Comparison between Rongorongo and the syllable sequence of ancient chants from the Easter Island

Fumihiko Yamaguchi

Keio University
Department of Information and Computer Science
Hiyoshi, Kanagawa, 3-14-1, JAPAN
yamagu@nak.ics.keio.ac.jp

Abstract. Rongorongo is sequence of symbols engraved on the surface of wooden artifacts made in the Easter Island. Rongorongo is considered to be script, but has not yet been deciphered. As Rongorongo is assumed to be the record of chant, we syntactically compare Rongorongo with chants in the Easter Island. A simple exhaustive search method is applied to find correspondence between the order of symbols in Rongorongo and that of syllables in the chants. However, the correct correspondence is unknown. Therefore, the proposed method is adapted to a similar problem of matching kanji-hiragana mixed text with katakana text in order to characterize the method. After confirming that the method produces higher precision when there are more kind of frequent symbols. As a result, some correspondences are found, whereas no correspondence is found in most pairs of Rongorongo lines and verses of the chants.

1   Introduction

In the Easter Island, sequence of symbols engraved in wooden artifacts are remained which is called Rongorongo. Rongorongo is considered that may be scripts, however it has not yet been deciphered.

The first European visitors to the Easter Island is Roggeveen, who is a Dutch explorer, arrived on Easter Day 1722. But they made no mention of written characters. The second European is Don Felipe arrived in 1770. There is a record that Easter Island people signed on the Spanish document with local symbols. However the first mention of the wooden artifacts covered with symbols is by Eyraud in 1864. Bishop Jaussen, who had been sent to Tahiti in the 19th century, owned some Rongorongo artifacts. He met a man named Metoro who is a native person of the Easter Island. As Jaussen let Metoro read the artifacts, Metoro read it in a singing voice. However, Rongorongo remained undeciphered because Metoro did not know what he said[6]. Emory surmised that Rongorongo was invented after that Easter Island people looked the Spanish document and for recording their chants[5].

It is remarkable that Metoro's reading was like chanting, though it was not significant to the content of his reading. Because even if he was actually not

Fig. 1. Rongorongo engraved on Mamari, side A, line 2.

able to read, there is a possibility of mimicking an original appearance in which Rongorongo is read. If Rongorongo is a document which record chants, some relations might be seen between the sequence of Rongorongo and the sequence of phoneme of the chant. Jorge Hotus Salinas collected the old chants in the Easter Island, and recorded them in Latin character[10].

If Rongorongo symbols are characters, each symbol may have its pronunciation which is a phonetic sequence. If a sequence of Rongorongo symbol represents a part of a verse of a chant, it is thought that the symbols appear in the sequence in the same order as their pronunciation of the verse. In this paper, only syntactic relation is considered, without considering the meaning of the Rongorongo symbol nor meaning of the word in the verse. When each of Rongorongo and pronunciation of verse is expressed by distinct sequences, the problem is to find a relation between two sequences. In this paper, lines of Rongorongo are expressed in sequences of codes called Barthel code, and the chants are expressed in sequences of syllables. However, even if there is a set of pairs (a, b) of symbols in sequences A and B such that a appears in A in the same order of b appearing in B, it's not always means that A corresponds to B. For example, assume $A = xy$ and $B = 1234$, there are many correspondence such as $\{(x, 1), (y, 2)\}$, $\{(x, 2), (y, 4)\}$ and so on. Its easy to see that both sequence needs some length. And a symbol which uniquely appears in one sequence can be paired to any symbol in the other sequence. Therefore, we should consider symbols which appears twice or more in each sequence. In this paper, an algorithm is proposed which finds a correspondence whether two sequence have a set of pairs of frequent symbols which appears in the same order in each sequence. Our interest is to apply the algorithm to Rongorongo and the chants, but we cannot evaluate any result directly, because Rongorongo is undeciphered. Thus, we apply the algorithm to known language. In this paper, Japanese texts are employed to characterize our method. And then we will show the output of the algorithm applied to Rongorongo.

Some natural language processing methods are expected to have little dependency on language specific features. There are some NLP researches about correspondence between a lost language and phonetic sequences, while lost languages are gaining increasing interest in NLP community [8]. Snyder et al investigated computer aided decipherment of lost language[12]. They extract pairs of Ugarit word and its cognate in Hebrew which is known language close to Ugarit geographically and temporally. Knight et al. investigated HMM-base decipherment of character substitution cipher which relates an unknown script that represents a known spoken language [7]. Their method find character-to-sound mappings from non-parallel character and sound sequences. Sidorov et al. investigated

searching tool for Mayan character [11]. They pointed out the importance of character identification, original scripts are the image data but computer readable texts.

Since remained Rongorongo is small and sparse corpus, we employ exhausitive search rather than statistical methods.

2 Rongorongo and Chants

Each Rongorongo artifact is named after local words or the city of which the possessing museum addressed. Most artifacts are tablets and its side is distinguished by side A/B or Verso/Recto. Rongorongo is written in boustrophedon style. Moreover the glyphs are upside down in each other line.

Barthel classified each symbol of Rongorongo in its shape and encoded them by three digits [2, 3]. This is called Barthel code. Barthel coding classify Rongorongo symbols in more detail than Mètraux's classification. By using Barthel code, Rongorongo can be treated as a computer readable text information. Yamaguchi et al. developed the tool that supports to encode from Rongorongo image data to sequences of Barthel codes [13]. In this paper, the sequences of Barthel code are obtained using this tool. The image of the Rongorongo symbol originally downloaded from rongorongo.org[1]. In this paper, the four major artifacts are examined that are named Aruku Kurenga, Tahua, Mamari and Keiti respectively. All these names come from local words. These artifacts were once owned by Jaussen[6]. They includes 83 lines total. There are 441 kinds of symbol counted in Barthel code, each line includes 60.5 Barthel codes in average.

The Salinas' archive of Easter Island chants includes 89 titles. Most titles include 3 or 5 variant of verses. Some chants also have another part of chorus. There are totally 372 verses.

Rongorongo symbol seem to be in the natural shape such as grass, fish and birds. The example of Rongorongo line is shown in Figure 1. Mètraux classified the Rongorongo symbols into about 120 kinds. He also pointed out that there are too few kinds of symbols to think of Rongorongo as ideographic script and too much as phonographic[9].

Therefore, Rongorongo is assumed to be syllabic or logo-syllabic in this paper, and thus each verse of the chant is treated as a sequence of syllables. According to Bellwood's family tree for the Fijian and Polynesian languages, Tahitian and Easter Island are closely related[4], such that both has the common ancestor Proto-Central Polynesian language. Moreover, the modern spoken language of the Easter Island is influenced by Tahitian language. A syllable of Tahitian language consists of one vowel or a consonant and a following vowel. And 'n' and 'm' can be appeared solitary. So, in this paper, each of them is considered as a syllable by itself. Each verse of the chant is successfully divided into syllables in the same way. There are few exception that solitary 'k' in a word 'kraverita' and solitary 's' in a word 'Mas'. However, each of these exceptions occurred only once in a verse. Thus, as descrived below, the exception does not influence to our

```
corres(i_R, i_C, H) :=
 1: if  i_R ≥ length_R then return true endif;
 2: if  i_C ≥ length_C then return false endif;
 3: if the frequency of R[i_R] ≤ 1 then return corres(i_R + 1, i_C, H) endif;
 4: if the frequency of C[i_C] ≤ 1 then return corres(i_R, i_C + 1, H) endif;
 5: if there exists X such that (R[i_R], X) ∈ H then
 6:   if  X = C[i_C] then
 7:     if  Rest_R[i_R] > Rest_C[i_C] then return false endif;
 8:     if corres(i_R + 1, i_C + 1, H) then return true endif
 9:   endif;
10: else
11:   if  Rest_R[i_R] ≤ Rest_C[i_C] then
12:     if corres(i_R + 1, i_C + 1, H ∪ {(R[i_R], C[i_C])}) then return true endif
13:   endif;
14: endif;
15: return corres(i_R, i_C + 1, H).
```

Fig. 2. Psudo code for judging correspondence between a Rongorongo line and a chant

experiment. There are 119 kinds of syllables used in the chant and each verse includes 35.0 syllables in average.

3   Correspondence

When a whole line of Rongorongo (sequence of symbols) represents a part of a verse (sequence of syllables), there exists a set of pairs of symbol and syllable such that they occur in the respective sequences in the same order. This relation is examined by exhausitive search. The algorithm is shown in Figure 2. In this code, R and C are the sequence which are intended a line of Rongorongo and the sequence of syllables of a verse of a chant respectively. The arguments of function corres are the index i_R and i_C of the sequences and hypothesis H of correspondence. $R[i_R]$ represents the i_R-th symbol of R, $length_R$ is the length of R, $Rest_R[i_R]$ is the number of frequent symbols in R after i_R-th symbol and so is in C. H is a set of pairs of symbols from R and C. The $Rest$ are used to cut search time. The call of corres$(0, 0, \emptyset)$ returns false when there is not possible correspondence. Otherwise, it returns true.

When R contains n kinds of frequent symbols, and C contains m frequent syllables, the algorithm checks two cases for each occurrence of frequent symbols whether it corresponds to the current frequent syllable or not. The pseudo code runs in $O(m2^n)$ in time.

As Rongorongo is unknown to be truly syllabic, it is assumed that the pronunciation of a symbol consists of zero or more syllables. Even if Rongorongo is ideogram, one symbol might corresponds to a sequence of syllables in some length (that is the pronunciation of the symbol) and thus the order of symbols

and the order of syllables will be allinged. However, when all association is enumerated, combinatorially many association will be found. The pseudo-code is to judge whether there is possible correspondence or not. Therefore, the algorithm does not enumerate all these association. Instead, it returns true when one association is found.
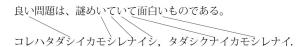
The correspondence of symbol and syllable is not exclusive: different symbols may correspond to the same syllable. Since unique symbols and its corresponding syllables do not constrain the check of the same order occurrence, only symbols and syllables which occur twice or more in the respective sequences are in concern. They are called frequent symbols and frequent syllables respectively. A frequent syllable corresponds to a symbol is thought as a part of the pronunciation of the symbol.

3.1 Experiments in known language

Our main purpose is to extract correspondence between lines of Rongorongo and verses of chants. However, because Rongorongo is undeciphered, we cannot evaluate the result directly. Therefore, we apply the proposed algorithm to known language in order to characterize the method. In this subsection, we describe about experiment on Japanese texts.

Japanese writing system have three distinct kinds of symbols, hiragana, katakana and kanji. Both hiragana and katakana are syllabaries and they have one-to-one correspondence, i.e. each syllable has two syllabic symbols. Kanji is logogram and each kanji character has some pronunciations which are able to be written in a sequence of one or more characters in katakana or hiragana. In usual modern Japanese writing, kanji and hiragana are used mixedly and katakana is used for loanwords etc.

For an experiment of a known language case, a set J of 20 kanji-hiragana mixed sentences is prepared each of which is extracted randomly from a Japanese math book. Each sentence $j \in J$ is in length of 21-41 characters. And a set K of 20 katakana sentence is prepared so that for each $j_i \in J$, $k_i \in K$ is the same pronunciation as j_i. Note that there is no intersection between the sets of characters in J and the set of characters in K.

良い問題は、謎めいていて面白いものである。

コレハタダシイカモシレナイシ，タダシクナイカモシレナイ．

Fig. 3. An example of misjudged case

The algorithm is applied to each of $20 \times 20 = 400$ pairs of all combination, to judge whether there is possible correspondence in which frequent symbols appears in the same order. The test set of 400 pairs contains 20 correct pairs

Table 1. Frequent hiragana and katakana correspondence in Figure 3

hiragana	katakana
い	シ
て	ナ

which are of kanji-hiragana mixed text and its pronunciation in katakana. As a result, for all of the 20 correct pairs, the algorithm returns true. And it returns false for 211 pairs of the 380 incorrect pairs. However, there are 169 patterns are misjudged. Those pairs are not correct pairs, nevertheless, there exists pairs of frequent symbols appearing in the same order in each sequence. An example of misjudged case is shown in Figure 3 and the correspondence between characters of the figure is shown in Table 1. In this example, there are only 2 frequent symbols in kanji-hiragana text, where katakana text contains 8 frequent symbols.

This is a typical characterization of many misjudged cases that the number of frequent symbols in j is inclinable less than that of correct case.

Table 2. The number of frequent symbols in kanji-hiragana sentence and the number of misjudged combination

n	# of misjudge	recall(%)	precision(%)	F-measure
2	74	100	11.9	0.213
3	67	100	13.0	0.230
4	56	100	15.2	0.263
5	23	100	30.3	0.465
6	16	100	38.5	0.556
7	9	100	52.6	0.690
8	6	100	62.5	0.769

Therefore, for another experiment, the sets J_n of 10 kanji-hiragana mixed sentences are prepared, for each $2 \leq n \leq 8$ where n indicates the number of frequent characters in kanji-hiragana mixed sentence. The algorithm is applied to all of 100 pairs for each n which contains 10 correct pairs. The result is shown in Table 2. For every n , the algorithm returns true for all correct pairs (thus recall is 100%). And the more n is, the less pairs are misjudged.

As a discussion, we found a rare case of misjudge for correct pairs. Because a kanji may have two or more pronunciation, there can be a case that a kanji is the frequent symbol in j and its different pronunciations are contained in the corresponding k. In this case, the proposed algorithm misjudged for this correct pair.

3.2   Experiment in Rongorongo and Chants

All of the 30,876 combination, that is 83 Rongorongo lines × 372 verses, are examined by the algorithm. As a result, 618 pairs are judged that there are possible correspondence. The rest 30,258 pairs are confirmed that they don't relate. As the proposed method is exhausitive search, when there is possible correspondence between frequent Rongorongo symbols and frequent syllables, they are expected to be found.

Table 3. Rongorongo and verse of chant which may have correspondence

Aruku Kurenga, Verso, line 1	
1. E Ira e, e Raparenga e #5	17. Ka u manu a roto #1
2. E Renga Mariki #3	18. Ko Tongariki #2
3. E mea tino mamahi rua e #4	19. Ko Tongariki #3
4. E nui te tamu #2	20. Kraverita #1
5. E nui te tamu #3	21. Kraverita #1 coro
6. I he a Hotu Matu'a e hura nei #4	22. Mai runga #3
7. Ka e'a te neru #2	23. Marìa Reina #1 coro
8. Ka haro au i vai a repa #1	24. Marìa Reina #2
9. Ka haro au i vai a repa #2	25. O mea o te hare #1
10. Ka memea(1) #1 coro	26. O mea o te hare #3
11. Ka memea(1) #3	27. Poio nuinui a Tuki #3
12. Ka memea(2) #1 coro	28. Renga mitimiti #3
13. Ka memea(2) #3	29. Renga varevare #4
14. Ka memea(2) #4	30. Tamaiti puai rahi #3
15. Ka tea mai te niho #4	31. Te pito #1
16. Ka tere te vaka(2) coro	32. Tuki horo pari #2
Mamari, side A, line 2	
1. E Ira e, e Raparenga e #5	4. Ka tea mai te niho #4
2. E nui te tamu #3	5. Ko Tongariki #2
3. Ka e'a te neru #2	
Mamari, side B, line 2	
1. Ka e'a te neru #2	2. Ka tea mai te niho #4

However, there are possibilities that the algorithm returns true even when the Rongorongo line is not the representation of the verse. As we shown by the Japanese case experiment, the number of frequent symbols will work as a good filter. Therefore, Rongorongo lines with many frequent symbols are remarkable. In each of 69 lines of Rongorongo out of 83 lines, there are 8 frequent symbols. There are only 39 pairs out of 618 pairs when the Rongorongo lines are selected to this 69 lines. These relatively significant pairs are shown in Table 3. In other words, Rongorongo lines of 579 pairs, in which the algorithm finds possible correspondence, contains 7 or less frequent symbols.
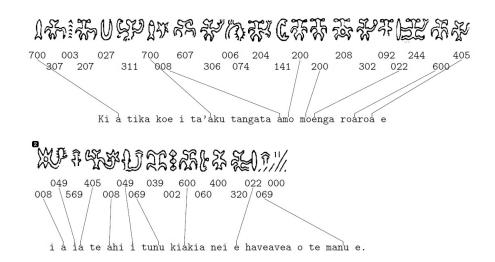
Fig. 4. Example of correspondence, Aruku Kurenga Verso 1 and "E Ira e, e Raparenga e" #5

Table 4. Symbol and syllable correspondence in Figure 4

Code	Syllable	Code	Syllable
700	a	008	a
200	mo	022	e
600	a	405	a
049	i	069	nu

In Table 3, list of possibly corresponding verse is shown for each Rongorongo line. Each verse is represented by its title and the variation. For example, "E Ira e, e Raparenga e #5" represents the 5th verse of chant titled "E Ira e, e Raparenga e". As is shown in this table, the result is not exclusive. There are 32 verses corresponding to Aruku Kurenga Verso line 1, and some verses such as "Ka tea mai te niho #4" correspond to 3 lines of Rongorongo.

The correspondences between each frequent symbol and each frequent syllable are extracted. An example of found correspondence is shown in Figure 4. In this figure,there are the image of Rongorongo (the first line of Aruku Kurenga, Verso), the sequence of Barthel codes and the 5th verse of chant titled "E Ira e, Raparenga e". Each straight line connecting a Barthel code and the verse

represents correspondence between symbols and syllables. The correspondence is shown in Table 4, which contains Barthel code, typical Rongorongo image of the code and corresponding syllable. As the judge algorithm returns when one correspondence is found, there may be another correspondence between each frequent symbols and syllables. The 2nd line of Mamari side-A which corresponds with the second most number of verses is shown above in the Figure 1.

4   Conclusion

It is examined whether the order of the codes in lines of Rongorongo corresponds to the order of the syllables in the chants in the Easter Island. As a result, 618 combination of correspondence are found. However, when Rongorongo contains few kinds of frequent symbols, the correspondence can be found accidentally. By selecting Rongorongo lines which contains many kinds of frequent symbols, relatively plausible 39 correspondences between lines of Rongorongo and verse of chant are listed. However, it might be a rashness to think the direct relation to the decipherment because the correspondence is not exclusive.

We cannot say that Rongorongo represents the verse even if the order of the symbols corresponds to the order of syllables. However, the Rongorongo line does not represent the verse when the orders do not meet, under the assumption that Rongorongo is syllabic or logo-syllabic and the whole line of Rongorongo represents the part of verse. The main contribution of this paper is that it is confirmed by a exhausitive search that the 30,258 combinations of Rongorongo line and verse do not relate, while there are few record or other corpus of old language in the Easter Island.

For the pairs of Rongorongo line and verses which is judged positively in this study, there are combinatorially many correspondence of symbol and syllable under the condition that the same symbol cannot corresponds to different syllable. As the future work, extracting plausible hypothesis about reading (or part of syllables) representing some symbols will be considered. There might be clues such as the distances between related symbols and syllables in the respect sequences. Once pronunciation of some Rongorongo symbols are assumed, the meaning will be inferred by comparing modern local language of the Easter Island.

Acknowledgment

References

1. Rongorongo or the hieroglyphs of the easter island tablets (accessed in Oct. 2003). retrieved from `http://www.rongorongo.org/`.

2. Barthel, T. S. (1958a). *Grundlagen zur Entzifferung der Osterinselschrift*, volume 36 of Abhandlungen aus dem Gebeit der Auslandskunde 64, Reihe B. Hamburg: Cram, de Gruyer & Co.

3. Barthel, T. S. (1958b). *Talking Boards of Easter Island*, Scientific American 198. pages 61–68.

4. Bellwood, P. (1987). *The Polynesians: Prehistory of an Island People.* Revised edition. London: Thames & Hudson.

5. Emory, K. P. (1972). *Easter Island's position in the prehistory of Polynesia.* Journal of Polynesian Society, vol. 81. pages 57-69.

6. Fischer, S. R. (1997). *RONGORONGO*, volume 14 of *Oxford Studies in Anthropological Linguistics.* Clarendon Press, Oxford.

7. Knight, K., Nair, A., Rathod, N. and Yamada, K. (2006) *Unsupervised analysis for decipherment problems*, In the Proceedings of the COLING/ACL, pp. 499–506, 2006

8. Knight, K. and Sproat, R. (2009). *Writing systems, transliteration and decipherment*, NAACL Tutorial, 2009

9. Mètraux, A. (1940). *Ethnology of Easter Island*, volume 160 of *Bernice P. Bishop Museum Bulletin.* Bishop Museum Press, Honolulu.

10. Salinas, J. H. Cantos antiguos de rapa nui. Archive of Museo Antropologico, Chile, Easter Island, (the year of production is unknown).

11. Sidorov, G., Pichardo-Lagunas, O., and Chanona-Hernandez, L. (2009). Search Interface to a Mayan Glyph Database based on Visual Characteristics, Lecture Notes in Computer Science, Vol. 5723, Springer-Verlag, pp. 222–29, 2009.

12. Snyder, B., Barzilay, R., and Knight, K. (2010). A statistical model for lost language decipherment. In *proc. ACL 2010.*

13. Yamaguchi, F., Nobesawa, S., and Sato, K. (2003). Rongorongo character listing tool. In *Proceedings of the 3rd IASTED International Conference on Visualization, Imaging, and Image Processing*, volume 1, pages 158–162, Benalmadena, Spain.