A Computational Evaluation of Syllabic Hypotheses for Rongorongo: Evidence from N-gram Analysis

Evgeniya Korovina
Institute of Linguistics, Russian Academy of Sciences
Moscow
jekorovina@gmail.com

Abstract

The study evaluates the hypothesis that the Rongorongo script of Easter Island functions as a syllabic substitution cipher where one symbol uniquely corresponds to one syllable. Using a genetic algorithm with a fitness function based on Rapa Nui n-gram statistics, we establish a performance baseline on controlled texts. Results show a strong correlation (0.75) between the algorithm's fitness score and decipherment accuracy, identifying a “noise threshold” at 2,500,000 points. We further demonstrate that cross-corpus genre variance significantly impacts recovery rates, with accuracy dropping by more than half when mismatched linguistic statistics are applied. Application to the CEIPP transliteration yields scores well below the noise threshold (1.0M–1.3M), suggesting a lack of a simple syllabic signal. However, testing the rongopy transliteration produces scores in the “gray zone” (up to 2.7M) and reveals stable mappings for high-frequency glyphs 200 and 006. The consistency of these results across independent inscriptions suggests that while a pure syllabic model is insufficient, specific structural simplifications of the script may capture latent linguistic patterns.

Keywords: Rongorongo, Rapa Nui, decipherment

1. Introduction

The decipherment of historical writing systems provides, on one hand, access to previously unknown aspects of the lives of their users (a subject of historical inquiry), and on the other, offers material for understanding how natural language can be represented in graphic form. One of the few such systems that remains undeciphered – at least in the sense of providing a verifiable and coherent reading of continuous text – is Rongorongo, the script of Easter Island, which fell out of use in the 1860s.

At the same time, there is little doubt among researchers that the language recorded in the Easter Island script is the ancestor of modern Rapa Nui or, at the very least, some Eastern Polynesian relative thereof (Horley 2021, Davletshin 2022). Furthermore, a relatively representative corpus of texts has come down to us – the total number of glyphs exceeds 12,000 (Horley 2021: 42), with nine inscriptions containing more than 500 glyphs each. These circumstances, it seems, create favorable starting conditions for the formal study of this writing system, particularly when compared to the Indus Valley (Harappan) script, the Epi-Olmec (Isthmian) script, or the Phaistos Disc.

A limiting factor for the direct comparison of Rongorongo material with Rapa Nui data is that modern Rapa Nui is a low-resource language with a relatively small corpus of published texts. Moreover, in the late 19th century, Rapa Nui underwent significant influence from Tahitian, which affected both its vocabulary and its grammar (Kieviet 2017: 15). Another obstacle consists of questions typical for undeciphered writing systems regarding the identification of glyphs: namely, determining which shapes should be considered variants of a single grapheme, and in which cases a shape represents a ligature of several glyphs.

Despite the enormous interest in attempts to decipher Rongorongo – a recent article by Wieczorek (2025) lists about 100 scholarly works published in the last 25 years alone, containing some 80 motivated suggestions for the reading of individual signs, varying widely in their degree of plausibility – the question of the script’s basic characteristics still seems to lack a satisfactory answer. For the most part, the signs that have been interpreted are described as logograms with nominal meanings, directly related to what the sign supposedly depicts. By contrast, signs that could potentially be interpreted as grammatical markers are almost entirely absent from such proposals. Moreover, since the late 1960s, the idea that Rongorongo exhibits a "telegraphic" style, largely devoid of grammar, has gained currency among researchers (cf. Кондратов 1968: 175; Pozdniakov & Pozdniakov 2007: 5). As a result, existing interpretations appear to be quite far removed from the structure of Polynesian languages, which makes their verification problematic and suggests the need for a different, more formal approach.

Unlike historical writing systems, whose successful decipherment has often relied on a combination of disparate factors (bilingual texts, iconographic context, and historical knowledge) and is therefore difficult to formalize, the principles underlying classical cryptography are far better formalized and algorithmically tractable. While various algorithms can efficiently solve simple substitution ciphers for alphabetic languages (see, e.g., Jakobsen 1995; Kang and Lee 2024), their application to syllabic systems presents a

Proceedings of LT4HALA 2026 @ LREC 2026, pages 177–183.

unique challenge. In such cases, the alphabet of potential referents is significantly larger than a standard phonemic alphabet yet remains more computationally tractable than a full logographic (word-based) system. In this paper, I treat Rongorongo as a large-alphabetic monoalphabetic substitution cipher, focusing on a one-to-one mapping between glyphs and syllables. It is important to clarify that this study does not account for homophonic substitution, where a single syllable could be represented by multiple distinct signs. While the script is widely considered to be logo-syllabic in nature, the 'purely syllabic' hypothesis frequently appears in literature. By modeling the script as a simple syllabic substitution system, I aim to provide a formal stress-test of this assumption. Testing the simble substitution hypothesis is a methodological necessity: if the script's structure is fundamentally more complex (e.g., homophonic or logo-syllabic), a simple substitution model should, by definition, fail to converge. Thus, the inability of the algorithm to find a stable solution serves as a formal indicator that the 'one-to-one' syllabic model is insufficient to describe the Rongorongo system. To do so, I will first analyze the feasibility of deciphering such a cipher using data from Rapa Nui – by artificially "encrypting" plain texts and testing whether the algorithm can recover the original – and then apply the same method to the transliterated Rongorongo data.

The paper proceeds as follows. Section 2 surveys prior formal and computational approaches to Rongorongo. Section 3 details the datasets, encompassing the Rongorongo corpus alongside texts in Rapa Nui and other Eastern Polynesian languages. Section 4 presents the methodology and core algorithm. Section 5 reports experimental results. Section 6 interprets their implications for Rongorongo decipherment and suggests avenues for future research. Section 7 concludes with a summary of key findings.

2. Previous Work

The first algorithmically oriented study of the Easter Island script was apparently undertaken in the 1960s by a research group led by Yuri Knorozov. The most comprehensive account of this work is presented in Kondratov (1968), where the author concludes, based on an analysis of the appearance of new glyphs, that Rongorongo falls formally between a logosyllabic and a purely logographic writing system. The study also identified glyphs that are evenly distributed throughout the texts, which were interpreted as potential grammatical markers. However, the technological limitations of the available computing technology at the time prevented further development of this line of research.

The next study of this kind was undertaken only in 2007, when Pozdniakov and Pozdniakov conducted an experiment based on the assumption that the number of distinct glyphs in the Easter Island script is substantially smaller than commonly thought, amounting to just over 50 signs. This hypothesis stemmed from the observation that published decipherment proposals – such as those by Fischer (1997) and Fedorova (2001), based on logosyllabic or logographic assumptions – remained far removed from the structure of Polynesian languages. Pozdniakov and Pozdniakov were able to demonstrate some statistical similarities, but their approach relied on heuristic comparisons of frequency distributions, which did not yield concrete readings or a reproducible mapping between signs and linguistic units, and the hypothesis that Rongorongo is predominantly syllabic in nature has not, overall, found broad acceptance among other researchers.

The next study of this kind was Harris (2010), who employed various computational methods – in particular, PCA and LSA – to uncover the structure of Rongorongo texts and the relationships between them. As a result, he obtained a formal clustering of the inscriptions that closely corresponded to the traditional classification.

Among the more successful attempts to identify formal similarities between Rapa Nui and the Easter Island script is the study by Horley, Davletshin, and Wieczorek (2018). The authors demonstrated linguistic parallels involving palindromes in Rongorongo, as well as various types of reduplication. However, this work was based on manual data analysis.

Finally, any brief overview must mention the work of Gregorio de Souza, who developed the rongopy repository and applied various algorithmic methods. In particular, he attempted to crack Rongorongo using a genetic algorithm scored by an LSTM language model. De Souza correctly identified that a direct frequency-based decipherment is hindered by the high variability of sign distributions across different subsets of the corpus. To mitigate this, he followed an approach like Aldarrab and May (2021), encoding sequences multiple times based on frequencies from random samples. Subsequently, in 2023, he sought to apply a Seq2Seq method for the same purpose. This work represents the most significant recent contribution to the computational analysis of the script.

As this overview demonstrates, while recent work has highlighted the challenges of frequency-based matching, there has been no systematic effort to formally measure the stability of these mappings under a controlled cryptanalytic framework. The present study addresses this gap.

3. Data

3.1 Rongorongo data

According to the counts of Horley (2021: 42), the total size of the Rongorongo corpus amounts to 12,289 glyphs (i.e., individually written elements). The majority of these – 10,879 glyphs – are found on twelve inscriptions, each exceeding 400 glyphs in length. For the purposes of this pilot study, only two of the longest and best-preserved inscriptions were selected: Tablet A (1,361 glyphs) and Tablet B (948 glyphs). These were chosen not only for their length and state of preservation, but also because, unlike, for example, I (Santiago Staff), they do not consist entirely of lists, which would preclude direct comparison with Polynesian language data. Facsimiles and drawings of these tablets are published in Barthel (1958), Fischer (1997), and Horley (2021).

However, for the purposes of the present analysis, it is not the glyph images themselves that are used, but rather transliterations of them. The primary catalogue for the study of the Easter Island script is the one developed by Barthel (1958), subsequently modified by CEIPP and made available online at http://kohaumotu.org/rongorongo_org/. While other catalogues exist (discussed below), their full transliterations have traditionally remained difficult to access. Recently, however, a machine-readable version based on Horley's structural analysis was made available via the rongopy repository (de Souza 2023).

In addition to the Barthel catalogue and the CEIPP version thereof – both containing around 500 glyphs – there exist catalogues that assume a significantly smaller number of signs. Notably, Pozdniakov & Pozdniakov (2007) identify 50 basic signs, while Horley (2021: 443) provides a list of 130 glyphs that "stand good chances of being independent". At the same time, Davletshin (2017) has shown that some graphically very similar signs are likely to be distinct based on collocation analysis, which may point to a larger number of signs in the writing system.

However, in addition to the usual debates about individual signs in the catalogue – typical for undeciphered writing systems – Rongorongo presents a specific problem: the presence of ligatured glyphs. The existence of such composite forms is evident from parallel passages, where two glyphs are written joined together in one text but separately in another. At the same time for many ligatures the nature of the script is such that such explicit demonstration of separateness is impossible in most cases. For example, see figure 1, glyph B206 could be interpreted as a combination of B200 and B006, as suggested by Horley and by Pozdniakov & Pozdniakov. However, in the case of B226, Horley following (Pozdniakov 2019) treats it not as a combination of B220 + B006, but as a distinct sign, whereas early he decomposes it even further, and in Pozdniakov & Pozdniakov (2007) he also breaks down B220 into B200 + B060.

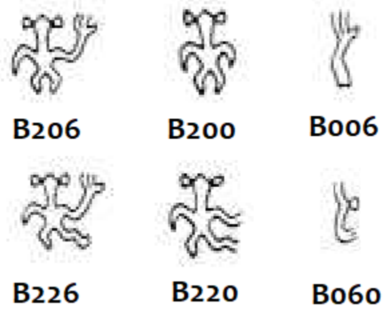

Figure 1: Appearance of the rongorongo glyphs mentioned above

Considering the issues outlined above, two types of transliteration were used for the purposes of this study: the original CEIPP encoding and an encoding from rongopy.

3.2 Rapa Nui data

The language of Easter Island is a low-resource language. The total size of the available corpus of texts in this language amounts to just over one million words (Kieviet 2017), approximately half of which consists of a recent translation of the New Testament (2018). Although the earliest transcriptions of Rapa Nui in the Latin alphabet date back to the late 18th century – when Rongorongo was still in use – the total volume of texts in the language remained around 130,000 words until the 1980s.

For the purposes of this study, five subcorpora of Rapa Nui texts were selected:

1. The text of Manuscript E (oral history, recorded by Rapa Nui speakers in the early 20th century).
2. Songs from the Campbell collection (1970).
3. The collection of texts published by Englert (1980), gathered in the 1930s (various genres: oral history, life stories, instructions).
4. Texts from Part 6 of the Rapa Nui textbook by Weber & Weber (1990) (Mai ki hāpī…).
5. The text of the Gospel of Matthew.

This selection, it is believed, provides a comprehensive representation of all the available Rapa Nui textual material. Each of the subcorpora exceeds 10,000 words in length, which is comparable to the volume of the Rongorongo corpus.

For the purposes of this experiment, all texts were converted to a unified format. All punctuation marks were removed, and orthography was standardized to a form that does not mark vowel length or the glottal stop, as these were used inconsistently in early texts.

	Syllables
Manuscript E	29,422
Campbell	23,274
Englert	67,146
Mai ki hāpī… 6	54,289
Matthew	70,394

Table 1: Length of the texts selected for analysis in syllables

4. Methodology

The research is structured according to the following logic: before applying the algorithm to the actual Rongorongo tablets, it is necessary to understand its behavior and limitations on controlled material. To this end, a series of experiments is conducted on Rapa Nui language texts, where the true correspondence between signs and syllables is known. Only after establishing which syllables and with what degree of reliability the algorithm is able to recover under "laboratory conditions" can the results obtained on the undeciphered script be interpreted.

An important characteristic of the experiments is the length of the analyzed fragments – 1000 signs. This parameter was chosen deliberately: it roughly corresponds to the length of the surviving Rongorongo tablets (the largest of them contain 800-1200 signs). Thus, the simulated conditions are as close as possible to the real ones.

In this work, Rongorongo is considered as a simple substitution cipher, where each graphic sign corresponds to one syllable of the Rapa Nui language. Under this assumption, the decryption task reduces to finding a mapping from the set of unique signs (graphemes) to the set of syllables. Since the total number of distinct signs on the tablets exceeds 50, the search space (the number of possible keys) is astronomical, making direct brute-force enumeration impossible.

To solve this problem, a genetic algorithm is employed – an optimization method that mimics natural selection. The algorithm operates on a population of candidate keys, gradually improving their quality through selection and mutation operators.

A key feature of the developed method is a specialized fitness function that, instead of directly calculating probabilities, uses a system of bonuses and penalties. This allows the algorithm to more quickly identify linguistically plausible solutions.

When evaluating a candidate, the algorithm decrypts the fragment using the tested key and scans the resulting syllable sequence, identifying bigrams and trigrams that are present in the reference frequency dictionary. To enhance search efficiency, a special class is distinguished – the "golden fund" of linguistic chains, which includes sequences with a frequency of occurrence exceeding 0.5% in the reference corpus. The first occurrence of such a chain in the text receives an increased bonus, encouraging the algorithm to seek keys that generate the most characteristic sequences for the language.

An additional constraining factor is a penalty for violating the rank correspondence between the frequency of signs in the encrypted text and the frequency of syllables in the language. The algorithm proceeds from the assumption that the most frequent signs should correspond to the most frequent syllables. The penalty is proportional to the rank discrepancy; however, its strictness is dynamically reduced towards the end of each training phase. This allows the algorithm in its final iterations to consider non-standard correspondences that might turn out to be correct, even if they do not fit into the rigid rank model.

The process of searching for the optimal key is divided into two sequential stages:

Phase 1 (Core). In the first stage, the algorithm works only with the top 20 most frequent signs occurring in the encrypted fragment. The goal is to create a reliable frequency framework for the decryption, matching the core graphemes with the core syllable inventory.

Phase 2 (Expansion). After a stable solution for the top of the frequency list is fixed, the remaining signs (up to the top 50) are introduced. The algorithm builds the overall picture, preserving the already found correspondences and selecting options for the rarer graphemes.

Each stage uses a population of 80-110 individuals, and evolution continues for 110 generations. The genetic search operators include elitist selection (preserving the best solutions) and swap mutations (randomly exchanging values between two positions in the key).

The experiment is structured as follows:

- Formation of reference frequency profiles. Based on several corpora of Rapa Nui texts (Campbell, Manuscript E, Englert, etc.), frequency dictionaries of unigrams, bigrams, and trigrams are computed.
- Generation of test fragments. A continuous fragment of 1000 syllables is randomly extracted from the source text. Then, a random mapping "syllable → numeric code" (simulating a cipher) is created, and the fragment is encoded. Thus, for each test, the true key is known, allowing for result verification.

- Algorithm execution and metric recording. For each fragment, the decryption algorithm is run multiple times. Three key indicators are recorded:
    - Score – the internal evaluation of key quality, used to guide the evolution.
    - Accuracy – the percentage of unique signs for which the algorithm proposes the correct syllabic correspondence.
    - Coverage – the proportion of the text (in signs, accounting for repetitions) that is correctly deciphered. This indicator is the most important for the subsequent interpretation of results on the actual tablets, as it reflects what portion of the message could potentially be considered readable.

This procedure allows not only for testing the algorithm's effectiveness but also for accumulating empirical data on which specific syllables and with what reliability are recovered given a fixed text length (1000 signs). Subsequently, this data serves as the basis for interpreting the results obtained when working with the actual Rongorongo tablets.

5. Results

The algorithmic performance is summarized in Table 2. When the model is tested using N-gram statistics derived from its own internal corpus, it correctly identifies an average of 26 syllables, resulting in a coverage of over 80% of the text. These figures demonstrate the robustness and viability of the proposed pilot algorithm for deciphering materials of this nature.

However, the performance drops significantly in cross-corpus scenarios. When using statistics from a different corpus, the average number of correctly identified syllables falls to 10, covering only 37% of the text – a result clearly insufficient for a successful “decipherment.” This drop highlights the high degree of genre-specific variance within Rapa Nui texts. The lowest cross-genre performance was observed when analyzing fragments from Manuscript E using statistics from the Gospel of Matthew, yielding an average of only 4.6 correct syllables.

When utilizing an aggregate statistical model based on all five sub-corpora, the algorithm correctly identifies an average of 17 syllables, achieving a text coverage of 62%.

Data regarding the most frequently correctly identified syllables for select corpus pairs are presented in Table 3. When utilizing statistics from the same text as the source fragment, many syllables are recovered in over 90% of cases. However, even in the most challenging scenario – analyzing Manuscript E using statistics from the Gospel of Matthew – the syllables “a” and “i” are correctly identified in 67.2% and 52% of cases, respectively.

The fitness score serves as a relatively accurate predictor of the success of the syllabic decipherment. Its correlation with the percentage of correctly identified syllables is 0.75 (as shown in Figure 2). Specifically, when the score exceeds 4,000,000, a high accuracy rate in syllable identification can be expected. Conversely, if the score falls below 2,500,000, the resulting solution is likely nothing more than statistical noise. The range between these two values represents a “gray zone,” where both successful and unsuccessful decipherment attempts are possible.

When testing the Rongorongo transliterations, specifically the CEIPP version, the scores obtained for both analyzed tablets ranged between 1,000,000 and 1,300,000. These values are extremely low, falling well below the previously established noise threshold of 2,500,000. This indicates that any syllabic “solution” generated in this context is, likely, merely a byproduct of the algorithm’s inherent tendency to suggest a mapping, regardless of the data's linguistic validity. Consequently, even if a few signs were coincidentally identified correctly in each run, the lack of statistical significance makes it impossible to advance any verifiable claims.

Furthermore, it is noteworthy that each of the five sub-corpora yielded approximately the same low score ranges when applied to the Rongorongo inscriptions. This consistency across diverse linguistic datasets (ranging from the Gospel of Matthew to folklore manuscripts) suggests that the failure to reach the convergence threshold is not a result of genre mismatch or lexical bias in the reference data. Instead, it points to a fundamental structural discrepancy between the CEIPP transliteration and the simple substitution syllabic model.

When testing the rongopy transliteration, the results were significantly higher, ranging between 2,000,000 and 2,700,000, effectively placing them within the previously defined “gray zone.” Notably, the mappings assigned by the algorithm for certain glyphs showed considerably more stability.

For Tablet A, glyph 200 was consistently interpreted as either the syllable “a” or “i”, while glyph 006 was identified in most cases as “i” or “te”. Other signs continued to exhibit essentially random interpretations. A similar pattern emerged for Tablet B: glyph 200 was again mapped to “a” or “i” (with “a” being assigned in 80% of runs), while glyph 006 received an “i” interpretation in 53% of cases (with alternative hypotheses being a, e, te, u).

	Campbell	Manuscript E	Englert	Mai ki hāpī	Matthew	Joined
Campbell	25.5 (80.5%)	6.6 (29.8%)	6.9 (30.6%)	7.4 (34.4%)	6.3 (31.2%)	11.9 (49.9%)
Manuscript E	4.8 (22.7%)	22.6 (70.6%)	15.3 (53.7%)	8.9 (35.8%)	4.6 (23.3%)	14.7 (53.8%)
Englert	5.0 (21.0%)	14.5 (51.1%)	23.0 (71.9%)	10.1 (37.5%)	6.2 (27.3%)	15.0 (52.2%)
Mai ki hāpī	7.5 (34.6%)	11.7 (47.6%)	15.8 (59.4%)	27.6 (85.4%)	13.6 (50.5%)	20.1 (69.4%)
Matthew	7.5 (36.3%)	9.0 (41.4%)	11.1 (45.1%)	20.1 (69.7%)	31.7 (93.8%)	27.2 (86.4%)

Table 2: Average number of correctly identified syllables by text (rows) / corpus (columns) pair, and (in brackets) the percentage of the text read correctly (coverage).

	Campbell	Manuscript E	Englert	Mai ki hāpī	Matthew
a	90,4	87,2	83,2	94,4	99,2
u	91,2	79,2	75,2	92,8	96,8
i	91,2	78,4	82,4	94,4	99,2
te	92,8	76	86,4	96	99,2
o	94,4	76	80,8	94,4	99,2
ma	88,8	76	76	94,4	95,2
na	86,4	75,2	67,2	85,6	100
ka	71,2	75,2	76	85,6	99,2
he	36	74,4	80	81,6	100
e	92	73,6	74,4	96	100

Table 3: The most “cracked” syllables when a fragment is deciphered based on the statistics of its own text.

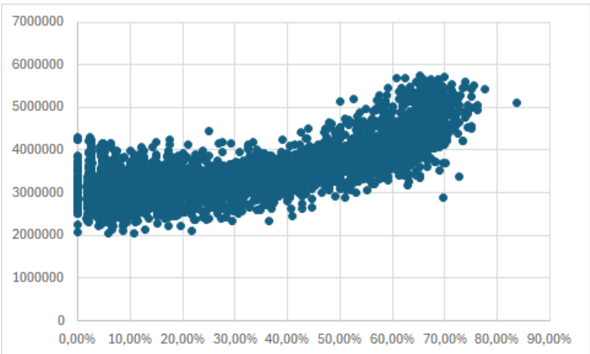

Figure 2: Correlation between the score value and the percentage of syllables deciphered

Such consistency across two independent texts suggests that these results might not be entirely accidental. However, given that the scores remain at the lower boundary of the “gray zone,” it stays possible that these patterns are merely a more structured form of statistical noise rather than a definitive decipherment.

6. Conclusion & Future Work

In this study, the syllabic hypothesis for the Rongorongo script was evaluated using a stochastic optimization algorithm. A fitness score metric was developed, which demonstrated a strong correlation (0.75) with actual decipherment accuracy in Rapa Nui control tests, establishing a clear threshold between linguistic signal and statistical noise.

While the results do not warrant a complete rejection of the syllabic hypothesis, the frequency characteristics of Rongorongo differ substantially from Latin-script Rapa Nui texts. This discrepancy suggests several possibilities: the texts may belong to an undocumented or lost genre; existing transliteration systems may be and likely is suboptimal; or Rongorongo may be a non-syllabic system incorporating numerous logograms or multi-glyph encoding. At the same time, the consistent mapping of glyphs 200 and 006 to the same values across two independent texts suggests that high-frequency signs might be identifiable through this method, though such interpretations remain speculative at this stage.

Future research will focus on testing the algorithm against logographic and logo-syllabic hypotheses, as well as models where a single Rapa Nui syllable corresponds to multiple signs. The research results should also be tested across different languages, for example, by comparing the outcomes of decryption attempts on certain Rapa Nui texts with texts in Maori or Tahitian to further refine the cross-linguistic reliability of the framework. Additionally, this algorithm should be tested on fragments of varying lengths.

7. Acknowlegement

This research was supported by project FMNE-2025-0003 conducted at the Institute of Linguistics, Russian Academy of Sciences.

8. Bibliographical References

Aldarrab and May (2021) “Can Sequence-to-Sequence Models Crack Substitution Ciphers?” In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 7226–7235, Online. Association for Computational Linguistics.

Barthel, T. (1958). Grundlagen zur Entzifferung der Osterinselschrift. Walter de Gruyter.

Davletshin, A. (2017). “Allographs, graphic variants and iconic formulae in the Kohau Rongorongo script of Rapa Nui (Easter Island).” The Journal of the Polynesian Society 126 (1), 61-92.

Davletshin, A. (2022). "The script of Rapa Nui (Easter Island) is logosyllabic, the language is East Polynesian: Evidence from cross-readings." The Journal of the Polynesian Society, 131(2), 185-220.

de Souza (2023) *rongopy*: Ideas for the decipherment of rongorongo using machine learning in Python (https://jgregoriods.github.io/rongopy/)

Fischer, S. (1997). Rongorongo: the Easter Island script: history, traditions, texts. Oxford University Press.

Harris, M. (2010). Corpus Linguistics as a Method for the Decipherment of Rongorongo. Unpublished MA thesis in Applied Linguistics, Birkbeck University, London.

Horley P. (2021) Rongorongo: Inscribed Objects from Rapa Nui. Viña del Mar: Rapanui Press.

Jakobsen, T. (1995). "A fast method for cryptanalysis of substitution ciphers". Cryptologia, 19(3), 265-274.

Kang, H., and Lee, W. (2024). "A Robust Decryption Technique Using Letter Frequency Analysis for Short Monoalphabetic Substitution Ciphers" Journal of Computing Science and Engineering, Vol. 18, No. 3, pp. 144-151

Kieviet, P. (2017). A grammar of Rapa Nui. Language Science Press.

Pozdniakov, I. & Pozdniakov, K. (2007) “Rapanui Writing and the Rapanui Language: Preliminary Results of a Statistical Analysis”, Forum for Anthropology and Culture, 3: 3–36.

Pozdniakov K. (2019). Correlation of Graphical ’Distinctive Features’ in Rongorongo as an Additional Resource for the Construction of the Sign Catalog. “Easter Island and the Pacific”, 9th International Conference on Easter Island and the Pacific (EIPC 2015), Ethnological Museum Dahlem, Berlin, Jun 2015, Berlin, Germany. pp.393-406.

Wieczorek, R. (2025). "Aspects of rongorongo decipherment." Waka Kuaka: The Journal of the Polynesian Society, 134(3), 343-392.

Кондратов А.М. (1968) Статистические методы дешифровки некоторых письмен древнего востока и средиземноморья [Statistical Methods for Deciphering Certain Scripts of the Ancient East and the Mediterranean]. PhD thesis.

Федорова И.К. (2001) «Говорящие дощечки» с острова Пасхи. Дешифровка. Чтение. Перевод. ["Talking Tablets" from Easter Island. Decipherment. Reading. Translation.] Spb: MAE RAS.