Igor Pozdniakov, Konstantin Pozdniakov

Rapanui Writing and the Rapanui Language: Preliminary Results of a Statistical Analysis

Igor Pozdniakov
St Petersburg
Konstantin Pozdniakov
INALCO, Paris

The writing of Easter Island has been the subject of dozens of monographs and articles, yet we know virtually nothing about it. There is no reliable catalogue of glyphs. We do not know the rules by which the glyphs are modified and combined. Nor do we know what language the texts attest. Moreover, there is no conclusive proof that we are in fact dealing with a writing system, and not with a pre-literary communicative code. The writing system of Easter Island remains undeciphered one hundred and fifty years after its discovery.

This assessment of the situation may come as a surprise. Many people are convinced that the Easter Island texts have been read and translated. Foremost among them are the decipherers themselves. Popular publications regularly inform the general reader of a sensational new decipherment, each one 'complete and definitive.'

It would be interesting to know why the decipherment of Rapanui writing exerts such a firm hold over readers' imagination. The general public is surely not interested in the content of the texts. Anthologies of Polynesian genealo-

gies, myths and songs sit gathering dust on library shelves, their pages uncut. These publications, which remain largely untouched even now, greatly outnumber the dozen or so pages of undeciphered texts which one day it may be possible to read. It is difficult to believe that any non-specialist would really want to learn, for instance, that *'The leader Taana a Xarai was the son of Xarai a Ataranga'*, or that *'The soul of Xau Maki came to Pei and gave this place the name Pei-a-Xau-Maka-o-Xiva'*. The reason for the intense interest in the written language of Easter Island lies elsewhere: even a reader without philological training senses that the very fact that this original writing exists at all requires thorough investigation and explanation. It is primarily the writing itself which is of general interest. One might venture the 'heretical' suggestion that the information the texts contain is of lesser concern. For this reason, any attempt to put forward a semantic interpretation of the texts on the basis of unsubstantiated 'revelations' is not simply unconvincing, but also unhelpful: paradoxical as it may seem, such attempts do not bring us any closer to deciphering Rapanui writing. In our view, the appearance of yet another monograph with a new 'translation' of the corpus of texts would be significantly less interesting than the appearance of an article presenting the results of a structural analysis of some specific aspect of the writing system — for example, an analysis of instances in which anthropomorphic glyphs are rotated to the left.

Unlike the decipherers, who have long since read and translated everything, professional philologists and cultural historians are very sceptical about the written texts of Easter Island. In studies of the history of writing it is often asserted that these texts do not attest a written language in the strict sense. It is not simply that the stream of 'decipherments' (particularly in recent years) has fundamentally discredited this area of academic research. The scepticism of philologists who are acquainted with the history of the development of the world's writing systems is entirely justified from the typological point of view: original systems of writing are not known to have come into being on distant islands, thousands of miles from the nearest shore, which have spent many centuries in complete cultural isolation. One of the main aims of this article is to present concrete arguments which will show that we are indeed dealing with an original writing system, and not with pictography or with some sort of mnemonic code for preserving information or with any other pre-literary form.

Let us set aside for a moment the fact that the translations of the Easter Island texts proposed by numerous decipherers lack substantiation. Let us assume that one of these translations — say, Fedorova's or Fisher's — is absolutely correct. In both cases that would mean that we are dealing with a logographic writing system in which the overwhelming majority of glyphs represents words. The so-called

Metoro readings (the famous attempt by a native of Easter Island to read the Rapanui texts for Bishop Jaussen) were the first attempt at a logographic interpretation of the Rapanui glyphs. The vast majority of decipherments treat the written texts of Easter Island as logographic.

In order to establish whether the logographic nature of the writing can be supported statistically, it is first necessary to compare the frequency of words in the Rapanui texts with the frequency of the glyphs. An attempt to do so was made in the late fifties by the Centre for Computation at the Academy of Sciences and the group led by Knorozov, but this huge project produced practically no results which are of use today. The problem lay not in the technological limitations of the first generation of computers; the frequencies can easily be worked out using pen and paper. The fundamental difficulty was that the researchers did not have at their disposal a reliable catalogue of glyphs. Their calculations were based on a catalogue of 600 glyphs similar to that made by Barthel [Barthel 1958]. Yet as they knew (and as Barthel himself remarked more than once), the catalogue contained not just individual glyphs, but also many ligatures, ligatures i.e. combinations of glyphs. However, no serious attempt to compile a catalogue of the basic glyphs has ever been undertaken.

Paradoxical as it may seem, despite the appearance in recent years of a whole series of monographs about decipherment, to this day there is still not a single publication in which the problems of cataloguing the glyphs are discussed systematically. Virtually all researchers base their work on Barthel's catalogue, even though virtually all researchers agree that the actual number of glyphs is significantly smaller. But by how many? 100? 300? Unless these questions are answered, it is impossible to make use of statistical data, for clearly the data will differ markedly depending upon the inventory of glyphs chosen. The Leningrad group remarked on *'the impossibility of identifying in the kohau rongorongo texts variable glyphs with a high relative frequency of appearance which could represent articles, prepositions, verbal particles etc.'* [Fedorova 1982: 30–1], but offered little by way of argument in support of this statement. It is important to note this lack of support, because the view of the 'telegraphic' ('agrammatical') nature of Rapanui writing which remains current to this day derives solely from calculations based on the frequency of Barthel's 600 glyphs.

1. The catalogue of glyphs

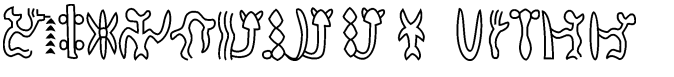
The catalogue published below contains not 600 glyphs, like Barthel's, but 52. For this reason, the statistics presented here differ fundamentally from the results of previous studies.

Limitations of space make it impossible to discuss in detail how each individual glyph was isolated. Instead, the briefest of commentaries will be given on the principles by which the catalogue was constructed.

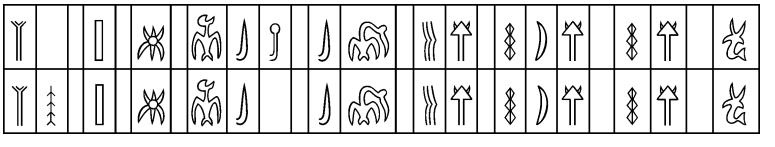
The basis for principled cataloguing of the glyphs was established by Kudryavtsev, who discovered that of the fifteen or so extant tablets three contain essentially the same text with only minor variations (tablets H, P, Q[1]). He later established that the text of K coincides closely with the text on the front of tablet G: possibly the most important discovery of all for the decipherment of the Easter Island writing. Thanks to the discovery of these parallel texts, there arose for the first time a real possibility of distinguishing significant variations in the graphemes from insignificant ones, of isolating glyphs and their alloglyphs. Barthel, whose monograph remains the most valuable work of all on the subject, subsequently identified a whole range of (fairly long) sequences which are repeated either partially or entirely in different texts or in different lines of the same text.

In actual fact there are many more such sequences than appear in Barthel's list. According to our data, even if the parallel complete texts are set aside, approximately half of the corpus comprises sequences of glyphs (from 10 to 100 glyphs) which are attested in several texts. There are more than one hundred such sequences in our card-index. They cannot be listed in full in an article, but the following example illustrates one of the many passages which have gone unnoticed until now:

Ab 5/6 

Er 4 

In the glyphs used in our catalogue a fragment of this sequence appears as follows:



Some texts (for example on tablets N, R, H/P/Q) consist more or less entirely of such sequences of glyphs. As soon as one sequence finishes, another one begins, though they appear in different orders in different texts. Some sequences are found in virtually all the extant Easter Island texts.

---

[1] The texts are referred to using Barthel's classification.

This type of structure, in which a text is made up of a number of micro-texts repeated in different combinations, shows that the extant texts are not integral. Strings of these relatively short sequences give mythologists, for instance, no grounds for hoping that decipherment will reveal an ancient myth recorded on one of the tablets. Were the writing to be deciphered successfully, the most that specialists interested in its content could count on would be texts of a brevity appropriate to a list of names of kaikai shapes (a kind of cat's cradle game, in which each shape requires a short text), genealogical (or other) lists, short poetic texts etc.

All these parallel sequences, together with the sequences identified by Barthel and the parallel texts discovered by Kudryavtsev, have been used in compiling the catalogue of glyphs presented here.

Possible principles for identifying alloglyphs (variants of a single glyph) have been discussed in [Pozdniakov 1996]. It was shown in particular that glyphs 6 and 64 should be considered variants of the same sign, since there are dozens of examples in which glyph 6 in one text corresponds to glyph 64 in a parallel text. The regularity of this correspondence rules out any other interpretation. If, following Fedorova, one takes these glyphs to be different (Fedorova reads glyph 6 as *mau* 'to take' and 64 as *tonga* 'the name of a sort of yam' [Fedorova 2001: 94, 98]), one is obliged to explain why in one and the same context but on different tablets the noun 'yam' turns into the verb 'to take', and vice versa, with depressing regularity.

The catalogue of glyphs compiled by analysing parallel texts and sequences is presented in Table of glyphs. Any of Barthel's glyphs not included in the catalogue are either ligatures or alloglyphs.

2. Statistics and the type of writing

Surprisingly little is known about the writing of Easter Island. The facts can be given in a single short paragraph. As discussed above, we know that some texts, and also many sequences within the texts, are repeated. In most instances we know what order the lines appear in (though for some texts that, too, is unknown). We can state reasonably confidently that a sequence from tablet C most probably contains a calendar [Guy 1990]. But that is the sum total of the knowledge accumulated over the many years in which this writing has been studied.

When faced with the task of extracting knowledge essentially from nothing, a researcher must be guided by statistics. Statistical data based on a reliable catalogue of glyphs not only produce concrete, if modest, results; crucially, they also rule out many erroneous approaches to decipherment.

As we will attempt to show, the principal merit of a statistical analysis is that it provides a whole host of parallel but independent arguments

Table of glyphs[1]

	1		15		47		69		380
	2		16		50		70		400
	3		22		52		71		530
	4		25		53		74		660
	5		27		59		76		700
	6		28		60		91		720
	7		34		61		95		730
	8		38		62		99		901
	9		41		63		200	?	999
	10		44		66		240		
	14		46		67		280		

in favour of the view that Easter Island writing script rather than a mnemonic ('telegraphic') code for recording information. These arguments will be discussed as they are presented. Statistical data also open up the possibility of establishing securely the type of writing system used in the Rapanui texts, the issue to which we now turn. This approach means that any decipherment of the texts which assumes a writing system for which there is no statistical support must necessarily be rejected.

Comparative analysis of the parallel texts and of repeated sequences leads to the conclusion that Rapanui writing comprises 52 glyphs (99.7% of the glyphs in the corpus of texts). Glyphs used very infrequently, which make up 0.3% of the glyphs in the corpus, were assigned the index 999. There is nothing to be gained by cataloguing

---

[1] The authors would like to express their gratitude to A. I. Azarov, who created a true type font (TTF) for representing this script.

these rare glyphs at the present stage of work, since interpreting them does not advance the decipherment. A complete catalogue should contain three or four more units to account for glyphs found in the text of I (a staff), which differs markedly from the rest of the corpus statistically and in other diverse respects. This text was excluded from the corpus because it produces a considerable distortion in the statistical averages.

One question which, strange to say, has never been discussed in the literature on Rapanui writing can be answered without recourse to calculations: if the writing system comprises approximately fifty glyphs, what could these glyphs represent: a) words? b) syllables? c) sounds?

Setting aside the fact that hypothesis c) is improbable from the point of view of cultural anthropology, let us evaluate it statistically. If we assume that the eastern Polynesian languages are the most likely linguistic basis for Rapanui writing, we immediately encounter a problem: these languages have strikingly few phonemes. Rapanui has 10 consonantal phonemes — /p/, /t/, /k/, /ʔ/, /v/, /r/, /h/, /m/, /n/, /ŋ/ (the glottal stop is not recorded in transcriptions published by Fedorova) — and five vocalic phonemes — /i/, /u/, /e/, /o/, /a/ (long vowels apparently do not have phonemic status and are to be regarded as combinations of two vocalic phonemes). Thus the number of glyphs needed to represent phonemes (15) is several times smaller than the number of glyphs identified in the written texts, which means that the hypothesis about the phonemic nature of the writing cannot be supported.

Let us now consider the hypothesis that the glyphs in the Rapanui texts denote words. Since nearly all the proposed 'decipherments' from Metoro's 'readings' onwards have assumed, wittingly or otherwise, a logographic type of writing, it is particularly important to assess the statistical probability of this hypothesis. A corpus of ten texts transcribed from the Rapanui language and used as a control sample for statistical comparison with the writing attested 335 different words which appeared three or more times (the word *te* occurred 665 times in the corpus), and a further 710 words which appeared once or twice. There is no escaping the fact that if the glyphs represent words, fifty glyphs represent fifty words, not one thousand.

The only way forward then is to limit the lexicon of the written texts to fifty words. The result is extremely monotonous. The frequent lexical repetitions have been taken to indicate 'the ritual importance of magic words'. Fedorova's translation [Fedorova 2001] illustrates this approach:

Text A (Aa8): *'... he cut a sugar-cane, yam, sugar-cane, taro, he cut, he cut, taha yam, he cut, sugar-cane, tea yam, taro, kahi yam, he*

dragged out, he took, he cut, sugar-cane, tea yam, he cut a taro yam, he cut, he took, he cut...' (p. 131) and so on ad infinitum. **Text B** (beginning): *'he cut a sugar-cane, he cut, lots, he took a taro, he cut lots, he took a taro tuber, a tuber he took, took...'* (p. 151) and so on to the end of the text and throughout the rest of the texts. Skipping over 200 pages, we come to the final lines of Fedorova's translation: **Text X** *'he cut hau batata, yam, yam, taro, taro, he cut a tuber of yam, he took a tuber of taro, a tuber, a tuber, he dug up, he cut, he cut, taro, turi sugar-cane'.*

So ends the translation. The translator's own commentary is of note: *'Why then did the Rapanui,* ***possessed by a doggedness worthy of a maniac*** [our emphasis, — I. P., K. P], *carve on to tablets [...] their complex glyphs? As shown by this decipherment, which required great effort, these carvings [...] are magic formulae intended to increase the harvest. They are, without doubt, ancient chants [...] upon which the growth of plants and a good harvest depended. [...] The texts are a series of songs made up of morphemic words which are repeated frequently throughout. [...] The texts [...] call to mind with the naïveté of their style and content the songs and incantations of other traditional societies'* [Fedorova 2001: 81].

Fedorova's catalogue contains approximately 130 glyphs. On occasion she includes 'homonyms' in the translation, bringing the total number of semes used in her translation to around 200. That is significantly more than in the lexicon of the Ellochka the Cannibal,[1] but significantly fewer than in folkloric texts of comparable length.

The strategy for deciphering the texts used by Fischer is typologically identical to Fedorova's, though his translation differs markedly from hers in meaning. In a recent monograph [Fischer 1997], he devotes a section entitled 'S. Fischer' to his place in the history of the decipherment of Rapanui writing. He describes in some detail how and when its mysteries were revealed to him. Essentially, Fischer noticed (though he was not the first to do so) that the text of I had an unusual structure: it was made up of groups of three graphemes that were separated from one another by Barthel's glyph 76. One example of many follows:


X Y Z

Fischer's 'first scientifically, as the author defines it himself, verifiable phonetic breakthrough' was to identify this group of three with an excerpt from the famous Rapanui cosmogonic text Atua mata riri,

[1] A character in Il'f and Petrov's *The Twelve Chairs*.

in which each period is constructed according to the formula: X copulated with Y, and they brought Z into the world. Fischer reads the cited example as, *'All the Birds joined with the Fish, and they gave birth to the Sun'*. Fischer went on to find this group of three in another text, and then in all the texts, which naturally took on a cosmogonic character.

The difficulty with this interpretation is that glyphs 700 (fish), 8 (star) and 400 (bird) taken together have a relative frequency of 8.3%. In combination with glyph 6 (which Fischer reads as *mau* 'all'), the overall frequency of the four glyphs is 19.6%, that is nearly one fifth of all the glyphs. Hence it is easy to find examples in which, on the contrary, 'the sun copulates with the fish', and sometimes also with the birds. Fischer does not mention the resulting chaos in which everything is copulating in all manner of unlikely combinations. Furthermore, it is by no means obvious in what sense this 'breakthrough' is 'phonetic'.

The table gives the variant readings for the three glyphs proposed by Fischer and Fedorova:

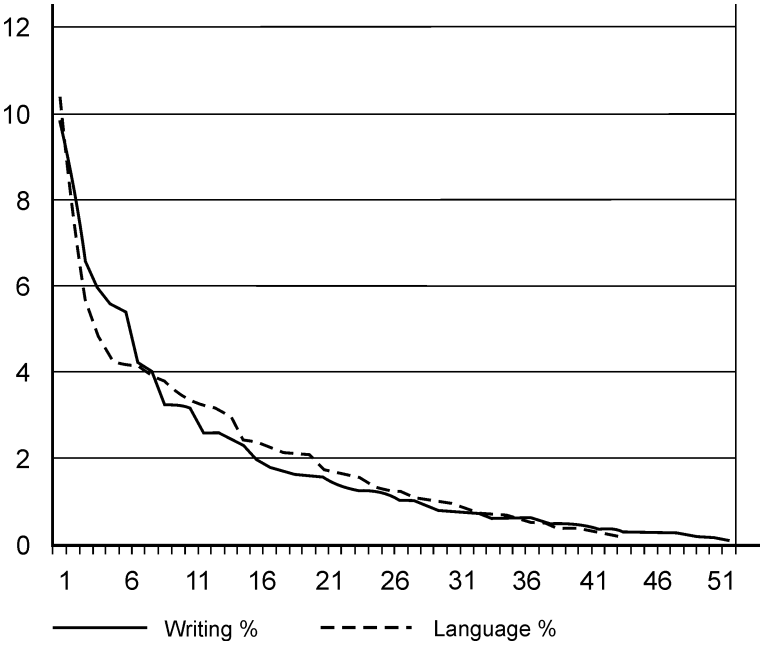
Glyphs (by Barthel)	Fischer	Fedorova
700, 701	*ika* 'fish'	*ika* 'plant' *uhi* 'yam tuber' *puku* 'a sort of batata' *ruhi* 'juicy'
8	*raa* 'sun'	*hetu* 'to move' *ketu* 'to dig'
400, 600	*manu* 'bird'	*turi* 'sugar-cane' *taha* 'frigate-bird'; 'yam'

The problem is not even that where Fischer sees a creation myth, Fedorova sees an agrarian ritual, and Ryabchikov a calendar. The problem is that all these versions which presuppose a logographic writing system conflict with the frequency distribution of the glyphs. Hence the inevitability of what Fedorova herself calls the 'maniacal' monotony of the texts' meaning. In general, in relation to the theory of decipherment, the situation may be stated as follows: in most instances, a jigsaw made up of a limited number of deciphered words indicates that the type of writing system has not been identified correctly. In connection with the Rapanui texts in particular, the probability that they use a logographic system is infinitesimally small.

If the writing is neither logographic not phonemic, only the third hypothesis remains to be tested: that Rapanui writing is syllabic. According to our data, Rapanui writing has 52 glyphs with a relatively high frequency of occurrence. The number of syllables in the Rapanui language can be determined easily, since consonant clusters are not permitted, no more syllables of VC structure. Rapanui has 10 consonantal phonemes and 5 vowels, meaning that there are 50 syllables with a CV structure and a further 5 syllables with the structure V: 55 syllables in total. Thus the inventory of glyphs and syllables is itself a weighty argument in support of the third hypothesis.

The coincidence between the inventory of glyphs and syllables does not prove that their relative frequencies coincide, however. In theory, syllables in language and glyphs in writing could be distributed quite differently. For example, some syllables could be very common, others very rare, yet the glyphs could all occur equally frequently (each of the 50 glyphs could make up 2% of the corpus). Graph 1 compares the frequency of glyphs in the written ten texts of the main corpus and the ten transcribed texts in the control corpus.

Graph 1. Frequencies in text (%)



The two curves are strikingly similar. The only important difference is that the curve for the language is a little shorter: there are fewer syllables than glyphs. That is entirely expected, because the texts in the control corpus use Fedorova's transcription [Fedorova 1978; 1988; 1993], which, as noted earlier, does not take account of the glottal

stop (syllables with a glottal stop are not distinguished from vocalic syllables). From the outset, the control corpus is five syllables short.

The most commonly used glyph (glyph 6) has almost the same frequency as the most common syllable (A). As will be shown below, this is by no means the only statistic which supports a link between the glyph and the sound. (Incidentally, the syllable A might appear a little more frequently that the glyph 6 because it has been combined in the transcription with the syllable ʔA.)

As noted already, in the fifties it was concluded that 'articles' (particularly the most frequent ones — *te* and *he*) could not be associated with any single glyph. That is not the case. The syllable TE has a frequency of 5.7%, which means that it could be associated with a number of common glyphs in Rapanui writing. The syllable HE has a frequency of 3.5%, which is close to the average frequency both of the syllables in the control texts transcribed from speech and of the glyphs in the written texts.

The number of units in the two corpora — language and writing — coincide, as do their distributions. These two facts (which are independent of each other) provide strong support for the hypothesis concerning the syllabic nature of Rapanui writing. Furthermore, there is no statistical support for the other hypotheses.

There is another criterion to consider, which again is independent of the other two. If the Rapanui writing is basically syllabic, it is legitimate to ask how boundaries between words are marked. The most natural markers are spaces left between groups of joined-up glyphs. It is interesting to compare the average length of a word in the control corpus with the average length of a 'graphic' word in the main corpus. If they differ markedly (which is entirely possible in principle, even if the frequencies coincide), the hypothesis about the syllabic basis of the writing system may be cast into doubt.

The table shows the relative frequencies of one-, two-, three- and four-syllable words in the control corpus and the corpus of 'graphic' words:

Type of word	**Language (control corpus)**	**%**	**Written texts (main corpus)**	**%**
1 syllable	2843	42	3006	45
2 syllables	2494	36	2178	32
3 syllables	1026	15	1193	18
4 or more syllables	484	7	354	5
Total	**6847**	100	**6779**	100

The average length of a word in the Rapanui language coincides almost exactly with the average length of a word in the written texts: 1.87 syllables and 1.85 glyphs respectively. This finding is important in many respects:

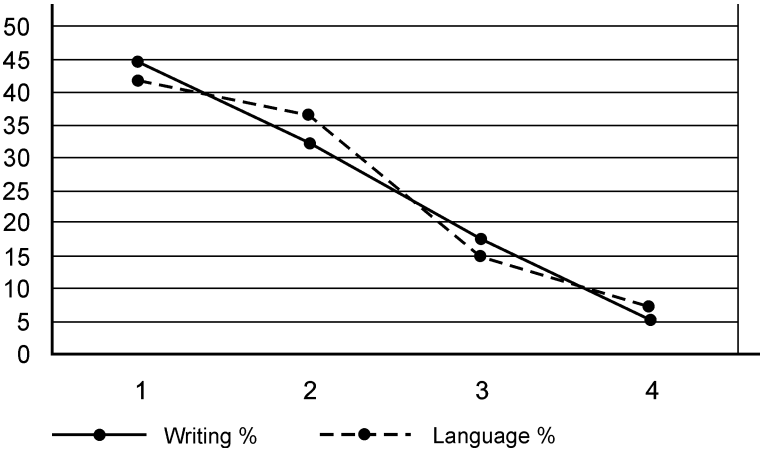
— Firstly, it provides a third parameter for the statistical description of Rapanui on the one hand and of the written texts on the other, and it greatly reduces the probability of chance similarities between them.

— It is a strong argument in support of the hypothesis that the writing system is syllabic.

— It makes it highly likely that the spaces between the groups of glyphs mark separate words, which is by no means obvious at first. It is no accident that those who assume that the writing system is logographic sidestep the 'difficult' question about the function of the spaces. If the glyphs that Fischer reads as 'bird', 'fish' and 'sun' are words, why do they often appear in combinations with other glyphs, without any spaces? If Fedorova often translates the ligature linking several glyphs as a phrase, why do these same glyphs often occur separately? The close correspondence between the length of words in the language and the length of sequences of glyphs in the written texts elucidates the function of the spaces.

The average length of words in the two corpora may coincide, but in principle it remains possible that the language and the texts could still differ in their percentages of one- and/or three-syllable words, say. As the table shows, the corpora are virtually identical with respect to this parameter, too, or rather with respect to four parameters: the percentages of words with one, two, three, and four or more syllables. The distribution of words (represented as a percentage of the total) is shown in Graph 2.

Graph 2: Word structure in the texts (%)



The frequency with which classes of words appear in the written texts is in an almost exact inverse proportion to their length. An increase of one syllable in length is associated with a 13% reduction in frequency. This conclusion is interesting in its own right, but it takes on a special importance when compared with the distribution of words in the Rapanui language: yet again the two corpora are virtually identical. The percentage of disyllabic words is a little higher in the texts transcribed from speech than in the written ones, while the written texts contain a few more words of three syllables, and possibly also of one syllable, but the nature of the relationship remains the same. The relationship between word length and word frequency is self-evident as far as the language is concerned, but it by no means obvious that it should hold for tablets containing unknown texts written using undeciphered glyphs. Furthermore, if we base our calculations on Barthel's catalogue of 600 glyphs, rather than 52, by definition we obtain quite different frequencies of graphic words: the percentage of monosyllabic 'words' is more or less doubled (because of the ligatures included in Barthel's catalogue), and there is a correspondingly sharp decrease in the percentage of polysyllabic words. We consider this an exceptionally important point: the fact that the results obtained obey Zipf's Law creates the impression of their predictability.

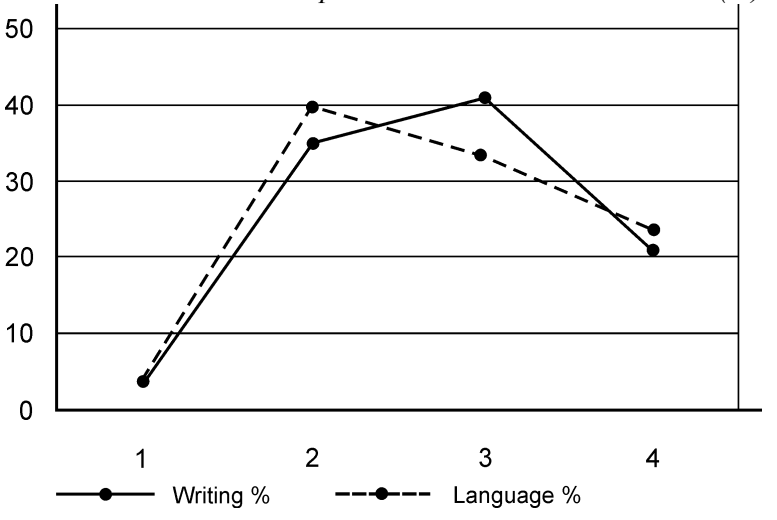
There have been many interpretations of Zipf's Law, according to which short words should be used in a text more often than long ones. Deviations from the classic curve (in our case the distribution of data in the corpus of written material is represented by a virtually straight line) depend among other things on the grammatical structure of the language: in Rapanui, grammatical morphemes, which for the most part are monosyllabic (articles, determinatives, prepositions), occur with a very high frequency. However, we are in a position to compare the number of different words in the samples of language and in the written documents, which is important because Zipf's Law applies to texts, but not to the lexicon. Frequency in the lexicon is yet another independent parameter that can be used to compare the language and the writing.

In the Rapanui control corpus there are 1047 words of varying length. In the written documents there are 1461 combinations of glyphs, separated by spaces. (This figure includes glyphs used on their own.) The distribution of the data presented in the table is plotted in Graph 3.

In the lexicon as in samples of text, the number of disyllabic words is a little higher in the Rapanui language than in the writing, while the number of three-syllable words is a little higher in the writing than in the language. But on the whole the curves are very similar: two- or three-syllable words make up 35–40% of the data and words

Type of word	Language (control corpus)	%	Written texts (main corpus)	%
1 syllable	39	4	51	4
2 syllables	416	40	511	35
3 syllables	347	33	596	18
4 or more syllables	245	23	303	21
Total	**1047**	100	**1461**	100

Graph 3: Word structure in the lexicon (%)



with four or more syllables a little over 20%; there are very few monosyllabic words in the lexicon, which is to be expected in a language with few syllables. Thus this parameter too provides clear evidence of the similarity between the two corpora.

In the lexicon of the control sample, the average length of word is absolutely identical to the average length of sequences of glyphs in the lexicon of the written texts: 2.8 syllables/glyphs. This statistic, too, is independent; it cannot be derived from any of the earlier ones. The average word length in the texts is completely different: 1.9 syllables in both corpora. The 'average' word is shorter in the samples because of the high frequency of monosyllables used there.

Thus according to all these general statistical criteria — over ten in total — there is a correlation between the Rapanui language and the Rapanui writing. The diversity of the criteria used makes the probability of a chance correlation close to zero. The following conclusions can be drawn:

a) we are dealing with a proper writing system;

b) this writing system is based upon the Rapanui language (or a language with similar statistical characteristics, for instance another East Polynesian language);

c) the writing system is primarily syllabic.

Nevertheless, statistical analysis of the samples of the language does suggest that some of the glyphs in the Rapanui writing may represent words rather than syllables. The frequency of some words with more than one syllable is exactly comparable to that of monosyllabic words, and indeed sometimes exceeds it. Examples of such high-frequency words in the control corpus include: *mai*, *oho*, *ariki*, *ai*, *hau*, *tuu*, *maka*, *ana*, *kuhane*, *tau*, *nei*, *ingoa*, *nape*, *noho*, *kiroto*, *hiva*, *era*, *matua*, *uta*, *kainga*, *atua*, *vaka*, *vai*, *rau*, *hotu*, *rua*. Their frequency varies between 0.4% and 1.7%, and this is the range within which the frequency of approximately twenty Rapanui syllables falls. For example, in the control sample the syllable TU comprises 0.2% of the total, which is less than the frequency of any of the words cited. There is no statistical reason why words such as *ariki* 'leader', *ingoa* 'name' and *rua* 'two' could not be represented by special glyphs. Furthermore, as already noted, the documents contain two dozen glyphs (assigned the index 999 in the catalogue) whose scarcity means that statistically they cannot be taken to represent syllables. We are not in a position to say whether the documents might attest to the presence of determinatives and phonetic complements, since we have no information about their possible frequency.

Our general conclusion is that it is highly likely that the glyphs of Rapanui writing represent syllables, and possibly also some common words.

3. Statistics and prospects for identifying the glyphs

It is entirely natural that an attempt should be made to use statistics in the decipherment of Rapanui writing, if only because the researcher has practically no other effective strategies at his disposal. At the beginning of this project, we thought a statistical approach to the task promising, particularly since just such an approach had demonstrated the fundamental similarity between the Rapanui language and the Rapanui writing. Nevertheless, it is clear that results obtained in the course of statistical analysis can vary greatly in accordance with a) the catalogue of glyphs; b) the methods of statistical analysis.

We present here the preliminary results obtained on the basis of the catalogue of glyphs listed earlier. In our view, some of the results open up real prospects for the decipherment of the Rapanui docu-

ments. Others, by contrast, make it possible to formulate clearly the difficulties of relating glyphs to syllables.

3.1. Relative frequency

Frequency is one of the most obvious criteria to draw on in a statistical approach to decipherment. How useful is it in relation to the Rapanui writing, and what complications does it bring?

The first problem is that the frequency of the glyphs varies from text to text. A number of glyphs which have a high frequency in one are not found at all in others. This is particularly true of the text found on the staff (I): the use of glyphs there is so unusual that it had to be excluded from the corpus of texts for statistical analysis. But even without I there is considerable instability in the frequency with which the glyphs appear. For instance, in the text on the front of G (text Gr) every seventh glyph is glyph 1 (frequency 14.1%), while in B this glyph has a frequency of just 3.4%. There is a difference of 10.7% between the minimum and maximum frequencies, against an average frequency of 5.6%. Glyph 200 has a frequency of 10.8% in B, but 4.5% in C. Glyph 41 (and its alloglyph 42) has a frequency of 5.8% in C, but it is hardly found at all in text Gr (frequency 0.5%). Thus there is a fundamental difficulty in using the frequency of occurrence as a criterion for comparing glyphs to syllables.

How is this variation in frequency to be interpreted? It can be explained by diverse factors — from the genre characteristics of a particular text to errors in compiling the catalogue of glyphs. Notably, however, there is no less instability in the frequency of syllables in the Rapanui language than in the frequency of glyphs in the writing. For example, among the texts making up the control corpus the difference between the minimum and maximum frequency of the syllable KI is 8.5%, of RA and HE 6.2%, and of E and A 6.1%.

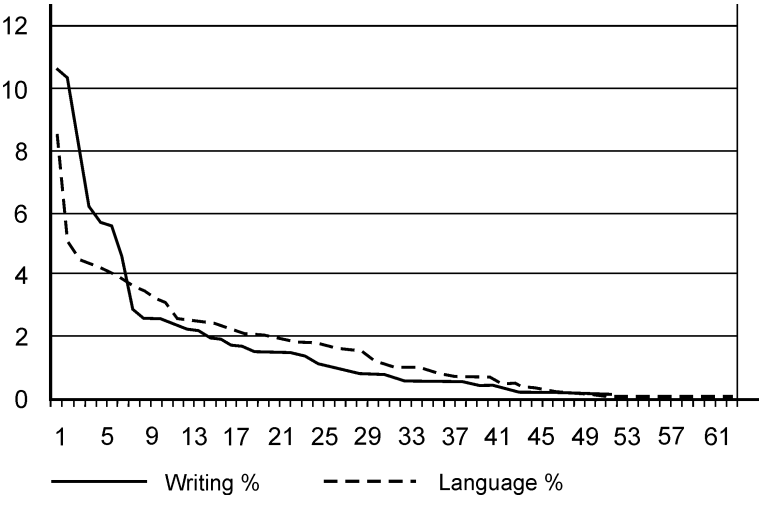
According to this parameter, too, there is a striking overall similarity between the language and the writing: there are 12 syllables and 12 glyphs whose maximum and minimum frequencies differ by 4% or more; for 13 syllables and 11 glyphs the figure is 2–4%.

The reasons for the variation in the frequencies of syllables can easily be explained: in the vast majority of instances it is caused by the repetition of one or more words. For example, in the famous text Atua mata riri, upon which Fischer based his decipherment, in almost every phrase (41 times) the words *kiroto* and *kapute* are repeated, since the entire text is of a single structure: each phrase means — *'X copulated (ki ai kiroto) with Y and in the world there appeared (ka pu te) Z'*. Clearly, the frequency of the syllables KI, RO, TO, KA, PU, TE here is going to be much higher than in the other texts.

It is reasonable to assume that the same consideration will apply in the written texts, which strongly resemble the language with regard to frequency variations. Yet this is precisely why we cannot base comparisons of syllables and glyphs on their frequencies in the texts: there are no grounds for supposing that those words which are repeated often in the transcribed Rapanui texts are the same as those attested in the extant Rapanui written texts.

In comparing syllables with glyphs, it is more effective to consider their frequency in the lexicon rather than in the texts. Graph 4 presents these data.

Graph 4: Frequencies of glyphs and syllables in the lexicon.



As the graph shows, this criterion is the first we have examined in which the Rapanui language and the Rapanui documents diverge substantially. The 6 most common glyphs in the 'lexicon' of the written texts have a higher frequency than the most commonly used syllables. The relative frequency of all the other glyphs turns out to be lower than the frequency of syllables. It is worth stressing that here the frequency of words in the texts does not influence the distribution: the point is that the most common syllables/glyphs are found in the largest number of words/graphic words. By way of example, the glyph 200 has a frequency of 10.7% in the lexicon of sequences which are isolated by spacing; approximately every tenth combination includes glyph 200 (in actual fact the number of combinations is slightly smaller, since the glyphs can appear more than once in the same graphic word). Other common glyphs in the lexicon include (in descending order of frequency) glyphs 6, 10, 3, 62, 400, 61.

This divergence, which is especially noticeable with the two most common glyphs, is compensated for by the fact that there are more

syllables than glyphs with a frequency of over 3%. The totals for both the language and the writing, rounded to the nearest 5%, are: syllables/glyphs used 3 times or more: 50%; syllables/glyphs used 2 or 3 times: 20%; syllables/glyphs which occur less frequently: 30%.

This confirmation of a statistical similarity between the syllables of the language and the glyphs of the written texts should no longer cause any surprise — their similarity across a diverse range of parameters is so great that it cannot be explained by chance factors. It also gives us grounds for preferring a decipherment strategy which seeks out correspondences (statistical, structural, graphic) between glyphs and syllables until such time as cogent counterarguments against the syllabic nature of the writing system are proposed. On the basis of the facts set out above, we intended to evaluate every syllable in the language and every glyph in the texts according to the largest possible number of statistical parameters, with the goal of ascertaining by reference to these statistical 'passports' the phonetic value of at least some of the glyphs.

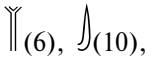
It is worth considering not only the overall approach to decipherment, but also each parameter in isolation, since an analysis of individual characteristics (including frequency analysis) provides us with new information about the written texts and sometimes about the language as well. The curious relationship between the length of a word and its frequency in the texts has already been discussed. There is another 'incidental' result, which to the best of our knowledge has not attracted linguists' attention until now. It is evident to specialists that that syllables of structure V have an especially high frequency in the lexicon of Rapanui. Less obvious is that syllables with a CV structure which include the phoneme /a/ also have a high frequency. There are 11 syllables in total which occur with a frequency of over 3%: a) I, E, A, O, U; b) TA, RA, KA, NA, MA; c) RI. The prominence of syllables with Ca structure in the group of the most commonly used syllables, and the markedly higher frequency of the syllable A than of other vocalic syllables, prompts systematic examination of the statistical distribution in the lexicon of syllables containing different vowels. The table records the incidence of syllables containing each of the 5 vowels:

i, Ci	u, Cu	e, Ce	o, Co	a, Ca
16%	15%	13%	18%	38%

The syllables of Rapanui are reasonably evenly distributed among the different degrees of height: the close vowels /i/ and /u/ are found in 31% of syllables, the mid vowels /e/ and /o/ in 31%, and the open vowel /a/ in 38%. The first two groups both contain two vowels, but even so they remain notably less frequent than syllables with /a/.

It is therefore highly likely that the most common glyphs in the lexicon of the Rapanui written texts represent either vowels or syllables containing the phoneme /a/.

What sort of support for this hypothesis can be sought in analysis of the writing? What sorts of glyphs might represent vocalic syllables, which all occur very commonly? (25% of the Rapanui lexicon comprises vocalic syllables, and the three most frequent syllables, /a/, /u/, /i/, form the three corners of the basic vowel triangle.) The most frequent glyphs in the documents share a common graphic feature. In the group of 6 glyphs whose frequency, like that of vowels in the language, exceeds 3% there are 4 which depict an arm in some form:



The table below shows various ligatures which in Barthel's catalogue are treated as individual glyphs. These similar combinations of glyphs illustrate that glyphs (6), (10), (61) and (62) do indeed depict an arm.

					
326	324	325	321	322	323
					
406	404	405	401	402	403

Commentary:

— The arm depicted in ligatures 326 and 406 is an alloglyph of glyph 6. The case for treating these graphemes as a single glyph has been made in [Pozdniakov, 1996].

— The table also includes the glyph  (63), which has a somewhat lower frequency: 2.4%. It is by no means agreed that this component of ligatures 323 and 403 represents an arm. The accepted interpretation is that it represents a ceremonial axe (*hoki*, in the readings of Metoro and, after him, Fedorova). Glyphs 6 and 61 have also been interpreted differently.

Thus the frequency group of the five vocalic syllables corresponds graphically to this group of four common glyphs (whilst there are six common glyphs in total). This coincidence cannot be ignored. The phonetic and graphic parallelism is in this instance so clear that it allows us to formulate the following working hypothesis: **glyphs whose prototype is a picture of an arm represent vocalic syllables**. It goes without saying that before it can be accepted this hypothesis must be tested against all the other parameters.

In that case it will be appropriate to add a further glyph to the group — 𓅓 (901). This glyph, which depicts a wing (fin? arm?) is the only glyph in our catalogue (apart from the 'technical' glyph 999) which has not previously been identified by researchers, including Barthel. (The problem with Barthel's catalogue is that it has too many glyphs, not too few.) We are not concerned with the wing which forms the leftmost element of the ligatures 406, 404 and 405, among others (see table above), but with the special form of wing which in Barthel is found notably in glyphs 407 and 408:

407 408

The hypothesis devised on the basis of frequency analysis finds strong support from the statistical data relating to positional criteria and to other criteria, which will be examined below.

Frequency analysis allows us to draw tentative conclusions regarding individual glyphs: it is highly likely that individual glyphs represent individual syllables; it is unlikely that relatively rare glyphs (e.g. glyph 16) represent common syllables (e.g. the syllable MA).

If a glyph corresponds (or does not correspond) to a syllable not only in frequency, but also on other statistical criteria, the identification of a glyph with a syllable becomes very much more probable. It is possible to model the generalised probability of the identification of each glyph and present possible readings as a scale of probabilities, as will be discussed further below. First, however, we will consider other statistical criteria which can be used alongside frequency statistics to help determine the phonetic value of the glyphs found in the Rapanui writing.

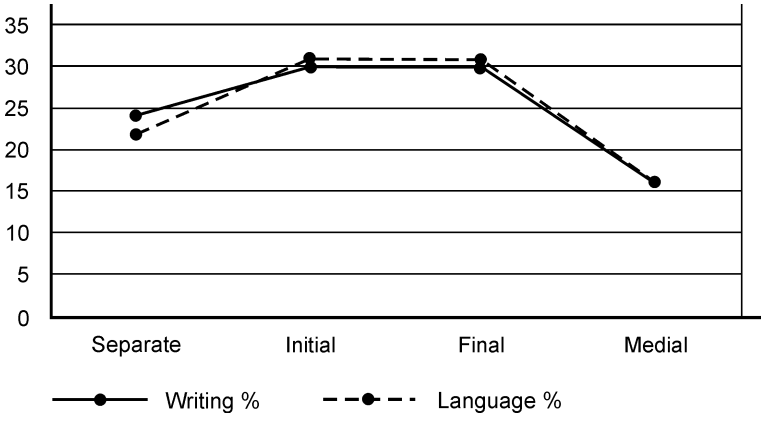
3.2. Position

Since it has been established that the spaces in the documents are very likely to separate graphic words, it becomes possible to compare the positions of each glyph and each syllable within words: some glyphs/syllables appear more often than others at the beginning of a word, some at the end, and others in a medial position. Some glyphs/syllables can be found separately, others cannot. A syllable in a separate position is a monosyllabic word, but that same syllable can also have a particular frequency of usage in each of the three possible positions (initial, medial and final) in polysyllabic words. Therefore, glyphs/syllables can be compared not only according to their frequency, but also according to four positional criteria.

In the positional distribution of glyphs the overall number of initial and final glyphs is necessarily the same (though any given glyph may

appear primarily in initial or primarily in final position). A comparison of the frequency of glyphs/syllables in the four types of position in the documents is presented in Graph 5.

Graph 5: Distribution of glyphs/syllables by position (%)



The positional distribution of syllables in the language and glyphs in the writing virtually coincides. The slightly larger proportion of syllables used separately in the writing may be explained by the presence of determinatives in the written language. This coincidence is yet another weighty argument in favour of the hypothesis about the syllable nature of the Rapanui writing system.

3.2.1. Glyphs used separately

The frequency of any given syllable found separately in the control texts is directly connected with the frequency of monosyllabic words: many monosyllabic morphemes have an extremely high frequency (for example, the syllable TE, which is also a monosyllabic verbal marker), whilst some syllables do not form words and correspondingly are not used separately (for example, the syllable RU).

It is on this parameter that the syllables and glyphs diverge most markedly:

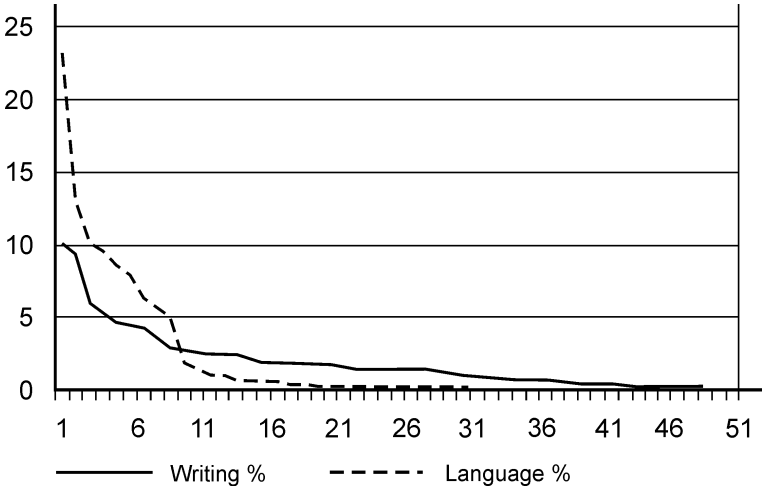
— Almost all the glyphs (apart from glyph 901) can be used separately, while in the Rapanui language only 30 syllables can function as monosyllabic words.

— In the control texts the frequency of the most common syllables (= separate words) is significantly higher than the frequency of the most common separate glyphs.

There are other differences to which we will return below.

Graph 6 plots the frequency of the glyphs and syllables found separately.

Graph 6: Frequencies of separate glyphs and syllables (%)



The first of these differences gives cause for hope rather than despair. As was shown above, the frequency distribution leads to the supposition that alongside the syllabic and also morphemic glyphs in the Rapanui writing there might also be a developed system of determinatives. If so, it is only natural that the curve for separate glyphs should be longer than the curve for separate syllables. Indeed, a coincidence between the curves would be much harder to explain, for then there would be no support for the hypothesis concerning the use of determinatives in the writing system.

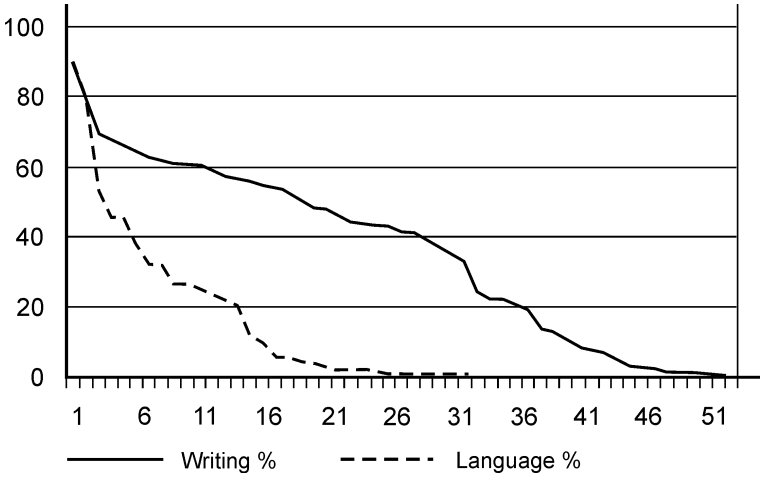
The second difference is entirely unconnected with the first. The curve plotting the frequency of syllables is steeper than the curve for the glyphs. No individual glyphs have such high frequencies as the first two syllables. Nearly every fourth syllable which stands alone (23.2%) is the morpheme TE. (For the morpheme HE the figure is 12.9%.) The glyph most commonly used in isolation (glyph 2) has an incidence of 10.2%. That would seem to indicate that there is no glyph in the writing system which can be identified with the morpheme TE on this criterion. But another factor must be taken into account. Analysis of the distribution of separate syllables in the control texts (separate syllables being basically grammatical morphemes and 'articles') shows that their frequency is significantly less stable than the frequency of 'non-morphemic' syllables. The text Apai provides convincing proof: in this text the article HE does not appear at all, whilst in the text Hotu Matua, of approximately equal length, it is found 119 times. On the one hand, this makes it possible to dismiss the relatively low frequency of separate glyphs in the Rapanui writing as quirks of particular texts: the complete absence of the morpheme TE from the undoubtedly archaic Apai is regarded

as just such a quirk. On the other hand, it casts doubt upon the usefulness of this criterion for the task of decipherment — as soon as the frequency of a separate syllable varies from 0 to 30% across the texts, it may be associated with any of the glyphs.

It is worth pointing out how this statistic works: it indicates what percentage of the total number of syllables/glyphs which occur separately is made up of any given syllable/glyph. For example, on average the syllable KI makes up 10.2% of all syllables used in isolation, i.e. every tenth monosyllabic word in the Rapanui control corpus is the word *ki*.

If we relate the number of uses of a syllable in a separate position to its total number of uses in all four positions (separate, initial, medial, final), we obtain another statistic for comparison, its index of separation. There are syllables (mainly grammatical morphemes) which occur predominantly in a separate position (so the syllable TE has an index of separation of 90.1%, whilst for TI the figure is 3.2%). The indices for two corpora are presented in Graph 7.

Graph 7: Index of separation for glyphs and syllables (%)



The two curves are strikingly different. This is the first major statistical divergence between the language and the writing to be observed in our work, and it is one of the very few marked ones. As the graph shows, the index of separation for the majority of glyphs exceeds 20%, whilst in the language, with the exception of the syllables TE, HE and KI, the index of separation does not lie in excess of 50 per cent. What this means is that the hypothesis about the syllabic nature of the writing system which was based on an analysis of frequency distributions stands in need of fundamental adjustment: in addition to their syllabic function, glyphs used separately (or many of them at any rate) have another function which

remains to be elucidated. It is possible that the high index of separation for the documents reflects the presence of a system of determinatives.

This hypothesis is confirmed by both the distribution of separate glyphs in the written texts and by the distribution of separate syllables in the language. In order not to overload this article with statistics, we will present just the most crucial finding. Our data show that in the language there is a fairly regular alternation between monosyllables and polysyllabic words, as might be predicted. It is natural that, for instance, a monosyllabic article should stand alongside the often polysyllabic noun which it determines. The distribution of separate glyphs in the writing is of a different kind: there, a glyph used in isolation is very likely to be found alongside several other separate glyphs (or indeed alongside the same one). In other words, the concentration of separate units in the written texts is significantly higher than in the language. Certain fixed groups of glyphs can be identified. What is denoted by these groups of glyphs which are unlikely to represent syllables? They could be complex determinatives, that is, complex glyphs which serve to indicate some particular value (possibly phonetic) of other glyphs.

It seems important to emphasise that we do not have enough data to make the case for this hypothesis, but on the basis of comparative statistical analysis we can at least draw attention to the divergence between the writing and the language with regard to the index of separation. None of the numerous logographic approaches to decipherment has even raised the issue, and it would be gratifying if their proponents would explain why, if the glyphs represent words, they sometimes stand in isolation from others (groups of separate glyphs) and at other times are joined up (groups of glyphs joined in ligatures).

Another unexpected finding may be of interest not only to specialists in Rapanui writing, but also to researchers in the field of Polynesian linguistics. As remarked earlier, the frequency of the most common monosyllabic morpheme in the Rapanui language, the morpheme *te*, is highly unstable, varying considerably from text to text. The same applies to other monosyllabic grammatical morphemes in Rapanui, foremost among them *he*, *ki*, *ka*, *ko*, and also to all the vocalic grammatical morphemes — *i*, *e*, *a*, *o* (Rapanui does not have a morpheme *u*). It turns out that whilst each of these morphemes is distributed very unevenly in its own right, the overall frequency of pairs of morphemes is in some instances remarkably stable. For example, it is possible to predict that if in a given text the verbal marker *te* has a high frequency of occurrence, then the article *e* will hardly ever be found, and vice versa. Moreover, not only in isolated positions, but also in general, the overall frequency of the syllables

TE and E is stable at 8–11% of the total number of syllables in any text. Judging by the statistics, this dependence is too clear to be accidental. Yet it is difficult to suggest a coherent linguistic interpretation of this relationship. The statistical distribution gives the impression that there exists in a Rapanui text a kind of 'balance of frequencies' between articles and verbal markers. A text may be 'verbal' ('he came, he saw, he conquered') or 'nominal' ('Friends, Romans, countrymen'); a text may be composed primarily of personal names (telephone directory, genealogy), but the proportion of articles in it remains constant, as the statistics show.

The most general conclusion that can be drawn with regard to the feature of separation is that the distribution of separated syllables (monosyllabic morphemes) in the language and of separated glyphs in the writing reveals not one, but two major differences between the Rapanui language and the Rapanui texts:

1) articles, which are distributed fairly evenly throughout the Rapanui language, are not indicated in writing by separate glyphs: either they are not indicated at all, or, as seems more likely, they are indicated by glyphs which form part of a single graphic word (it is entirely possible, for instance, that in polysyllabic words glyph 200 indicates the article);

2) groups of separate glyphs are not phonetic glyphs and may represent complex determinatives.

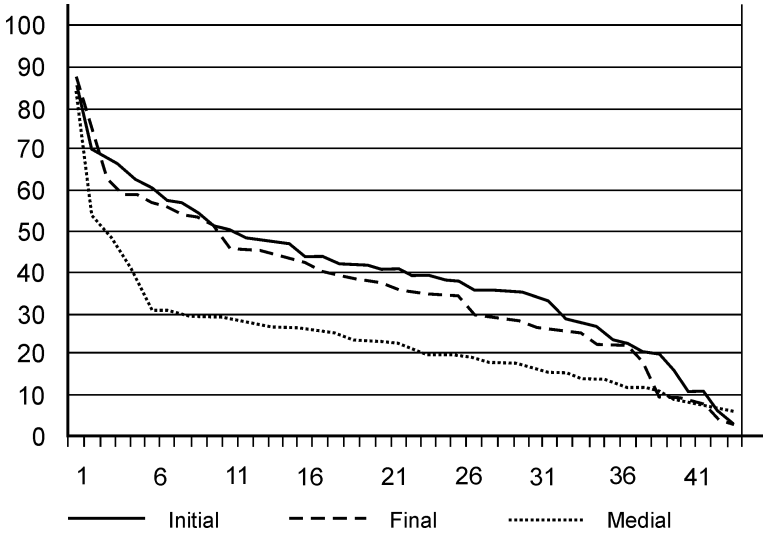
For practical purposes this means that attempts to establish the phonetic value of the glyphs must focus on combinations of glyphs which are not separated from one another; as possible compound determinatives, the fixed groups of separated glyphs must be studied in their own right.

3.2.2. Positions within polysyllabic words

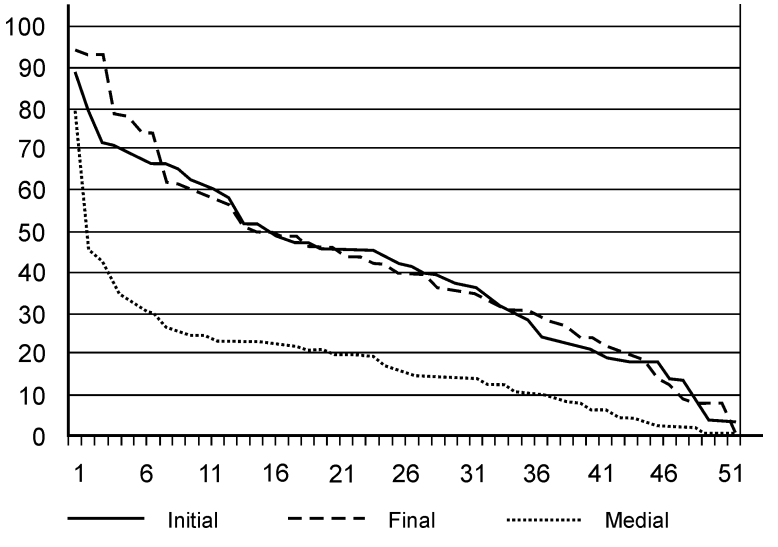
Despite the almost complete correspondence between the percentage of initial/final and medial syllables/glyphs in the language and the writing, the frequency distributions in the two corpora could in principle be radically different. For example, all the syllables in the language could have approximately the same frequency in medial position, whilst glyphs in that same position could be distributed unevenly. There could also be important differences in the steepness of frequencies of syllables/glyphs in initial position. But on this parameter, too, the data for the language and for the writing are virtually identical, as Graphs 8 and 9 illustrate clearly.

In both corpora, the frequency distribution for medial position differs markedly from the distribution for the other positions. Only a few syllables/glyphs appear predominantly in the middle of words. Most glyphs appear in medial position with an incidence of less than 25%.

Graph 8: Frequency distribution by position in the language (%)



Graph 9: Frequency distribution by position in the writing (%)



The frequency distributions for initial and also for final position are almost identical for the language and the writing. There are around a dozen syllables which are predominantly initial, a dozen which are predominantly final.

Once again we stress that Graphs 8 and 9 compare two absolutely distinct sets of data: syllables in the Rapanui language and glyphs in the undeciphered writing. The likelihood of a chance statistical

correspondence between them is so small as to be negligible. Their similarity between the syllables and the glyphs across a whole range of independent statistical parameters is an incontrovertible indication of the syllabic nature of the writing system and shows that their linguistic basis is the Rapanui language. Any other hypothesis conflicts with the statistics.

The following syllables and glyphs, arranged in descending order of frequency (from 90% to 50%), are found mainly in initial position:

MA, ME, MO, HA, NO, PO, PI, TA, HE, KU, O;

60, 380, 4, 91, 67, 200, 15, 99, 22, 720, 41, 45, 25, 700, 50.

The following syllables and glyphs are virtually never found in initial position (frequency from 10% to 3%):

NGA, NGO, NGI

9, 76, 74, 52.

This distribution demonstrates that in the Rapanui language there is indisputably a correlation between the type of consonant and its frequency in initial position: the three most frequent and the three least frequent initial syllables all contain a nasal consonant: the most frequent all contain /m/, the least frequent /ŋ/.

The following syllables and glyphs, arranged in descending order of frequency (from 95% to 45%), are found mainly in final position:

NGI, NGA, I, PE, NA, U, NE, KE, KI, HO, E, A, TI, HU, VA, RE, TO;

74, 76, 52, 53, 71, 3, 9, 63, 901, 16, 7, 48, 70, 660, 27, 69, 59, 44, 2, 6, 62, 61, 10.

A high frequency in final position is yet another characteristic which consistently links vocalic syllables and 'arm' glyphs. Thus positional frequencies serve to confirm the hypothesis about their correspondence.

The following syllables and glyphs are virtually never found in final position (frequency from 10% to 1%):

PO, NGO, HA, MA, PI, ME;

200, 99, 60, 380, 240.

The following syllables and glyphs have the highest frequencies of appearance in medial position (from 84% to 40%):

NGO, NI, RI, RO;

240, 38, 61.

The following syllables and glyphs have the lowest frequencies of appearance in medial position (from 10% to 0%):

NGI, NE, NO, KI, MA;

16, 2, 700, 4, 3, 45, 27, 7, 52, 67, 60, 76, 74, 53, 71.

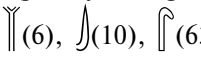
There is a notable divergence between the language and the documents: significantly more glyphs than syllables appear only rarely in medial position.

3.3. Repetition of syllables/glyphs

3.3.1. It is said that in Rapanui, as in other Polynesian languages, the lexicon contains an unusually high frequency of words in which a syllable is repeated, for example: *haha* 'mouth', *mamari* 'egg', *hehehehe* 'earth', *hihi* 'eyebrow', *hohonu* 'deep', *rarara* 'to condemn'. In the lexicon of Rapanui repeated syllables are found approximately 1.5 times more frequently than would be predicted by a chance distribution. For instance, the combination RA–RA occurs six times in the lexicon of the control corpus (in the words *rara*, *rarama*, *raraku*, *ngarara*, *pararara*), rather than the three times that would be expected on the basis of the frequency of the syllable RA.

In the lexicon of graphic words, the number of repeated glyphs is very close to the predicted norm, exceeding it by only 8%. In other words, on this criterion the language and the writing diverge substantially, although the proportion of combinations of repeated syllables/glyphs to the total number of pairs of syllables/glyphs is absolutely identical for the language and the writing at 5.3%.

An important point emerges in connection with the hypothesis which links glyphs depicting arms with vocalic syllables. In contrast to most syllables, which are repeated in the lexicon more often than expected, there are four syllables that are repeated altogether less often than their overall frequency would suggest likely: I, A, U, MA. In the lexicon of Rapanui compiled on the basis of the control corpus of ten texts one would expect to find 4 combinations of the syllables I–I, but there is not a single such combination. The combination A–A is found 6 times instead of the expected 14. In the lexicon of graphic words in the writing, only three combinations of repeated glyphs are found less frequently than predicted:



For example, the combination 6–6 glyphs is found 15 times rather than the expected 29.

Thus once more we find that there is a correlation between the statistical characteristics of vowels (or at least the three basic vowels in the triangle) and of glyphs which depict various forms of arm.

3.3.2. In addition to syllables repeated in immediate succession, the lexicon also contains an extremely large number of words in which repeated syllables alternate (e.g. *ro-ngo-ro-ngo*, *ka-i-ka-i*, *a-ku-a-ku* etc). In the vast majority of cases this word structure is determined by the functionally marked reduplication of words with a CVCV structure (this conclusion, arrived at on the basis of an analysis of the words in the Rapanui lexicon, makes it reasonable to suppose that words of the type *hehehehe* 'earth' should also be interpreted as reduplications of roots with a CVCV structure — *hehe-hehe*).

Statistical analysis shows that the word structure $CV^1–CV^2–CV^1–CV^2$ (henceforth, 'interval 2') is even more characteristic of the Rapanui language than the structure $CV^1–CV^1$ (henceforth, 'interval 1'), and furthermore that it is characteristic not only of the language, but also of the Rapanui writing. In the lexicon of Rapanui the proportion of words with the structure $CV^1–CV^2–CV^1–CV^2$ is exceptionally high. There are 227 such words in the lexicon, 756% more than the 30 that would be expected in our control corpus.

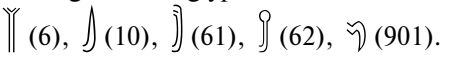
On the 'interval 2' criterion, as with most other statistical criteria, the Rapanui language and the Rapanui written texts share an evident similarity. In our corpus there are 137 combinations of the form $glyph^1–glyph^2–glyph^1–glyph^2$ (predicted figure: 64 combinations), that is, 214% of the norm.

Nevertheless, this characteristic remains very much more prominent in the language than in the writing: the frequency of words with the structure $CV^1–CV^2–CV^1–CV^2$ is not simply high, but exceptionally high. Unlike 'interval 1', in the language such words form a significantly higher percentage of the total number of four-syllable combinations: they form 10.6% of all such combinations in the writing, 27.1% in the language. This is a fundamental difference between the corpora. It is possible that it reflects structural peculiarities of the Rapanui language. A working hypothesis which would be worth testing is that the altogether lower percentage of reduplicated graphic words could suggest the presence of a special glyph indicating reduplication. In that case, words with the interval 1 structure $CV^1–CV^1$ and words with the interval 2 structure $CV^1–CV^2–CV^1–CV^2$ would be given in the writing by the formulae $glyph^1$ + reduplicator and $glyph^1–glyph^2$ + reduplicator respectively. The statistical characteristics of the glyphs 3  (in postposition) and 200  (in preposition) in particular make them the most promising candidates for such reduplicator glyphs; they differ markedly on a number of important parameters from the possible syllabic glyphs.

The results of a structural analysis of variants of glyphs show that reduplication of the root could in principle be indicated not by a special glyph, but rather by the orientation of a glyph in the text. For instance, it is not known what function is served by a mirror-image

orientation of glyphs (and in particular why the head on anthropomorphic glyphs should sometimes be turned to the left, when most of the time it faces right). The possibility that structural elements of this sort could be used to indicate reduplication cannot be ruled out.

On the 'interval 1' criterion, the characteristics of vocalic syllables and arm glyphs coincide: it is these syllables/glyphs which are most rarely found next to one another. Comparing these groups on the interval 2 criterion again reveals their systematic similarity. In the lexicon, interval 2 combinations of identical vocalic syllables and combinations of identical glyphs depicting arms have the highest frequencies of all. The syllable A appears in 23 combinations (predicted frequency: 8), the syllable O in 12 combinations (predicted frequency: 1), and the syllable U in 13 combinations (predicted frequency: 3). These are the sharpest deviations in the entire data set. In the written texts the greatest deviations from the predicted norm involve the following familiar glyphs:

 (6), (10), (61), (62), (901).

Glyph 6 appears in 29 combinations against the predicted 14 (a deviation of +15, which coincides exactly with the deviation shown by the syllable A, and which is the greatest deviation in both language and writing). Glyph 10 appears in 22 combinations (predicted: 9), glyph 61 in 15 combinations (predicted: 3), glyph 62 in 14 combinations (predicted: 4), and glyph 901 in 5 combinations (predicted: 0). There are virtually no other glyphs other than those depicting arms which are comparable to the vocalic syllables on this criterion.

4. Prospects for using statistics to interpret the glyphs

The texts and the lexicon of the two Rapanui corpora have been systematically compared according to a dozen statistical criteria, some of which (for instance, the criterion of combinability of glyphs/syllables) have not been discussed in the present article. Each glyph and each syllable has been described with reference to a whole range of statistical parameters, which provide concrete arguments for (and also against) the decipherments that have been proposed. For example, it is quite likely that the glyph  (7) could be read as PU, MO or TO; it is not impossible that it could be read as RE; but its statistical characteristics do not coincide at all with those of syllables such as TA or RI.

As has been mentioned on more than one occasion already, the characteristics of vowels, and especially the cardinal vowels I, U and A, virtually coincide with the characteristics of glyphs which depict arms: they occur very frequently; they very often appear in final position in a word; they are rarely combined with each other; they

are often repeated in combinations in which two identical glyphs/syllables are separated by another glyph/syllable. It is highly likely that the glyph  (6) is to be read as A, and the glyph  (10) as I.

Of course, statistical criteria alone cannot be used to determine the phonetic value of the glyphs. The following fundamental difficulties were encountered in the present statistical approach to the decipherment of Rapanui writing:

— There are some glyphs whose statistical characteristics can be compared to those of twenty or more syllables, which precludes even tentative hypotheses regarding their phonetic value. Some glyphs do not correspond statistically to any syllable.

— The statistical characteristics of a number of glyphs correspond not only to syllables, but also to certain monosyllabic words.

— Any way in which the writing system deviates systematically from the phonetic principle (for example, using a special glyph to mark reduplication instead of repeating the syllabic glyphs) greatly reduces the possibility of being able to use statistics to identify which glyphs correspond to which syllables.

— The statistical characteristics of syllables differ markedly amongst the diverse texts in the control corpus taken from the Rapanui language. This presents an insurmountable obstacle to the reliable interpretation of the glyphs on a statistical basis. Suppose that the characteristics of a particular glyph coincide with the characteristics of a particular syllable in just one of the ten texts in the control corpus. Does that rule out even the possibility of identifying the sign with the glyph? Probably not. It is not known what sorts of texts the extant Rapanui writing records. It is possible that they are all of the same genre as that one control text in which there is a correspondence between the statistical characteristics of the glyph and the syllable. Moreover, if the statistical characteristics of syllables are so heavily dependent on the meaning of the text, it becomes very likely that the characteristics of glyphs and syllables will fail to correspond in a single text simply because the texts recorded on the tablets are of different sorts from the Rapanui texts used in the control corpus.

— In the many years of work on this project the catalogue of glyphs has changed, and has changed substantially. There have been corresponding changes in the statistical characteristics of the glyphs, and therefore in their probable phonetic values. It is worth stressing that there can be no guarantee that the catalogue of glyphs published in the present article is free from serious errors.

It is important to stress this point so that the reader does not think that the hypotheses about the interpretation of the Rapanui glyphs which have been presented in this article are yet another pretence

at a decipherment, of which there have been so many. Nonetheless, our work gives us hope that the complications can be resolved:

— It is unfortunate if a single glyph corresponds statistically to many syllables. However, there are other glyphs whose statistical characteristics are highly distinctive. If it is possible to identify the phonetic value of even five glyphs, that constitutes an undoubted breakthrough in decipherment.

— Some glyphs do not correspond statistically to any syllable. But perhaps this feature makes it possible to identify determinatives and other 'functional' glyphs in the Rapanui writing?

— Some glyphs coincide statistically not with syllables, or not only with syllables, but also with words. And why not? It would be naïve to suppose that the writing system discovered on Easter Island is purely syllabic.

— Of course, if the writing system does contain a special 'functional' glyph which indicates the reduplication of a root, that affects the statistical characteristics of every glyph. Yet it is the statistics themselves which allow for the presence of such glyphs to be posited in the first place. The entirely plausible hypothesis regarding reduplicator glyphs arose because the sets of data from the Rapanui language and the Rapanui writing diverge sharply on the 'interval 1' and especially on the 'interval 2' criteria, whilst coinciding on all the other criteria.

— The considerable variation in the statistical characteristics of syllables depends on the genre, or, more specifically, on the particular words which occur with a high frequency in a given text. As mentioned several times already, we can attempt to overcome this problem by working with statistics based on the lexicon of the texts rather than on the texts themselves. In this situation, however, there is an insufficient range of data to produce some statistics, and it is no longer possible to determine in what number of texts the characteristics of the writing and the language coincide.

— Mistakes made in compiling the catalogue will indeed greatly reduce the possibility of identifying the glyphs. At the same time, anomalies in the statistical characteristics of glyphs can also indicate that a glyph has been identified incorrectly and that the catalogue needs to be corrected.

By way of an example, suppose that we were trying to decipher the writing system of Latin. A structural analysis of its glyphs might suggest the following graphical proportion: **O : Q = P : R**. We may identify 3 glyphs: **O**, **P** and the oblique line which is found in the 'ligatures' **Q** and **R**. How would our mistake be reflected in the statistics? On the criterion of 'combinability', the oblique line would have an exceptionally high frequency of occurrence in combination

with the glyphs **O** and **P**. But precisely this abnormally high incidence of co-occurrences, taken together with the absence of combinations of oblique lines with any other glyphs, would most probably lead us to the conclusion that we had made a mistake in identifying the oblique line as a glyph in its own right.

This invented example illustrates sufficiently clearly the problems which arise when compiling a catalogue of Rapanui glyphs. Is the glyph  (99) a glyph in its own right, or is it a combination of two glyphs, perhaps  (14) and  (95), or (14) and  (200)? In our previous version of the catalogue, glyph 99 was treated as a combination of two glyphs, 14 + 95. Statistical analysis showed that the index of combinability of these two glyphs was abnormally high, and therefore that a mistake had most probably been made when compiling that catalogue.

On the whole, in developing a statistical approach to decipherment, we have proceeded on the assumption that investigation into an unknown writing system ought to be geared not only towards a search for 'solutions', but also towards elucidating as many restrictions as possible. These restrictions make it possible to discard solutions which are statistically improbable and in particular to refute 'logographic' decipherments, which conflict with the whole range of statistical data. The results of the statistical analysis (together with the results of the structural analysis, which has not been discussed in the present article) make it possible to narrow down the interpretation of most glyphs to a few alternatives, and to give arguments for and against each of them. Amassing concrete facts of this sort seems to us to be the most important task in the careful work needed to establish a reliable basis for the decipherment of the Rapanui writing.

Our views on the decipherment of the Rapanui writing system were first published 10 years ago [Pozdniakov 1996]. One of the critical responses posted on the internet [Sproat 2003] posed the legitimate question: if there is such a remarkable statistical correspondence between the syllables of the Rapanui language and the glyphs in the writing, why can the texts still not be read?

The problem is not simply that the authors of the present article cannot propose a reading of the texts, but also that, judging from the works published in recent years, we are evidently the only people in the world who cannot.

That means that nowadays there is no possibility of a constructive discussion about the decipherment of the Rapanui writing system; research in this area is the preserve of lone individuals. In this sense, the field is in a much worse state now than it was fifty years ago, when intensive work was being conducted in a number of centres of learning around the world.

References

Barthel, T. S. *Grundlagen zur Entzifferung der Osterinselschrift.* Hamburg, 1958.

Fedorova, I. K. *Mify, predaniya i legendy ostrova Paskhi.* [Myths, Lore and Legends of Easter Island]. M., 1978.

_____ 'Issledovanie rapanuiskikh tekstov' [Research into the Rapanui Texts] // Knorozov, Yu. V. (ed.). *Zabytye sistemy pisma* [Lost Writing Systems], M., 1982.

_____ *Mify i legendy ostrova Paskhi.* [Myths and Legends of Easter Island]. L., 1988.

_____ *Ostrov Paskhi: ocherki kultury XVIII–XIX vv.* [Easter Island: Essays on the Culture of the Eighteenth and Nineteenth Centuries]. SPb., 1993.

_____ *"Govoryashchie doshchechki" s ostrova Paskhi.* [The "Talking Tablets" of Easter Island]. SPb., 2001.

Fischer, S. R. *Rongorongo, the Easter Island Script: History, Traditions, Texts. Oxford Studies in Anthropological Linguistics 14.* Oxford, 1997.

Guy, J. On the Lunar Calendar of Tablet Mamari. *Journal de la Société des Océanistes* 1991. Issue 2. Pp.135–49. See also: [http://rongorongo.org], the sole website in which material about Rapanui writing is collected and annotated to a professioonal standard.

Pozdniakov, K. 'Les bases du déchiffrement de l'écriture de l'île de Pâques'// *Journal de la Société des Océanistes* 1996. Issue 2. Pp. 289–303.

Sproat, R. 'Approximate String Matches in the *rongorongo* Corpus'. http://serrano.ai.uiuc.edu/rws/ror.

Translated by Sarah Turner