

未解読言語に挑戦する人工知能

山口 文彦¹

概要: イースター島で木片に刻まれた記号の列であるロンゴロンゴを中心に、失われた言語を計算機を用いて解析しようとする試みについて概観する。

Artificial Intelligence applied to Undeciphered Languages

FUMIHIKO YAMAGUCHI¹

Abstract: Experiments upon computer aided lost languages analysis are subject to this presentation, especially that of Rongorongo of the Easter Island.

1. はじめに

人が使っていた言葉や文字が、一度は忘れ去られ、年月が経ってから、その文字を記した遺物が発見されることがある。そうした文字の中には未解読のまま残されているものもある。矢島 [1] は、古代文字 (言語) の解読においては次のような場合がありうると述べている。

- (1) 文字と使用言語が、ともに不明な場合
- (2) 文字は不明だが、使用言語は推定できる場合
- (3) 文字は判明しているが、使用言語が不明の場合

このうち、使用されている言語の推定の確からしさの度合いはさまざまであって、1と2の境界は曖昧である。3は少し不思議に思えるかも知れないが、文字は言語とは独立に他の文化圏に伝わることもある。近隣の文化圏などにおいて文字が忘れられずに後世に伝わった場合には、文字は読めるが内容が分からないということが起こる。

これまで、考古学者らの努力によって、未知の文字の解明が進められてきた。解読の手掛りとしては、ロゼッタストーンと呼ばれる石碑が有名である。この石碑には、古代エジプトの神聖文字と民衆文字およびギリシア文字の3つの文字と言語で同じ内容が書かれており、このうちギリシア語が既知であったことが解読の手掛りとなった。このように、複数の言語で同じ内容を記した文書の組を、対訳

コーパスと呼ぶ。しかし未解読言語を含む対訳コーパスが見つかることは稀である。また、使用言語が推定できる場合には、言語の特徴が解読の手掛りとなることもある。楔形文字であるウガリットの解読においては、年代や地理的な近さからヘブライ語と近縁の言語が使われていたと推定され、ヘブライ語の語尾の変化に対応して記号列の末尾が変化していることが解読の手掛りとなった。

計算機の発達と普及にともなって、歴史研究にも計算機が導入されるようになった。例えば、膨大な量の遺物を整理する必要があり、遺物の写真とともにアノテーションをデータベース化しようとする研究などが進んでいる。

一方、解読そのものを計算機を用いて行おうとする研究もある。しかし未解読の言語を対象とする研究では、正解が分からないために、機械学習における訓練データを用意することができず、また再現率や適合率といった正解と比較した定量的な評価は得られないという問題がある。そこで、対象の問題と本質的に同じ問題を既知の言語を使って設定し、その正解が分かっている問題を使って、手法そのものを評価するといった研究方法が使われることがある。

Knight ら [2] は、同じ言語だが内容の異なる文字のコーパスと音声のコーパスから、期待値最大化法を用いて文字の音価を見つけている。この研究は使用言語が推定できて文字が未知であるような未解読言語に対して有効であるとしながらも、特定の未解読言語には適用せず、既知の言語を用いた実験を報告している。Snyder ら [4] は、ウガリッ

¹ 長崎県立大学 / University of Nagasaki,
Manabino 1-1-1, Nagayo, Nishisonogi, Nagasaki, 851-2195,
JAPAN / yamagu@sun.ac.jp

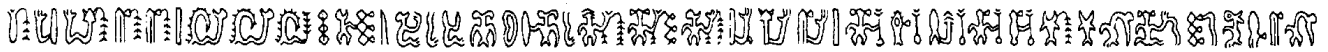


図 1: ロンゴロンゴの例, Aruku Kurenga の Verso 面 4 行目 [3]

トとヘブライ語のコーパスから、ノンパラメトリック・ベイズモデルを用いて同根語の対を見つけている。

イースター島で木片に刻まれた記号列であるロンゴロンゴ [5] も未解読文字の一つである。筆者らは、ロンゴロンゴを対象に計算機を用いて解読に資する情報を抽出しようと試みている。

2. ロンゴロンゴ

イースター島に到達した最初の西欧人は、オランダ人探検家 Roggeveen で、1722 年の復活祭の日のことである。しかし、記号が彫りつけられた木製品の存在が報告されたのは、1864 年の宣教師 Eyraud によるものが最初である。この記号は、ロンゴロンゴと呼ばれ、文字であるかも知れないと言われているが、未解読である。ロンゴロンゴについて、Emory は、島民がスペイン人の書類を見たことから刺激を受けて作り出したもので、歌を明瞭な形で残すために工夫・創造されたものであると推察している [6]。もしそうだとすれば、ロンゴロンゴの成立は 18~19 世紀ということになる。

図 1 にロンゴロンゴの例を示す。ロンゴロンゴの記号のそれぞれは、植物、魚、鳥などの自然物を象形しているように見える。Mètraux はロンゴロンゴの記号が約 120 種類あるとし、文字の種類が表意文字にしては少な過ぎ、表音文字にしては多過ぎると指摘している [7]。

Barthel はロンゴロンゴの記号を、その形状によって分類し、3 桁の数字からなる符号を付けている [8]。また、Barthel は Mamari と呼ばれる板の一部に月齢カレンダーと思われる記号の並びがあることを指摘している。

3. 計算機を用いたロンゴロンゴへの取り組み

筆者らはロンゴロンゴが歌を記憶する補助として作り出された音節文字であるとする説に基いて、イースター島の歌の音節とロンゴロンゴの各行の Barthel 符号の間で出現順序を比較した [9]。結果として、いくつかの行と歌の間に対応する可能性があることが分かった。

Barthel 符号のように、未解読文字によっては考古学者の努力によって、記号にコードが付されている。しかし、これは文字の異なりによって分類しているとは限らない。そこで、筆者は記号の画像を文字クラスに分類する作業を支援ないし自動化しようと試みている [10]。もちろん同じ文字であれば、画像としても似ていると考えられるが、どの程度似ていれば同じ文字であるかの基準がないことが問題となる。未解読の文字に対しては結果の評価ができない

ので、既知の言語を用いた実験によって手法そのものを評価する必要がある。そこで評価手法を提案した上で日本語の手書き文字に対して適用してみた [11] が、現在のところ良い結果は得られていない。

4. まとめ

残念ながら現在までのところ、未解読言語を対象に人工知能の技術を用いて得られた結果が、考古学上の新発見につながった事例を筆者は知らない。しかし、人間の考古学者による発見の手順を計算機で模倣することは行われている。そして、そのような人工知能を作ることによって、同種の発見がなされることを期待しても良いように思う。

情報機器を使うことの利点はネットワークを介した情報の共有や、膨大な計算を遂行できることだけでなく、明確に定義された計算手順を対象を変えても同じように遂行できることにもある。同種の問題であることを慎重に検討する必要はあるが、既知言語で有効性が示された手法を未解読の言語に適用することで、解読に資する情報が得られるのではないかと期待している。

参考文献

- [1] 矢島 文夫, “解読 古代文字”, ちくま文芸文庫, 1999
- [2] Kevin Knight, Kenji Yamada, “A computational approach to deciphering unknown scripts”, ACL Workshop on Unsupervised Learning in Natural Language Processing, pp. 37–44, 1999
- [3] Kohaumotu.org, <http://kohaumotu.org/>, last accessed: 2016/10/26
- [4] Benjamin Snyder, Regina Barzilay, Kevin Knight, “A Statistical Model for Lost Language Decipherment”, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1048–1057, 2010
- [5] Steven Roger Fischer, “RONGORONGO — the Easter Island script —”, Clarendon Press, Oxford, Oxford Studies in Anthropological Linguistics, vol. 14, 1997
- [6] K.P. Emory, “Easter Island’s position in the prehistory of Polynesia”, Journal of Polynesian Society, vol. 81, pp. 57–69, 1972
- [7] Alfred Mètraux, “Ethnology of Easter Island”, Bishop Museum Press, Bernice P. Bishop Museum Bulletin 160, Honolulu, 1940
- [8] Thomas S. Barthel, “Talking Boards of Easter Island”, Scientific American, Vol. 198, pp. 61–68, 1958
- [9] 山口 文彦, 中川 正樹, “ロンゴロンゴとイースター島の古謡のシラブル列との統語論的対応づけ”, 情報考古学会誌, Vol. 18, No. 1・2(合併号), pp. 25–37, 2012
- [10] 山口 文彦, “Zipf 則を指標とするロンゴロンゴ記号の分類”, 日本情報考古学会第 35 回大会講演論文集, pp. 80–85, 2015
- [11] 山口 文彦, “文字画像分類手法の評価指標の提案”, 日本情報考古学会第 36 回大会講演論文集, pp. 60–65, 2016