'autres morts; puis, au cas où il serait passé outre à cette défense, une phrase rigide énonçant que le sacrilège sera mené devant les magistrats et subira une peine pécuniaire: assez souvent le chiffre des 'drachmes sacrées' est indiqué». And in the epitaphs in the Lycian language »rien n'est changé à ce formulaire que la langue». Moriz Schmidt began the decipherment by confronting Greek and Lycian epitaphs.

The first interpreted and translated Hittite texts were the famous 'Arzawa letters', found at el-Amarna. The stereotyped introduction known from other el-Amarna letters served as the starting-point for Knudtzons decryptment: 'Me (the sender), I am well, my house, my wives, sons, officers, soldiers etc. are well. You

(the addressee) may be well, your house, your wives etc. may be well'. (J. Friedrich, *Hethitisch. Gesch. der idg. Spr. wiss.* II, 5[1] p. 7 ff.)

4. Grammatical and etymological comparison has been utilized with great success in the interpretation of an unknown language. For example Oscan and Umbrian, having first been identified with Etruscan, were interpreted and translated by comparing them with the nearly related Latin. Knudtzon too, having found the schematic formula in the Arzawa letter [1] (J. A Knudtzon, *Die zwei Arzawa-Briefe*, Leipzig 1902, p. 35.) concluded that the suffix *-mi* in the first part must correspond to »my» and accordingly the *-ti* must be »thy». In fact, the interpretation of *e-eš-tu* as »may be» was even supported by Skr. *astu*, Lat.*esto*, Gr. *ἔστω*,

When Hrozný definitely solved the Hittite question, he seems to have had even etymological support for his interpretations although he declares that he observed a purely combinatory method.[2] He had the following sentence *nu* BREAD-*an e-iz-za-at-te-ni ua-a-tar-ma e-ku-ut-te-ni* and knew that the form on *-an* was an acc. sg. and the following word a verb, probably pr. pl. 2.: he guessed *ezzatteni* 'you will eat' (conferring with OHG *ezzan* 'food'), *uatar* 'water' (cfr. Old Saxon *watar*), *ekutteni* 'you will drink' (cfr. Latin *aqua*) and translated 'now will you eat bread, you will drink water'.

---

[1] 3. kat-ti-*mi* dmq-in bit-zun-*mi* dam-meš-*mi* tur-meš-*mi*
4. amēlu-meš gal-gal-aš zab-meš-*mi* imēru kúr-ra-zun-*mi*
5. bi-ib-bat-*mi* kúr-kúr-zum-*mi*. — — —
8. bit-zun-*ti* dam-meš-*ti* tur-meš-*ti* amēlu-meš gal-gal-aš
9. zab-meš-*ti* imēru kúr-ra-zun-*ti* bi-ib-bi-it-*ti*
10. kur-zun-*ti* ḫu-u-ma-an dmq-in e-eš-tu

[2] *Die Sprache der Hethiter*, Leipzig 1917, p. VI. »Hier möchte der Verfasser mit grösstem Nachdruck betonen, dass er sich bei seinen Bedeutungsfeststellungen in den ersten Stadien seiner Arbeit immer von dem Kontext allein leiten liess.» The author, however, has not consistently applied the method of combination; he has often been misled by etymologies, since he compared Hittite words with those of various languages from widely different ages, being unsatisfactorily acquainted with the Indo-European philology.

It is only natural, that in interpreting Old Turkish inscriptions Thomsen also sought support from grammatical and etymological comparison with the present Turkish dialects, since Old Turkish itself was not known from any other source.

Etymological comparison, however, is to be used with the utmost caution. There are many example of too courageous, and therefore erroneous, etymological conclusions. So in commenting the Arzawa letters Torp interpreted *kuiš* (= Lat. *quis*) = 'deiner' and *kuin* = 'dich' (Knudtzon *o. c.* p. 116) connecting it with the Lycian pronoun as derived from **tvē*, while Bugge (Knudtzon *o. c.* p. 65) translated *kuin* 'die Prinzessin' and gave this word also an Indo-European etymology »Ich möchte **kui-* 'Prinzessin' und das Lykische *χbi* 'Prinz' als eine Ableitung von **ku*' Königin» erklären . . . Gehört das von mir vermutete **ku-* 'Königin' zu dem indogerm. **gou-* 'Kuh'?»[1]

Cryptographic codes are either alphabetic or non-alphabetic: in alphabetic the code groups as well as the true readings are side by side in alphabetic order, so that every group, if correctly interpreted, illustrates also the meaning of the surrounding groups, and it can be said that the decipherment of an unknown language is very analogous to the solution of such a code: instead of alphabetic order we can use here etymological comparison.

4. Ideograms, proper names etc. occurring in an unintelligible text afford valuable help in penetrating into the meaning of the sentence. It is, therefore, an absolute rule in cryptography, that there shall be no word in plain language among the ciphered elements, since these always can give the eventual analyst some hints about the contents of the text.

In deciphering the inscriptions of the Hittites, their practice of putting in the text Babylonian and Sumerian ideograms as well as phonetically written Babylonian words gave rich information about the contents. In parallel texts ideograms and phonetic writing often cover each other and thus make possible the discovery of

[1] It may be noted that an English scholar once connected the Lycian *lada* 'wife' with the English *lady!*

the meaning of the phonetic word. Most useful was the Hittite manner of writing the stems ideographically and terminations phonetically. In this way Hrozny solved the above problem, where the ideogram NINDA 'bread' gave reason to think of eating.

Many times there are many more Sumerian and Babylonian than Hittite words in a sentence, e. g. in VAT 7492 = KBo IV N:r 9. Col. II l. 15 ff.: 2 TUR ·ÊGAL *A-NA* LUGAL·SAL·LUGAL *ME-E QA-TI bi-e-da-an-zi* LUGAL·SAL LUGAL·ŠU(^{MEŠ})-*SU-NU ar-ra-an-zi*, translated by J. Friedrich *ZDMG* N. F. 1 (1922) p. 158 »2 Palastbeamte *reichen* dem Könige und der Königin Handwasser, König und Königin *waschen* ihre Hände».

6. 'The method of combination' (»die kombinatorische Methode»), which, progressing from the known to the unknown, by means of analogy, parallelism, resemblance of words etc., endeavors to elucidate the unknown words. It is thus to some extent the contrary of the comparative method related above. In fact, combination must always go before comparison, and the latter can not start before the former has given at least some certain results. As to Etruscan, Eva Fiesel says *l. c.* p. 33 ». . . es fehlt an einer relativen Chronologie der etruskischen Denkmäler und ihrer Inschriften. — — — Die chronologisch-historische Deskription ist bei den indogermanischen Sprachen der Vergleichung vorausgegangen; für das Etruskische begann man mit der Vergleichung, ehe jene nur versucht wurde.»

Most valuable help can be derived from all existing informations about the language in question, which can be obtained outside of it, e.g., the Etruscan words cited by Greek and Roman authors: Hesych.: *αἰσοι ·θεοὶ ὑπὸ τυρρηνῶν* etc.

The decypherer must also know all about texts with supposedly similar contents in other, known languages; that is, routine. Historical information is of great value, and in some cases it has even solved the whole problem. H. Bauer says in *ZDMG* N. F. IX, 1930 p. 251, that if the Kings of Karkemisch were as well known as are those of Persia, the Khattish hieroglyphs were as easily deciphered by the same method as the Persian cuneiforms.

Texts, then, must be treated statistically, so completely as possible, in preparing for the decipherment itself. As soon as the meaning of some words has been established, by the means related above in paragraphs 1—5, the deciphering goes on systematically. Some frequent word, which occurs near to a known word — or rather between two known words — is to be studied in all details as to its own form and place as well as the details of all the surrounding words. A similar investigation must be made in regard to all combinations in which this word occurs, and it is naturally very advantageous to begin with the most frequent words.

How all possible criteria are taken into account in investigating the meaning of one or more words, is well elucidated by the efforts made by several scholars to solve the correct order of the six first Etruscan numerals known from dice. F. Skutsch in *IF* V, p. 256 ff. made evident the impossibility of the hypothesis of Pauli who had resulted in three probable orders with the commen value of *zal* = 2, since according to it a person, dead at the age of twenty-seven should have been eleven times a *zilath*. Skutsch himself started by examining which of the numerals can *not* be = 1. He regarded it as unlikely that the numeral *one* be mentioned in cases, when the deceased had *once* held an office. Consequently only *maχ*, *huϑ* and *śa* can come in question as = 1. Of these, *śa* occurs some times in connection with a noun provided with the suffix for plural, e. g. *avils* XX *tivrs śas* '20 years *śa* months'. As to a decision between *huϑ* and *maχ*, Skutsch thinks on the basls of homology between the dice that *maχ* must have been engraved first and be = 1. *Zal* being opposite to *maχ* can only be = 6 (cfr. above). The sum of numerals on two opposite sides of the dice is consequently = 7 thus:

I. *ci* + *śa* = 7 and
II. *ϑu* + *huϑ* = 7.

The inscription on the Cippus Perusinus gives the following equation: *ci* + *huϑ* + *z(a)l* = 12, but *zal* being = 6.

III. *ci* + *huϑ* = 6

These three equations have only two solutions

2	3	4	5
ci	*ϑu*	*huϑ*	*śa*
huϑ	*śa*	*ci*	*ϑu*

Referring to the assertion of Fabretti, that an age under 4 years was not mentioned in the epitaphs, Skutsch cites the epitaph of Larϑi Ceisi (Fabretti 2104) *avils śas amce* 'lived *sas* years', thus *śas* $> 4 = 5$ and the order of numerals

1	2	3	4	5	6
maχ	*ci*	*ϑu*	*huϑ*	*śa*	*zal.*[1]

Slotty resumed this question in *Archiv.Or.* 1937 p. 381—. He had studied the dice and thought that having begun with *maχ* the artisan carried on his work by entering the numerals in the inverse order, i. e. *śa* = 5, *huϑ* = 4, *zal* = 3, *ci* = 2, *ϑu* = 1. He also criticized the prejudice of Skutsch that if the artisan had begun with *maχ*, it should be = 1. and *zal* = 6.[2]

H. L. Stoltenberg, *Glotta* 30, 1943, p. 234 ff. criticizes — »als Soziologe» — the theory of Slotty departing from the teens *ϑunem*, *eslem* and *ciem*, which he declared to be composed like the Latin *duodeviginti*, *undeviginti*, so that *ϑu* = 1, since the others are followed by plurals. He then studied the numerals on epitaphs in light of the statistics on deathrate, which have 3 tops: in the first, third and seventh decade. This argumentation seems him to ascertain the following order[3]:

1	2	3	4	5	6
ϑu	*zal*	*ci*	*huϑ*	*maχ*	*śa*

[1] Some arguments are not very convincing.

[2] But Slotty himself too seems to have had some prejudices.

[3] *ϑu* = 1, since there is no decade derived from it; *ci* = 3, since it occurs in the Agr. mum. and 3, not 2, is the numeral used in religious texts (!); *maχ* = 5 and *śa* = 6 are based on the death rate.

which he then adjusts to correspond to the dice[1] and results of comparative philology.[2]

When a probable translation for the examined word has been found, the meaning of surrounding words must be studied in the same way. Should it not be possible to give them a satisfactory and natural interpretation, the first established meaning must accordingly be revised. Every interpretation thus implies a long series of trials (and errors too!), before it can possibly be regarded as verified.

Very instructive as to method is Torp's study on the Lycian words *esedeñneve* and *χñna* (*Lykische Beiträge* I, *Christiania Videnskabsselskabs Skrifter, Hist.-Fil. Klasse* N:r 4, 1898): *χñna* occurs often in sequence *hrppi esedeñnevi χñnahĭ ehbiyehi*, but even alone. Its meaning depends on that of *esedeñnevi*, which can not be 'brother', as stated by Imbert *MSL* 8, p. 468, since in an epitaph strangers are forbidden to bury their *esedeñnevi* in the tomb. Torp thus concluded that its significance must be larger, either 'descendants or 'relatives'. Regarding to Xanth. I. »He destined the upper sepulchre for his wife and *meñne teidehe esedeñnevi* and destined the lower sepulchre for his own house» Torp interpretes' (für seine Frau) und Meñne Teides' Nachkommenschaft'. This translation will be verified, if *χñna* now can be correctly interpreted. Referring to St. X. S. 24. »er schrieb (mit Angabe) des Vaters und *χñnahe* (Namen): Harpagos' Sohn, Cheriga's (Bruder?), Kuprlli's Schwiegerson, Cheziga's Vetter» Torp — contrary to the opinion of Imbert and Bugge — concluded that *χñnahe* means 'relatives' and proved that this meaning conforms with the other quotations of this word. By

---

[1] The order on the dice is

1 : 4 2 : 5 3 : 6

θu : *huθ* *zal* : *maχ* *ci* : *śa*, but according to the author »die Gesetze des Lebens» give more reliable information (!).

[2] Hyttenia = Tetrapolis, thus *huθ* = 4; *cezp* = 7 since Basque *zazpi* = 7 etc.

As to the comparison of the Etruscan with other languages, Cortsen has published interesting studies on the inscription from Lemnos in *Glotta* 18 (1930) p. 101 ff.

a similar, most careful analysis of all authoritative passages Torp proved *ti* to be the relative pronoun.[1]

Etymological study and comparison with other languages can not start until many grammatical facts and verified interpretations of words have been assured by this method of combination. Otherwise the results will be most uncertain, sometimes even ridiculous.

III. Both Script and Language Are Unknown

In the foregoing we compared the solution of an unknown language with that of a code. If even the script is unknown to us, we are entitled to regard the problem as identical with the solution of a recyphered code. Cryptology has paid to this question very little public attention, but it is only natural that there are reasons why successful methods for breaking recyphered codes must be kept secret. The following means, however, may be regarded as the most useful in such a task.

1. A bilingual text affords the most valuable help for the decipherer, like the famous Rosetta stone, which, being composed — according to its own wording — *ἱεροῖς, ἐγχωρίοις καὶ Ἑλληνικοῖς γράμμασιν*, gave a good starting-point.

Also the Achaemenidae inscriptions were composed in three languages: in Persian, Elamite and Akkadian, while the so-called Table of Dalion was in Greek (but nobody knew that it was Greek!) and Phoenician. But a translation only — albeit a verbatim one — is not enough to make a solution possible: there must occur in

---

[1] How difficult it sometimes can be to ascertain the sense of a word, is illustrated by Fiesel *l. c.* p. 49 ff.: Krall, Torp and Cortsen interpreted Etruscan *etnam* as 'Opfergabe' since the following word is 'vinum' (in the Agr. mum.); Torp changed later his mind and regarded it as a particle 'deinde', 'post'; Vetter translated it 'item', Trombetti 'ferner' and Goldman *etn-am* 'Idustag'. The cup-inscription of Narce begins with *ipas ikam*, which Vetter and Torp interpret 'cuius ego' while Cortsen and Bugge read 'Weihung (*ikam*) der Schale (*ipas*)', all referring to their own interpretation of *etnam*.

both texts words which have similar phonetis, and thus help us to *read* the unknown script phonetically, for instance

2. Proper names. James Prinsep had found on bilingual Bactrian coins Greek proper names, *e. g.*, ΜΕΝΑΝΔΡΟΥ, searched then for the same names in the unknown Pehlvi script and through them unraveled the value of some signs of the alphabet, which he in turn applied to the translated epithets and titles, and was thus led to knowledge of the language employed. (J. PRINSEP, *Indian Antiquites* I—II, 1858, London. I, p. 179).

The Numidian alphabet was able to be read with the aid of Numidian-Punic bilinguals. (Jensen o. c., p. 106). —The Lydian alphabet was deciphered on basis of a longer Lydian and Aramean inscription by Kahle, Sommer and Brandenstein. This language, however, seems to have no relation to Etruscan. (J. FRIEDRICH in *Die Welt als Geschichte* 1937, p. 58 ff.).

Champollion, who was very familiar with all the known ancient alphabets and languages, started the decipherment of the hieroglyphs by re-transscribing Greek proper names from demotic script in hieratic and then in hieroglyphic. The result was, as he had expected, similar to the group of hieroglyphs, which he supposed to be Ptolmis:

α	*β*	*γ*	*δ*	*ε*	*ζ*	*η*
p	*t*	*o*	*l*	*m*	*i*	*s*

These values he then applied to an other group, which he also considered as a proper name, and read

ϑ	*δ*	*ι*	*γ*	*α*	*ϰ*	*λ*	*μ*	*ν*
	l		*o*	*p*				
k		*e*			*a*	*t*	*r*	*a*

The Egyptian alphabet consists of mere consonants. Thus the only ways to get information about the vowels are through foreign loan-words as well as Egyptian loan-words in foreign languages, and by even conclusions based on Coptic.

Norris fixed the values of most of the Elamite syllabic characters with the aid of the already known Persian parallel texts. In the solution of the Babylonian alphabet and language these Persian parallels played almost the same part as the Greek portion of the Rosetta stone in the decryptment of hieroglyphs. The task was most difficult since the script consists of over 500 signs expressing even whole words, as *e. g.*, in the Achaemenidae inscription the signs 'king' and 'son'. From the proper names which were written phonetically — Xerxes with six characters — the phonetic value of some signs could be derived. In 1847 de Longpérier interpreted a longer inscription with the aid of the already known cuneiform signs »Palace of — — —, of the Great King, of the Mighty King, of the King of Kings, of the King of the Land Assur», but he *could not read a single word* phonetically. Later on he was able to read the name of Sargon, the first syllable of which was the character 'King', comparing it with the Hebrew *sar* 'Prince'. Botta found several parallel inscriptions, of which some had phonetic readings instead ot the ideograms as in the others. From these many equations could be formed. The values of Sumerian cuneiform characters and ideograms were explained only after the dictionaries were found at the library of Nineveh.

The solution of the Cypriote syllabic script was found in a very strange way (the result itself was strange too!): The text — the so-called Bronze-Table of Dalion — was written in Cypriote and Phoenician.[1]

In the Cypriote part the words were separated by points. G. Smith paid attention to the words *melek*, *Melekyaton*, *Citium*, and *Idalion* in the Phoenician part, and found without difficulty the word corresponding to *melek* in the Cypriote part, since it was the only word occurring twice in the text. A word was longer than all the others and he conceived it to be *Melekyaton* (in fact *Μιλιχιάθωνος*). Two others had a repetition at the end and he regarded these as *Idalium*

[1] A German Professor Roth interpreted and translated the text as a declaration addressed to the Cypriotes by King Amasis by using some Hebrew and much imagination.

and *Citium* in gen. and read the stems *Ki-ti-* (pro *Κετίων*) and *E-da-li-* (pro *'Ηδαλίων*) The sign *-li-* occurred as the second one in the group supposed to be *Melekyaton* and strengthened thus his readings. Since the *K* in *Kition* was unlike to that in *Melekyaton*, he read the other as *ka*, the other as *ki*. The words corresponding to *melek* had a different letter as penultimate. According to their position in the text Smith concluded that the first was a nom., the second a gen. then he looked for a language in which the penultimate letters in nom. and. gen. of the word 'king' were different, and found the Greek *βασιλεύς* ~ *βασιλεῶς* (in fact we must read *βασιλεύο[ντος]*). By substitution he solved the values of eighteen characters and interpreted with the aid of them legends on coins. But he could not complete the decipherment, since he was a former engraver and did not know the Greek language. But because of this, he had no preconceived ideas: »quel homme ayant reçu l'éducation du collège aurait osé chercher du grec sous ces signes hétéroclites?» (BRÉAL, *Journal des Savants* 1877, pp. 503—513 and 551—566.)

As to the Hittite hieroglyphic script, the first decipherer JENSEN (*ZDMG* 48, 1894, p. 235 ff.) did not even try to read the words phonetically, but interpreted the contents of some repeated passages. Comparing the characters and groups in various texts he could with the aid of repeated groups discover several words, determinants and suffixes. The first sign occurring was often a head with arm, in which the hand points to the head. Jensen interpreted it 'me' and regarded the following co-ordinated groups as a royal title. These groups are then repeated, but with another suffix and followed by a new ideogram, apparently the title in gen. followed by 'son'. The derminative for 'God' Jensen deduced in the following way (p. 277): $\alpha\ \beta$ is a very common sequence and is once followed by $\gamma\ \delta$; $\alpha\ \beta$ never has the nominative-suffix, while γ has it very often and thus the sequence is to be read 'the $\alpha\ \beta$ of $\gamma\ \delta$'; as α occurs always in beginning of a group, it must be a determinative.[1]

[1] Jensen believed it means 'land', while in fact it is 'god', and β is the name of god Šantu. His later attempts to read the signs phonetically were most uncer-

Later Forrer has with great succes continued the decipherment of hieroglyphic texts. As a starting-point the took the routine interdictory formulas at the end of an inscription and with the aid of the proper names and titles of kings be cleared up the construction of these forms. He found also the relative pronoun, its correlation *-tu* 'him' and the indefinite, composed of the doubled relative. Also the forms of the present, the preverbs *arḫa* 'out' and *rata* 'to', the possessive *mi(a)s* 'my' as well as the suffix of nom. and acc. plur. *-ai* were identified by him.[1] — Hrozny and Meriggi have greatly contributed to the decryptment and interpretation of this language.

P. Meriggi has published most interesting studies on decipherment of the Indus-script in *ZDMG* N. F. XII, p. 198 ff. Here too it is impossible to find the phonetical reading of the hieroglyphs, but the author has with the greatest sagacity interpreted the contents of several passages, and even made available some information about the grammatical structure of the language.[2]

As soon as the script has been deciphered, the interpretation of the language goes on along the methodical lines summarized above in paragraph II.

Much human ingeniousness and work has been spent in efforts to decipher the so-called Disk of Phaistos. Mr. Gleye, (*Kretische*

---

tain (the ideogram for '*Šantu*' he read '*Ḫilik*' etc.) and were even abandoned by himself in *ZA* N. F. 1 p. 245 ff. See also Jensen, *Hittiter und Armenier*, Strassburg 1898.

[1] See *e. g.* his preliminary note in AJSL 48, 1932, pp, 137—169, which is a proof of the most ingenious cnyptologic analysis.

[2] About his starting-point he states (p. 199 f.) »... ein sehr günstiges Moment ist die unendliche Zahl der Variationen, in denen die Legenden auftreten. Es stellt sich dabei bald heraus, dass die kürzesten oft nur eine Aneinanderreihung einiger Ideogramme darstellen, die Längeren aber, und viele der kurzen, auch komplementierte Ideogramme und wohl ganz phonetisch ausgeschriebene Wörter enthalten. Das zweite günstige Moment ist das Verhältnis der Legenden zu dem sie tragenden Gegenstand — — — und der Schlüssel zum ganzen ist m. E. der nach dem archäologischen Befund naheliegende Grundgedanke, dass diese Legenden in der Mehrzahl der Fälle Verwaltungs-stempel ohne Personennamen sind.»

Studien I—II, Tomsk 1912—1914 and *Die Sprache der Eteokreter*, Riga 1928) has stated that the language of this document was — Finnish! In *IF* 47 (1929) p. 1 ff. G. Ipsen analyzed the text with the greatest sagacity and carefulness making available some information on the construction of forms of the language by stems and formants. With the aid of repetitions of the syllabic signs he also identified the determinative for person, some prefixes and formants of nominal and verbal flexion. His study must be considered as being methodically most instructive. But a Finnish scholar, V. H. JUVELIUS (in Finnish year-book »Suomalainen» IV) has proposed a very interesting hypothesis concerning the origin and interpretation of the Disk. As an archaeologist he had some experience on falsifications and on the basis of confused styles and types of signs and writing he points out that the Disk is perhaps nothing but a counterfeit (!)[1], which the 'Pleti' dug in just before the 'Kreti' dug it out.

Not even a bilingual text is always a reliable help: the legends on Tartessian coins remained for a long time undecipherable though there are also transcriptions in Latin. For instance, on coins from Asido this name is written with three characters, which were read

by de Zangroniz as	*n*	*tz* *a*
by Schulten	*i*	*s* *a*
by Meinhof	*d*	*s* *χ*

(MEINHOF, *ZDMG* N. F. IX p. 239). In *ZDMG* N.F. XII (1934), p. 50 ff., E. Zyhlharz has in all probability identified the values

[1] Vide Pauly-Wissowa, *Realenzycl. Suppl.* VI c. 205, where W. Brandenstein states: »Die Schrift, eine Silbenschrift, ist nicht von der kretischen abhängig, sondern weist nach Ägypten, wobei die Bilder nicht die ägyptischen Gegenstände, sondern die ägäischen wiedergeben. Die Worttrennung weist auf kleinasiatischen Einfluss, die Anwendung von Zeichenstempel nach Mesopotamien Ein Schriftzeichen zeigt einen Kopf mit Federkrone: diese Tracht ist hauptsächlich von den Philistern bekannt, die aber um die Zeit des Diskos (17 Jhdt.) noch unbekannt sind .... Der Diskos ist das einzige Dokument in dieser Schrift.»

of *ten* Tartessian characters by confronting always two proper names, occurring on coins, which have one or more common letters, thus establishing more certain relations than the former decipherers e. g.

$$\gamma \cdot \beta \cdot \alpha = \text{Bailo}$$
$$\varepsilon \cdot \alpha \cdot \delta = \text{Oba}$$

thus $\alpha = b$ (the script is read from right to left). The author regards the alphabet as derived from cursive elements of the Punic-Phoenician alphabet.[1]

In the foregoing most of the methods successfully used in the deciphering of unknown scripts and languages have been somewhat critically presented. It may be useful for everyone having to handle such problems to possess some knowledge about the results as well as about the methods of his great predecessors in order to be able to appreciate and criticize them, to be successful and avoid mistakes. In several connections coordinations have been made between true cryptography and the philological study of unknown writings and languages. It is highly recommendable that persons working with the latter be also familiar with the former. On the other hand, the philological and linguistic analysis of questions to be solved in the above cases II and III offers very instructive illustrations on such problems, which have been treated only superficially in cryptographic literature, and may thus be useful for persons interested in ciphering and deciphering.

It is clear, however, that the most profound knowledge and observation of the possible methods with all their advantages and disadvantages do not alone guarantee success, since success needs wide and many-sided knowledge, a vivacous imagination, never-failing patience and — plenty of good luck too!

---

[1] Strabo writes in 139 about the Tartessians: *σοφώτατοι δ'ἐξετάζονται τῶν Ἰβήρων, οὗτοι καὶ γραμματικῇ χρῶνται καὶ τῆς παλαιᾶς μνήμης ἔχουσι, συγγράμματα καὶ ποιήματα καὶ νόμους ἐμμέτρους ἑξακισχιλίων ἐπῶν, ὡς φασι · καὶ οἱ ἄλλοι δ' Ἴβηρες χρῶνται γσαμματικῇ, οὐ μιᾷ δ'ἰδέᾳ, οὐδὲ γὰρ γλώττῃ μιᾷ.*

Utilized cryptographic literature:

FIGL: Systeme des Chiffrierens, Graz 1926.
GIVIERGE: Cours de cryptographie, Paris 1925.
SANDLER: Chiffer, Stockholm 1943.
STÅLHANE: Hemlig skrift, Stockholm 1934.
YARDLEY: The American Black Chamber, Indianapolis, USA, 1931.