

日本語 AI書庫版

AIが古代文字を解読する

AIがよみがえらせた古代の記録

金京鎮弁護士

kimkj.com • AI書庫 • 2026

AIが古代文字を解読する

AIがよみがえらせた古代の記録

金京鎮弁護士

火に焼けて炭となった巻物、泥に埋もれた木片、砕け散った粘土板は、長い間読むことのできない物であった。本書はその沈黙を破る12の事件を追う。火山灰に埋もれたヘルクラネウムの巻物を広げずに読み解いた仮想展開、X線で蘇らせたエン・ゲディのレビ記、筆跡から書記の人数を見分けたイザヤ巻物の鑑定、数万もの破片をつなぎ合わせるデジタル接着剤、ギルガメシュ叙事詩の再建、アッカド語の帳簿の翻訳、甲骨文字の3次元仮想編綴、新羅木簡の草書体の判読、線文字Aとインダス文字、ロンゴロンゴを扱う計算言語学、欠損したギリシャ碑文を埋める言語モデル、原本が失われても残るマルチスペクトル・アーカイブへと続く。各章は、事件ファイル、科学捜査隊の追跡、明らかになった真実の順で構成した。背景知識がなくても読めるよう用語をわかりやすく解説し、サブチャプターごとに確認された参考資料を付した。付録には、読者が自宅で直接開いて見ることができる公開資料の窓口をまとめた。本書は、NotebookLMのノート「AI基盤古代言語および古文獻解読研究コレクション（2024-2026）」に集めた資料251件をもとに執筆した。

公開研究資料保管庫

本書は人工知能によって整理・制作された。全世界の主要言語による資料を収集・総合したものである。NotebookLMは、資料がどの言語で作成されているかに関わらず、質問の言語を制限しない。読者は、本書の準備に用いた同じ研究資料保管庫に対し、自身の言語で質問することができる。一部の資料が韓国語、英語、その他の言語であっても同様である。NotebookLMで設定を開き、出力言語を希望の言語に指定すれば、回答や生成結果がその言語で出力される。

本書の準備に用いた公開NotebookLM保管庫（AI基盤古代言語および古文獻解読研究コレクション 2024-2026、資料251件）：<https://notebook.google.com/notebook/cf32a6a1-66cc-473d-99d8-ed5a51596eb3>

本書の韓国語原版はkimkj.comのAI書齋にある：<https://kimkj.com/ai-library/?mod=document&uid=12641>

目次

各章のタイトルを押すと該当の本文へ移動する。最初から読むには第1章を選択する。

- 第1章. 火山灰が呑み込んだ哲学者の書斎：ヘルクラネウムの炭化巻物・4
- 第2章. 焼け焦げた聖櫃の中で花開いた御言葉：エン・ゲディの封印されたレビ記・12
- 第3章. イザヤ巻物に隠された二人の影：書記鑑定・19
- 第4章. 数万片の粘土・羊皮紙パズルをつなぎ合わせるデジタル接着剤・26
- 第5章. 粘土板の折れた断面を合わせる：ギルガメシュ叙事詩の計算機的復元・34
- 第6章. 古代メソポタミアの領収書と商人たちのささやき：AI翻訳・41
- 第7章. 亀の腹甲に眠っていた帝国の未来：甲骨文字と3D仮想綴合・49
- 第8章. 泥の穴から生還した筆跡：新羅木簡とくずし字判読人工知能・57
- 第9章. エーゲ海の失われた謎：線文字Aと幾何学的格子・65
- 第10章. 太陽の周期と動物印章の暗号をハックする：インダス文字とロンゴロンゴの音・73
- 第11章. 人工知能、破損したギリシャ語碑文を読み直す：ファウンデーションモデルとIthaca・81
- 第12章. 永遠に燃えない仮想図書館：超高解像度マルチスペクトルアーカイブと未来・89
- 付録. 私だけのAI考古学旅行記：一般向けグローバル探索ポータル・97

第1章. 火山灰が呑み込んだ哲学者の書斎： ヘルクラネウムの炭化巻物

1.1. 事件ファイル

火山が書斎を覆った日

西暦79年にイタリア南部のヴェスヴィオ火山が噴火した。火山は熱いガスと火山灰を一気に噴き出した。山の麓の海岸にはヘルクラネウムという都市があった。ローマの富裕層が夏を過ごす保養都市であった。この都市も一日で地中に埋もれた[1]。

都市の外には大きな邸宅が一軒建っていた。今日の学者たちはここをパピルス荘と呼ぶ。邸宅の主はルキウス・カルプルニウス・ピソ・カエソニヌスと見られる。彼はユリウス・カエサルの義父であった。この屋敷にはギリシャの哲学者フィロデモスが入り出していた。フィロデモスはエピクロス派の学者であった。邸宅の一角には巻物を集めた書斎があった。彼の著作が際立って多く出土したため、彼と関係の深い書庫と見なされている[1]。

パピルスは古代地中海世界の紙である。エジプトのナイル川のほとりに生える草で作る。茎の芯を薄く削ぎ、横にずらりと並べる。その上に再び縦方向に一層を重ねる。水を吹きかけて圧力をかけると、草の粘液が接着剤の役割を果たす。乾燥させると一枚の紙になる。そのためパピルスの表面には縦横に交差した繊維の筋が残る。この筋は後に重要な手がかりとなる[7]。

その日、書斎に熱いガスが押し寄せた。巻物は炎を伴わずにそのまま焼かれた。紙の役割を果たしていたパピルスは真っ黒な炭へと変わった。このように燃えて炭になることを炭化と呼ぶ。巻物は縮み、表面が膨れ上がった[2]。その上には火山堆積物が20メートル以上も積もった。火山堆積物とは、火山灰と石の粉が積もって固まった層である[1]。

地中深くに埋もれたため、空気が触れることはなかった。空気が触れなければ植物繊維は腐りにくい。災厄が遺物を守る蓋の役目を果たした。巻物はそうして2,000年を耐え抜いた[2]。

地下トンネルから出てきた黒い棒

1738年にナポリ王室がこの一帯の発掘を始めた。1750年には井戸を掘っていた作業員が色とりどりのモザイクの床に突き当たった。邸宅の場所はこうして明らかになった。発掘の責任者はカール・ヴェーバーが務めた。彼はスイス出身の軍事技術者であった[1]。

ヴェーバーは固まった火山堆積物にまっすぐ下に向かって穴を掘った。このような穴を垂直坑道と呼ぶ。そして狭いトンネルを迷路のように掘り進めた。トンネル内は暗く、空気も悪かった。壁が崩落する危険も常にあった。そのような条件の下でも、ヴェーバーは邸宅の平面図を綿密に描いた。その図面のおかげで、どの部屋から何が出土したのかが分かる[1]。

1752年から作業員たちは黒い棒のような物を取り出し始めた。最初はそれが木炭の塊だと思われていた。しかし、実際にはパピルスの巻物であった。この邸宅から出土した巻物は約1,100本と知られている。古代世界の書庫が丸ごと残された唯一の事例として知られている[1]。

この書庫にはフィロデモスの著作が際立って多かった。完成本や写本、草稿と一緒に置かれていた。学者たちは彼がこの屋敷で執筆し、整理していたと見ている[1]。苦労して開いた巻物には、哲学や修辞学、倫理に関する文章が記されていた。修辞学とは、言葉や文章で他者を説得する方法を扱う学問である。広げられなかった巻物に何が記されているのかは、誰にも分からなかった[2]。

無理に広げて損壊した巻物

学者たちはその中の文章を読みたいと願った。問題は、巻物が触れれば崩れてしまうことだった[2]。1753年にアントニオ・ピアッジョ神父が一台の機械を作った。糸と重りを使って巻物を少しずつ引っ張りながら広げる装置であった[3]。この機械は何本かを救い出したが、内側の層を引き裂いてしまうこともあった[2]。

19世紀にはイギリスの化学者ハンフリー・デーヴィーが乗り出した。彼は塩素ガスやヨウ素蒸気、酸を巻物に用いた。薬品によって層を剥がそうとする手法であった。結果は思わしくなかった。薬品はパピルス脆くし、崩壊させてしまった[3]。

物理的に広げようとする試みは、大半が遺物を損なう結果となった。文字の書かれた破片が床に散らばって失われることもあった。結局、学界は開封の試みを中止した。現在も開かれていない巻物が数百本残されている[2]。これらの遺物はナポリ、パリ、オックスフォードの所蔵機関に分かれて保管されている[4]。

X線を阻む炭素の壁

科学者たちは開かずに読む方法を模索した。まず思い浮かんだ手段はコンピュータ断層撮影であった。一般にCTと呼ばれる技術である。CTは物体を切断することなく内部の密度差を測定する装置だ。密度の高い物質はX線をより多く遮る。そのため断面画像には明るく写る[5]。

金属成分が含まれるインクは、CT画像において白く輝く。中世ヨーロッパでは鉄成分を含んだインクが広く使われた。イスラエルのエン・ゲディで出土した炭化巻物も同様であった。この遺物は古代イスラエルのヘブライ語写本である。インクに金属が含まれていたと見られ、画像に痕跡が残った。そのため2016年に開かれることなく判読された[5]。

ヘルクラネウムの写字生たちの事情は異なっていた。写字生とは文字を手作業で書き写していた人々である。彼らは木を燃やした煤に水と樹脂を混ぜて墨を作った。墨の主成分は炭素である。しかし、火山の熱で焼かれたパピルスもまた炭素となっていた。紙と文字が同じ物質になってしまったのである[6]。

密度が同じであれば、X線は両方を同じように透過する。断面画像には境界線が現れない。学者たちはこの壁を炭素インク問題と呼んだ。2015年にイタリアとフランスの研究チームが別の手法を試みた。位相コントラスト撮影という技法である。X線が物質の境界を通過する際、進行方向がわずかに屈折する性質を利用する。密度が同じであっても境界面では信号が変化する。研究チームはこの手法により、丸まった巻物の内部からいくつかの文字の輪郭を捉えることに成功した。しかし、文章を読めるほどではなかった[6]。長い間、人々はこの壁を越えられないと信じていた[2]。

参考資料

- [1] Herculaneum Society. The Villa of the Papyri. University of Oxford.
<https://www.herculaneum.ox.ac.uk/index.php/the-villa-of-the-papyri/> (2026年9月3日閲覧)
- [2] Vesuvius Challenge. FAQ. scrollprize.org. <https://scrollprize.org/faq> (2026年9月3日閲覧)
- [3] Villa of the Papyri. Wikipedia. https://en.wikipedia.org/wiki/Villa_of_the_Papyri (2026年9月3日閲覧)
- [4] University of Kentucky. (2026, 6月 25日). The day the Herculaneum scrolls began speaking again. UKNow.
<https://uknow.uky.edu/research/day-herculaneum-scrolls-began-speaking-again> (2026年9月3日閲覧)
- [5] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science Advances, 2(9), e1601247.
<https://doi.org/10.1126/sciadv.1601247>
- [6] Mocella, V., Brun, E., Ferrero, C., & Delattre, D. (2015). Revealing letters in rolled Herculaneum papyri by X-ray phase-contrast imaging. Nature Communications, 6, 5895. <https://doi.org/10.1038/ncomms6895>
- [7] Papyrus. Encyclopaedia Britannica. <https://www.britannica.com/topic/papyrus-writing-material> (2026年9月3日閲覧)

1.2. 科学捜査隊の追跡 1: 仮想のタマネギの皮むき

開かずに読むという発想

米国ケンタッキー大学のブレント・シールズ教授は別の道を選んだ。遺物を開く代わりにコンピュータの中で広げようという発想であった。実物はそのままにして3次元画像だけを取得する。その上で画面上で層を一枚ずつ剥がしていく。タマネギの皮をむくように内側へと入っていく手法である[1]。

研究チームはこの手法を仮想展開と名付けた。英語ではバーチャル・アンラッピングと呼ぶ。2016年にこの手法が初めて成功を収めた。対象は前節で触れたエン・ゲディの巻物であった。この遺物は1970年に古い会堂跡から出土した。外見は真っ黒に焦げた塊であった。研究チームが画面上で展開してみると、ヘブライ語の文字が現れた。旧約聖書のレビ記の冒頭であった[1]。

仮想展開は5つの段階から成る。3次元画像の取得、層の分離、インク値の付与、平坦化、貼り合わせである。エン・ゲディの写本ではこの手順がそのまま通用した。インクが画像にすでに見えていたからである[1]。

ヘルクラネウムの巻物はエン・ゲディの写本よりも条件が悪かった。火山の熱に押し潰されて断面が歪んでいた。円形が崩れ、層と層が互いに密着してしまっていた。さらにインクが画像に写らなかった。5つの段階のうち2番目と3番目がともに足かせとなった[2]。

そこでヘルクラネウムでは順序を変更した。まず層を分離して平らに広げる。インクの検出はその後で個別に行う。1.3節はそのインク探しの物語である[2]。

粒子加速器が撮影した3次元画像

最初の段階は巻物の内部を丸ごと撮影することである[1]。病院のCTは通常1ミリメートル前後まで識別する。パピルスの層はそれよりもはるかに薄い。ひとつの層の厚みは髪の毛の太さにも満たない。病院の設備では層と層が潰れて写ってしまう[3]。

研究チームは粒子加速器を訪ねた。最初に訪れたのは英国のダイヤモンド・ライト・ソースであった[3]。後にはフランス・グルノーブルのヨーロッパ放射光施設も併用した[4]。これらの施設は極めて明るく直進性の高いX線を生み出す。光が明るいほど、短時間で鮮明な画像を得ることができる[3]。

研究チームは巻物を筒に立ててしっかりと固定した。撮影中にわずかでも揺れが生じれば画像が不鮮明になるからである。こうして得られたデータの解像度は、1ボクセルあたり約8マイクロメートルであった。ボクセルとは深さの値まで持った立体画素を指す。お絵描きソフトの1つの点を小さな立方体に置き換えたものと考えればよい[3]。

データの量は膨大であった。巻物1本を撮影すると約300テラバイトに達する。スマートフォン数千台分の保存容量を埋め尽くす規模である。このデータを扱うこと自体が新たな課題となった[4]。

フランス・グルノーブルの施設は第4世代放射光源と呼ばれる。2020年に新設備が稼働を開始した。欧州21カ国が1億5000万ユーロを投じて性能を向上させた。この施設は本来、物質や生命現象を調べるために建設された場所である。古代の巻物を読むために使われるとは、当初は誰も予想していなかった[4]。

層を切り分ける体積分割

3次元画像を取得した後は、層を分離する段階に入る。この作業を体積分割と呼ぶ。立体画像から層を一枚ずつ切り分ける作業である。コンピュータは断面画像ごとに紙の表面が通った軌跡をたどる。その線を繋ぎ合わせると、紙1枚の立体モデルが出来上がる。このモデルをメッシュと呼ぶ[1]。

当初は人間がマウスで境界をひとつずつ指定していた[2]。巻物1本に含まれる紙の長さは1メートルを超える[4]。手作業でたどるには長すぎ、あまりに細かすぎた。さらに層同士が押し潰されて密着しており、境界が曖昧であった[2]。

大会の参加者たちが自動化ツールを提案したことで速度が上がった。スイスのユリアン・シュリガーが作成したプログラムがその代表例である。このプログラムは隣接する断面画像の模様を追尾して層を繋げていく。人間が数日かけていた作業を数時間に短縮した[2]。

この作業は一つの研究室が単独で担える規模ではなかった。主催者側は撮影データをインターネット上にそのまま公開した。参加者たちは自作したプログラムのソースコードを公開した。他者が作ったツールの上に次の参加者が機能を追加していった。判読テキストとコードは、現在も誰もがダウンロードできるようになっている[5]。

平らに広げて繋ぎ合わせる

層を分離しても、紙のモデルは依然として丸まったままである。このままでは文字が歪んで見えてしまう。そのため、モデルを平らに広げる計算を行う。この段階を平坦化と呼ぶ[1]。

コンピュータはモデルの頂点を小さな球に置き換える。球と球の間はバネで結ばれていると仮定する。その上で、この網を床の上に平らに寝かせる計算を繰り返す。バネの伸び縮みの度合いを最小限に抑えるように調整する。そうすれば文字の形状が大きく崩れることはない[1]。

最後は繋ぎ合わせである。巻物は一度にすべてを広げられるわけではない。コンピュータは複数の区画に分割して個別に広げる。広げた区画を再び一枚に張り合わせなければならない。この際、パピルス特有の格子状の繊維パターンを利用する。2つの区画の繊維パターンが合致すれば、もともと隣り合っていたことを意味する。この照合作業まで完了すれば、仮想展開は完成する[1]。

5つの段階は一度で終わるものではない。どこかの段階でミスが生じれば、その後の段階がすべて狂ってしまう。層の分離を誤れば、別のページの文字が混ざり込んでしまう。そのため研究チームは同じ巻物を何度も繰り返し処理する。結果を目視で確認し、設定を変更して再び実行する[5]。

参考資料

- [1] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. *Science Advances*, 2(9), e1601247.
<https://doi.org/10.1126/sciadv.1601247>
- [2] Vesuvius Challenge. (2024, 2月 5日). Vesuvius Challenge 2023 Grand Prize awarded: we can read the scrolls! scrollprize.org. <https://scrollprize.org/grandprize> (2026年9月3日閲覧)
- [3] Vesuvius Challenge. FAQ. scrollprize.org. <https://scrollprize.org/faq> (2026年9月3日閲覧)
- [4] University of Kentucky. (2026, 6月 25日). The day the Herculaneum scrolls began speaking again. UKNow. <https://uknow.uky.edu/research/day-herculaneum-scrolls-began-speaking-again> (2026年9月3日閲覧)
- [5] Vesuvius Challenge. (2026, 6月 25日). An entire Herculaneum scroll has been read for the first time. scrollprize.org. <https://scrollprize.org/firstscroll> (2026年9月3日閲覧)

1.3. 科学捜査隊の追跡 2: インクの微細な指紋探し

炭素の上に書かれた炭素

仮想展開が終わっても、画面は依然として空白のままであった。紙は広がったものの文字が見えなかった[1]。原因は前述の炭素問題であった。パピルスも炭素であり、墨も炭素であった。X線は両者を区別することができない[2]。

シールズ教授の研究チームはひとつの仮説を立てた。墨がまったく痕跡を残していないはずはない、というものであった[1]。墨は液体の状態で紙に触れる。液体は繊維の隙間に染み込む。そして乾燥するにつれて表面をわずかに引っ張る[3]。

そうなると、紙の表面の形状が微細に変化する。密度は同じでも起伏は異なる。研究チームはこの起伏を探し出すことにした[3]。目視では確認できない大きさであるため、機械に委ねる必要があった[1]。

ここで用いられたツールが機械学習である。人間がルールを一から記述して教える手法ではない。正解ラベルを付与した事例を大量に見せ、自らルールを見つけ出させる。猫の写真数千枚を見せれば猫を識別できるようになるのと同じ原理である。インク検出においては、墨がある場所とない場所を示したサンプルを提示する[1]。

クラックルという手がかり

2023年3月、ヴェスヴィオ・チャレンジという大会が開催された[1]。シールズ教授とナット・フリードマン、ダニエル・グロスが共同で立ち上げた[4]。主催者側は撮影データをインターネット上で無料公開した。賞金を懸けて世界中のプログラマーを呼び集めた[1]。

最初の数か月間は誰もインクを見つけられなかった。その年の夏、ケイシー・ハンドマーという参加者が手がかりをつかんだ。彼は展開された表面データをただ目でじっと見つめた。そうするうちに特異な模様を発見した。細かくひび割れた痕跡だった[1]。

日照りで干上がった田の地面を思い浮かべればよい。乾いた泥は網の目のようにひび割れる。墨が付いた場所でも同様の細かいひび割れが生じていた。参加者たちはこの模様をクラックルと呼んだ。ハンドマーはこの発見によって最初のインク賞を受賞した[1]。

ハンドマーは発見後すぐにオンラインチャットルームに知らせた。この共有が大会の流れを変えた。他の参加者たちが直ちにこの模様を学習材料とした[1]。

小さな窓からのみ見せる

アメリカの大学生ルーク・ファリターがいち早く動いた。彼はクラックル模様を機械学習モデルに学習させた。モデルはどこに墨があるかを確率で答える。学習が進むにつれて答えは正確になっていった[1]。

問題は学習データが少ないことであった。データが少ないと機械は答えを暗記してしまう。これを過学習と呼ぶ。試験問題を理解せずに答えだけを覚えた生徒と同じである[1]。

参加者たちはデータを回転させたり反転させたりして量を増やした。画像を回転させてもひび割れ模様の性質は変わらない。機械はそのおかげで、模様自体の特徴だけを学習することができた[1]。

入力ウィンドウの大きさにも手を加えた。最初は一度に広い領域を見せていた。広く見ると文字の形が丸ごと目に入ってくる。すると機械が文字を勝手に作り出し始めた。墨がない場所にもっともらしい画を描き加えた。これを人工知能の幻覚（ハルシネーション）と呼ぶ[1]。

研究者たちはウィンドウを64×64ピクセルに縮小した。その程度の広さでは文字の形は見えてこない。機械はひび割れの質感だけを見て判断しなければならない。作り出しは大幅に減った。代わりに、本物の筆跡が画面にそのまま現れた[1]。

この計算を担うモデルは、2種類をつなぎ合わせて作られた。一つは3次元畳み込みニューラルネットワークである。小さな立方体の中の模様を調べて特徴を抽出する構造だ。もう一つはTransformerである。離れた信号同士が互いにどれほど関連しているかを検討する構造だ。2つの構造を併用することで、薄い層の中のかすかな模様を見逃さなくなる[1]。

機械が出力した画面は、直ちに正解として受け入れられるわけではない。古文書の専門家たちがその画面を個別に読み解く。ギリシャ語の単語の並びが文法に合っているかを調べ、続く文章の意味が通じるかも確認する。機械と人間の判読が食い違えば、その部分は再計算される[1]。

表面の起伏だけで

その後もインクを見つける方法は進歩し続けた。参加者のユセフ・ナデルは事前学習の方法を用いた。正解データのない巻物データを先に大量に見せる方式である。機械はパピルス繊維の形状を自ら学習する。その上で、少量の正解データを使ってインクの検出だけを学習させるのだ[1]。

2026年3月には、起伏そのものを扱った研究が発表された。ジョルジョ・アンジェロッティ、フェデリカ・ニコラルディ、ポール・ヘンダーソン、ブレント・シールズが共同で執筆した。研究チームはすでに広げられたパピルス標本の表面を光学機器で測定した。高低差だけが記録されたデータを機械に学習させ、明るさや密度の情報は入れなかった[3]。

それにもかかわらず、機械は墨がある場所を見分けた。表面の高低差だけでインクを検出できるということだ。研究チームはどれほど細かく測定すべきかもあわせて計算した。この数値は次の撮影の解像度基準となる。この論文は同年、学術誌『サイエンティフィック・リポーツ』に掲載された[3]。

それでもインク検出はまだ完成された技術ではない。墨が薄く残っている箇所は依然としてかすれて現れる。層が圧着して癒着した部分では、撮影データ自体がぼやけてしまう。異なる2枚の紙が1枚にくっついてしまうこともある。ある巻物でうまくいったモデルが、別の巻物では空回りすることもある。そのため人間が随時目で確認し、修正を加える。研究は今も一歩ずつ前進している[5]。

参考資料

- [1] Vesuvius Challenge. (2024, 2月 5日). Vesuvius Challenge 2023 Grand Prize awarded: we can read the scrolls! scrollprize.org. <https://scrollprize.org/grandprize> (2026年9月3日閲覧)
- [2] Mocella, V., Brun, E., Ferrero, C., & Delattre, D. (2015). Revealing letters in rolled Herculeanum papyri by X-ray phase-contrast imaging. Nature Communications, 6, 5895. <https://doi.org/10.1038/ncomms6895>
- [3] Angelotti, G., Nicolardi, F., Henderson, P., & Seales, W. B. (2026). Ink detection from surface topography of the Herculeanum papyri. Scientific Reports. <https://doi.org/10.1038/s41598-026-58467-1>
- [4] University of Kentucky. (2026, 6月 25日). The day the Herculeanum scrolls began speaking again. UKNow. <https://uknow.uky.edu/research/day-herculeanum-scrolls-began-speaking-again> (2026年9月3日閲覧)
- [5] Vesuvius Challenge. (2026, 7月 10日). Open problems: why reading every Herculeanum scroll is still a challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026年9月3日閲覧)

1.4. 明らかになった真実

2023年に取り戻した最初の単語

最初の成果は一つの単語であった。2023年10月12日に発表が行われた。主人公はアメリカの大学生ルーク・ファリターであった。当時の年齢は21歳だった。彼が読み取った単語はポルピュラスであった。紫色を意味するギリシャ語だ。古代の文章でよく使われる言葉ではない。ファリターは最初の文字賞として4万ドルを受け取った。2位のユセフ・ナデルは1万ドルを受け取った。開いていない巻物から単語を読んだ最初の事例であった[7]。

2024年2月5日に大会の結果が発表された。大賞は3人が共同で受賞した。エジプトのユセフ・ナデル、アメリカのルーク・ファリター、スイスのユリアン・シュリガーである。賞金は70万ドルであった[1]。

彼らが読み解いた遺物は、パリ学士院が所蔵する巻物である。名称はPHerc. Paris. 4だ。3人はこの巻物から2,000文字を超えるギリシャ語を蘇らせた。15段の文章であった。段とは巻物を縦に区切って書かれた文章の区画のことである。巻物全体の約5パーセントに相当する[1]。

審査基準は厳格であった。140文字の節を4か所提出しなければならなかった。各節で文字の85パーセント以上を判読できれば合格であった。古文書の専門家たちが応募者の名前を伏せた状態で審査した[1]。

内容は快樂に関する議論であった。希少なものが、ありふれたものよりも必ずしもより大きな快樂をもたらすのかを考察している。本文には食べ物为例に挙げた一節がある。希少だからといって、直ちにより楽しいとは信じないという趣旨である。エピクロス学派の思想がそのまま表れた文章だ[1]。

題名を取り戻した巻物

次の成果は題名であった。2025年5月5日に最初のタイトル賞が発表された。受賞者はマルセル・ロートとミハ・ノヴァクであった。賞金は6万ドルであった[2]。

対象となった遺物は、オックスフォード大学ボドリアン図書館のPHerc. 172であった。画面に現れた文字は、フィロデモスの『悪徳について』であった。第1巻と判読された。ナポリ・フェデリコ2世大学のフェデリカ・ニコラルディ研究チームが確認した。巻かれた状態の巻物の題名を、開くことなく読み解いた初の事例である[2]。

この巻物からは、引き続いて70を超える段が読み解かれた。強欲や傲慢、へつらいといった性格の欠陥を扱った文章である。これで議論の流れを最初からたどることができるようになった[3]。

初めて最後まで読み解かれた一巻

2026年6月25日にさらに大きなニュースが報じられた。巻物一巻を最初から最後まで読み切ったという発表であった。遺物はナポリ国立図書館所蔵のPHerc. 1667である。大会参加者たちはこれを4番巻物と呼んでいる[4]。

撮影はフランスのヨーロッパ放射光施設（ESRF）で行われた。その後、仮想展開とインク検出を順次経た。蘇った文章の長さは約1.4メートルであった。段数にして22段ほどである。ヘルクラネウムの巻物を丸ごと読み解いたのは、この時が初めてである[4]。

内容は倫理と人間の品性を扱ったギリシャ語の論考である[3]。本文にはストア派の哲学者クリュシッポスの甥への言及がある。その名はアリストクレオンである[4]。文章が書かれた時期は紀元前3世紀末から2世紀頃とみられている。著者が誰であるかはまだ確定していない[3]。

研究成果は論文としても公開された。ジョルジョ・アンジェロッティ、スティーブン・パーソンズ、フェデリカ・ニコラルディを含む27人が名を連ねた[5]。データとコード、判読テキストもあわせて公開された[4]。

研究チームは検証の手順もあわせて明らかにした。2023年に読み解いたPHerc. Paris. 4を、より高い解像度で再撮影したのだ。新たなデータから得られた判読テキストは、以前の結果と正確に一致した。機械が捏造した文章ではないという証拠である。同じ方法を他の巻物にも適用できるという意味でもある[4]。

まだ残された600巻

同じ発表には別の知らせも含まれていた。PHerc. 139という巻物の題名が読み解かれたのだ。フィロデモスの『神々について』第8巻であった。それまでこの著作は第1巻しか知られていなかった。この作品が8巻以上にわたって続いていたという事実が初めて明らかになった[3]。

未解読の巻物はまだ600巻以上残っている[3]。大会は2027年のグランプリを掲げている。賞金総額は100万ドルだ。1位には80万ドルが贈られる。対象は指定された13巻の巻物である。各段に残っている文字の70パーセント以上を読み解かなければならない。締め切りは2027年6月25日である[6]。

この物語には、まだ多くの空白がある。文字を読み解いたとしても、その意味を深く吟味する作業は別に残る。判読テキストを1段分学術的に校訂するだけでも長い時間がかかる。著者が判明していない巻物もまだ残っている[3]。

それでも方向性は明確になった。2,000年の間、誰も開くことができなかった書物が画面の上で開かれた。灰となって消えたと思われていた書斎が、少しずつよみがえりつつある[4]。

この偉業を成し遂げたのは一人ではない。撮影を担当した物理学者や、層を分離したプログラマーがいた。古代ギリシャ語を解読した古文書学者たちもいた。大会に参加した学生や会社員も力を合わせた。2,000年ぶりに書物が開かれたのは、この人々が共に成し遂げた成果である。残された巻物も、同じ方法で開くことができるはずだ[4]。

参考資料

[1] Vesuvius Challenge. (2024, 2月 5日). Vesuvius Challenge 2023 Grand Prize awarded: we can read the scrolls! scrollprize.org. <https://scrollprize.org/grandprize> (2026年9月3日閲覧)

- [2] Vesuvius Challenge. (2025, 5월 5일). \$60,000 First Title Prize awarded! scrollprize.substack.com. <https://scrollprize.substack.com/p/60000-first-title-prize-awarded> (2026年9月3日閲覧)
- [3] University of Kentucky. (2026, 6월 25일). The day the Herculaneum scrolls began speaking again. UKNow. <https://uknow.uky.edu/research/day-herculaneum-scrolls-began-speaking-again> (2026年9月3日閲覧)
- [4] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. <https://scrollprize.org/firstscroll> (2026年9月3日閲覧)
- [5] Angelotti, G., Parsons, S., Nicolardi, F., Nader, Y., Johnson, S., Josey, D., et al. (2026). Complete virtual unwrapping and reading of a rolled Herculaneum papyrus. arXiv:2606.29085. <https://arxiv.org/abs/2606.29085> (2026年9月3日閲覧)
- [6] Vesuvius Challenge. Open Prizes. scrollprize.org. <https://scrollprize.org/prizes> (2026年9月3日閲覧)
- [7] Vesuvius Challenge. (2023, 10월 12일). First word discovered in unopened Herculaneum scroll by 21yo computer science student. scrollprize.org. <https://scrollprize.org/firstletters> (2026年9月3日閲覧)

第2章. 焼け焦げた聖櫃の中で花開いた御言葉： エン・ゲディの封印されたレビ記

2.1. 事件ファイル

死海の西のオアシス

エン・ゲディは死海の西岸にあるオアシスである。エンゲディとも記される。周囲は乾いた砂漠だが、ここには湧き水が出る。水があるため人が集まった。人々は何千年もの間、この水辺に集落を形成して暮らした。旧約聖書のサムエル記上にもこの土地の名が登場する。ダビデがサウル王から逃れてエン・ゲディの荒野に身を隠したという話である。

遺跡調査によると、この丘には紀元前8世紀後半から人が住んでいた。そして西暦600年頃の大火災で集落は崩壊した[1]。千三百年を超える歳月が土の中に幾重にも積み重なっている土地である。

死海の周辺は雨が少なく、空気が乾燥している。このような気候は古い品を長く残すのに役立つ。木や革、布が他の場所よりも腐りにくい。死海文書がこの地域から出土した理由もここにある。

1970年にこの地で大規模な発掘が始まった。発掘はダン・バラグとエフド・ネツェルが率いた。二人はエルサレム・ヘブライ大学考古学研究所に所属していた。ヨセフ・ポラトは1970年5月5日、この巻物を掘り出した[1]。発掘団が見つけたのは、古代のユダヤ教会堂の跡であった。会堂はユダヤ人共同体が集まって祈り、聖書を読んでいた建物である。床には色の異なる小さな石を組み合わせたモザイクが敷き詰められていた。

会堂の中には聖櫃を置く空間がある。聖櫃は神聖な聖書の巻物を納めておく箱である。ユダヤ教共同体はこの箱を会堂の中で最も神聖な場所とみなしている。発掘団はその聖櫃から焼け焦げた残骸を取り除いた。残骸の中から黒い塊が現れた[1]。

聖櫃から出土したという事実そのものが重要な手がかりである。会堂の聖櫃には、どんな文書でも入れるわけではない。共同体が読み、守ってきた聖書の巻物を納める。黒い塊の正体を推測する根拠がここにあった。

古代の聖書写本は、それ自体が貴重な資料である。印刷機がなかった時代には、人が手で書き写していた。書き写すたびに文字が変わる可能性が生じる。古い写本が見つければ、本文がどのように伝承されてきたかを確認できる。聖櫃から出た黒い塊に学者たちが熱中した理由がここにある。

火が残した黒い塊

火は記録を消し去る力を持つ。紙も革も火の中では灰になる。しかし、エン・ゲディでは事情が少し異なっていた。残骸に押しつぶされた巻物は、酸素が不足した状態でじゅくじゅくと蒸し焼きにされた。灰となって散ることなく、炭のように固まった。このように炭のように固まる変化を炭化と呼ぶ。空気が不足した火の中で燃えると、このような現象が起きる。

炭焼き窯で木が炭になる過程と原理は同じである。木をそのまま燃やせば灰しか残らない。土で覆って空気を遮断して蒸し焼きにすると、黒い炭が残る。炭化巻物もそのようにして作られた。

炭化した塊は、崩壊した建物の灰と土の下に埋もれた。空気が届かないため、腐敗することも止まった。革は本来、腐りやすい素材である。そのまま地表に置かれていたなら、とうの昔に消滅していただろう。遺物を破壊した火が、遺物を残した力でもあった。

発掘団が取り出した塊は、一見するとただの炭の塊に見えた。しかし、切り取られた側面には丸く巻かれた層が残っていた。考古学者たちはこれを、巻かれた状態の革の巻物だと判断した[1]。遺物には発見された土地の名が付けられた。エン・ゲディの巻物という名はこうして生まれた。

この巻物は、動物の皮を加工して作った羊皮紙である[1]。イタリアのヘルクラネウムで出土した巻物はパピルスで作られていた。パピルスは草の茎を圧着させた植物性の紙である。二つの遺物は支持体となる材料からして異なる。この違いが、後に解読方法を分ける分岐点となる。

羊皮紙は製造に多大な手間がかかる素材である。動物の皮から毛や油分を取り除いて洗浄する。木枠に張ってピンと引っ張った状態で乾燥させる。そうして整えた薄い面に文字を記す。紙よりも丈夫で長持ちするが、高価であった。

発見された状態も珍しいものであった。古代の巻物は通常、細かく砕けた破片として出土する。1947年からクムラン洞窟で見つかった死海文書も破片が多い。エン・ゲディの遺物は、巻かれた形状を保ったまま出土した。内側に何層もの層が残されていることを意味していた。

手を触れられない遺物

文字を読むためには巻物を広げなければならない。しかし、この遺物は広げることができなかった。熱によって革が収縮し、層同士が癒着してしまっていたからである。外側は石のように硬く、内部は脆く崩れやすかった。指先で触れれば粉々になってしまうような状態であった[1]。

方法がなかったわけではない。ナイフで層を切り離してみる道がある。薬品を塗って革を柔らかくする道もある。しかしどちらの方法も遺物を損壊させてしまう。一度壊れてしまった巻物は二度と元には戻せない。

表面を見るだけでは、文字があるかどうかすら分からなかった。炭化した表面は一面真っ黒で何の模様もない。墨も黒く革も黒いため、肉眼では見分けがつかない。内側に文字があるということも、重なり合う層の形状から推測したものにはすぎなかった。

文化財を扱う人々は、この点を常に最優先に置く。遺物は世界に一つしかなく、たった一度の失敗も許されない。そのため保存科学には、手を触れずに調査するという原則がある。このような調査方法を非侵襲調査と呼ぶ。侵襲とは内部に入り込むという意味であり、非侵襲はその反対である。エン・ゲディの巻物は、非侵襲調査のみが許される遺物であった。

45年の沈黙

巻物はイスラエルの遺物収蔵庫へと移された。のちにイスラエル考古学庁がこの遺物を引き継いで保管した[1]。箱の中で待つことだけが、当時の唯一の保存法であった。収蔵庫は温度と湿度を一定に保つ空間である。空気や光、害虫を遮断し、遺物がそれ以上劣化しないようにする。解読できない遺物であっても、その状態を守る取り組みは続けられた。

死海文書の保存を担当するプニナ・ショアが管理を引き継いだ。彼女は後に、この巻物を解読した研究にも名を連ねている[1]。

待ち時間は1970年から2015年まで続いた[1][2]。45年間、誰もその内部の文字を見ることができなかった。表面が真っ黒なだけで、文字があるかどうかすら判別が難しかった。学者たちの間では、永遠に読めない遺物ではないかという声も上がっていた。

待つことには理由があった。1970年の技術では、この塊を扱うことができなかった。物体の内部を細かくスライスして撮影する装置も一般的ではなかった。仮に撮影できたとしても、その膨大なデータを計算処理できるコンピューターが存在しなかった。遺物はそのまま保全し、技術の進歩を待ったのである。

その間、世界各地で同様の苦悩が続いていた。イタリアのヘルクラネウムの巻物も同じ壁にぶつかっていた。物理的に開こうとした過去の試みは、遺物を破壊するだけに終わった。二つの遺物は出土地も材質も異なっていたが、問題は一つであった。手を触れずに内側を読み解く方法が必要だった。

沈黙を破ったのは、新たな手先の職人技ではなかった。物体を切断せずに内部を透視する撮影技術と、コンピューターによる計算処理であった。シールズ研究チームは、手を触れずに古代の文書を読む方法を長年研究

してきた。この研究は米国国立科学財団とGoogleの支援を受けていた[2]。長年蓄積してきた手法を試す遺物が必要であり、エン・ゲディの巻物がその対象となった。

2015年、米国ケンタッキー大学のブレント・シールズ研究チームがこの任を引き受けた。研究結果は2016年9月21日、学術誌『Science Advances』に掲載された[1][2]。

残された問いは一つであった。手を触れずにどのようにして内部の文字まで読み解いたのか。答えは撮影装置と計算の手順にある。

参考資料

[1] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science Advances, 2(9), e1601247.
<https://doi.org/10.1126/sciadv.1601247>

[2] University of Kentucky Research. (2016, 9月 22日). Virtually Unwrapping the En-Gedi Scroll. research.uky.edu.
<https://research.uky.edu/news/virtually-unwrapping-en-gedi-scroll> (2026年9月3日閲覧)

2.2. 科学捜査隊の追跡：X線デジタル浄化

切断せずに内部を見る撮影

研究チームが選んだ最初の道具はマイクロCTである。CTはコンピュータ断層撮影を意味する言葉だ。病院で体内を検査する際に用いられる撮影と原理は同じである。X線を様々な角度から照射し、物体内部の密度を測定する。コンピューターはその値を集めて立体地図を作成する。マイクロCTは、これを極めて微小な目盛りで行う装置である。

エン・ゲディの巻物の撮影には、ブルカー社のSkyScan 1176という装置が使用された。撮影を担当したのはイスラエルのメルケル・テクノロジーズ社である。同社は撮影データをイスラエル考古学庁に寄贈した。撮影の目盛りは1コマあたり18マイクロメートルであった[1]。1マイクロメートルは1ミリメートルの1000分の1である。髪の毛の太さの数分の一に相当する大きさだ。

この目盛りのサイズには理由がある。層を分離して識別するには、革の1層よりもはるかに細かく測定しなければならない。目盛りが層の厚さよりも大きければ、二つの層が一つに潰れて塊になってしまう。撮影の目盛りは、読み取ることができる層の限界を決定づける条件である。

撮影によって生成されたデータは、単なる1枚の写真ではない。微小な立方体の粒子がびっしりと積み重なった立体的な塊である。この粒子をボクセルと呼ぶ。写真の画素が平面の粒子であるならば、ボクセルは立体的な粒子である。粒子の一つひとつには、その場所の密度を表す数値が記録されている。

この段階で、遺物は手を触れられることなくコンピューターの内部へと移行された。ここから扱う対象は革ではなく数値である。数値はいくら触れても壊れることはない。計算を間違えたとしても、最初からやり直せばよい。非侵襲調査の強みがここにある。

金属の混ざった墨が残した明るい点

立体データを得たからといって、すぐに文字が見えるわけではない。文字が見えるためには、墨と下地の密度が異なっていなければならない。この条件が、ヘルクラネウムの巻物では崩れていた。あちらの墨は木炭の粉から作られており、炭化したパピルスも炭素の塊である。両者の密度が酷似しているため、X線では容易に見分けがつかない。

エン・ゲディの巻物は条件に恵まれていた。研究チームはスキャンデータから、文字が書かれた場所の密度が周囲よりも高いという事実を確認した。そのため、この墨には鉄や鉛のような金属が含まれていると推測した。ただし、表面が露出している箇所がなかったため、成分を直接分析することはできなかった。研究チームはこの限界を論文にそのまま明記した[1]。

金属はX線を多く吸収する。吸収が多い部分は画面上で明るく表示される。炭化した革はX線の吸収が少ないため、暗く表示される。文字と下地の明暗差はこのようにして生じる。この差異こそが、コンピューターが文字を検出するための最初の手がかりとなった。エン・ゲディの巻物がヘルクラネウムよりも先に解読された理由もここにある。

ただし、この墨の成分はまだ確定されていない。研究チームが提示したのは、密度が高いという観察結果と、それに基づく推定にすぎない[1]。古代の書記が何を混ぜたのかは、試料を採取して初めて明らかになる。確定していない事柄を確定した事実であるかのように扱わないことが、この分野の原則である。

仮想展開の4段階とムラの除去

明るい点を見つけたからといって、文章が読めるわけではない。文字は幾重にも丸まった層の上に散らばっている。この重なりをコンピューターの中で平らに広げなければならない。この作業を仮想展開と呼ぶ。シールズ研究チームは仮想展開を4つの段階に区分した[1]。

第1段階は、層の表面を特定する作業である。コンピューターは立体データを薄くスライスしながら、革の層の座標を検出していく。こうして集めた点を結び合わせ、薄いシート状の面を構築する。層同士が癒着していると、この段階で迷走してしまう。2つの層を1つの層として誤認して結合させてしまうと、文字も混ざり合ってしまう。

第2段階は、その面に輝度値をマッピングする作業である。面が通過する座標の撮影値を読み取り、面上に転写して描画する。密度が高かった墨が、この時に明るい筆跡として現れる。

第3段階は、歪んだ面を平らに展延する作業である。立体を平面に押し広げると、形状が伸びたり歪んだりする。地球儀を紙の地図に展開すると極地方が過大に広がるのと同じ問題である。計算処理により、この歪みを可能な限り抑える方向で面を平坦化する。

第4段階は、個別に展開した複数の面をつなぎ合わせる作業である。面と面の境界を整合させ、1枚の大きな画像へと合成する[1]。

4つの段階の中で、第1段階が最も難易度が高い。層の特定を一度でも誤ると、その後の3つの段階がすべて狂ってしまう。誤って展開された画像には、本来存在しない文字が現れてしまうこともある。研究者たちがこの段階を手作業で確認する理由がここにある。

これら4つの段階を経ることで、黒い円筒が1枚の平面画像となる。研究チームはこの方法により、革の5つの層を展開した。展開された面の面積は約94平方センチメートルと推計された[1]。ノート1ページ分よりやや狭い広さである。

展開された画像にも課題は残る。革の厚みが箇所によって異なり、しわも残っている。撮影の過程で生じたノイズも混入する。そのため、ある部分は過度に明るく、ある部分は暗くなってしまう。文字と明暗のムラが混ざり合うと、判読が困難になる。

研究チームはこのムラを計算処理によって取り除いた。まず、周囲の画素の平均輝度を測定し、下地の輝度マップを作成する。このマップは広範囲に広がる明暗の傾向のみを反映する。次に、元の画像をこの下地マップで除算する。除算処理を経ることで、大域的な輝度の偏りが除去される。狭い範囲に集中している墨の信号だけが鮮明に残る。このプロセスを経た画像から、ヘブライ文字が浮かび上がった[1]。

2026年にも続く課題

仮想展開は、その後ほかの遺物にも使われた。2026年6月25日、ヴェスヴィオ・チャレンジが知らせを発表した。ヘルクラネウムの巻物1巻を最初から最後まで読んだという内容である。遺物番号はPHerc. 1667である。撮影はフランス・グルノーブルの欧州シンクロトロン放射光施設で行われた。展開されたギリシア語本文は、約1.4メートルの長さの22段である。紀元前2世紀の倫理学著作であることが確認された[2]。

シンクロトロンは、高速で走る電子を磁場で曲げる装置である。電子が曲がる時、きわめて強いX-rayが出る。この光で撮影すると、小さな密度差も現れる。実験室の撮影機では得にくい鮮明さである。エン・ゲディで使った装置より一段上の道具である。

この成果には撮影方法の改善があった。新しい撮影法では、炭素インクが資料の内部で直接見えた。密度差が小さくて見えなかったインクを捉えたのである。エン・ゲディで自然に与えられていた明るさの差を、撮影技術で作り出したのである。残る約600巻の巻物にも同じ方法を使えると伝えられた[3]。

課題はなお残っている。2026年7月10日に公開された整理では、残る問題を6つに挙げる。層が押しつぶされた区間ではX-rayが散乱し、画面がぼやける。層を見つける計算は穴を開けたり、2つの層を貼り合わせたりしてしまう。ボクセル資料には、どの層がどの層につながるかが入っていない。人が作った学習用の印は、粒子単位では正確でない。ある巻物で学んだモデルが、別の巻物ではうまく効かない。インクが層のどの深さにあるのかも明確でない[4]。

解決方法もあわせて示された。正解の印が足りなければ、資料そのものから規則を学ばせる道がある。自己教師ありの学習モデルが、その候補に挙げられる。層を見つける計算にも、新しいニューラルネットワーク構造が使われている[4]。

新しい手がかりも出た。2026年3月29日に公開された研究は、インクが表面の形を変えるとという点に依拠した。研究陣は、すでに開かれているパピルス表面の起伏を精密に測り、機械学習に教えた。その結果、起伏だけでも文字のある場所を選び分けることができた。この研究は、起伏を読むにはどの程度の撮影目盛りが必要かも検討した[5]。エン・ゲディで密度から得た手がかりを、別の性質から再び見つけた研究である。

参考資料

- [1] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science Advances, 2(9), e1601247. <https://doi.org/10.1126/sciadv.1601247>
- [2] Vesuvius Challenge. (2026, 6月 25日). An entire Herculaneum scroll has been read for the first time. scrollprize.org. <https://scrollprize.org/firstscroll> (2026年9月3日閲覧)
- [3] ASNT. (2026, 8月 13日). X-Ray Imaging and AI Recover First Complete Text from Sealed Ancient Scroll. Materials Evaluation. <https://www.asnt.org/me/26/8/x-ray-ai-recover-text-sealed-ancient-scroll> (2026年9月3日閲覧)
- [4] Vesuvius Challenge. (2026, 7月 10日). Open Problems: Why Reading Every Herculaneum Scroll Is Still a Challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026年9月3日閲覧)
- [5] Angelotti, G., Nicolardi, F., Henderson, P., & Seales, W. B. (2026, 3月 29日). Ink Detection from Surface Topography of the Herculaneum Papyri. arXiv:2603.27698. <https://arxiv.org/abs/2603.27698> (2026年9月3日閲覧)

2.3. 明らかになった真実

画面に浮かび上がったレビ記

コンピューター画面に浮かび上がった文字はヘブライ文字だった。研究陣は革の5層を広げ、その中から2段の文章が出た。本文は旧約聖書の第3の書であるレビ記だった。読み取られた箇所はレビ記1章と2章である[1]。

合成された画像には、巻物1枚分の文章35行が収められていた。このうち18行には文字が残っていた。残りの17行は学者たちが復元したものである。復元は、コンピューターが文章を作り出すことではない。残った筆画と知られている本文を照合し、欠けた箇所を埋める作業である。1行に入った文字と語間の空白は33から34だった[1]。

論文が残った行と復元した行を分けて明らかにした点は重要である。どこまでが確認された事実なのかを読者が区別できるからである。残ったものと補ったものを混ぜて発表すれば、後の人が検証できない。そのおかげで、ほかの学者たちが同じ資料でもう一度検討できた。

レビ記1章は焼き尽くす献げ物の規則を扱う。焼き尽くす献げ物とは、祭壇の上で供え物を完全に焼いてささげる祭儀である。2章は穀物でささげる献げ物の規則を扱う。上質の粉に油を混ぜ、香を載せる方法がここに出てくる。火で焼けた聖櫃から、火で焼いてささげる祭儀の規則が出てきたのである。

復元された文章にはレビ記の最初の一節も含まれていた[1]。主が会見の天幕からモーセを呼び、語ったという文である。書の冒頭が残っていたことは、この巻物がレビ記1巻の始まりだったことを意味する。

2段という言葉にも意味がある。古い巻物は、文章を縦長の欄に分けて書く。1つの欄をすべて埋めると、隣の欄へ移る。2段が出たということは、そのような欄2つが続いて出たという意味である。

読み取られたのは文字だけではない。書記が筆を押しつけて書いた筆画の太さまで画面に残った。人が手で写した写本であるという証拠である。

千年の空白を埋める

研究陣は巻物の年代も測った。放射性炭素年代測定により、西暦3世紀から4世紀という値が出た[1]。この方法は、遺物に残った炭素の変化を測って年代を調べる。これらの値は、古い聖書写本研究にあった長年の空白を埋める。

まず空白が何かを見てみよう。クムラン洞窟から出た死海文書の聖書写本は、紀元前3世紀から紀元後1世紀のものである。これに対して、完全な形で残った標準ヘブライ語聖書写本は、西暦1000年ごろのものである。2つの群の間には、ほぼ千年の空白がある。この期間に本文がどのように写されたのかを示す写本は少なかった。

空白が大きければ疑いも大きくなる。中世写本が古い本文をそのまま写したのか、確認する道がないからである。学者たちの間には、中世の写字生が本文を整えたのだらうという見解もあった。

エン・ゲディ巻物は、この空白の真ん中から出てきた。そして読み取られた本文は、中世ヘブライ語聖書の子音本文と完全に同じだった[1]。ヘブライ語は子音文字だけを書き、母音は書かない。その子音だけで記された本文を子音本文と呼ぶ。この中世本文をマソラ本文と呼ぶ。ユダヤ教の写字生たちが規則を定めて守ってきた標準本文である。段落を分けた位置まで同じだった[2]。この結果は、長い年月の間、本文がそのまま写されてきたことを示す。

ユダヤ教の写字生たちは聖書を写すとき、規則を厳格に守ったと知られている。文字を数えて確認する方法も、そのような規則の一つである。エン・ゲディ巻物は、その規則が実際に守られていたことを示す物証となった。写本1点が、長年の論争に一つの資料を加えたのである。

この写本には、もう一つの記録が付く。これまでに聖櫃の中から出たモーセ五書の写本の中で、年代が古い[1]。モーセ五書とは、旧約聖書の最初の5巻をまとめて指す言葉である。

確定したことと残ったこと

明らかになったことと、なお議論中のことは分けて見る必要がある。

明らかになったことは次のとおりである。読み取られた本文はレビ記1章と2章である。炭素年代測定値は西暦3世紀から4世紀を指す。子音本文は中世の標準本文と同じである。インクの密度が革より高く、金属成分が混じっていたとみられる[1]。

残った問題も明確である。インクの成分は直接分析できなかった[1]。外に現れた面がなく、試料を得られなかったからである。筆跡を根拠に年代を別に見る見解もある。古文書学者アダ・ヤルデニは、西暦1世紀後半または2世紀初めと見た。古文書学者とは、古い文字の形を比べて年代を測る学者である。これに対してドリュウ・ロンゲネッカーは、炭素年代測定のほうが信頼できると反論した[2]。2つの見解はなお整理されていない。

保管期間も解けていない問題である。巻物は西暦3世紀から4世紀に作られた。会堂は西暦600年ごろに焼けた[1]。そうであれば、この写本は200年以上にわたり聖櫃に置かれていた。共同体がなぜ古い写本を置き続けたのかは明らかになっていない。受け継いだ写本を大切にしていたのだろうという推定だけがある。

読み取られた範囲も、巻物の一片である。広げられたのは外側の5層である[1]。内側の深いところは、層がさらに激しく押しつぶされている。押しつぶされた区間で層を分けられない問題は、2026年にも残っている。ヘルクラネウム巻物を扱う研究者たちが今も取り組んでいる課題である[3]。

内側を読むには、2つの条件がともによくななければならない。層をさらに細かく分けて見る撮影目盛りが必要である。貼りついた層を分ける計算もよくななければならない。2つの条件が整えば、エン・ゲディ巻物は再び撮影台に上がることができる。遺物がそのまま残っているからこそ可能なことである。

年代測定そのものを助ける方法も増えている。2025年には、ヘブライ語の古い写本の筆跡を機械学習で学び、年代を推定するモデルが発表された。このモデルは、炭素年代測定値と筆跡資料をあわせて学んだ。人の目では捉えにくい筆画の曲がりまで測る方式である[4]。こうした道具が蓄積すれば、エン・ゲディ巻物の年代論争も再び検討できる。

仮想展開が開いた道は、聖書写本だけにとどまらない。折りたたまれて封印された古い手紙にも同じ方法が使われた。壊れやすい本や文書にも撮影と計算が入っている。手を触れずに読むという原則が、資料を扱う基本になりつつある。

コンピューターが出した画像は、それ自体が最終解答ではない。2016年の研究でコンピューターがしたことは、層を広げて画像を作るところまでである。その画像から文字を読み、意味を付けたのは人である。だから古文書学者が一文字ずつ改めて調べなければならない。この出来事が残したものは、本文一つだけではない。触れれば壊れる遺物を扱う方法が一つ生まれた。撮影で数値を得て、計算で広げ、人が検討する順序である。この順序は、後に出てくる多くの遺物にもそのまま使われる。

エン・ゲディで読み取られたレビ記も、その検討を経て学界の記録となった。45年の沈黙を破ったのは、機械と人がともにに行った仕事である。撮影装置は密度を測り、計算は層を広げた。そして人が文字を読み、意味を付けた。次の章では、文字の形そのものを調べる話が続く。

参考資料

- [1] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. *Science Advances*, 2(9), e1601247. <https://doi.org/10.1126/sciadv.1601247>
- [2] En-Gedi Scroll. (2026, 9月 3日). Wikipedia. https://en.wikipedia.org/wiki/En-Gedi_Scroll (2026年9月3日閲覧)
- [3] Vesuvius Challenge. (2026, 7月 10日). Open Problems: Why Reading Every Herculaneum Scroll Is Still a Challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026年9月3日閲覧)
- [4] Popović, M., Dhali, M. A., Schomaker, L., van der Plicht, J., Rasmussen, K. L., La Nasa, J., Degano, I., Colombini, M. P., & Tigchelaar, E. (2025). Dating ancient manuscripts using radiocarbon and AI-based writing style analysis. *PLOS ONE*, 20(6), e0323185. <https://doi.org/10.1371/journal.pone.0323185>

第3章. イザヤ巻物に隠された二人の影：書記鑑定

3.1. 事件ファイル

洞窟から出た革の巻物

1947年に死海の西側にあるクムラン地域の洞窟で、古い革の巻物が発見された。ヤギを飼っていたベドウィンの人々が、崖の洞窟の中で壺と巻物を見つけた。これらの遺物は死海文書という名で呼ばれる。その後、いくつもの洞窟で発掘が続いた。これまでに集まった死海文書は、約930点の写本として整理されている。手で写して書いた本を写本と呼ぶ。紙片ごとに数えると、革とパピルスの断片は数千枚を超える[1]。記録された時期は紀元前3世紀から紀元後1世紀の間と見られている[2]。

死海は塩分の濃い湖で、周囲は乾いた砂漠だ。雨はほとんど降らず、空気は乾いている。洞窟の中は温度も大きく変わらない。革とパピルスが腐らずに残った理由はここにある。これほどの規模の古い文書群が残った場所は珍しい。

この遺物の中に、研究者の目を引きつけた巻物が一つある。大イザヤ巻物だ。学界で使う記号は1QIsaaである。クムラン第1洞窟から出たイザヤ書の第一写本という意味だ。紀元前2世紀に動物の革の上にヘブライ語で書かれた写本である。長さは約7.3メートルだ。旧約聖書イザヤ書の内容をほぼ丸ごと含んでいる。本文は縦長の54の段に分かれている[2]。この段を学術用語でコラムと呼ぶ。死海の洞窟から出た聖書写本の中で、保存状態が最もよい。

巻物は何枚もの革を糸でつなぎ合わせて作った。1枚にいくつかの段を書き、次の革へ移る。書き終えた巻物は丸めて保管する。読むときは片側をほどき、反対側を巻きながら見る。巻物という名はここから出た。

この写本の価値は、その古さにある。死海文書には、現在残っているヘブライ語聖書写本の中で最も古いものが含まれている[3]。大イザヤ巻物もその一つだ。聖書本文が長い歳月の中でどのように伝わったのかを比べる資料が生まれたのである。ところが、巻物を作った人物については知られていなかった。

誰が書いたのかをめぐって分かれた学界

巻物を読むことと、誰が書いたのかを明らかにすることは別の仕事だ。文字はすべて読めるが、この写本には写した人の名前がない。文を書き写す仕事を担った人を書記と呼ぶ。この巻物には書記の署名もなく、日付もない[2]。そのため研究者は、文字の形だけを見て人数を数えなければならなかった。

この問題をめぐり、学界は二つの道に分かれた。一方は、書記一人が最初から最後まで一人で書いたと見た。巻物全体の字形がそろっていることが根拠だった。もう一方は、二人以上が分担して書いたと見た。中ほどを過ぎるにつれて、筆画の形が少しずつ変わることが根拠だった。

二つの主張はいずれも難しい問題を抱えていた。一人が書いた字でも、最初と最後がまったく同じになるわけではない。手が疲れれば筆画はゆるみ、ペンを替えれば太さが変わる。この揺れの大きさが分からなければ、人物が変わったのかを判断できない。二つの主張はいずれも、目で見た印象に頼っていた。どちらが正しいのかを量る重りがなかった。

クムランには、巻物を作り守っていた共同体があったと見られている。人々は巻物を壺に入れ、洞窟に隠した。戦争が近づいたため、文書を守ろうとしたと見られる。隠した人々の名前も残っていない。巻物を作った手を探すことは、その人々に近づく道でもある。

この論争は、文字の話だけにとどまらない。写本を複数の人が分担して作ったのなら、書記たちが共同で働く組織があったという意味になる。聖書本文がどのような人々の手を経て今日まで来たのかを明らかにする仕事につながる[2]。死海文書の研究者がこの問いを長く抱えていた理由はここにある。

目による鑑定の限界

古い筆跡を研究する学問を古文書学という。研究者が拡大鏡を手に文字の形を調べる方法だ。文字の高さと傾き、筆画の太さを測り、図に写し取る。複数の写本の文字を時代順に並べ、変化の流れをつかむこともある。長い間、この方法が唯一の鑑定道具だった。

この方法には弱点がある。研究者の目と経験に頼る幅が広い。同じ文字を見ても、研究者ごとに異なる評価を出すことが多かった[2]。人の目はごく小さな数値の差を測れない。筆ペンを押した力の強さや、インクがにじんだ厚みは目で見積もりにくい。記憶にも限界がある。数メートル離れた二つの文字を頭の中で並べて比べるのは簡単ではない。結局、意見の対立だけが繰り返された。

もう一つの壁は、その時代の教育方法だった。書記になろうとする人は、師の字をそのまま写す訓練を受けた。自分だけの個性を出さないように学んだ。その結果、同じ学校から出た書記たちの字は、見た目にはほとんど同じだった。人間の研究者にとって、これは濃い霧だった。巻物の前後にある細かな違いが、一人の体調の変化なのか、人物が変わったのかを見分けられなかった[2]。

鑑定がずれると、その後に続く判断もずれる。書記が一人なら、一つの写本には一人の時間が込められる。二人なら、仕事を分担する組織があったということになる。写本を作った場所の規模と性格が変わる。一つの筆跡鑑定に、いくつもの結論がぶら下がっていた。

教室で同じ手本を見て書き取りをした子どもたちを思い浮かべると、事情は似ている。みな同じようにきちんと書こうとしたノートは、見た目にはなかなか区別できない。それでも、筆画を折る癖は子どもごとに少しずつ違う。問題は、その差を測る物差しがなかったことだ。

状況が変わったのは、写本写真がデジタルで公開されてからだ。実物の巻物は光と湿気に弱く、頻繁に広げることにはできない。写真があれば、遺物に触れずに何度でも調べられる。イスラエル考古学庁はGoogleとともに死海文書のデジタル資料集を作った。MegaVisionという装置で複数の波長の光を当て、遺物を撮影した。人の目には見えない光である近赤外線まで含め、薄れた文字をよみがえらせた。実物を直接のぞき込むのに近い画質を目標にした[1]。

コンピューターが読める高解像度写真が蓄積されると、新しい道具が入る道が開かれた。写真は数値の集まりだ。画面を構成する小さな点を画素と呼ぶ。画素ごとに明るさの値が付いている。筆画の太さと角度を数値で測れるという意味だ。目で行っていた鑑定を計算に変える準備が整ったのである。ここからAI探偵が乗り出す。

参考資料

- [1] Israel Antiquities Authority. (n.d.). The Digital Library. The Leon Levy Dead Sea Scrolls Digital Library. <https://www.deadseascrolls.org.il/about-the-project/the-digital-library> (2026年9月3日閲覧)
- [2] Popović, M., Dhali, M. A., & Schomaker, L. (2021). Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsaa). PLOS ONE, 16(4), e0249769. <https://doi.org/10.1371/journal.pone.0249769>
- [3] University of Groningen. (2021, 4月 21日). Cracking the code of the Dead Sea Scrolls. <https://www.rug.nl/news/2021/04/cracking-the-code-of-the-dead-sea-scrolls?lang=en> (2026年9月3日閲覧)

3.2. 科学捜査隊の追跡：字体に残された生体指紋

捜査チームが組まれる

オランダのフローニンゲン大学の研究チームがこの事件を担当した。古文書学者Mladen Popovićがチームを率いた。AI研究者Lambert SchomakerとMaruf Dhaliも加わった[1]。彼らは2017年に、死海文書の筆跡をコンピューターで測る方法を先に発表した[2]。その方法を磨き、大イザヤ巻物に向けた結果が2021年に出た[1, 5]。

研究チームが立てた目標は明確だった。人の印象評価を数値に変えることだった。同じ人が書いた字の中で生じる揺れの大きさをまず測る。次に、巻物の前後の差がその揺れより大きいかを調べる。差が大きければ、別の人が書いたという意味になる。

2017年に作られた特徴計算法は、写本同士を比べるために使われた[2]。研究チームはその計算法を大イザヤ巻物一つに集中して再び適用した。一つの写本の中で書記が変わった場所を探す作業は、写本同士を比べる作業より難しい。文字の差がはるかに小さいからだ。

この方法にはよい点がある。計算手順を書き残しておけば、誰でも同じ写真から同じ答えを得られる。研究者の眼力に頼らなくてもよい。結果が間違っていれば、どの段階で間違ったのかをたどれる。鑑定結果をめぐって再び争うとしても、争う根拠が数値として残る。

染みを消す最初の段階

最初の作業は、写真からインクだけを残すことだ。この作業を二値化と呼ぶ。すべての画素をインクと背景のどちらかに分ける処理である。写真を黒い文字と明るい背景に分ける作業だ。

この段階が必要な理由がある。巻物は約2千年間、洞窟にあった。革の表面には染みとカビの跡ができ、細かいひびも入った。コンピューターから見ると、こうした跡とインクの筆画はいずれも暗い画素だ。背景をそのままにすると、機械は文字ではなく革の目を学習する。そうすると、書記ではなく革の種類を区別することになり、鑑定がずれる[1]。

研究チームは、この作業を担うニューラルネットワークを別に作った。名前はBiNetである。ニューラルネットワークは、人の脳の神経細胞のつながりをまねて作った計算構造だ。BiNetは、手でインクと背景を示した見本を見て、その判断基準を学ぶ。学習を終えたBiNetは、新しい写真から染みを取り除き、筆画だけを残す。

ここで守るべき規則がある。元の筆画の太さと縁の形を変えてはならない。書記の癖が入っている場所が、まさにその縁だからだ。汚れを消そうとして筆画まで滑らかに整えると、証拠が消える。BiNetはインクの痕跡をそのまま残す方向で設計された[1]。この技術は2019年に公開プレプリントとして先に出た[3]。設計と性能を改めて整理した論文は、2026年の国際文書認識学術論文集に掲載された[4]。

文字全体の癖を測る物差し

きれいになった文字から二種類の特徴を取り出す。一つ目はテクスチャに関する特徴で、二つ目は形に関する特徴である。

テクスチャに関する特徴をHingeと呼ぶ。Hingeは扉の蝶番を意味する言葉だ。筆画が曲がる場所ごとに、二方向の角度を測る。その角度の組がどれほど頻繁に出るかを表にする。この表は、字を書くときに手首が回る癖を数値として収める。文字を一つ一つ認識しなくても計算できる[1]。

形に関する特徴はFragletと呼ぶ。筆画を短く切った断片を意味する言葉だ。筆画を切って集め、一覧にする。研究チームは、複数の写本から取り出した断片で標準一覧を作った。新しい文字が入ると、その断片が一覧のどの項目に似ているかを数える。人が文字を学ぶときに身につけた筆画の基本形が、ここに現れる[1]。

この二つの特徴には共通点がある。文字の意味を知らなくても計算できることだ。ヘブライ語を読めないコンピューターでも、筆画の角度と曲がった形は測れる。傷んだ文字が混じっていても、残った筆画から値を取り出す。古い文書を扱うときに役に立つ性質だ。

二つの特徴は異なるものを見る。Hingeは手の動きを見て、Fragletは学んだ字体を見る。研究チームは両者をそれぞれ使い、合わせても使った。異なる物差しが同じ答えを出すかを確認するためだ[1]。

アレフ一字に残った痕跡

前の二つの特徴は、文字全体をまとめて測る。今度は一つの文字を大きく拡大してのぞき込む番だ。

研究チームは、ヘブライ語の字母一つを別に選んで精密検査を行った。アレフである。アレフはヘブライ語の最初の文字で、巻物に最も頻繁に出る。斜めの筆画一つに短い筆画二つが付いた形だ。

アレフを選んだのには理由がある。標本が多いほど統計は安定する。よく出る文字でなければ、前後の区域から偏りなく集められない。巻物全体にアレフは5,011個ある。研究チームはそのうち758個をコンピューターで選び出し、分析に使った[1]。筆画が途切れず、完全に残ったものだけを選んだ。

選んだアレフは二つの群に分けた。巻物の前側から出たものと、後ろ側から出たものだ。各群の画像を同じ位置に重ね、平均の形を作った。何枚もの半透明の紙を重ね、上から見下ろす方法と同じだ。よく重なる筆画は濃くなり、人ごとに異なる筆画は薄くなる。二つの平均画像を互いに引くと、残るのは差だけだ。その差がどの筆画に集中しているかを色で見ることができる[1]。

二つの群に分かれた点

数値に変わった特徴は、値が数百個ずつ並ぶ長い一覧だ。人はこのような一覧を目で比べることができない。そのため研究チームは、一覧を二つの値に縮め、平面上に点として打った。値が似た文字は近くに集まり、異なる文字は遠ざかるように縮める。点の位置を少しずつ動かす計算を繰り返して作った図である。

54の段のうち、写真が完全なものをすべてこの平面に載せた。16番と46番の段は、使える写真がなく除外された。点が打たれると、二つの群が現れた。前側の段が一つの群を作り、後ろ側の段が別の群を作った[1]。二つの群の間には、目に見える隙間があった。

研究チームは、この分裂が偶然かどうかを検証した。カイ二乗検定やt検定のような統計手法を用いた。カイ二乗検定は、差が偶然かどうかを調べる計算である。t検定は、二つの群の平均差を調べる計算である。どちらの方法も、偶然にこのような形が現れる確率を求める。その確率がきわめて小さければ、偶然とは見なさない。一人の筆跡の中で生じる揺れだけでは、これほどの差は出なかった[1]。

複数の物差しが同じ答えを出したことも重要である。ヒンジで測っても、フラグレットで測っても、同じ場所で境界が捉えられた。アレフの平均図でも同じ場所が分かれた。境界は第27欄と第28欄の間に置かれた[1]。互いに異なる方法が一つの場所を指したなら、その境界は写真の汚れから生じたものではない。人の目が見落とした線を、機械が引いてくれたのである。

研究チームは、反対の可能性も検討した。写真の明るさや革の状態が前後で違うために生じた結果かもしれない。しかし、二値化で背景を取り除いた後も、境界はそのまま残った[1]。原因が革ではなく、文字そのものにあるという意味である。

参考資料

- [1] Popović, M., Dhali, M. A., & Schomaker, L. (2021). Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsaa). PLOS ONE, 16(4), e0249769. <https://doi.org/10.1371/journal.pone.0249769>
- [2] Dhali, M. A., He, S., Popović, M., Tigchelaar, E., & Schomaker, L. (2017). A digital palaeographic approach towards writer identification in the Dead Sea Scrolls. Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods, 693-702. <https://doi.org/10.5220/0006249706930702>
- [3] Dhali, M. A., de Wit, J. W., & Schomaker, L. (2019). BiNet: Degraded-manuscript binarization in diverse document textures and layouts using deep encoder-decoder networks. arXiv:1911.07930. <https://arxiv.org/abs/1911.07930> (2026年9月3日閲覧)
- [4] Dhali, M. A., de Wit, J. W., & Schomaker, L. (2026). BiNet: A deep encoder-decoder network for binarizing degraded ancient manuscripts. In Document Analysis and Recognition - ICDAR 2025 Workshops (Lecture Notes in Computer Science, pp. 166-193). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-09371-4_11
- [5] University of Groningen. (2021, 4월 21일). Cracking the code of the Dead Sea Scrolls. rug.nl. <https://www.rug.nl/news/2021/04/cracking-the-code-of-the-dead-sea-scrolls?lang=en> (2026年9月3日閲覧)

3.3. 明らかになった真実

第27欄で替わった手

分析結果は、複数書記説を支持した。イザヤ大巻物は二人が分担して書いた。第一の書記が第1欄から第27欄までを書いた。第二の書記が第28欄から第54欄までを書き継いだ[1]。

この境界は、他の証拠ともかみ合う。第27欄の下部分は、三行分の高さだけ空いている。第28欄は、新たに継ぎ足した革のシートから始まる。そこはイザヤ書の新しい章が始まる場所でもある[1]。仕事を引き継ぐのに適した区切りだった。コンピュータが見つけた統計上の境界と、巻物の物理的な境界が同じ場所にあった。

第27欄の後で書記が替わっただろうという推測は、以前にもあった。革のシートがつながる場所と本文が分かれる場所を見て出た見解だった。ただし、学界の合意には至らなかった。AI分析が、この推測に数値を与えた[2]。

この発見の意味は大きい。聖書写本の一つを複数の人が続けて作ったという根拠が、数値で示された[1]。二人の書記は名も残さず、署名も残さなかった。残したものは、手が動いた癖だけだった。2千年後の機械が、その癖を読み取った。ポポヴィッチは、名前は最後まで分からないだろうが、筆跡を通じて彼らと握手するような感覚だと述べた[2]。

同じ学校で学んだ二人

二人の書記の文字は、なお互いに似ている。機械が測った後でも、差はわずかだった。研究チームは、この類似そのものをもう一つの証拠として読んだ。二人の書き方が互いに近いという意味である[1]。研究チームは、二人の書記が同じ場所で訓練を受けたか、同じ系統から出たと見た[2]。

ここから古代の書記学校の姿が浮かぶ。訓練生は公認された字体を一画ごとに身につけた。自分の個性を抑えなければ、よい書記にはなれなかった。その規律のおかげで、一つの巻物の中で文字が均一に保たれた。

それでも、体の癖まで消すことはできなかった。筆を握る角度は人によって異なる。手を動かす速度や押す力も異なる。こうした差は、画が折れる角度や画の縁の形に残る。人の目は、この細かな差を集めて判断できない。コンピュータは数千個の値を集め、平均とばらつきを計算する。規律が覆い隠していた個人の痕跡が、そのように現れた[1]。

この結果は、写本を作る仕事が共同作業だったことを示す。長い巻物一つを作るには何か月もかかる。一人が最後まで抱え込む理由はない。仕事を分担し、引き渡した可能性がここから出てくる[1]。聖書本文は、このような手から手を経て今日まで伝わった。

機械がまだできないこと

この成果は、AIの万能を意味するものではない。同じ研究チームが作った次の道具が、その限界を示す。

研究チームは2025年に、Enochという年代推定モデルを発表した。Enochは文字の形を見て、写本が書かれた時期を当てるモデルである。放射性炭素年代測定は、古い物に残った炭素で年代を測る方法である。Enochは、その方法で年代を測った標本24点を見て学習した。年代が分からない写本135点に適用した結果もあわせて発表した。予測誤差は平均27.9年から30.7年の間だった。古文書学者たちがこの予測を検討した結果、79パーセントが妥当だという評価を受けた[3]。

残りの21パーセントで論争が起きた。古文書学者クリストファー・ロールストンは、一つの事例を挙げて反論した。Enochは、ダニエル書を含む4Q114写本を紀元前230年から160年の間と見た。ロールストンは、ダニエル書7章から12章が一つの事件を指すと見た。紀元前167年にアンティオコス4世がエルサレム神殿を汚した出来事である。その本文が事件より前に書かれたはずはない。ところがEnochが出した範囲は、大部分が167年より前である。ロールストンは、Enochの年代が早すぎると指摘した。彼は、Enochは道具箱に入った複数の道具の一つであるべきだと述べた[4]。

この論争が教えることがある。機械は文字の形だけを見る。文章に記された歴史的イベントや文の意味は読めない。だから機械の答えは確定ではなく候補である。候補を絞った後、歴史学者が文脈と照らし合わせて判定しなければならない。どの文字を測るべきかを決めた人も古文書学者だった。アレフを選んだ判断は、古い文字を長く見てきた経験から出た。

学習資料の数も障害である。Enochは年代が分かっている標本24点で学習した[3]。古い遺物は標本を増やしにくい。年代を測るには、遺物を少し切り取って燃やさなければならないためである。少ない標本で学んだモデルは、学んだ範囲を外れると揺らぐ。書記鑑定も事情は似ている。よく出る文字は標本が十分にあるが、まれな文字は数個しかない。標本が少ない場所では、機械の判断は力を失う。

そのため研究チームも、AIを唯一の審判にはしない。革が継がれた場所と本文が分かれた場所、学者たちが積み上げてきた見解がともに置かれる。複数の証拠が一つの場所を指すとき、結論が立つ。イザヤ大巻物で答えが出たのも、そのような方法だった。機械は、人が立てた問いに数値で答えただけである。

次の捜査の標的

イザヤ大巻物の研究は、より大きな計画へと続いた。欧州研究会議は2026年6月23日に、ポポヴィッチへ Advanced Grantを与えることを決めた。研究費は250万ユーロで、期間は5年である[5]。新しい計画の名は Tracing Scribes and Scrollsである[6]。書記と巻物の痕跡をたどるという意味である。

イスラエル考古庁は2026年6月30日に、この計画へ参加すると発表した。ピサ大学、ナポリ・フェデリコ2世大学、南デンマーク大学、ルーヴェン大学も加わる。ベルリンとトリノのエジプト博物館も名を連ねた[6]。

今回の問いは、誰が書いたのかから、どこで作られたのかへ移る。研究チームは死海文書の標本約250点を調査する。革、パピルス、インクの化学成分を測る。その成分を古代エジプトのパピルスの成分と突き合わせる。材料がどこから来たのかが分かれば、写本を作った場所を絞ることができる。AIは、この化学資料を整理し、文字に隠れた規則を探す仕事を担う[7]。ポポヴィッチは、実験室分析と古い文字の研究を組み合わせれば、以前は届かなかった問いに近づけると述べた[6]。

この方法は、イザヤ大巻物にだけ使われるわけではない。死海文書には、まだ筆写者が分からない写本が多く残っている。同じ手が書いた写本を束ねれば、一人の書記の作業目録が作られる。そのような目録が積み重なれば、書記たちが互いにどのようにつながっていたのかを描いてみることができる[1]。二人を見つけ出したこの研究が、その出発点になった。

この研究が残したもう一つの成果は、方法そのものである。写真から汚れを取り除き、画を数値に変え、統計で検定する手順が整理された。他の言語の古い文書にも、同じ手順を適用できる。この本の後の章で出会う粘土板と木簡の研究も、似た道を歩む。

イザヤ大巻物から始まった捜査は、まだ終わっていない。二人の影を見つけ出した道具が、今度は巻物を作った場所を探しに出る。

古い文書を読む仕事は、文字を見分けることで終わらない。その文字を誰がどこで書いたのかまで問わなければならない。イザヤ大巻物は、何人が書いたのかという問いに数値で答えた最初の事例である。巻物は2千年の間、何も語らなかった。人は今、文字の震えからその答えを読む。

参考資料

- [1] Popović, M., Dhali, M. A., & Schomaker, L. (2021). Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsaa). PLOS ONE, 16(4), e0249769. <https://doi.org/10.1371/journal.pone.0249769>
- [2] University of Groningen. (2021, 4月 21日). Cracking the code of the Dead Sea Scrolls. rug.nl. <https://www.rug.nl/news/2021/04/cracking-the-code-of-the-dead-sea-scrolls?lang=en> (2026年9月3日閲覧)
- [3] Popović, M., Dhali, M. A., Schomaker, L., van der Plicht, J., Rasmussen, K. L., La Nasa, J., Degano, I., Colombini, M. P., & Tigchelaar, E. (2025). Dating ancient manuscripts using radiocarbon and AI-based writing style analysis. PLOS ONE, 20(6), e0323185. <https://doi.org/10.1371/journal.pone.0323185>

- [4] Steinmeyer, N. (2025, 6월 20일). Can AI date the Dead Sea Scrolls? Bible History Daily, Biblical Archaeology Society. <https://www.biblicalarchaeology.org/daily/biblical-artifacts/dead-sea-scrolls/can-ai-date-the-dead-sea-scrolls/> (2026年9月3日閲覧)
- [5] University of Groningen. (2026, 6월 23일). Mladen Popović awarded ERC Advanced Grant to trace the origins of the Dead Sea Scrolls. rug.nl. <https://www.rug.nl/about-ug/latest-news/news/archief2026/nieuwsberichten/mladen-popovi-awarded-erc-advanced-grant-to-trace-the-origins-of-the-dead-sea-scrolls?lang=en> (2026年9月3日閲覧)
- [6] The Jerusalem Post. (2026, 6월 30일). New AI-powered research project aims to uncover the origins of the Dead Sea Scrolls. jpost.com. <https://www.jpost.com/archaeology/article-900938> (2026年9月3日閲覧)
- [7] Archaeology News Online Magazine. (2026, 6월). Where the Dead Sea Scrolls come from: AI and chemical analysis trace origins. archaeologymag.com. <https://archaeologymag.com/2026/06/dead-sea-scrolls-origins-studied-with-ai-and-chemical-analysis/> (2026年9月3日閲覧)

第4章. 数万片の粘土・羊皮紙パズルをつなぎ合わせるデジタル接着剤

4.1. 事件ファイル

ヤギを探していて発見した壺

1947年の春、死海の西側の崖で出来事が起きた。ベドウィンの少年たちが逃げたヤギを探していた。少年たちは岩の間に狭い洞窟の入り口を見つけた。洞窟の中には背の高い粘土の壺が立ち並んでいた。壺の中には革に文字が書かれた巻物が入っていた[1]。

これらの巻物が死海文書である。死海文書は紀元前3世紀頃から紀元後1世紀の間に書かれたユダヤ文献である。その後9年間にわたり、近くの洞窟がさらに10カ所見つかった。巻物が出土した洞窟は計11カ所である。ここから出土した文書は900編を超える。ヘブライ語、アラム語、ギリシャ語で書かれている[1]。大部分は羊やヤギの革を薄く伸ばした羊皮紙に書かれていた。パピルスという植物の茎で作った紙に書かれた写本もある[7]。

巻物には、ヘブライ語聖書の写本の中で年代の古いものが含まれていた。共同体の規律を記した文書や暦を記した文書も出土した[1]。乾いた荒野の洞窟が、この文字を2,000年間守り続けた。

洞窟の近くにはクムランという古い集落の跡がある。多くの学者は、ここに暮らしていたユダヤ人共同体が巻物を残したとみている。彼らは聖書を書写し、自分たちの規則を書き留めておいた。ローマ軍が迫ってくると、巻物を洞窟に隠したと推測されている。

第4洞窟の破片の山

洞窟ごとに遺物の状態は大きく異なっていた。第1洞窟の巻物は、壺の中でかなり完全な状態で残っていた。第4洞窟は事情が異なっていた。この洞窟の巻物は壺に入れられないまま床に積まれていた。ここ1カ所からだけでも、巻物約500編の破片数千点が出土した[1]。

洞窟内の湿気が革を徐々に劣化させた。虫や獣が革をかじった。長い歳月が流れるうちに、巻物は細かく碎けた。残されたのは爪ほどの大きさの破片の山だった。

イスラエル考古学庁が保管する破片は25,000点を超える[2]。破片の大部分には、どの巻物から剥がれ落ちたものかが記されていない。破片1点に文字が2、3文字しか残っていないものも珍しくない。学者たちはこれらを大きなガラス板の上に並べ、目で見てつなぎ合わせた。

つなぎ合わせる手がかりはわずかしかなかった。文字の形が似ているかを見た。革の色が似ているかを見た。破損した縁が噛み合うか突き合わせてみた。この方法では一度に2、3点しか比較できない。破片が万単位に増えると、比較すべき組み合わせの数は途方もなく膨れ上がる。

手作業でのパズル合わせには長い時間がかかった。人間の目では数万個の破片の縁を一度に比較することはできない。その上、破片の縁はすでに摩耗し縮んでいた。何十年も学者たちが取り組んでも、元の位置を見つけれない破片が多く残った。

文書を最初に読む権限が少数の学者にしか与えられなかったことも問題だった。外部の研究者は写真すら見ることが難しかった[1]。1991年9月に米国のハンティントン図書館がこのかんぬきを外した。図書館は所蔵していた写真フィルムを無条件で公開した[3]。写真が公開されると、コンピュータで破片をつなぎ合わせようとする試みが始まった。

遺物が複数の国に散逸したことも障害となった。発見の初期、巻物の一部はベツレヘムの骨董商を経て売り払われた[1]。発掘費用を負担した国の機関が破片を分配されることもあった。そのため、1つの巻物から剥がれた破片が別々の都市に置かれることになった。2つの破片がつながるかを確認するには、人が飛行機に乗らなければならなかった。

市場に現れた偽物の破片

死海文書の破片の価格が高騰するにつれ、別の問題が生じた。2002年頃から骨董品市場に死海文書の破片とされる品物が出回った。聖書の句が記された断片が約70点現れた[4]。学者たちはこれらの品物を「ポスト2002破片」と呼んだ。これらの品物には、どの洞窟からいつ出土したかを知らせる記録がなかった[6]。

アメリカ・ワシントンの聖書博物館もこうした破片16点を買って取り展示した。多くの学者が文字の形が不自然だと疑った。博物館は美術品偽造調査会社に精密鑑定を依頼した。調査を率いたのはコレット・ロール調査官である[4]。

2020年3月に結果が出た。16点すべてが現代に作られた偽物だった[4][5]。調査チームは3次元顕微鏡で表面を観察した。素材は本物の死海文書のような薄い羊皮紙ではなかった。少なくとも15点は厚くきめの粗い革だった。インクはすでにひび割れていた隙間に溜まっていた[4]。革がひび割れた後に文字を書いた証拠であった。

この事件は聖書考古学界全体を揺るがした。ジョージ・ワシントン大学のクリストファー・ロールストン教授は、偽物が今も市場に出回り続けていると語る。彼は自分の所持品が本物だと信じている人によく会うが、そのほぼすべてが偽物だと明かした。発掘記録が残っている遺物と市場で購入した遺物の違いがここに現れた[6]。

砕けた破片をつなぎ合わせる 것과偽物をふるい落とすことは、根底が同じである。どちらも人間の目では捉えられない微細な物理的痕跡を読み取らなければならない。人工知能がこの作業に導入された。2026年6月に欧州研究会議が新たな研究費を割り当てた。死海文書の出所を突き止める5年間の研究に250万ユーロを交付した。研究名は「書記と巻物の追跡」である。この研究は素材の成分分析と人工知能、古文字の研究を併用する[7]。

本章で追う出来事は2つある。1つは散逸した破片を再びつなぎ合わせることであり、もう1つは偽物をふるい落とすことである。この2つの作業は、同じ資料を前にして異なる問いを投げかける。前者の問いは「この2つはもともと1つだったのか」であり、後者の問いは「これはもともと古代のものなのか」である。

参考資料

- [1] Israel Antiquities Authority. (n.d.). Discovery and publication. The Leon Levy Dead Sea Scrolls Digital Library. <https://www.deadseascrolls.org.il/learn-about-the-scrolls/discovery-and-publication> (2026年9月3日閲覧)
- [2] Israel Antiquities Authority. (n.d.). The digital library. The Leon Levy Dead Sea Scrolls Digital Library. <https://www.deadseascrolls.org.il/about-the-project/the-digital-library> (2026年9月3日閲覧)
- [3] Deseret News. (1991, 9月 22日). Research library will give scholars access to Dead Sea Scroll photos. Deseret News. <https://www.deseret.com/1991/9/22/18942605/research-library-will-give-scholars-access-to-dead-sea-scroll-photos/> (2026年9月3日閲覧)
- [4] Greshko, M. (2020, 3月 13日). "Dead Sea Scrolls" at the Museum of the Bible are all forgeries. National Geographic. <https://www.nationalgeographic.com/history/article/museum-of-the-bible-dead-sea-scrolls-forgeries> (2026年9月3日閲覧)
- [5] Solly, M. (2020, 3月 16日). All of the Museum of the Bible's Dead Sea Scrolls are fake, report finds. Smithsonian Magazine. <https://www.smithsonianmag.com/smart-news/all-museum-bibles-dead-sea-scrolls-are-fake-report-finds-180974425/> (2026年9月3日閲覧)
- [6] Govier, G. (2026, 4月 8日). New Dead Sea Scrolls exhibit is the real deal. Christianity Today. <https://www.christianitytoday.com/2026/04/new-dead-sea-scrolls-exhibit-is-the-real-deal/> (2026年9月3日閲覧)
- [7] The Jerusalem Post. (2026, 6月 30日). New AI-powered research project aims to uncover the origins of the Dead Sea Scrolls. The Jerusalem Post. <https://www.jpost.com/archaeology/article-900938> (2026年9月3日閲覧)

4.2. 科学捜査隊の追跡1: 幾何学的エッジ親和性と物理的指紋

破れた縁を数値に変える

砕けた破片をつなぎ合わせる最初の手がかりは縁の形状である。紙を手で破ると、鋸の歯のようにギザギザした線ができる。隣り合う2つの破片の線は互いに噛み合う。人間はこの線を目測で比較する。コンピュータはこの線を数値に変換する。

コンピュータは縁をたどりながら曲がり具合を測定する。曲がり具合を順序通りに並べると波型の数列になる。2つの破片の数列を互いに反転させて重ね合わせた際によく合致すれば、結合の候補となる。この手法をエッジ親和性計算と呼ぶ。韓国語ではteduri manmullim jeomsu（縁の噛み合いスコア）である。

このスコアは、韓国語ではteduri manmullim jeomsu（縁の噛み合いスコア）と呼ぶことができる。2つの破片の縁がどれほどよく噛み合うかを示す数値である。

縁だけを見る方法には明らかな限界がある。古い破片の縁はすでに摩耗している。乾燥して縮み、元の形状を失った箇所も多い。そのため研究者たちは、縁の外にある手がかりも併せて観察する。

素材そのものが残した指紋

縁だけでは不十分である。破片の縁はすでに摩耗している場合が多い。そのため研究者たちは、素材そのものに目を向ける。

パピルスは植物の茎を細かく裂き、縦横に重ねて作る。そのため表面に茎の模様が筋のように残る。もともと1枚だった2つの破片は、この筋がつながっている[1]。

イスラエル・ハイファ大学のロイ・アビトボル研究チームがこの筋を読み取った。研究チームは破片の縁の付近を小さな四角形に切り取った。ニューラルネットワークが2つの四角形を見て、筋がつながっているかのスコアを算出した。そのスコアを集約し、破片全体のペアを判定した[1]。

テスト対象は状態の悪い死海文書の破片群であった。正解がわからない破片群であったため、誤った候補も多く出た。それでも、一致するはずのないペアの98パーセントを排除した[1]。人間が確認すべき分量が50分の1に減った。

この結果の意味は明白である。コンピュータがペアを確定したわけではない。コンピュータは見直す必要のないペアを取り除いたのだ。破片合わせにおいて人工知能が担う役割は、大抵こうしたふるい落としである。

フランス・ボルドーのアントワヌ・ピロン研究チームは、ここにシャムネットワークを用いた。シャムネットワークは2枚の画像を並べて入力し、同じ出所であるかを判定する。研究チームはこのモデルの名をPapy-S-Netと名付けた。モデルはパピルスの破片約500点で学習した。文字ではなく紙の筋だけを見てペアを選び出した。実際のデータでペアの79パーセントを的中させた[2]。

羊皮紙にも似たような指紋がある。革には毛が生えていた痕跡である毛穴の跡が残る。同じ動物の革から取られた破片は、この跡の配列がつながる。虫が巻物を貫いて通り抜けた穴も手がかりだ。巻かれていた巻物を広げると、虫食い穴が並んで繰り返される。巻物は内側の層が短く、外側の層が長い。そのため、穴同士の間隔は外側へ行くほど少しずつ広がる。この間隔を測れば、破片が何層目にあったかを推測できる。

こうした物理的指紋を扱うには、まず写真を整理しなければならない。イスラエル考古学庁の死海文書の写真には、定規やカラーチャート、ナンバープレートと一緒に写り込んでいる。背景は黒いためインクと混同しやすい。ブロンソン・ブラウン＝デヴォスト研究チームは、この障害物をあらかじめ除去する4段階の前処理手法を構築した。研究チームは正解ラベル付きのデータセットも併せて公開した[3]。

正解ラベル付きのデータセットを公開することには大きな意義がある。他の研究者が同じデータを用いて自らの手法を試すことができる。2つの手法のうちどちらが優れているかも数値で比較できる。古文書研究は、このように互いに検証し合う分野へと変わりつつある。

粘土板と骨、そして紙へと広がった技術

同様の発想は他の遺物にも波及した。ドイツ・ミュンヘン・ルートヴィヒ・マクシミリアン大学にエンリケ・ヒメネス教授がいる。彼は電子バビロニア図書館（eBL）プロジェクトを率いている。このプロジェクトの破片接合ツールがフラグメンタリウム（Fragmentarium）である[4]。

メソポタミアの人々は湿った粘土に葦のペンで楔形文字を押し付けて書いた。この粘土板は壊れやすい。研究チームは大英博物館とイラク博物館で粘土板の写真を撮影した。2018年から約22,000点の破片をデータとして扱った[5]。コンピュータは破片に残された文字を書き写した文字列同士を比較する。その際、生物学で遺伝子配列を比較する際に用いられる手法を借用した。古代の書記たちが地域ごとに異なる綴りを用いていたためである[4]。

成果は数字に表れた。このプロジェクトは5年間で約1,500件の新たな接合を発見した。それ以前の150年間に学者たちが発見した接合は約5,000件であった[4]。

粘土板には羊皮紙にはない手がかかりがもう1つある。割れた断面の立体形状である。粘土板は大半が天日干しされ、一部だけが火で焼かれた。固まる過程で内部が不均一に硬化する。そのため、割れた面ごとに異なる起伏が生じる。この断面を立体的に測定しておけば、噛み合うペアを見つけ出すことができる。

中国では甲骨の破片に同様の技術を用いている。甲骨は商（殷）の人々が占いをしながら亀の腹甲や牛の骨に刻んだ記録である。河南大学の張重盛研究チームは甲骨接合のデータセットを作成した。研究チームの接合アルゴリズムは上位10位以内の正解率98.39パーセントを達成した[6]。張展研究チームも甲骨破片の写真を用いてDeep Rejoiningモデルを訓練した。この研究チームもまたデータセットを併せて公開した[7]。

上位10位以内の正解率という言葉は、かみ砕いて理解する必要がある。機械は正解を1つだけ提示するのではない。可能性の高い候補10個を順番に提示する。その10個の中に正解が含まれている確率が98.39パーセントなのだ[6]。最終的な判断は人間が行う。

データを1カ所に集約する取り組みも並行して行われた。安陽師範学院は「殷契文淵」という甲骨データプラットフォームを運営している。2023年4月時点で甲骨の写真23万枚余りが登録されている。甲骨文字関連の著作152部と論文約3,000編も併せて収集された[8]。

敦煌莫高窟から出土した仏經の破片も国ごとに散逸している。陳明昆研究チームは2026年に新たな手法を発表した。この手法は人間が正解を与えなくても自律的に学習する。こうした学習方式を自己教師あり学習と呼ぶ。研究チームは完全な写本とペアが判明している破片を共に集めた。ペアが不明な孤立した破片もデータに含めた。コンピュータはこのデータから自ら練習問題を作り出した。研究チームは縁の方向を併せて保持する新たな符号化方式を設計した。その後、シャムネットワークによって2つの破片の類似度をスコアリングした[9]。

死海文書の分野では、欧州とイスラエルの研究者たちが共同でオンライン作業台を構築した。名称はScripta Qumranica Electronicaである。この作業台において研究者は、破片の写真を画面上で移動させたり回転させたりしながら配置する。配置した結果は他の研究者と共有することができる[10]。破片合わせは、1人の頭の中での作業から、複数の人々が共に検討する作業へと変化した。

参考資料

- [1] Abitbol, R., Shimshoni, I., & Ben-Dov, J. (2021). Machine learning based assembly of fragments of ancient papyrus. *Journal on Computing and Cultural Heritage*, 14(3), 1-21. <https://doi.org/10.1145/3460961>
- [2] Pirrone, A., Beurton-Aimar, M., & Journet, N. (2019). Papy-S-Net: A Siamese network to match papyrus fragments. *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing*, 78-83. <https://doi.org/10.1145/3352631.3352646>
- [3] Brown-deVost, B., Kurar-Barakat, B., & Dershowitz, N. (2024). Segmenting Dead Sea Scroll fragments for a scientific image set. *arXiv*. <https://arxiv.org/abs/2406.15692>
- [4] Jiménez, E. (2023, 5月 25日). AI is helping us read ancient Mesopotamian literature. *The Conversation*. <https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091> (2026年9月3日閲覧)

- [5] Phys.org. (2023, 2月 2日). Researcher uses AI to make texts that are thousands of years old readable. Phys.org. <https://phys.org/news/2023-02-ai-texts-thousands-years-readable.html> (2026年9月3日閲覧)
- [6] Zhang, C., Zong, R., Cao, S., Men, Y., & Mo, B. (2020). AI-powered oracle bone inscriptions recognition and fragments rejoining. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 5309-5311. <https://doi.org/10.24963/ijcai.2020/779>
- [7] Zhang, Z., Zhang, H., Guo, A., Jiao, Q., Liu, Y., Li, B., & Gao, F. (2025). Deep rejoining model and dataset of oracle bone fragment images. npj Heritage Science, 13(1), 66. <https://doi.org/10.1038/s40494-025-01651-9>
- [8] 王者. (2023, 4月 8日). "殷契文淵"数据平台为甲骨文研究提供助力. 人民日报. https://news.hangzhou.com.cn/gnxw/content/2023-04/08/content_8509027.htm (2026年9月3日閲覧)
- [9] Chen, M., Xiang, Y., & Zheng, Y. (2026). A self-supervised learning approach for pairwise matching of ancient Dunhuang fragments. Journal on Computing and Cultural Heritage, 19(1), 1-19. <https://doi.org/10.1145/3776642>
- [10] Scripta Qumranica Electronica. (n.d.). Scripta Qumranica Electronica. University of Haifa. <https://megillot.haifa.ac.il/index.php/en/research/scripta-qumranica-electronica> (2026年9月3日閲覧)

4.3. 科学捜査隊の追跡2: 偽作探偵AI

偽物16点を見破った方法

聖書博物館が購入した死海文書の断片16点は、すべて偽物だった[1][2]。調査チームは、目で文字だけを見たのではない。材料とインクを、物理と化学の方法で細かく調べた。この鑑定に人工知能は使われなかった。人が顕微鏡と成分分析で見抜いた[1]。

聖書博物館の鑑定は一度で終わらなかった。2018年10月に、ドイツ連邦材料試験研究所が5点を先に調査した。この5点は古代遺物の特徴と合わず、展示から外された[5]。その後、残りもすべて調査した結果が2020年に出た[1][2]。

第一の証拠は基材だった。本物の死海文書は、薄く伸ばした羊皮紙に書かれた。聖書博物館の断片のうち、少なくとも15点は革だった。この革は羊皮紙より厚く、きめが粗い[1]。

第二の証拠はインクが置かれた場所だった。調査チームは3次元顕微鏡で表面をのぞき込んだ。割れ目の中にインクがたまっていた。裂けた端では、インクが下へ流れ落ちた跡が見えた[1]。

この跡がなぜ決定的なのかは、順序を考えればわかる。本物の写本は、平らな羊皮紙に先に文字を書いた。その後、長い年月の間に革が乾いて割れた。だから本物では、割れ目で文字の筆画が途切れる。偽物は順序が逆だ。偽造者は、すでに割れている古い革を手に入れ、その上にインクを載せた。インクは開いた割れ目の中へ流れ込んだ[1]。

第三の証拠は、表面に塗られた物質だった。調査チームは、断片が琥珀色の液体に浸されていた痕跡を見つけた。この液体は、動物の皮から作った膠とみられる。偽造者は、本物の死海文書の表面にあるつやのある質感をまねようとしたのだ[1]。

三つの証拠は、別々の方向から同じ答えを指していた。一つだけなら偶然かもしれない。三つが重なれば偶然ではない。遺物鑑定で異なる種類の証拠を合わせて集める理由はここにある。

手が残す痕跡

偽物を作る人は、本物の写真を見て文字をなぞって描く。このなぞり描きが痕跡を残す。

人が慣れた文字を書くとき、手は止まらずに流れる。見慣れない文字の形を写すときは違う。筆画を描く途中で止まり、次の筆画の位置を確認する。筆先の速度が落ち、線が細かく震える。この差は、拡大して見ると筆画の外郭の形に現れる。

コンピューターは、こうした差を数値で捉える。2026年6月に、英国ブラッドフォード大学のハサン・ウガイル研究チームが関連研究を発表した。研究チームは、古い素描作品が本物かどうかを判定する人工知能を作った。このモデルは、絵から取り出した五つの数値だけを見る。模様が繰り返される度合いと、明るさの値が散

らばる度合いを測る。明るい部分と暗い部分の差も測る。質感が均一な度合いと、線が小枝のように枝分かれする度合いを加える[3]。

研究チームは、画家10人の真作素描でモデルを訓練した。作品の年代は1510年ごろから1917年までにわたる。資料は、公開資料を提供する美術館6館から集めた。メトロポリタン美術館とアシュモレアン博物館がここに含まれる[3]。

判定900件で試験した結果が論文に掲載された。均衡精度は87.6パーセントだった。真作を真作として受け入れた割合は77.8パーセントだった。他人の作品をその画家のものと誤って受け入れた割合は2.6パーセントだった[3]。

この研究には、注目すべき結果がもう一つある。広く使われる大きなディープラーニングモデルは、この仕事ではかえって成績が悪かった。大きなモデルは、本物の作品を偽物として退けることが多かった。資料が少ないときは、問題に合わせて設計した小さなモデルのほうがよかった[3]。

この結果は古文書研究にも参考になる。残っている古代写本は数が少ない。大きなモデルを十分に学習させるほどの資料がない。何を見るかをあらかじめ決めてやる設計のほうがよいという意味だ。

書体を測る機械

同じ技術は、本物の遺物を研究するときにも使われる。オランダのフローニンゲン大学のムラデン・ポポヴィッチ研究チームは、死海文書の書体を人工知能で分析した。研究チームは、ニューラルネットワークでインクが通った跡だけを抽出した。その後、文字の曲がり具合と傾きを数値に変えた[4]。

研究チームは、炭素年代測定を終えた写本24点でモデルを訓練した。モデルの名前はエノクだ。エノクは、年代のわからない写本135点の年齢を推定した。平均誤差は約28年から31年だった。複数の写本が、これまで知られていたより古いものだとして出た[4]。

この技術は真偽判定にも役立つ。本物の書記の文字は、時代ごとに一定の統計的特徴を持つ。偽物はこの分布から外れる。人の目は二つの文字が似ていると感じるが、機械は差を数値で出す。

ただし、この助けは参考資料を一つ増やす程度のものだ。エノク研究には注意点がある。このモデルは年代を当てるように訓練された。本物と偽物を見分けるように訓練されたわけではない。年代推定が妙に出たからといって、すぐに偽物だとは言えない。機械の結果は、その機械が学んだ仕事の範囲の中でだけ意味を持つ。

機械が代わることのできないもの

機械の判定にも限界がある。聖書博物館の断片の真偽を最終判定したのは、人工知能一つではなかった。顕微鏡観察と化学試験と文献学の知識が合わせて使われた[1][2]。

来歴記録の重みはそのまま残る。来歴記録とは、遺物がどこから出たのかを書いた記録だ。ロールストン教授は、発掘記録のある遺物と市場で買った遺物を分けて見るように言う。発掘で出た巻物には、どの洞窟からいつ出たのかが残っている。市場に出た断片には、その記録がない[5]。

機械は人が見落とす微細な信号を読んてくれる。人はその信号が何を意味するのかを判断する。偽作を見分ける仕事は、この二つがそろって初めて終わる。

参考資料

[1] Greshko, M. (2020, 3月 13日). "Dead Sea Scrolls" at the Museum of the Bible are all forgeries. National Geographic. <https://www.nationalgeographic.com/history/article/museum-of-the-bible-dead-sea-scrolls-forgeries> (2026年9月3日閲覧)

[2] Solly, M. (2020, 3月 16日). All of the Museum of the Bible's Dead Sea Scrolls are fake, report finds. Smithsonian Magazine. <https://www.smithsonianmag.com/smart-news/all-museum-bibles-dead-sea-scrolls-are-fake-report-finds-180974425/> (2026年9月3日閲覧)

- [3] Ugail, H., Ritch-Frel, J., Matuzava, I., & Stork, D. G. (2026). Verification of historical sketches via one-class learning on compact feature representations. PLOS One, 21(6), e0344796. <https://doi.org/10.1371/journal.pone.0344796>
- [4] Popović, M., Dhali, M. A., Schomaker, L., van der Plicht, J., Lund Rasmussen, K., La Nasa, J., Degano, I., Perla Colombini, M., & Tigchelaar, E. (2025). Dating ancient manuscripts using radiocarbon and AI-based writing style analysis. PLOS One, 20(6), e0323185. <https://doi.org/10.1371/journal.pone.0323185>
- [5] Museum of the Bible. (2018, 10월 22일). Museum of the Bible releases research findings on fragments in its Dead Sea Scrolls collection. Museum of the Bible. <https://www.museumofthebible.org/newsroom/museum-of-the-bible-releases-research-findings-on-fragments-in-its-dead-sea-scrolls-collection> (2026年9月3日閲覧)

4.4. 明らかになった真実

確定した接合と候補接合を分ける線

コンピューターが二つの断片はくっつくと言っても、それで終わりではない。機械は確率を出すだけだ。研究者たちは、確定した接合と候補接合を厳格に分ける。

確定した接合は、複数の種類の証拠が同時に合うときだけ認める。まず、折れた面の形が隙間なくかみ合わなければならない。次に、材料の目が続かなければならない。パピルスは茎の模様が続かなければならない[1]。羊皮紙は毛穴の跡が続かなければならない。その次に、境界をまたぐ文字の筆画が自然につながらなければならない。最後に、つなぎ合わせた文章がその言語の文法に合わなければならない。

候補接合は、これらの条件の一部だけが合う場合だ。虫食い穴の間隔から断片の位置を推測した結果がここに入る。インクの成分が同じなので一つの巻物群としてまとめた結果も候補に入る。こうした結果は研究の出発点にはなるが、結論にはならない。コンピューターが選別した候補を人が確認する順序だ[1]。

この区分を守らないと危険だ。確率が高いという理由で二つの断片をくっつけると、存在しなかった文が生まれる。その文が引用されると、誤りは次の研究へ広がる。だから資料記録には、確定と候補を別々に書いておく。

機械が出す数値も、この区分を前提にしている。絞り込み率が高いからといって、接合が確定するわけではない[1]。上位候補の中に正解が入る確率も同じだ[2]。どちらであれ、最後の判定は人が行う。

よみがえった文章

この方法で実際の文献がよみがえった。フラグメンタリウムは2018年から粘土板断片約22,000点を扱った。このプログラムは、新しい接合と新しい写本をいくつも見つけた。研究チームは2023年2月にフラグメンタリウムを一般公開した。『ギルガメシュ叙事詩』のデジタル版も同時に公開した[3]。

よみがえった文章の中に『ギルガメシュ叙事詩』がある。『ギルガメシュ叙事詩』はメソポタミアの英雄物語だ。2022年11月に、プログラムが一つの断片を見分けた。この断片は紀元前130年ごろに写された版から出た。これまで確認されたギルガメシュ写本の中では時期が遅い。その時代まで人々がこの物語を写していたという意味だ[3]。

楔形文字は地域ごとに綴りが違った。書記たちが自分の方言に合わせて文字を書き換えたからだ。人が目で探するとき、この差は大きな壁だった。研究チームは、遺伝子配列を比べる方法を借りてこの壁を越えた。少し違う文字列同士でも比べられるようになった[4]。

死海文書の側でも資料基盤が広がった。イスラエル考古庁は、25,000点を超える断片遺物を高解像度写真で撮影して公開した[5]。研究者たちは、この写真の上で断片を動かしながら配置する[6]。定規やカラーチャートのような邪魔なものを自動で消す処理方法も公開された[7]。

2026年に開いた次の扉

2026年6月に始まった「書記と巻物の追跡」研究が次の扉を開く。この研究が投げかける問いは、断片合わせより一歩先にある。死海文書をどこで誰が書いたのかを問う[8]。

研究チームは、羊皮紙とパピルスとインク試料約250点を調査する。試料とは、検査に使う小さな標本をいう。ここに人工知能による書体分析と古い文字の研究を重ねる。イスラエル考古庁とピサ大学とナポリ大学が協力する。南デンマーク大学とベルギーのルーヴェン大学も参加する。ベルリンとトリノの博物館が所蔵するエジプト・パピルスとも比べる[8]。

試料の成分が同じなら、同じ工房から出た可能性が高くなる。この情報は断片合わせにも使われる。材料の出所が同じもの同士を先にまとめれば、比べる相手の数が減る。

残った断片と残った仕事

成果が積み重なっても、まだ本来の場所を見つけていない断片は多い。考古庁の保管箱には、どの巻物から出たのかわからない断片が残っている[8]。粘土板の側にも、まだ見つかっていない接合が多い[4]。

装置の違いも問題として残っている。立体スキャナーで遺物を測るには、遺物を装置の前へ運ばなければならない。遺物を外へ出しにくい博物館も多い。だから、ある遺物には写真資料だけがあり、立体資料がない。こうした遺物には、折れた面を比べる方法を使えない。

技術が残した課題もある。断片写真は国ごとに違う装置で撮られた。照明と背景が違えば、同じアルゴリズムでも違う答えを出す。そこで研究者たちは、正解が付いた公開資料セットを作った。互いの結果を同じ物差しで比べるためだ[7]。敦煌断片研究も、人が正解を付けなくても学ぶ方向へ向かった[9]。

この技術が人を押しのけるわけではない。機械は比べる相手を減らし、人は意味を判断する。よみがえった文がその時代の言い回しに合うかどうかは、人が読まなければわからない。二つの役割は重ならないから、一緒に働ける。

公開のあり方も変わった。以前は少数の学者だけが原本写真を見た。今は粘土板資料と死海文書の写真がインターネットで開かれている[4][5]。学生も画面上で断片を動かしてみることができる[6]。

デジタル接着剤は断片を物理的にくっつけない。遺物は箱の中にそのままある。コンピューター画面の上でだけ、巻物が再びつながる。だから判定が間違っていれば、いつでも戻せる。2,000年の間散らばっていた文字が、こうして少しずつ文へ戻っている。

参考資料

- [1] Abitbol, R., Shimshoni, I., & Ben-Dov, J. (2021). Machine learning based assembly of fragments of ancient papyrus. *Journal on Computing and Cultural Heritage*, 14(3), 1-21. <https://doi.org/10.1145/3460961>
- [2] Zhang, C., Zong, R., Cao, S., Men, Y., & Mo, B. (2020). AI-powered oracle bone inscriptions recognition and fragments rejoining. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 5309-5311. <https://doi.org/10.24963/ijcai.2020/779>
- [3] Phys.org. (2023, 2月 2日). Researcher uses AI to make texts that are thousands of years old readable. *Phys.org*. <https://phys.org/news/2023-02-ai-texts-thousands-years-readable.html> (2026年9月3日閲覧)
- [4] Jiménez, E. (2023, 5月 25日). AI is helping us read ancient Mesopotamian literature. *The Conversation*. <https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091> (2026年9月3日閲覧)
- [5] Israel Antiquities Authority. (n.d.). The digital library. The Leon Levy Dead Sea Scrolls Digital Library. <https://www.deadseascrolls.org.il/about-the-project/the-digital-library> (2026年9月3日閲覧)
- [6] Scripta Qumranica Electronica. (n.d.). Scripta Qumranica Electronica. University of Haifa. <https://megillot.haifa.ac.il/index.php/en/research/scripta-qumranica-electronica> (2026年9月3日閲覧)
- [7] Brown-deVost, B., Kurar-Barakat, B., & Dershowitz, N. (2024). Segmenting Dead Sea Scroll fragments for a scientific image set. *arXiv*. <https://arxiv.org/abs/2406.15692>
- [8] The Jerusalem Post. (2026, 6月 30日). New AI-powered research project aims to uncover the origins of the Dead Sea Scrolls. *The Jerusalem Post*. <https://www.jpost.com/archaeology/article-900938> (2026年9月3日閲覧)
- [9] Chen, M., Xiang, Y., & Zheng, Y. (2026). A self-supervised learning approach for pairwise matching of ancient Dunhuang fragments. *Journal on Computing and Cultural Heritage*, 19(1), 1-19. <https://doi.org/10.1145/3776642>

第5章. 粘土板の折れた断面を合わせる： ギルガメシュ叙事詩の計算機的復元

5.1. 事件ファイル

4千年を越えてきた物語

ギルガメシュ叙事詩は、現存する物語文学の中でも古い部類に入る。これより古いシュメールの歌も伝わっている。主人公はメソポタミアの都市ウルクを治めた王ギルガメシュだ[1]。彼は友人エンキドゥとともに遠い杉の森へ行く。二人はその森を守る怪物フンババと戦う[2]。後にエンキドゥが死ぬと、ギルガメシュは死なない道を探して旅立つ[1]。

この物語を書き記した土台は紙ではなく粘土板だ。古代の書記たちは葦を斜めに削り、尖った棒を作った。この棒を尖筆と呼ぶ。尖筆で湿った粘土を押して刻むと、くさび形の跡が残る。線を引くのではなく、押して作った文字だ。粘土が乾くと、その跡はそのまま固まる。そのため、この文字を楔形文字と呼ぶ。書記になるうとする学生たちは、文字の練習としてこの叙事詩を書き写した。そのおかげで写本の断片が中東の各地の遺跡から出てくる。エジプトのアマルナ遺跡からも、この叙事詩の断片が出た。メソポタミアから遠く離れた場所にまで物語が広がっていたという意味だ[1]。

伝わっている版は大きく二つある。一つは紀元前2千年紀初めに書かれた古バビロニア版だ。もう一つは後に整理された標準バビロニア版だ。標準版は十二枚の粘土板で構成されている。そのうち十一枚が本編だ。十二枚目の板は後に加わった付録と見なされる。今日私たちが読む翻訳本は、この標準版を基礎にしている。空白部分は古バビロニア版と別の断片で埋める。標準版を整えた人物は、シン・レキ・ウンニンニという書記として知られている。彼は散らばっていた古い歌を一つの長い作品に編んだ。板ごとに収める内容を決め、行数を合わせた。だが、この標準版にも空白がとても多い。最初から最後まで完全に残った写本は一つもない[1]。

物語の内容は、今読んでなじみがないものではない。ギルガメシュは力が強いが、民を苦しめる王だった。神々は彼の相手となる者としてエンキドゥを作って送る。二人は戦った末に友人になる。ともにフンババを倒し、天の牡牛も打ち倒す。その代償としてエンキドゥが病にかかり死ぬ。友人の死を見たギルガメシュは、自分も死ぬという事実気づく。第11板には、大洪水から生き残った人を訪ねて話を聞く場面が入っている[1]。

この物語は2千年以上忘れられていた。楔形文字を読める人がいなくなったためだ。19世紀に学者たちが楔形文字を再び解読し、叙事詩は世に戻った。ニネヴェから運ばれてきた粘土板の箱を調べていた研究者が洪水の物語に気づいたことが始まりだった。その後150年にわたり、学者たちは散らばった写本を集め、本文を少しずつ埋めてきた[1]。

ニネヴェと紀元前612年の火

標準版の写本は、ほとんどが一つの古代図書館から出た。紀元前7世紀のアッシリア王アッシュルバニパルがニネヴェ王宮に設けた蔵書庫だ。この図書館は占いの書、医学書、辞書、文学作品を幅広く集めた。王は帝国の各地から粘土板を集めさせた。臣下に手紙を送り、足りない本を探して持ってくるよう命じた。アッシュルバニパルは楔形文字を自ら読める王だった[3]。

ニネヴェは紀元前612年ごろ、大火に包まれた。王宮が崩れ、図書館もともに焼けた。この火は遺物に二つのことを同時にもたらした。日陰で乾かしただけだった粘土板が、炎の中で硬く焼かれた[4]。焼かれた粘土板は、窯から出たレンガのように長く耐える体を得た。同じ炎の中にパピルスや革があったなら、灰になって消えていただろう。

しかし、崩れる建物の重みは、その硬い板を粉々に砕いた。現在、大英博物館が保管するニネヴェ出土の粘土板と断片は3万点を超える[3][4]。19世紀半ば、発掘隊が土の山から集め、船に積んで運んできた遺物だ[3]。一つの物語が数十個の断片に散った状態で、私たちのところへ来た。

断片の大きさはまちまちだ。手のひらほどのものもあれば、親指の爪ほどのものもある。一つの断片に残った文字は、数行にすぎない場合が多い。その数行が叙事詩十二板のどこに入るのか、誰も教えてくれない。文字を読み取る作業と位置を探す作業は別の仕事だ。

断片の一つが本来の場所を見つけると、意味が変わることもある。前後がつながって初めて、誰の言葉なのか、どの場面なのかがわかる。離れている時には意味のわからない単語の羅列だった文字が文になる。一つの接合が物語の一場面を丸ごとよみがえらせることもある。

35万行を手で合わせることはできない

壊れた断片を合わせる作業は、長い間、人の手と記憶に任されていた。学者は収蔵庫で断片を一つ手に取る。そして別の断片の写真や手写し図を思い浮かべながら相手を探す。二つの断片が同じ板から出たと判断すれば、それを接合と呼ぶ。

この方法は人の記憶容量で行き詰まる。アッシリア学が始まってから150年の間に、このようにつなぎ合わせた接合は約5,000件だ[5]。残る断片の数はそれよりはるかに多い。大英博物館はニネヴェから出た粘土板だけで3万点以上を所蔵している[3]。その他の所蔵品まで加えると、数はさらに増える。そのうち多くは古代以降、一度も読まれたことがない[5]。世界の博物館に残る楔形文字粘土板をすべて数えると、50万点ほどと見られる[6]。

作業が難しい理由がもう一つある。断片が複数の国に散らばっている。大英博物館、ベルリン、イエール大学、フィラデルフィアのペン博物館が、それぞれ所蔵品を持っている[7]。二つの断片を実際に合わせてみるには、二つの大陸を行き来しなければならない。

楔形文字そのものにも手間がかかる。同じ単語を書記ごとに少しずつ違って記した。地域と時代によって綴りが揺れる[5]。人が目で二つの断片の文字を比べても、同じ作品なのか気づきにくい。

さらに、楔形記号一つが複数の音で読まれる。同じ跡が文脈によって別の値を持つ。このような文字を数十万行並べて相手を探す作業は、人の目だけでは終わられない。

そこで、ドイツ・ミュンヘンのルートヴィヒ・マクシミリアン大学にいるエンリケ・ヒメネス教授は別の道を選んだ。2018年、彼はバビロニア文献をできるだけ完全によみがえらせる事業を始めた。世界に散らばった粘土板の読解文を一か所に集める作業だ[5]。この事業の名前が電子バビロニア文献図書館、略してeBLだ。eBLが作った断片合わせの作業室がFragmentariumである。2024年までに、この作業室には約2万5千点の粘土板読解文が蓄積された。行数で数えると35万行を超える[7]。

この作業が急がれる理由は時間にある。粘土板は火で焼かれていても永遠ではない。塩分と湿気が表面を膨らませ、文字を消す。博物館の収蔵庫で読まれるのを待つ粘土板は、今も数十万点ある[5]。その中にどれほど多くの物語が入っているのかは、誰にもわからない。

楔形文字を読む人の数も問題だ。世界にアッシリア学の研究者は多くない。残る粘土板のうち、翻訳されたものは一部にとどまる[6]。一人が一生で読める粘土板には限界がある。機械の助けを探した理由はここにある。

参考資料

- [1] George, A. R. (2003). The Babylonian Gilgamesh Epic: Introduction, Critical Edition and Cuneiform Texts. Oxford University Press.
- [2] Al-Rawi, F. N. H., & George, A. R. (2014). Back to the Cedar Forest: The beginning and end of Tablet V of the Standard Babylonian Epic of Gilgamesh. Journal of Cuneiform Studies, 66, 69-90.
<https://doi.org/10.5615/jcunestud.66.2014.0069>
- [3] The Ashurbanipal Library Project. (n.d.). What is the Library? Oracc, University of Pennsylvania.
<https://oracc.museum.upenn.edu/asbp/whatisthelibrary/index.html> (2026年9月3日閲覧)
- [4] British Museum. (n.d.). What was Ashurbanipal's Library? britishmuseum.org.
<https://www.britishmuseum.org/research/projects/what-was-ashurbanipals-library> (2026年9月3日閲覧)

- [5] Jiménez, E. (2023, 5월 25일). AI is helping us read ancient Mesopotamian literature. The Conversation. <https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091> (2026年9月3日閲覧)
- [6] Wang, Z. (2026). TabletCraft: Bridging a 4,000-year cultural gap with bidirectional Akkadian NMT and cuneiform rendering. arXiv:2608.02609. <https://arxiv.org/abs/2608.02609> (2026年9月3日閲覧)
- [7] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Hättinen, A., Heinrich, A., Mitto, T., Rozzi, G., Sáenz, L., & Jiménez, E. (2024). Transliterated cuneiform tablets of the electronic Babylonian Library platform. Journal of Open Humanities Data, 10, 19. <https://doi.org/10.5334/johd.148>

5.2. 科学捜査隊の追跡: クレイエッジ・デジタル融合とFragmentarium

文字をアルファベットに移し書きする最初の段階

Fragmentariumの出発点は文字だ。研究者は粘土板の写真を見て、楔形文字をローマ字に移し書きする[1]。この作業を音訳という。日本語で言えば、漢字を仮名の音に移し書きする作業に近い。楔形記号一つ一つを音価に合わせてアルファベットに変える作業だ。壊れて読めない箇所は空白記号で表示しておく。

こうして作った音訳文は、コンピューターが扱える文字列になる。eBLは2018年5月から2023年8月まで、この音訳文を蓄積した。その結果、約2万5千点の粘土板、35万行を超える文字が検索可能な資料になった。eBLが管理する粘土板記録の一覧は26万件を超える。大英博物館、イェール・バビロニア・コレクション、ペン博物館の所蔵品がここに含まれる[1]。

音訳は人が行う。機械が写真から文字を完全に取り出すことはできないためだ。eBL研究陣は5年を超えてこの音訳作業を続けた。その成果物は誰でもダウンロードできる公開資料として出された[1]。機械に仕事をさせるには、まず人が資料を作ってやらなければならない。

遺伝子検査に似た文字列合わせ

中核アルゴリズムは意外な分野から来た。生物学者が遺伝子配列を比較するときに使う文字列整列の方法だ[2]。遺伝子はA、T、G、Cの四文字が長く続く鎖だ。二つの生物の遺伝子を比べるときは、文字が少し違っていても同じ位置に合わせて見る。

楔形文字でも同じ問題が起きる。書記ごとに同じ単語を違って書いたためだ。ヒメネス教授は、この揺れが遺伝子の突然変異に似ていると考えた。そこで生物学の整列アルゴリズムを楔形文字に合うように直して使った[2]。

コンピューターが実際に数えるのは短い文字片だ。音訳文を数文字ずつ切り出した片をエングラムと呼ぶ。システムは二つの粘土板が共有するエングラムの数を数える。長い片を共有する組み合わせほど高い点数を得る。点数の高い組み合わせが人の前に候補として上がってくる。

Fragmentariumはウェブ画面で開く。世界のどの国の研究者でも接続して同じ資料を見る。ロンドンの断片とミュンヘンの断片を一つの画面に並べて比較できる[8]。荷物をまとめて飛行機に乗る必要がなくなった。

この方法は綴りが揺れても耐える。一つの単語が違って書かれていても、周囲の片が重なれば点数は残る。eBLはこのマッチングプログラムを公開リポジトリに上げている。別の研究者も同じ道具をダウンロードして使える[6]。

コンピューターは文字を読むこと自体も助ける。楔形記号一つが複数の音で読まれることは多い。eBLが学習させたモデルは、この複数の意味の中から正しいものを選ぶ問題で98パーセントの精度を出した[2]。

この作業がなぜ必要なのかは、例を挙げるとわかりやすい。同じ楔形の跡を、ある文では一つの音として読む。別の文では単語全体を意味する記号として読む。読みを誤ると、マッチングの点数が乱れる。読みを正す作業が、そのまま接合の精度を高める作業になる。

成果は数字に表れた。アッシリア学が150年間に手をつなぎ合わせた接合は約5,000件だ。eBLは5年で約1,500件の接合をさらに見つけた[2]。音訳文資料集を作る過程だけで、約1,250件の接合が明らかになった[1]。直接接してはいないが同じ作品に属する断片も数千点確認された[2]。

壊れた箇所を埋める言語モデル

断片をつなぎ合わせても、文字はなお所々で空いている。この空白部分をラクナと呼ぶ。ここからは言語モデルが登場する。

2023年、テルアビブ大学とアリエル大学の研究陣がアッカド語翻訳モデルを発表した。このモデルは楔形文字の符号を入れても、アルファベット音訳文を入れても英語に移す。二種類の入力をどちらも扱ったのだ。英語をアッカド語に戻す作業は扱わなかった。学習に使った資料は、すでに出版された粘土板読解文だ[3]。

2026年には反対方向まで扱う道具が出た。TabletCraftという名前の公開ソフトウェアだ。この道具はアッカド語から英語へ、英語からアッカド語へ、双方向の翻訳を行う。学習資料は11万6千対の文だ。楔形記号変換表は1万4千個を超える対応を含んだ。開発者は、残る50万点の粘土板を読む道具が必要だと述べた[4]。

このようなモデルは空白の前後の文脈を見て、入りそうな単語を確率で出す。バビロニアの詩には、同じ構造の二行を並べる規則がある。モデルはこのような形式も手がかりにする。モデルが出すものは確定した答えではなく、候補の一覧だ。

筆跡まで読む2026年の道具

2026年5月18日、ドイツのヴュルツブルク大学が新しい道具を公開した。マインツ学術院とドルトムント工科大学が共同で作った。名前はPaleographicumだ。この道具は粘土板の写真から楔形記号一つ一つの形を学ぶ。そして同じ方式で押して書かれた記号を、所蔵品全体から探してくれる[5]。

この道具が扱う写真は7万枚で、そこに含まれる記号は500万個を超える。5つの断片の筆跡を比較する作業には、以前なら3日かかった。この道具を使えば5分で終わる。人によって筆跡が違いうように、書記ごとに楔を押す角度と間隔が違う。その差を手がかりに、同じ板から出た断片を選び出すことができる[5]。

この道具が扱うヒッタイト楔形文字の記号は375個である。これまでに知られているヒッタイト粘土板断片は3万点に達する。研究を率いたダニエル・シュヴェーマー教授は、この道具で数千時間を節約できると述べた[5]。

研究チームの次の目標は、書記個人を識別することだ。同じ人物が書いた粘土板をすべて探し出すという意味である。それが実現すれば、1人の書記がどこで働き、どこへ移っていったのかを描くことができる。3,500年前の人々の職業生活が資料として姿を現すのである[5]。

折れた粘土の縁を合わせる別の道

文字ではなく形で断片を合わせようとする研究も長く続いてきた。割れた粘土板には折れた面が残る。この面を破断面という。2つの断片の破断面が互いにかみ合えば、同じ板から出たものだ。

この道を進んできた代表的な事業が、仮想楔形文字粘土板復元事業である。英国キール大学のサンドラ・ウーリーがこの事業を率いる。ティム・コリンズとエルレント・ゲルケンが加わっている。研究チームは断片を3次元スキャナーで走査する。そして2つの断片の折れた面がかみ合うかどうかをコンピューターに判断させる。古い断片は縁がすり減っている。この摩耗まで計算に入れる[9]。

この方法の成果の一つは、大洪水物語から出た。アトラハシスという洪水叙事詩の第3板断片である。一つの断片はロンドンに、もう一つの断片はジュネーブにあった。両地は1,000キロメートルほど離れている。実物を突き合わせることができず、接合を確認できなかった組み合わせだった。3次元資料を重ねてみると、2つの断片はつながった[9]。この方法は2014年に論文として発表された[10]。

3次元資料を扱う研究はいまも続いている。2026年3月に一つの論文が公開された。粘土板の3次元点群資料で遺物の情報を分類するニューラルネットワークを扱ったものだ。点群とは、物体表面を多数の点の座標として記録した資料である。3次元スキャナーで粘土板を走査すると、このような資料が得られる。表面の起伏と厚さが数字として残る。この研究は、先行するトランスフォーマー基盤モデルより良い成績を出した[7]。

ただし、現在Fragmentariumが実際に使う主な武器は3次元形状ではない。写真と翻字文、つまり文字と画像である[1][2]。破断面合わせは断片を一つずつスキャンしなければならず、手間が大きい。二つの道は互いを代替せず、並んで進む。

写真から記号を見つけ出す作業も急速に進んだ。2026年6月にeBL研究チームが新しい論文を出した。粘土板写真から楔記号の位置を自動で見つける方法である。対象となった断片写真は8万7千枚を超える。見つけ出した記号は290万個に達する。先行する方法より成績が大きく上がった[11]。

参考資料

- [1] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Härtinen, A., Heinrich, A., Mitto, T., Rozzi, G., Sáenz, L., & Jiménez, E. (2024). Transliterated cuneiform tablets of the electronic Babylonian Library platform. Journal of Open Humanities Data, 10, 19. <https://doi.org/10.5334/johd.148>
- [2] Jiménez, E. (2023, 5月 25日). AI is helping us read ancient Mesopotamian literature. The Conversation. <https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091> (2026年9月3日閲覧)
- [3] Gutherz, G., Gordin, S., Sáenz, L., Levy, O., & Berant, J. (2023). Translating Akkadian to English with neural machine translation. PNAS Nexus, 2(5), pgad096. <https://doi.org/10.1093/pnasnexus/pgad096>
- [4] Wang, Z. (2026). TabletCraft: Bridging a 4,000-year cultural gap with bidirectional Akkadian NMT and cuneiform rendering. arXiv:2608.02609. <https://arxiv.org/abs/2608.02609> (2026年9月3日閲覧)
- [5] Julius-Maximilians-Universität Würzburg. (2026, 5月 18日). AI reads cuneiform: A milestone for Ancient Near Eastern Studies. EurekAlert!. <https://www.eurekalert.org/news-releases/1128630> (2026年9月3日閲覧)
- [6] Electronic Babylonian Literature. (n.d.). ngram-matcher [ソフトウェア]. GitHub. <https://github.com/ElectronicBabylonianLiterature/ngram-matcher> (2026年9月3日閲覧)
- [7] Hagelskjær, F. (2026). A novel network for classification of cuneiform tablet metadata. arXiv:2603.03892. <https://arxiv.org/abs/2603.03892> (2026年9月3日閲覧)
- [8] electronic Babylonian Library. (n.d.). Fragmentarium. ebl.lmu.de. <https://www.ebl.lmu.de/fragmentarium> (2026年9月3日閲覧)
- [9] International Association for Assyriology. (2022, 10月 22日). In the spotlight: The "Virtual Cuneiform Tablet Reconstruction Project". iaassyriology.com. <https://iaassyriology.com/in-the-spotlight-the-vctr-project/> (2026年9月3日閲覧)
- [10] Collins, T., Woolley, S. I., Hernandez Munoz, L., Lewis, A., Ch'ng, E., & Gehlken, E. (2014). Computer-assisted reconstruction of virtual fragmented cuneiform tablets. In Proceedings of the 2014 International Conference on Virtual Systems and Multimedia (pp. 70-77). IEEE. <https://doi.org/10.1109/VSMM.2014.7136691>
- [11] Che, W., Garcés Arias, E., Niaz, A., Bender, A., & Jiménez, E. (2026). Automated sign detection across the Electronic Babylonian Library: A large-scale dataset and end-to-end cuneiform OCR pipeline. arXiv:2606.22608. <https://arxiv.org/abs/2606.22608> (2026年9月3日閲覧)

5.3. 明らかになった真実

ギルガメシュ叙事詩に戻ってきた行

Fragmentariumはギルガメシュ叙事詩の断片20点を新たに発見してくれた。そのおかげで100行を超える本文が本来の位置に戻った。失われていた場面もいくつかよみがえった。エンキドゥがギルガメシュにフンババを殺さないよう止める場面がその一つだ。二人がフンババを殺したあと、ニップルへ向かう場面も新たに繋がった。ニップルは神エンリルをまつていた宗教の中心都市である[1]。

この方法がどれほど速いかを示した事例が2025年に出た。バグダード大学のアンマル・ファディルとヒメネス教授が、バビロンを称賛する長い詩をよみがえらせた。この詩は1,000年以上忘れられていた。研究チームはeBLプラットフォームで、同じ詩に属する写本30点を見つけた。手作業なら数十年かかった仕事である。散らばった写本を集めると、250行に及ぶ詩の大部分が再びつながった[3]。

この詩はバビロンという都市の一日を見せてくれる。川の水、城門、人々の営みが出てくる。女司祭たちが担っていた仕事に関する新しい記録も含まれている。写本が30点も出たという事実は、この詩が広く読まれていたことを意味する。学校で学生たちが書き写していた文章だった可能性が高い[3]。

この作業にかかった時間が方法の力を示している。写本30点を手で探すには、世界各地の博物館の目録を一つずつ調べなければならない。eBLプラットフォームでは、同じ作業が1回の検索で終わる[3]。

この方法はギルガメシュ叙事詩だけに使われるわけではない。バビロニアの占術書、医学書、辞書、祈祷文にも同じ方法が入っていく。eBLが見つけた接合は、複数の種類の文章にまたがっている[5]。

人の手が先に解いた杉の森

機械がすべてを解いたわけではない。ギルガメシュ叙事詩で広く知られる新発見は、人の手から出た。

2011年にイラクのスレイマニヤ博物館が粘土板1点を購入した。ファルーク・アルラウィ（Al-Rawi）とアンドリュー・ジョージがこの粘土板を読んだ。二人は2014年に結果を発表した。板には標準版第5板の前半と末尾が含まれていた。ここから約20行の新しい本文が出た[4]。

新しい本文は杉の森を音で描く。猿たちが騒がしく鳴き、蝉が合唱する。あらゆる鳥がさえずる。これらの音が、森を守るフンババを毎日楽しませていた。森を初めて見た二人が驚いて立ち止まる場面も出てくる。フンババを殺したあと、ギルガメシュが木を切ったことを後悔する一節もある[4]。メソポタミアの平原には猿の住む森がない。そのため、この描写がどこから来たのかが学者たちの問いとして残った。

この粘土板が出てきた経路はきれいではなかった。博物館は遺物商からこれを買入れた。戦争後のイラクで盗掘された遺物が散逸するのを防ぐための措置だった。どの遺跡から出たのかという記録はない[4]。このような遺物は、断片の相手を探す手がかりがさらに少ない。

この発見は人工知能がなくても出た。しかし、今後同じ種類の発見をどれほど頻繁にできるかは道具にかかっている。スレイマニヤ板が世に知られるまでには偶然と人の目が必要だった。eBLはその偶然を検索に変えようとしている[5]。

丸い箱舟の物語

大洪水物語も事情は似ている。ギルガメシュ叙事詩第5板が杉の森を含んでいたなら、第11板は大洪水を含んでいた[6]。下で見る粘土板はギルガメシュ叙事詩の写本ではない。同じ洪水伝承を含む、より古い板である[8]。

大英博物館のアーヴィング・フィンケル博士は2009年に一つの粘土板を調べた。個人収集家が博物館へ持ち込んだ板である。彼は紀元前1750年ごろに書かれたものと見た。大きさは横11.5センチメートル、縦6センチメートルほどである。ここに60行が書かれている[8]。内容は、神が一人の人物に船を作れと指示する物語である。動物を二匹ずつ乗せよという言葉も出てくる[7]。

目を引くのは船の形である。この船は細長い箱ではなく丸い船だ。川で使われていたコラクルという丸い籠船と同じ形である[7]。粘土板には縄と木杵が記されている。水が漏れないように塗る瀝青の量も出てくる。瀝青は黒く粘る天然の油である。フィンケル博士はこの解読を2014年に本として刊行した[8]。

この粘土板は人工知能が読んだものではない。人が一行ずつ読み解いた。人工知能の役割は別のところにある。世界各地に散らばった洪水物語の写本を一か所に集め、互いに重なる行を見つけてくれることだ[5]。

機械がすることと人がすること

この分野は結果を三つの層に分けて扱う。

第一の層は物理的接合である。2つの断片の折れた面が実際にかみ合う場合だ。このとき、2つの断片は一つの粘土板から出たと確定される。

第二の層は文献的接合である。2つの断片が直接接してはいないが、同じ作品の異なる部分であることが明らかになる場合だ。Fragmentariumが見つけてくれるものの大半はここに入る[5]。同じ作品に属する断片が、このように数千点確認された[2]。

この二つの層は性格が違う。物理的接合は目で確認されれば、そのまま事実になる。文献的接合は確率に頼るため、新しい断片が出れば変わることがある。eBLは二つの層を分けて記録する[5]。

第三の層は解釈である。よみがえった本文が何を意味するのかを考える仕事だ。杉の森の猿の声がどこから来た記憶なのか、機械にはわからない。丸い船がどのように聖書の四角い箱舟になったのかも、機械は答えられない。この問いは、古い文字を研究する学者と歴史学者の役割である。

機械が出した候補は、人の検討を経て初めて正式な記録になる。学者は候補リストを開き、2つの断片の写真を並べて見る。文字の形と行間、板の厚さまで調べたうえで接合を確定する[5]。機械がする仕事は、数多い組み合わせの中から見る価値のあるいくつかを選んでくれるところまでだ[2]。

2026年6月に出たある整理記事も同じ線を引いている。現在の人工知能は楔形文字を自力で解読できない。人の研究を助ける道具である[9]。ヴュルツブルク大学のPaleographicumも同じだ。3日かかっていた筆跡比較を5分に縮めてくれるだけである。何を比較するかを決め、結果を判断するのは人である[10]。

変わったのは速度である。150年間で5,000件だった接合が、5年で1,500件増えた[2]。ギルガメシュが探し求めた永遠の命は、結局、彼の物語が得た。土の中で砕けたまま2,600年を待った文章が、いま一つずつ本来の位置に戻っている。

残った仕事はまだ多い。世界の粘土板50万点のうち、読まれたものは一部である[11]。eBLの翻字資料に入った粘土板は2万5千点である[5]。ヒッタイト断片3万点は、いま写真で整理され始めたばかりだ[10]。折れた面を合わせる3次元作業はいま始まった段階である。次の世代の学生がこのリストを引き継ぐだろう。そのとき、ギルガメシュ叙事詩の空白は今より減っているはずだ。ニネヴェの火が残した断片は、まだすべて読まれていない。読み解くための道具は年ごとに増えている。

参考資料

- [1] Ofgang, E. (2024, 8월 12일). Piecing together an ancient epic was slow work. Until AI got involved. The New York Times. https://edition.pagesuite.com/tribune/article_popover.aspx?guid=235bb893-7f2d-41cd-9857-139c7d1e8355 (2026年9月3日閲覧)
- [2] Jiménez, E. (2023, 5월 25일). AI is helping us read ancient Mesopotamian literature. The Conversation. <https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091> (2026年9月3日閲覧)
- [3] Fadhil, A. A., & Jiménez, E. (2025). Literary texts from the Sippar Library V: A hymn in praise of Babylon and the Babylonians. Iraq, 1-58. <https://doi.org/10.1017/irq.2024.23>
- [4] Al-Rawi, F. N. H., & George, A. R. (2014). Back to the Cedar Forest: The beginning and end of Tablet V of the Standard Babylonian Epic of Gilgamesh. Journal of Cuneiform Studies, 66, 69-90. <https://doi.org/10.5615/jcunestud.66.2014.0069>
- [5] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Härtinen, A., Heinrich, A., Mitto, T., Rozzi, G., Sáenz, L., & Jiménez, E. (2024). Transliterated cuneiform tablets of the electronic Babylonian Library platform. Journal of Open Humanities Data, 10, 19. <https://doi.org/10.5334/johd.148>
- [6] George, A. R. (2003). The Babylonian Gilgamesh Epic: Introduction, Critical Edition and Cuneiform Texts. Oxford University Press.
- [7] PBS NewsHour. (2014, 1월 24일). New evidence suggests Noah's ark may have been round. pbs.org. <https://www.pbs.org/newshour/world/new-evidence-suggests-noahs-ark-may-have-been-round> (2026年9月3日閲覧)
- [8] Finkel, I. (2014). The Ark Before Noah: Decoding the Story of the Flood. Hodder & Stoughton.
- [9] Centre for the Study of Manuscript Cultures. (2026, 6월 14일). Can artificial intelligence decipher cuneiform tablets? Universität Hamburg. <https://www.csmc.uni-hamburg.de/publications/mesopotamia/2026-06-14.html> (2026年9月3日閲覧)
- [10] Julius-Maximilians-Universität Würzburg. (2026, 5월 18일). AI reads cuneiform: A milestone for Ancient Near Eastern Studies. EurekAlert!. <https://www.eurekalert.org/news-releases/1128630> (2026年9月3日閲覧)
- [11] Wang, Z. (2026). TabletCraft: Bridging a 4,000-year cultural gap with bidirectional Akkadian NMT and cuneiform rendering. arXiv:2608.02609. <https://arxiv.org/abs/2608.02609> (2026年9月3日閲覧)

第6章. 古代メソポタミアの領収書と商人たちのささやき： AI翻訳

6.1. 事件ファイル

粘土に刻んだ領収書

古代メソポタミアの人々は、湿った粘土板に文字を書いた。葦の先を角ばるように削り、鉄筆を作った。鉄筆はインクを使わず、跡だけを残す筆記具である。この鉄筆を粘土に押しつけ、すぐに離した[1]。押された跡はくさびに似た三角形になった。この文字の名前はくさび文字である。学界では楔形文字という名前も併用する。楔形とは、くさびの形という意味である。

粘土は乾くと石のように固まる。紙や皮と違い、時間が過ぎても腐らない。火事になっても灰にならない。むしろ熱が粘土をレンガのように固く焼き締める。戦争で都市が燃える時、粘土板はさらに丈夫になった[3]。そのおかげで、4千年前の帳簿が今日まで残った。

粘土板には手のひらほどの大きさのものが多く、ポケットに入れて持ち歩ける大きさである。書記は湿った粘土をこねて板を作った。書記とは、文字を代わりに書き、読んでくれる人である。文字を書き終わると、日に干した。わざわざ窯で焼くことはまれだった。現在、博物館で焼くのは保存のための措置である[2]。直す必要が生じると、表面をこすって消した。

これまで発掘された粘土板は約50万枚と推定される[4]。地中には、まだ掘り出されていない粘土板がさらに埋まっている。そこに書かれた文字は、王の自慢だけではない。穀物を何袋受け取ったという領収書がある。銀を貸し、利息を定めた契約書がある。遠くへ出た夫に妻が送った手紙もある[6]。

領収書1枚は短い。物の名前と数量がまず出る。次に、渡す人と受け取る人の名前が来る。最後には日付と証人の名前を書く。今日、店で受け取る領収書と順序が似ている。このような文書1枚の中身は小さい。しかし数万枚が集まると、話は変わる。どの年に穀物価格が上がったのかが見えてくる。どの都市が何を売り買いしたのかも見える。

ほとんどは翻訳されていない

問題は、この文字がほとんど読まれていないことにある。粘土板のうち現代語に移されたものは一部だけである[5]。残りは博物館の箱の中で眠っている。

文字を書く人には誰でもなれたわけではない。書記になるには、学校で長く学んだ。学生たちは記号を粘土板に繰り返し書いた。書き損じた板は水で練り直し、また板にした。残っている練習板も遺物として発掘される。練習板は、その時代の教室を見せてくれる。

文字を知る人は少なかった。ほとんどの人は書記に代筆させた。手紙を受け取ると、書記が読んでくれた。手紙の文が丁寧な型を備えている理由はここにある。話す人と書き留める人が違ったからである。

粘土板1枚を読むにも、いくつもの段階を経る。まず記号を一つ一つ見分ける。次に、その記号がどんな音なのかを決める。その次に単語を区切り、意味を探す。最後に文を組み立てて翻訳する。慣れた書式なら早く終わる。見慣れない文書は長く向き合わなければならない。

理由は、読む人が足りないからである。くさび文字を扱う学問をアッシリア学という。学者1人を育てるには数十年かかる。古い言語と古い歴史と一緒に身につけなければならないからである。そのため、この文字を読む人は世界にごく少ない[5]。

新しく出てくる粘土板の数は、学者が読む速度を超えている。人が増えないので、箱だけが積み上がる。2026年6月、英国ヨーク大学が新しい研究事業を始めた。事業名は「くさび文字遺産への普遍的アクセスに向けて」である。くさび文字資料をアラビア語にも移し、誰もが見られるようにする計画である。この事業は2029年まで3年間続く[7]。

資料を一か所に集める作業も進んでいる。くさび文字デジタル図書館計画がそのような事業である。世界に散らばる粘土板の情報を集め、インターネットで公開する。写真、音訳、翻訳を並べて置く。音訳とは、くさび記号の音をアルファベットで書いた文である。これまでに40万点を超える遺物が目録に載った[4]。資料が集まってこそ、機械も学ぶ材料を得る。

粘土板はいま、多くの国に散らばっている。イラクとトルコの博物館に多い。ロンドン、パリ、ベルリンの博物館も大きな割合を持っている。米国と日本の大学にも分かれている。1つの文書の表と裏が別々の国にあると、合わせるのには難しい。写真をインターネットに一緒に載せて初めて、照合できる。

王の記録に押しのけられた帳簿

日常文書が後回しにされたのには、古くからの事情がある。19世紀と20世紀初めの発掘隊は、王宮と神殿を先に掘った。大きな石碑と金製品が人々の目を引いた。学者たちも王室碑文と法典を先に読んだ。神話と叙事詩を移すことに、先に時間を使った。

その間、商人の領収書と手紙は順番から押し出された。穀物配給の一覧は雑用の記録と見なされた。1枚ずつ切り離して見ると、中身がないように見えた。学者たちは、この退屈な文書を手で写す作業を後回しにした。そうして普通の人々の声が長く埋もれた。

この偏りのために失ったものがある。価格がどのように上下したのか分かりにくくなった。女性たちが何をしていたのかが見えなくなった。子どもと召使いの名前は一覧にだけ残った。王の碑文は自慢を書く。領収書は自慢しない。取引した事実だけを書く。だから領収書の方が、より正直な資料になることもある。

1つのくさびが複数の意味を持つ

機械で読もうとしても、くさび文字は手ごわい。記号1つが複数の役割を重ねて担うからである[8]。ある場所では音を表す音節記号になる。別の場所では単語の意味を丸ごと表す。さらに別の場所では音を出さない。後に来る単語が木なのか神なのかを知らせる標識としてだけ使われる。このような標識を限定符という。

くさび文字は1つの言語だけを書いた文字ではない。シュメール語とアッカド語が同じ文字と一緒に使った。後にはヒッタイト語とエラム語もこの文字を借りて使った。同じ記号が言語ごとに違う音を出す。読む人は文書の言語から決めなければならない。

木を意味するシュメール語の記号gishが良い例である。家具の名前の前に付くと、材料が木であるという標識になる。単独で置かれると、木という単語そのものになる。別の場所ではgishという音だけを表す。読む人は前後の文を見て、役割を選ばなければならない。

数字を書く方式も現在とは違う。メソポタミアの人々は60を1つのまとまりとした。時計の60分と60秒はここから来た。領収書の数量もこの方式で書かれた。穀物と銀は互いに違う単位を使った。同じ記号でも、穀物の時と銀の時では値が変わる。機械はこの単位の違いまで一緒に学ばなければならない。

単語を続けて書いた点も障害である。昔の書記たちは単語の間を空けなかった。コンピューターはどこで単語が切れるのかを見つけにくい[8]。粘土板が割れ、すり減っていることも問題である。くさびの跡が浅くなると、人の目でも区別しにくい。アッカド語の単語の形がよく変わることも難しい。1つの語根に複数の語尾が付く。主語なのか目的語なのかによって末尾の音が変わる。時制と人称によってまた変わる。辞書で探す基本形を先に見つけなければならない。

粘土板が割れる理由もいくつもある。地中で圧迫され、水に濡れる。発掘中に角が欠け落ちることもある。昔の遺物取引によって、割れた粘土板が複数の国に散らばった。地面から取り出した後は、塩分が染み出して文字を覆うこともある[2]。文字が消えた空白部分を、学者たちは欠損部と呼ぶ。

この幾重もの壁のために、数十万枚の帳簿が沈黙してきた。

参考資料

- [1] Truman State University. (2026). How it works. Truman State University Cuneiforms. <https://exhibits.truman.edu/cuneiform/how-it-works/> (2026年9月3日閲覧)
- [2] Guinan, A., Oller, G., & Ormsby, D. (1976). Nippur rebaked: The conservation of cuneiform tablets. Expedition, 18(3). <https://www.penn.museum/sites/expedition/nippur-rebaked/> (2026年9月3日閲覧)
- [3] Bouscaren, D. (2023). Assyrian women of letters. Archaeology, 76(6). <https://archaeology.org/issues/november-december-2023/features/assyrian-women-of-letters/> (2026年9月3日閲覧)
- [4] Cuneiform Digital Library Initiative. (2026). About the CDLI. cdli.earth. <https://cdli.earth/about> (2026年9月3日閲覧)
- [5] Gutherz, G., Gordin, S., Sáenz, L., Levy, O., & Berant, J. (2023). Translating Akkadian to English with neural machine translation. PNAS Nexus, 2(5), pgad096. <https://doi.org/10.1093/pnasnexus/pgad096>
- [6] Michel, C. (2020). Women of Assur and Kanesh: Texts from the Archives of Assyrian Merchants (Writings from the Ancient World 42). SBL Press.
- [7] Cuneiform Digital Library Initiative. (2026). Research Programme "Towards Universal Access to Cuneiform Heritage" 2026-2029. cdli.earth. <https://cdli.earth/postings/239> (2026年9月3日閲覧)
- [8] Gordin, S., Gutherz, G., Elazary, A., Romach, A., Jiménez, E., Berant, J., & Cohen, Y. (2020). Reading Akkadian cuneiform using natural language processing. PLoS ONE, 15(10), e0240511. <https://doi.org/10.1371/journal.pone.0240511>

6.2. 科学捜査隊の追跡: アッカド語AI翻訳

翻訳の通路を2つ開く

2023年、テルアビブ大学の研究陣が先に道を開いた。ガイ・グテルツと同僚たちは、アッカド語を英語に移す翻訳機を作った[1]。アッカド語はメソポタミアで長く使われた言葉である。くさび文字で書かれ、現在は使う人がいない。

研究陣は翻訳の通路を2つ開いた。第1の通路の名前はC2Eである。コンピューター文字に写し取ったくさび記号を入れ、英語を受け取る道である。粘土板の写真を直接読む方式ではない。学者の音訳を経ないので、段階が短い。第2の通路の名前はT2Eである。学者がくさび記号の音をアルファベットに移して書いた文を入れる道である。このアルファベットへの移し替えを音訳という。音訳の段階で記号の意味が一度整理される。そのためT2E側の翻訳の方が正確である[1]。

音訳を作る作業も簡単ではない。学者は記号を見て音を決める。音が複数あれば文脈を見て選ぶ。この選択がそのまま判断である。判断が一度終わった文は、機械が扱いやすい。T2Eの点数が高い理由はここにある。

翻訳機の中は2つに分かれる。前側は原文を数字のかたまりに変える。後ろ側はその数字を見て英語の単語を一つずつ選ぶ。前側をエンコーダー、後ろ側をデコーダーと呼ぶ。エンコーダーは単語の意味と順序を一緒に入れる。デコーダーは前に選んだ単語を見ながら、次の単語を決める。

写真から記号を直接読む研究は別に続いている。2026年6月、電子バビロニア図書館の研究陣が結果を発表した。割れた粘土板8万7千点余りの写真を機械が調べた。見つけた記号の跡は290万個に近い。従来の方法より検出成績が最大37パーセント上がった。大きく割れた粘土板と乱れた行の配列には、まだ弱い[6]。

8千篇の文書で学んだ機械

機械は例文を見て翻訳を学ぶ。研究陣はORACCという公開資料集から文書を集めた。ORACCは、くさび文字文献を整理して公開するインターネット資料集である。ここで集めた文書は8,056篇だった。文の数では5万個を超えた[1]。

集めてきた文書には、いくつもの種類があった。王の業績を刻んだ碑文が約3千篇だった。行政書簡が約2千篇だった。裁判記録と占星報告書も入っていた[1]。文書ごとに、学者が作った英語訳が添えられていた。機械はこの対を数万回見つめながら、規則を自ら見つけた。

機械が学ぶ過程は試験勉強に似ている。まず問題と答えを一緒に見る。答えを隠して自分で解いてみる。間違えれば、どこで間違えたのかを振り返る。この振り返りを数百万回繰り返す。人と違う点は速度と粘り強さである。

機械は単語を丸ごと扱わない。単語をさらに小さな断片に分けて扱う。この断片をトークンという。珍しい単語も断片に分ければ扱える。語尾が多く付くアッカド語には、この方法が合う。

翻訳品質はBLEUという点数で測る。機械の翻訳と人の翻訳がどれほど重なるのかを数える方法である。重なる単語のまとまりの比率で点数を出す。0点から100点までで表す。30点を超えると、意味が通じると見ることが多い。しかし絶対基準ではない。文書の種類と文の長さによって解釈は変わる。

研究陣は畳み込みニューラルネットワークという構造を選んだ。近くにある情報同士をまとめて学ぶ構造である。資料が少ない時に学習が速く安定するからである。検証の結果、C2Eは36.52点を得た。T2Eは37.47点で、より高かった。以前使われていた自動検索方式は27.09点と23.51点にとどまった。文が短く、書式が決まった文書で成績が良かった[1]。

資料を分ける方法も重要である。学ぶ分と試験する分を別に取り分けておく。試験する文書を先に見せれば、点数は無意味になる。人の試験で問題を先に知っているのと同じである。研究陣はこの区分を守って点数を出した。

同じ原文でも、人によって翻訳は違う。そのため点数だけを見て良し悪しを決めない。点数のほかに人の評価も一緒に使う。翻訳文を学者に見せ、等級を付けさせる。意味が通じるかを別に尋ねる。間違っただ所があるかも別に尋ねる。学者が使えたと見た翻訳は、T2Eで約44パーセントだった。C2Eは約38パーセントだった[1]。点数と人の評価が食い違う時がある。その時は人の判断を優先する。

アッカド語翻訳は低リソース課題に属する。低リソースとは、学ぶ例文が少ないという意味である。英語とフランス語の翻訳機は数千万文を学ぶ。アッカド語が持つ対訳文は、その千分の一にも及ばない。資料が少ないと、大きなモデルはかえって迷う。

2025年には、複数の研究チームが競う共同課題も開かれた。課題は単語の基本形を見つけ、抜けた文字を当てることだった。資料は電子バビロニア図書館とArchibabから持ってきた。同じ資料で成績を測るので、どの方法が良いのかがすぐに見えた[2]。

基本形探しは翻訳の下作業である。この作業を学界ではレンマ探しという。語尾が付いて長くなった単語を辞書形に戻す作業である。下作業が終わって初めて、単語別の検索と統計が可能になる。

もっともらしい嘘を見つけ出す

機械翻訳には危険な癖の一つある。根拠がないのに、もっともらしい文を作り出すことだ。この癖を幻覚と呼ぶ。研究陣は幻覚を二つに分けて調べた。一つは原文を見ても意味を誤って組み立てる場合だ。もう一つは原文を無視して別の話を作る場合だ[1]。

壊れた粘土板では幻覚が起きやすい。文字が消えた空白を機械が想像で埋めるからだ。だから翻訳文は必ず学者が見直さなければならない。機械が出した文は下訳であり、最終判断ではない。

幻覚を減らす方法はいくつかある。壊れた箇所を空欄表示ではっきり知らせる。機械にその箇所は分らないと答えさせる。同じ文を何度も翻訳してみることもある。答えが毎回変わるなら信じない。学者はこうした表示を見て、どこを見直すかを決める。

楔形記号をアルファベットに移し、語を区切る作業が翻訳より先に来る。同じ研究の流れで、この作業も機械に任せてみた。双方向LSTMというニューラルネットワークの成績が優れていた。精度は最高97パーセントに達した[3]。

二つの方向に開いた扉

これまでの翻訳機は一方向にだけ流れていた。古い文を現代英語に移すだけだった。2026年5月、米国南カリフォルニア大学のワン・ジャオフィが反対方向を開いた。彼が作った道具の名前はタブレットクラフトだ[4]。

タブレットクラフトは英語をアッカド語にも移す。移した結果を楔形記号で描き、粘土板の画像まで作ってくれる。学習には双方向の文11万6千組を使った。記号対応表は1万4千余りを備えた。試験成績はアッカド語から英語へが49.1点だった。英語からアッカド語へは48.5点が出た[4]。

この道具は博物館の体験機器と大学の授業で使われた。作った人がはっきり示した限界もある。機械が作り出したアッカド語を、本物の古記録のように引用してはならない。この道具が得意とする文章は、新アッシリア時代の王室碑文だ[5]。

双方向翻訳には使い道がある。学生は自分の文を楔形記号に変えてみることができる。記号を自分で描いてみると、文字の組み立てが目に入る。博物館では来館者が自分の名前を刻んでみる。古い文字が見物の対象から、扱う道具へと変わる。

前の研究は資料とプログラムをともに公開した[1][5]。公開すれば、別のチームが同じ試験を繰り返してみることができる。誤りがあれば早く表に出る。古代言語研究は参加する人が少ない。公開が発展の速さを大きく左右する。

翻訳機の使い方も整理されつつある。学者は機械の下訳を画面に表示する。原文と翻訳を並べて比べる。疑わしい語に印を残す。印がたまると、次の学習資料になる。

参考資料

- [1] Gutherz, G., Gordin, S., Sáenz, L., Levy, O., & Berant, J. (2023). Translating Akkadian to English with neural machine translation. PNAS Nexus, 2(5), pgad096. <https://doi.org/10.1093/pnasnexus/pgad096>
- [2] Gordin, S., Sahala, A., Spencer, S., & Klein, S. (2025). EvaCun 2025 shared task: Lemmatization and token prediction in Akkadian and Sumerian using LLMs. In Proceedings of the Second Workshop on Ancient Language Processing (pp. 242-250). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.alp-1.33>
- [3] Gordin, S., Gutherz, G., Elazary, A., Romach, A., Jiménez, E., Berant, J., & Cohen, Y. (2020). Reading Akkadian cuneiform using natural language processing. PLoS ONE, 15(10), e0240511. <https://doi.org/10.1371/journal.pone.0240511>
- [4] Wang, Z. (2026, 5月 14日). TabletCraft: Bridging a 4,000-year cultural gap with bidirectional Akkadian NMT and cuneiform rendering. arXiv:2608.02609 (2026年 5月 14日 제출, 2026年 8月 공개). <https://arxiv.org/abs/2608.02609> (2026年9月3日閲覧)
- [5] Buyukyildirim, K. (2026, 8月 11日). TabletCraft opens a two-way AI translation channel for archaeologists. Arkeonews. <https://arkeonews.net/tabletcraft-opens-a-two-way-ai-translation-channel-for-archaeologists/> (2026年9月3日閲覧)
- [6] Che, W., Garces Arias, E., Niaz, A., Bender, A., & Jimenez, E. (2026, 6月 21日). Automated sign detection across the Electronic Babylonian Library: A large-scale dataset and end-to-end cuneiform OCR pipeline. arXiv:2606.22608. <https://arxiv.org/abs/2606.22608> (2026年9月3日閲覧)

6.3. 明らかになった真実

カネシュ商人たちの手紙

トルコ中部のカイセリ近くにキュルテベという丘がある。4千年前、この場所の名前はカネシュだった。アッシリア商人たちはここに交易拠点を置いた。商人たちの拠点共同体をカルムと呼んだ。カルムはもともと川辺の波止場を意味する言葉だ。荷物を下ろしていた場所が、商人たちの区域になった。

カネシュのカルムはアッシリア商人たちの自治組織だった。彼らは自分たちで規則を作り、争いを裁いた。現地の王に税を納め、安全を保障してもらった。品物は口バの背に載せて運ばれた。アッシルからカネシュまでは数週間かかる遠い道のりだった。

1948年から続く発掘で粘土板が大量に出た。これまでに出土した文書は約2万3千5百枚だ。発掘を率いるフィクリ・クラコールが示した数字だ[1]。内容は王の自慢ではない。借用証と貨物送り状がある。貨物送り状は送り出した荷の一覧表だ。裁判記録と家族の手紙がある。

手紙には催促と不満がそのまま残っている。金をいつ送るのかと尋ねる。品物が遅いと責める。体に気をつけるよう挨拶を添える。息子に勉強しなさいと言うこともある。4千年前の人の心配が文の中にそのままある。

この文書から、古い経済の骨組みが明らかになった。商人たちはアッシュルから織物と錫を運んできた。アナトリアでそれを銀と金に換えた。税と通行料をどう払うかも記されていた。争いが起きたとき、どう判決したかも残った。

錫は青銅を作るのに欠かせない金属だ。青銅は銅と錫を混ぜて作る。アナトリアでは錫を手に入れにくかった。錫は遠い東から運んでこなければならなかった[8]。この不足が長い交易路を生んだ力だった。

アッシュルに残った女性たちの役割も明らかになった。商人の妻と娘たちは家で織物を織って送った[2][8]。彼女たちは生産量を調整し、値を交渉した。夫に送った手紙で相場を問いただすこともあった。セシル・ミシェルはこうした文書300余編を集めて英語に訳した。ミシェルはフランス国立科学研究センターとハンブルク大学で働く[2]。

封筒を開けずに読む

古い商人たちは重要な粘土板を粘土の封筒に入れて封じた。中を読むには封筒を壊さなければならなかった。ところが封筒にも印章の跡が押されている。封筒を壊すと、その情報が消える。

封筒は偽造を防ぐ装置だった。内側の粘土板に契約内容を書く。その上に粘土をかぶせて封筒を作る。封筒の外面に同じ内容をもう一度書く。証人たちが円筒印章を転がして押す。外と中が違えば、書き換えた事実が明らかになる。

封筒を残したまま中を読む方法が必要だった。2026年6月、カイセリ考古学博物館が答えを示した。封筒に入った粘土板一つを開けないまま読み取ったのだ。フランス国立科学研究センターとハンブルク大学がともに行った作業だ。物を切らずに中をのぞくCT撮影と同じ方法を使った。中に記された内容は小麦と大麦の取引約束だった。クラコールは、この撮影は大麦の粒一つまで見せてくれると述べた[1]。こうして得た文字写真は翻訳機の入力資料になる。読み取る文書が増えてこそ、機械も学ぶ材料を得る。

5万ドルがかかった翻訳大会

翻訳大会も開かれた。ディーブ・ファスト・イニシアチブという団体が大会を準備した[3]。大会は人工知能競技サイトKaggleで開かれた[4]。課題はカネシュの商業文書を英語に移すことだった[3]。大会は2025年12月16日に始まった。参加締切は2026年3月23日だった。賞金は5万ドルだった。運営費はある金融会社が出した[7]。

大会資料は誰でもダウンロードできるよう公開された[4]。公開資料は大会が終わった後にも残る。次の研究者がその資料の上で再び始める。大会一つが資料一つを残す構造だ。

キュルテペ文書の中にも、まだ移せていないものが多い[3]。発掘は今も毎年続いている[1]。新しい粘土板が出ると目録に加わる。読む人が増えなければ、目録だけが長くなる。機械の助けが必要な理由はここにある。

同じ状況は一つの遺跡で終わらない。世界各地の博物館に似た文書が積まれている。まだ読めていない箱がそこにもある。書式の決まった文書はどこにでも多い。穀物帳簿と税台帳がそうだ。カネシュで通じた方法は、ほかの文書庫にも使える。

大会案内文は文書に出てくる実在の人物を前面に出した。アッシュルナダとアッシュルイディのような商人の名前だ。イシュタルラマシとエンナムアッシュルの名前もともに掲げられた[3]。4千年前のある商人一家の名前が大会の主題になったのだ。

大会方式には利点がある。いくつかの研究室ではなく、世界の開発者が同じ問題に取り組む。同じ物差しで成績を測るので、どの方法がよいかすぐに分かる。賞金は人を集める力になる。

この文書群は機械翻訳に合う資料だ。書式が決まっており、文が短い。同じ表現が多く、文書に繰り返される。機械は繰り返される規則をよく見つける。学者がすでに移しておいた文書も少なくない。その翻訳文が機械の教材になる。

機械ができないこと

成果は積み上がったが、残る壁もはっきりしている。第一の壁は資料の不足だ。近ごろの言語の翻訳機は数億の文を食べて育つ。アッカド語は対になった文が数万しかない[5]。資料が少ないと、まれな地名と人名で誤りやすい。

第二の壁は記号の複数の意味だ。同じ記号が音なのか、語なのか、標識なのか紛らわしい[6]。前後の文脈が短いと、機械も迷う。

第三の壁は機械の判断方式だ。機械はどの語が来る確率が高いかを計算する。その文が歴史的に正しいかは判断できない。だから学者の検証が最後の段階に必ず必要になる[5]。機械と人が組になったとき、翻訳は信頼できるものになる。

作業を分けて担う方式が根づきつつある。機械は数万枚を見渡して、およその意味を出す。学者はそこから重要な文書を選び出す。選んだ文書だけを手で丹念に読み直す。こうすれば読める文書の数が大きく増える。

翻訳文にはタグが必要だ。どの箇所を機械が作ったのかを書いておく。人が直した箇所も表示する。そうしてこそ、後の人が根拠をたどれる。歴史研究は根拠をたどることから始まる。

翻訳で名前は扱いにくい箇所だ。人の名前と土地の名前は意味を移せない。音をそのまま写して書かなければならない。ところが同じ名前が文書ごとに違って記されることもある。機械は初めて見る名前を見当違いの語に変えやすい。だから名前の一覧を別に作り、機械に知らせる。

それでも方向は明らかだ。以前は学者一人が一生で数百枚を読んだ。今は機械が下訳を先に作る。学者はその下訳を直すことに時間を使う。箱の中で眠っていた4千年前の領収書が順番を待っている。王ではない普通の人の声が聞こえ始める。

カネシュの商人たちは、自分の手紙が長く残るとは思っていなかった。彼らはその日の勘定を書いただけだった。その勘定が今、一つの時代の経済を見せてくれる。残るのは自慢より記録だ。

参考資料

- [1] Türkiye Today. (2026, 6월 13일). 4,000-year-old clay tablet sealed in mud envelope goes on display in central Türkiye. Türkiye Today.
<https://www.turkiyetoday.com/culture/4000-year-old-clay-tablet-sealed-in-mud-envelope-goes-on-display-in-central-turkiye-3221841> (2026年9月3日閲覧)
- [2] Michel, C. (2020). Women of Assur and Kanesh: Texts from the Archives of Assyrian Merchants (Writings from the Ancient World 42). SBL Press.
- [3] Deep Past Initiative. (2026). Deep Past Challenge: Introduction. deeppast.org.
<https://www.deeppast.org/challenge/intro/> (2026年9月3日閲覧)
- [4] Kaggle. (2026). Deep Past Challenge: Translate Akkadian to English. kaggle.com.
<https://www.kaggle.com/competitions/deep-past-initiative-machine-translation> (2026年9月3日閲覧)
- [5] Gutherz, G., Gordin, S., Sáenz, L., Levy, O., & Berant, J. (2023). Translating Akkadian to English with neural machine translation. PNAS Nexus, 2(5), pgad096. <https://doi.org/10.1093/pnasnexus/pgad096>
- [6] Gordin, S., Gutherz, G., Elazary, A., Romach, A., Jiménez, E., Berant, J., & Cohen, Y. (2020). Reading Akkadian cuneiform using natural language processing. PLoS ONE, 15(10), e0240511.
<https://doi.org/10.1371/journal.pone.0240511>

- [7] Deep Past Initiative. (2025, 12월 16일). Deep Past Initiative partners with Kaggle on prize challenge to translate 4,000-year-old tablets. EIN Presswire.
<https://www.einpresswire.com/article/875885564/deep-past-initiative-partners-with-kaggle-on-prize-challenge-to-translate-4-000-year-old-tablets> (2026年9月3日閲覧)
- [8] Bouscaren, D. (2023). Assyrian women of letters. *Archaeology*, 76(6).
<https://archaeology.org/issues/november-december-2023/features/assyrian-women-of-letters/> (2026年9月3日閲覧)

第7章. 亀の腹甲に眠っていた帝国の未来： 甲骨文字と3D仮想綴合

7.1. 事件ファイル

薬房で発見された3千年前の文字

1899年、清朝の学者である王懿榮が古代の文字を見出した。この出来事には、古くから伝わる逸話がある。王懿榮が病を患い、竜骨という生薬を受け取ったという話だ。竜骨は古い獣の骨を削って作った粉薬であった。昔の人々はこの骨を竜の骨だと信じていた。彼が受け取った骨は、まだ削られていない状態だったという。その表面に見慣れない模様を見たというのだ[1]。

この話は事実として確定してはいない。中国の学界はこれを裏付ける記録がないと見ている。古文字学者の李学勤は2007年の論考でこの話を否定した。薬材として売られた竜骨はすでに砕かれており、文字を見ることはできなかったというのだ。実際には骨董商が安陽から出土した骨を買い取って運び、その骨が金石学者である王懿榮のもとへ渡って鑑定を受けたと見られている[2]。

王懿榮は翌年に世を去り、骨の出所を突き止めることはできなかった。友人の劉鶚が1903年にこの文字を集めた最初の本を出版した。1908年には学者の羅振玉が骨の出所を突き止めた[1]。骨は河南省安陽の畑から出土していた。

安陽のその地こそが殷墟である。殷墟は商の最後の都の跡であった。商は紀元前13世紀から紀元前11世紀までここを治めた[3]。正式な発掘は1928年に始まった。それ以前に農民たちが掘り出した骨は薬材として売られていった。文字が刻まれた面を削り落として売ることもあった[1]。3千年を耐え抜いた記録が、薬材として消えてしまうところだった。

王が神に問うた国

殷墟の骨に刻まれた文字を甲骨文字と呼ぶ。甲骨文字より古い文字の記録は、東アジアでは知られていない[3]。

商の人々は、大事を控えると神に問うた。雨が降るかを問うた。豊作になるかを問うた。出陣してよいかを問うた。王妃が子を産む日も問うた。このように神の意を問う儀式を占卜と呼ぶ。占卜は王室が直接管理する国家の事業であった[3]。

商の王妃である婦好は、軍を率いる司令官でもあった。婦好の名は甲骨文字の記録に何度も登場する。1976年、殷墟から婦好の墓が完全な状態で発掘された[4]。墓の中の青銅器や玉器は、彼女の地位を物語っていた。甲骨文字は、このように墓とともに昔の人々の生を教えてくれる。

甲骨文字に記された問いは、国家の政事にとどまらなかった。王の歯痛を心配する問いもあった。狩りに出かけてよいかを尋ねる問いもあった。そのため、これらの骨は商の人々の日常を映し出す記録となる[3]。

火で熱した骨、刃で刻んだ文字

商の人々は占卜に二つの材料を用いた。一つは亀の平らな腹甲である。もう一つは牛の広い肩甲骨である。二つの材料を合わせて甲骨と呼ぶ[3]。

準備の過程は入念であった。書記たちは骨の表面を滑らかに整えた。その上で裏面に小さな溝を列状に彫った。彫り込んだ溝に熱した棒を当てた。熱を受けた骨は表面がピシリと割れた。王と卜者は、割れ目の向きや角度を観察した。そしてその割れ目を神の答えとして読み取った[3]。

読み取りが終わると、書記たちが刃を握った。彼らは占卜の過程を骨の表面に一画ずつ刻んだ。この刻まれた文こそが甲骨文字である。

記録は決められた順序を守っていた。まず占った日付と担当者の名を記した。次に神に投げかけた問いを記した。続いて王が下した吉凶の判断を記した。最後に実際に起きた出来事を記した。四つの欄を順番に埋める形式である。ただし、四つの欄がすべて残っている記録は稀である。骨が破損して一、二欄が失われていることもあった。書記が初めから空欄のままにしておいた場合もある[3]。それでも、記す順序そのものは一定であった。

この規則正しさは、後になって大きな助けとなる。書式が決まっていれば、空欄を絞り込めるからだ。前の欄に日付があれば、後ろの欄には問いが来る。コンピュータもこの順序を学び、欠落した文字を推測する[3]。

日付を記す方法も決まっていた。商は十個と十二個の記号を組み合わせて日を数えた。この組み合わせは六十日ごとに一巡した。骨に日付の記号が二文字残っているだけでも順序を割り出すことができる。欠損した記録の時期を絞り込む手がかりがここから生まれる[3]。

15万片に砕かれた記録

商が滅んだ後、殷墟は土の中に埋もれた。その長い年月の間、骨が無事であるはずはなかった。土の重みが薄い腹甲を押し潰して砕いた。水と微生物が骨のきめをむしばんだ。本来は一枚であった大きな腹甲が、数十片に割れてしまった。

これまでに世に知られた甲骨の破片は約15万片に上る[5]。これらの遺物は一箇所に集まっていない。19世紀末に発見された後、少なからぬ数が骨董市場へと流出した[1]。そのため、一つの骨から割れた破片が、それぞれ異なる国の博物館に収蔵されることになった[5]。中国の安陽や北京にある遺物もあれば、ヨーロッパや韓国の機関が所蔵する遺物もある[6]。

散らばった破片を元の一枚に戻す作業を綴合と呼ぶ。綴合は長きにわたり、もっぱら人の目に頼っていた。学者たちは何万枚もの白黒の拓本を拡大鏡でのぞき込んだ。拓本とは、骨の上に紙を当てて墨を叩きつけて写し取った複製である。硬貨の上に紙を置き、鉛筆でこすって模様を浮き上がらせるのと似ている。学者たちは骨のきめや割れ目の線の屈曲を見比べた。途切れた線の端がつながるかどうかも一つ一つ確認した[5]。

このやり方は遅く、過酷であった。一対を見つけるのに何年もかかることさえあった。相手の見つからない破片は箱の中にそのまま残された。対になる破片がなければ、文も途切れたまま残る。問いだけが残ри、結果が失われた記録が積み重なった。

作業の規模を考えれば、難しさはさらに際立つ。破片が15万片あれば、二つを選ぶ組み合わせの数は100億対を超える。人間が1日に100対を見比べたとしても、一生かかっても見切れない。その上、破片は世界各国に散らばっている[5]。二つの破片を並べて見る機会すら訪れないことが多かった。

3千字がいまだ沈黙している

破片をすべてつなぎ合わせても問題は残る。文字そのものを読むことができないからだ。甲骨文字全体で確認されている異なり文字数は約4,500字である。しかし、学界で意味の合意が得られている文字は約1,500字にすぎない。残りの3,000字ほどはいまだ意味が分かっていない[7]。

解読が難しい理由はさまざま。刃物の痕が摩耗して輪郭が薄れている。骨が割れて文字の半分が失われている。書記ごとに刀子を握る癖が異なり、同じ文字を違った形に刻んだ。そのため、意味は同じなのに形の異なる文字がいくつも生じた。こうした文字を異体字と呼ぶ[3]。現代の漢字辞典は、この3千年前の字形を説明してはくれない。

この沈黙は商の歴史にも空白を残す。祭祀の名が読めなければ、祭祀の種類を数え上げることができない。地名が読めなければ、国の境界を描くことができない。人名が読めなければ、誰が何を司っていたのかが分からない。文字一つの沈黙が、歴史の一場面の沈黙へとつながる。

意味の分からない文字3千字と、相方を失った破片数万点が残された。この沈黙を破るため、人工知能が殷墟へと足を踏み入れた。

参考資料

- [1] Mark, E. (2016, 2月 26日). Oracle bones. World History Encyclopedia. https://www.worldhistory.org/Oracle_Bones/ (2026年9月3日閲覧)
- [2] 姚小鸥. (2020, 7月 16日). 谁是甲骨文的最早发现者. 澎湃新闻. https://www.thepaper.cn/newsDetail_forward_8297008 (2026年9月3日閲覧)
- [3] Chen, Z., Hua, W., Li, J., Zhu, Y., Zhi, X., Liu, Z., Chen, T., Zhang, W., & Zhai, G. (2026). Oracle bone inscriptions information processing: a comprehensive survey. npj Heritage Science, 14, 220. <https://doi.org/10.1038/s40494-026-02511-w> (2026年9月3日閲覧)
- [4] Smarthistory. War and sacrifice: The tomb of Fu Hao. Smarthistory. <https://smarthistory.org/tomb-of-fu-hao/> (2026年9月3日閲覧)
- [5] Zhang, C., Wang, B., Chen, K., Zong, R., Mo, B., Men, Y., Almpandis, G., Chen, S., & Zhang, X. (2022). Data-driven oracle bone rejoining: A dataset and practical self-supervised learning scheme. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 4482-4492). ACM. <https://doi.org/10.1145/3534678.3539050>
- [6] 梁文艳. (2026, 5月 26日). 以科技赋能, 让甲骨文"活"起来. 中国青年报. https://zqb.cyol.com/pc/content/202605/26/content_426168.html (2026年9月3日閲覧)
- [7] Liu, Y., Guan, H., Wang, P., Wang, X., Wan, J., Zhang, K., Zheng, H., Liu, X., Kuang, Z., Yang, H., Li, B., Liu, Y., Jin, L., & Bai, X. (2026). AlphaOracle: Oracle bone script decipherment via human-workflow-inspired deep learning. The Innovation, 7(11), 101462. <https://doi.org/10.1016/j.xinn.2026.101462>

7.2. 科学捜査班の追跡：甲骨専用マルチモーダルモデルと3D仮想綴合

散らばった資料を一つの画面に集める

捜査は資料を集めることから始まる。甲骨研究も同様であった。安陽師範学院と中国社会科学院の研究陣は、2019年に殷契文淵というデータプラットフォームを開設した。殷契文淵は甲骨の拓本、文字情報、論文を一箇所に集め、無料で公開している[1]。

このプラットフォームは写真だけを掲載しているのではない。甲骨の破片を高解像度で撮影し、3次元モデルとして構築して掲載している。3次元モデルとは、実物をさまざまな方向からスキャンし、コンピュータ内に立体として復元したデータである。画面を回転させれば、骨の表面、裏面、割れた側面まで見ることができる。研究陣はミリメートル単位の精度でこのモデルを作成している。浅く刻まれた刃物の痕の深さまでもが、このデータに記録されている[1]。

海外へ流出した甲骨を取り戻す取り組みも、このプラットフォームと結びついた。研究陣は実物の代わりに3次元データを取り込む手法を選んだ。ドイツ、韓国、フランスなどの所蔵先と協力し、900点余りのデータを集めた[1]。実物は元の所蔵先に置いたまま、その姿だけが安陽へと戻ってきたのである。

2025年10月には、このプラットフォーム上に殷契行止というツールが追加された。このツールは文字を識別し、拓本を見比べ、関連文献を検索してくれる[1]。資料を蓄積しておくだけの倉庫が、問いに答える助手へと変わったのだ。

資料がどれほど蓄積されたかは、数字からも見て取れる。2026年に発表された総合レビュー論文は、甲骨関連の公開データセット36種を紹介している。文字認識用のデータセットもあれば、綴合研究用のデータセットもある[2]。

多様な種類のデータを組み合わせて用いる手法をマルチモーダルと呼ぶ。甲骨研究においてこの言葉は、三つのデータを併せて観察することを意味する。一つは墨で写し取った白黒の拓本である。一つは色彩と質感が保たれたカラー写真である。残る一つは深度情報を含んだ3次元スキャンデータである。拓本は筆画が鮮明だが深さが失われる。3次元データは深さを捉えるが画がぼやけて見える。三者を重ね合わせることで、互いの欠落を補い合うことができる[2]。

人間が文字を読む手順をなぞる

2026年に登場したAlphaOracleは、これらのデータの上で稼働する人工知能である。このモデルは華中科技大学や安陽師範学院などが共同で開発した。設計の基本方針は、人間の学者の作業手順をそのまま再現することであった[3]。

人間の学者は四つの段階を踏む。まず拓本を観察して筆画を整理する。次に文字の字形を分析する。続いて同じ文字が現れる他の文を探索する。最後に古代文献の知識に照らして意味が妥当かを検証する。AlphaOracleは、これら四つの段階をそれぞれ担当するモジュールに分かれている[3]。

この構造には理由がある。結論だけを唐突に出す機械を学者が信頼することは難しい。段階を分けることで、どの段階で判断が分かれたのかが見えてくる。学者はその段階に立ち返り、機械と議論を交わすことができるのだ[3]。

このモデルが読み込んだデータの量は、人間が生涯で見られる量をはるかに超えている。拓本73,883点を処理した。スキャンデータ45,364点と研究論文23,755編も併せて扱った。研究陣は86名の専門家にこのツールを使用してもらった。その結果、1文字を分析する時間が64パーセント短縮された。参加者の79パーセントが、このツールを有用であると回答した[3]。

文字を構成要素に分解して見る

別の研究陣は、文字を丸ごと記憶させない手法を選択した。甲骨文字は、繰り返し現れる絵画的な構成要素から成り立っている[4]。手の形、足の形、器の形といった部品が、様々な文字に分散して含まれている。部品の意味が分かれば、初めて目にする文字であっても推測することができる。

この着想に基づき2026年4月に発表されたのがOB-Radixデータセットである。OB-Radixには、互いに異なる934字の文字が収録されている。それらの文字を構成する部品画像は1,853点に及ぶ。部品の種類は478通りある。各部品には専門家が検証した説明が付与されている[4]。

人工知能は見慣れない文字に出会うと、まず部品へと分解する。次に部品同士がどのような関係で配置されているかを精査する。二つの部品が上下に並んでいるのか左右に並んでいるのかによって意味が分かれる。モデルはこの位置関係まで吟味した上で意味を提案する[4]。

同年6月にはOracleAnalyzerというモデルも登場した。このモデルは、30億パラメータ規模の小規模言語モデルを甲骨文字のデータで再学習させたものである。研究陣は甲骨文字の推論データと嗜好性データを新たに作成して学習に組み込んだ。嗜好性データとは、人間がより優れていると選んだ回答を集めたデータである。その結果、小規模なモデルがはるかに巨大なモデルよりも優れた成果を収めた[5]。データを適切に選別すれば、規模の大きさだけに頼る必要はないことを示している。

かすれた筆画を復元して描くモデル

解読する前に、まず見えるようにする研究もある。2024年に発表されたOBSDがそうしたツールである。このツールは、画像を新たに生成する拡散モデルを用いている[6]。拡散モデルとは、かすれた斑点から次第に鮮明な画像を生成していく手法である。

OBSDは摩耗してかすれた甲骨文字の画像を入力として受け取る。そしてその文字がどの現代漢字につながるかを画像として提示する。この研究は2024年の国際学会で最優秀論文賞に選ばれた[6]。文字の解読を画像生成の問題へと転換した着想が高く評価されたのである。

割れた破片の断面をコンピュータがつなぎ合わせる

文字の解読と並行して、破片をつなぎ合わせる研究も進められた。2022年に公開されたOB-Rejoinは、綴合研究用のデータセットである。実際の甲骨拓本998枚が収録されている。同研究陣はS3-Netという接合ニューラルネットワークも同時に発表した[7]。

このニューラルネットワークは、破片の縁の曲線が互いにどれほど似ているかを計算する。二つの縁がかみ合うと点数が高くなる。機械は点数が高い順に候補を並べる。同じ研究チームは2020年に先行研究も発表した。そのときは候補10個の中に正解を入れる割合が98.39パーセントだった[11]。

パズルを合わせるとき、人は絵より破片の形を先に見る。機械も同じ方法を使う。ただし機械は一日に数十万回比べることができる。

機械が出すものは正解ではなく候補一覧だ。一覧の先頭が間違える場合もある。そのため研究チームは、10個の中に正解が入っているかを性能基準にした[7]。人が10個だけ確認すればよい状況と、10万個を確認しなければならない状況は違う。

首都師範大学甲骨文字研究センターは、河南大学とともに編綴プログラムを作った。名前は綴多多で、2020年から使われた。このプログラムは拓本18万枚を一度に比べて組み合わせを探す。研究センターはこの方式で、甲骨破片約1,500件の記録を整理した。以前は一つの破片の相手を探すのに8年から10年かかることもあった。今は数日で候補が出る[8]。

文単位へ移る試験

一つの文字を読むことと、文を理解することは別のことだ。2026年6月末に公開されたS-OBiは、この違いを測る試験問題集だ。研究チームは、専門家が確認した原本拓本95点から問題を選んだ。問題は全部で695個で、文単位で答える形になっている。そして複数の大規模AIに同じ問題を解かせた。試験結果では、モデルごとの成績差がはっきり出た。文全体の流れをつかむことがまだ難しい事実も明らかになった[9]。

この試験が教えた事実がもう一つある。甲骨文字の問題は、一般のAIがうまく解けないという事実だ[9]。世界の文章を広く学んだモデルも、3千年前の文の前では揺らぐ。学んだことのない文字と語法だからだ。そのため甲骨専用で再訓練する研究が続いている[5]。

割れた断面を越えて文脈をつなぐ研究も出た。ある研究チームは、二つの文がもともと同じ骨にあったかを判定する課題を作った。彼らはOBID-ACRという資料集を公開した。そして文字の形の情報と文の情報を一緒に使う方法を提案した[10]。破片を実際につなぐ前にも、文脈をつないでみる道が開かれたのである。

参考資料

- [1] 梁文艳. (2026, 5월 26일). 以科技赋能, 让甲骨文"活"起来. 中国青年报. https://zqb.cyol.com/pc/content/202605/26/content_426168.html (2026年9月3日閲覧)
- [2] Chen, Z., Hua, W., Li, J., Zhu, Y., Zhi, X., Liu, Z., Chen, T., Zhang, W., & Zhai, G. (2026). Oracle bone inscriptions information processing: a comprehensive survey. npj Heritage Science, 14, 220. <https://doi.org/10.1038/s40494-026-02511-w> (2026年9月3日閲覧)
- [3] Liu, Y., Guan, H., Wang, P., Wang, X., Wan, J., Zhang, K., Zheng, H., Liu, X., Kuang, Z., Yang, H., Li, B., Liu, Y., Jin, L., & Bai, X. (2026). AlphaOracle: Oracle bone script decipherment via human-workflow-inspired deep learning. The Innovation, 7(11), 101462. <https://doi.org/10.1016/j.xinn.2026.101462>
- [4] Zhang, J., Li, R., Pang, H., Xia, D., Zhu, Z., Zhang, Q., Li, C., & Yang, X. (2026, 4월 8일). Specializing large models for oracle bone script interpretation via component-grounded multimodal knowledge augmentation. arXiv:2604.06711. <https://arxiv.org/abs/2604.06711> (2026年9月3日閲覧)
- [5] Song, Z., Wang, Y., Ma, Z., Yu, Z., Wang, T., Zhang, J., Wang, T., & Yu, K. (2026, 6월 24일). OracleAnalyser: Analysing implicit semantics of oracle bone scripts through MLLMs with post-training. arXiv:2606.25906. <https://arxiv.org/abs/2606.25906> (2026年9月3日閲覧)
- [6] Guan, H., Yang, H., Wang, X., Han, S., Liu, Y., Jin, L., Bai, X., & Liu, Y. (2024). Deciphering oracle bone language with diffusion models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics. <https://aclanthology.org/2024.acl-long.831/> (2026年9月3日閲覧)
- [7] Zhang, C., Wang, B., Chen, K., Zong, R., Mo, B., Men, Y., Almpantidis, G., Chen, S., & Zhang, X. (2022). Data-driven oracle bone rejoining: A dataset and practical self-supervised learning scheme. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 4482-4492). ACM. <https://doi.org/10.1145/3534678.3539050>

- [8] 北京日报. (2026, 8月 28日). 高校团队"唤醒"三千年前文明碎片. 新浪教育.
<https://edu.sina.com.cn/l/2026-08-28/doc-inipwhrw0020843.shtml> (2026年9月3日閲覧)
- [9] Li, Z., Chen, Z., Chen, T., & Zhai, G. (2026, 6月 30日). Beyond single character: Evaluating MLLMs for sentence-level oracle bone inscription understanding. arXiv:2606.31169. <https://arxiv.org/abs/2606.31169> (2026年9月3日閲覧)
- [10] Zhang, H., Wang, T., Zhang, Z., Li, B., Sun, H., Hou, C., Wang, N., Yu, Y., Jiao, Q., Xiong, J., & Liu, Y. (2026). A multi-modal dataset and method for bone-level association prediction in oracle bone inscriptions. npj Heritage Science, 14, 19. <https://doi.org/10.1038/s40494-025-02282-w> (2026年9月3日閲覧)
- [11] Zhang, C., Zong, R., Cao, S., Men, Y., & Mo, B. (2020). AI-powered oracle bone inscriptions recognition and fragments rejoining. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20), 5309-5311. <https://doi.org/10.24963/ijcai.2020/779>

7.3. 明らかになった真実

一文字が変わると地図が変わった

AlphaOracleが出した最初の成果は一文字だった。研究チームは、長く議論されてきた一文字をこのシステムにかけた。後の漢字の劣へつながら文字だ[1]。システムは筆画の形と部品の構成を整理した。そしてその文字が出る別の文を探して並べた。続いて古い文献知識と照らし合わせ、意味を絞った。

システムが提案した答えは、この文字が地名または氏族名だというものだった。氏族とは、一人の祖先から分かれて出た親族の家を意味する。この文字を文を修飾する語として読んできた解釈とは違う結果だった。研究チームは、この解釈が古文字学者の検討を経たものだとして明らかにした[1]。

一文字の意味が変わると、その文字を含む文がすべて変わる。人名なら、その文は一個人の話だ。地名なら、同じ文が地方行政の記録になる。殷がその土地をどのように治めたのかも描き直す必要がある。研究チームは、この結果が殷の行政と社会を見る視点を見直させるものだとして報告した[1]。

この研究で注目すべきものは、答えそのものではない。システムが根拠を一緒に出した点だ。AlphaOracleは、どの文とどの部品を見てその答えを選んだのかを示す[1]。学者はその根拠をたどりながら、同意するかどうかを決めることができる。人が検討できる形で答えを出すことが、この研究の核心だった。

数年かかっていた組み合わせが数日に短くなった

編綴の側でも結果が出た。前に見た綴多プログラムが出した成果だ。このプログラムは、学者たちが長く見落としてきた組み合わせを300件以上見つけた[2]。100年以上にわたり手で探してきた組み合わせに加わった数だ。

人がする仕事はそのまま残った。学者は機械が出した一覧を一つずつ確認する。確認は実物の照合で終える。

安陽師範学院の研究室も独自の成果を出した。2026年5月までに新しく接合した甲骨は155件だ。同じ研究室が重複記録を見分けた事例は7,000件近くになる[3]。一つの骨が複数の図録に別々に載ることがあった。すると一つの骨が、互いに異なる二つの遺物のように数えられる。新しくつないだ数より、重複を見分けた数のほうがはるかに多い。機械がする仕事は新発見だけでなく、古い記録の点検でもあるという意味だ。

破片がつながると文が長くなる。片方には質問だけが残し、結果はもう片方にある場合が多い。二つの破片が会ってこそ、占った日と実際に起きたことが一行につながる。短い記録がいくつもあつたのではなく、一つの事件記録になるのである。

この違いは歴史叙述にすぐ影響する。切れた記録だけを見ると、事件が散らばって見える。つながった記録を見ると、前後の順序が現れる。王がいつ尋ね、いつ結果を受け取ったのかがつかめる。

重複を見分けた記録7,000件も同じ重みを持つ[3]。同じ骨を二つに数えると、資料の数からずれる。同じ文を互いに異なる二つの記録だと思って引用することもある。機械が拓本を大量に改めて比べながら、こうした重複を取り除いた。目立つ発見ではないが、基礎資料を正す仕事だ。

海を越えて散らばった骨が画面で出会う

編綴を妨げていた壁は距離だった。一つの骨から外れた破片が互いに別の国にあると、並べて見ることはできなかった。写真をやり取りしても、厚みと曲がり具合は分からなかった。3次元資料はこの壁を低くした。

前に見た海外資料900点余りが、この壁を越えた事例だ[3]。実物はもともとあった場所にそのままある。移ってきたものは表面の形の情報だ。

安陽の研究者は、画面でその破片を回しながら見ることができる。北京の破片と欧州の破片を同じ画面に置き、断面を合わせてみることもできる。

公開の方式も変わった。以前は所蔵機関を訪ねなければ遺物を見ることができなかった。許可を受けるだけで何か月もかかることがあった。今は学生も家で甲骨資料を開いて見ることができる[3]。資料を扱う人の数が増えれば、確認する目も増える。目が増えれば、誤ってつないだ組み合わせを取り除く機会も増える。

機械がまだできないこと

成果とともに限界もはっきりした。2026年6月末に公開されたS-Obi試験が、その限界を数字で示した。複数の大規模AIが同じ甲骨文の問題を解いた。一つの文字を認識する問題より、文全体を理解する問題で成績が下がった。原因も明らかになった。モデルは、無事に残っている文字から誤って見た。その誤りを抱えたまま推論を続けるため、答えがずれた[4]。3千年前の文の流れをつかむ仕事は、まだ人の役割が大きい。

さらに慎重でなければならない問題は、作り出した。AIは確率で答えを選ぶ。甲骨文字のように学習資料が少ない分野では、もっともらしいが実際にはなかった文字を作り出すことがある。殷の言葉は、比較できる周辺言語の資料も十分ではない[5]。辞書を調べて答えを検算しにくい状況だ。

資料の偏りも残っている。これまで集まった甲骨資料は、特定の時期と特定の発掘地に偏っている。偏った資料で学んだ機械は、なじみのない地域の甲骨に出会うと間違えやすい。2026年に出た総合整理論文も、資料の品質と均衡を今後の課題に挙げた[5]。

数字で見ると、道のりはまだ長い。意味が分からない文字は、なお3,000字ほど残っている。AlphaOracleが根拠をそろえて解いた文字は、そのうちの一つだった[1]。一文字を解くのにかった手間を考えると、残された仕事は大きい。それでも方法が整理されれば、次の文字は少し速くなる。

人と機械が分け合う仕事

そのため現在の研究方式は、人と機械を並べて置く方向に固まった。機械は候補を広く出し、根拠を整理する。人はその候補を文献と実物で確認する。

前に見た分析時間の短縮も、この分業の結果だ。機械が判断を代わりにしたのではなく、人が見る資料を先に絞ってくれた。専門家たちがこの道具を役に立つと答えた理由もここにある[1]。

編綴でも同じ方式が固まった。以前は何年もかかっていた照合を、プログラムが数日で終える[2]。学者はその候補を持って収蔵庫へ行く。実際の骨を突き合わせ、断面がつながるかを確認する。最後の判断は人の指先で行われる。

この分業には一つの条件が付く。機械が出した答えの根拠が人に見えなければならないという条件だ。根拠が見えなければ、学者は確認する方法がない。確認できない答えは、歴史叙述に入れることができない。AlphaOracleが段階を分けた理由もここにある[1]。

3千年前の殷の王は、骨に入ったひびを見ながら未来を問うた。今日の研究者は、同じ骨の割れた断面を見ながら過去を問う。問いの向きは反対だが、のぞき込む痕跡は今も同じだ。

参考資料

- [1] Liu, Y., Guan, H., Wang, P., Wang, X., Wan, J., Zhang, K., Zheng, H., Liu, X., Kuang, Z., Yang, H., Li, B., Liu, Y., Jin, L., & Bai, X. (2026). AlphaOracle: Oracle bone script decipherment via human-workflow-inspired deep learning. *The Innovation*, 7(11), 101462. <https://doi.org/10.1016/j.xinn.2026.101462>
- [2] 北京日报. (2026, 8월 28일). 高校团队"唤醒"三千年前文明碎片. 新浪教育. <https://edu.sina.com.cn/l/2026-08-28/doc-inipwhrw0020843.shtml> (2026年9月3日閲覧)
- [3] 梁文艳. (2026, 5월 26일). 以科技赋能, 让甲骨文"活"起来. 中国青年报. https://zqb.cyol.com/pc/content/202605/26/content_426168.html (2026年9月3日閲覧)
- [4] Li, Z., Chen, Z., Chen, T., & Zhai, G. (2026, 7월 1일). Beyond single character: Evaluating MLLMs for sentence-level oracle bone inscription understanding. arXiv:2606.31169. <https://arxiv.org/abs/2606.31169> (2026年9月3日閲覧)
- [5] Chen, Z., Hua, W., Li, J., Zhu, Y., Zhi, X., Liu, Z., Chen, T., Zhang, W., & Zhai, G. (2026). Oracle bone inscriptions information processing: a comprehensive survey. *npj Heritage Science*, 14, 220. <https://doi.org/10.1038/s40494-026-02511-w> (2026年9月3日閲覧)

第8章. 泥の穴から生還した筆跡： 新羅木簡とくずし字判読人工知能

8.1. 事件ファイル

木に文字を書いていた時代

紙が貴重だった時代があった。古代東アジアの人々は木を薄く削って整え、その上に文字を書いた。文字を記したこの木の板を木簡と呼ぶ。木簡は通常、指を2、3本合わせたほどの幅である。長さは手のひらを広げた幅ほど（約20センチ）のものが多い。木を柱のように削ったものもある。このような木簡は4面すべてに文字を書くことができる。複数の面に文字を書いた木の板という意味で、多面木簡と呼ぶ。

木簡の用途は様々であった。ひとつは品物に取り付ける荷札である。地方から王京へ穀物や塩辛を送る際、木簡を紐で結び付けた。そこには送った村の名や品名、数量を記した。もうひとつは官庁の文書である。役人の名前や等級、受け取った物品や日付を記した木簡が出土している[1]。文字を習得するために同じ漢字を何度も書いた練習用の木簡もある。

木は紙よりも硬いため再利用することができた。役目を終えた木簡は表面を小刀で薄く削り落とした。新しい面が現れると、その上に再び文字を書いた。このとき削り落とされた薄い木屑も遺跡から一緒に出土する。この削り屑を削屑と呼ぶ。文字を消すために削り取った木の屑という意味である。扶余官北里遺跡から出土した木簡類329点を見ればそれがわかる。そのうち247点が削屑であった。文書を作成し書き直す施設が近くにあったという意味である[1]。

歴史学者が木簡を重宝するのには理由がある。歴史書は通常、事件が終わった後に書かれる。執筆者の判断や王朝の都合が入り込む。木簡はその日の業務をその日のうちに記した記録に近い。後世に見せるために飾られた文章ではない。穀物何石を誰がいつ送ったかがそのまま残されている。事件が起きた当時に作られたこのような記録を一次史料と呼ぶ。木簡にも書き間違いや書き直した跡はある。そのため研究者は何点もの木簡を並べて比較対照する。

木簡は朝鮮半島と日本列島の双方から出土する。両地域ともに漢字で文書を作成していた社会であった。その一方で自国の言葉を書き記そうと工夫を凝らした。新羅の人々は漢字の意味と音を交えて自国語の語順で記した。この表記を吏読と郷札と呼ぶ。日本の人々も漢字の音を借りて自らの言葉を記した。木簡はこうした試みがいつ始まったのかを教えてくれる実物証拠である。紙の文書はほとんど残っていないからだ。両国の木簡を並べて比較する研究が続けられている理由がここにある。

泥が守り抜いた千年

木簡が出土する場所はおおむね決まっている。宮殿の傍らにある池の底である。城壁の内側に水が溜まった低湿地である。溝を埋めた泥の層である。人々が使って捨てた木が水中に沈み、そのまま埋もれたのだ。

水中の泥には空気がほとんど入り込めない。木を腐らせるのは主に空気を吸って生きる微生物である。空気が遮断されると、この微生物は活動できない。空気なしで生きる微生物が残ってはいるものの、木が朽ちる速度は著しく遅くなる。そのため千年を超えても木は形を保ち続けた。革や紙が跡形もなく消え去る土壌にあって、木簡は残った。墓ではなくゴミ捨て場が記録を守ってくれたのである。

その代わり、別の変化が生じた。木の内部へ水が染み込み、細胞壁を満たした。土の中の鉄分や多様な鉱物も一緒に入り込んだ。木は水をたっぷりと吸い込んだ状態になった。このような木を水を吸った木、すなわち飽水木材と呼ぶ。

飽水木材は取り出してそのまま乾燥させてはならない。水分が抜けるにつれて急激に収縮し、ひび割れてしまう。発掘現場で木簡を直ちに水桶に浸す理由がここにある。保存処理室では木の中の水分を薬品へとゆっくり置き換えていく。この工程だけでも数か月から数年を要する。判読はその前に済ませておかねばならない。処理を経ることで表面の色が変わってしまう恐れがあるからだ。

黒い木の上の黒い文字

鉱物が染み込んだ木は色が濃くなる。褐色を乗り越えて真っ黒に変色するものも多い。木が黒く変色するこの現象を黒化という。問題は文字の材料である。昔の人々は松を燃やした煤を膠で練り合わせて墨を作った。墨の主成分は炭素であり、色は黒い。

黒く変色した木の上に、黒い墨で書かれた文字がある。明るい照明をいくら当てても両者は見分けがつかない。その上、水中で長い年月を過ごすうちに墨の粒子がかなり洗い流されてしまった。木目の暗い縦線が筆跡と入り混じる。泥や腐植物が筆の通った溝を埋める。人間の目では文字があるのかどうかさえ判別が難しい。

判読が難しければ学説も分かれる。同じ木簡を巡って学者ごとに異なる漢字として読み解く事態が生じる。一文字変われば文章の意味が変わり、歴史の解釈も変わる。そのため木簡研究者たちは古くから新しい目を探し求めてきた。人間の目が見落とすものを捉えてくれる道具が必要だったのだ。

解読すべき木簡は増え続けている。慶南咸安の城山山城からは新羅の木簡が絶えず出土してきた。2025年8月にも銘文木簡2点が新たに確認された[2]。銘文とは遺物に記されている文字を意味する。忠南扶余の官北里遺跡からは百済泗沘期の木簡類が大量に出土した[1]。京畿楊州の大母山城からは5世紀のものと推定される木簡が出土した。事実であれば、これまでに知られている百済の文字資料の中で最も初期のものとなる[3]。

日本でも事情は同じである。奈良文化財研究所は各地から出土した木簡の情報を集めて公開している。このデータベースの名称は「木簡庫」である[4]。ここでは木簡に記された文字だけで検索するのではない。用途や年代、材質、出土地などからも検索できる[5]。解読すべき木がこのように積み重なっているのだ。

参考資料

- [1] 파이낸셜뉴스. (2026, 2월 5일). 부여 관북리 '백제 피리' 첫 확인..목간 329점 무더기 발견. 파이낸셜뉴스. <https://www.fnnews.com/news/202602051448564101> (2026年9月3日閲覧)
- [2] 경남일보. (2025, 8월 7일). 함안 성산산성서 새로운 목간 2점 명문 확인. 경남일보. <https://www.gnnews.co.kr/news/articleView.html?idxno=616028> (2026年9月3日閲覧)
- [3] 뉴시스. (2025, 11월 20일). 1500년 전 '백제 문자' 깨어났다...양주 대모산성서 5세기 목간 출토. 뉴시스. https://www.newsis.com/view/NISX20251120_0003410107 (2026年9月3日閲覧)
- [4] 奈良文化財研究所. 木簡庫(MOKKAN-KO). <https://mokkanko.nabunken.go.jp/ja/> (2026年9月3日閲覧)
- [5] カレントアウェアネス・ポータル. (2018). 奈良文化財研究所、木簡に関する情報を検索するシステム「木簡庫」(MOKKAN-KO)を公開. 国立国会図書館. <https://current.ndl.go.jp/car/35714> (2026年9月3日閲覧)

8.2. 科学捜査隊の追跡1：デジタル墨汁ルーペ

目に見えない光を用いる

人間の目は狭い範囲の光しか見ることができない。虹の赤の外側にも光が存在する。この光を赤外線と呼ぶ。波長が長い目には捉えられないが、カメラでなら撮影できる。リモコンがテレビに送る信号も赤外線である。

木簡の判読に赤外線を用いる理由は材料の性質にある。墨の炭素粒子は赤外線をよく吸収する。黒化した木目は赤外線を多く跳ね返す。人間の目には同じように黒く見える二つの物質が、赤外線カメラの前では分かれるのだ。吸収した側は暗く写り、反射した側は明るく写る。消え去ったと思われていた筆跡が写真の上に浮かび上がる。

主に使用する光は波長800ナノメートル前後の近赤外線である。ナノメートルは1メートルを10億で割った長さだ。人間が見る赤色光は700ナノメートル付近で終わる。そのすぐ外側の光を用いるという意味である。撮影は外光を遮った部屋で行う。光は一方向に偏らないよう均一に拡散させて照射する。片側からだけ照らすと、木目の筋ごとに長い影ができてしまうからだ。

赤外線撮影は木簡研究の基本手順となった。発掘報告書に掲載される木簡写真の隣には、大抵赤外線写真が並べて掲載される。奈良文化財研究所は赤外線カメラを用いて木簡の文字画像を鮮明にする装置を早くから開発してきた[1]。

赤外線からハイパースペクトルへ

赤外線写真1枚ですべての文字が蘇るわけではない。染みの位置や木目の状態によってよく見える波長が異なる。そこで登場した手法がハイパースペクトル画像である。

通常の写真は赤、緑、青の3つの値のみを保存する。ハイパースペクトル画像は異なる。ひとつの場面を数十から数百もの狭い波長帯に細かく分割して撮影する。光を虹の色のように細かく分けて別々に撮影すると考えればよい。画面の1点ごとに波長別の輝度値がずらりと記録される。そのため、1枚の写真の代わりに分厚い写真の束が作られることになる。

これらの値を比較すれば、墨が残っている場所と泥の染みが付着している場所を見分けることができる。墨と泥では波長ごとに明るさが変化するパターンが互いに異なるからだ。ある文字は900ナノメートル付近でよく見え、別の文字は他の波長帯で蘇る。研究者は波長帯を選びながら最も判読しやすい組み合わせを探す。コンピューターが複数の波長帯を自動で比較し、候補を絞り込むこともある。人間が1枚ずつ目で確認していた作業を計算によって削減したのだ。

国立伽耶文化遺産研究所は2025年にこの手法を初めて導入した。咸安城山山城木簡の判読にハイパースペクトル画像を採り入れたのである。同研究所はこの手法が従来の赤外線撮影よりも精密であると明らかにした。肉眼や通常の撮影では識別が困難な文字まで蘇らせることができるという説明だ[2]。

字形から文字を探す

写真が鮮明になっても次の問題が残る。不鮮明な筆跡がどの漢字であるかを特定せねばならない。草書体で崩して書かれた文字は形状が大きく変化する。画が数本しか残っていない文字も多い。漢字は種類が数万字にも及ぶため、人間が辞書を引くのにも限界がある。

奈良文化財研究所と東京大学史料編纂所はこの作業を支援するシステムを開発した。名称は「MOJIZO（もじ蔵）」である。利用者が解読したい文字画像を切り抜いてアップロードする。システムは木簡や古文書から収集した文字画像を検索し、その中から形の似ているものを提示する[3]。人間が候補を見て判断するのだ。正解を決める機械ではなく、類似の用例を提示する検索ツールである。

こうしたツールが変えたのは作業の手順である。以前は学者が記憶と字引に頼って候補を思い浮かべていた。現在は機械が数万件の中から似た文字をあらかじめ選び出してくれる。人間は絞り込まれた候補の中から文脈を吟味して選択する。判読に要する時間は大幅に短縮された。

機械が描いた画を見分ける

昨今の生成AIは、ぼやけた画像を鮮明にしてくれる。途切れた線をつなぎ、欠けた部分を補ってくれる。木簡の写真にこの技術を用いれば、見栄えの良い結果が得られる。だがここに落とし穴がある。

機械が補完した画は、木の上に実際に残っていた墨ではない。学習した数万点もの文字の平均から導き出された推測にすぎない。木目の模様が丸く残っていれば、機械はそれを丸みを帯びた画として認識してしまうことがある。そうなれば、本来存在しなかった文字がそれらしく生成されてしまう。史料が汚染され、歴史の解釈が歪んでしまうのだ。

そのため研究現場には明確なルールがある。原本の写真と補正結果を必ず並べて掲載する。実際に確認できる墨の跡と機械が生成した画素を区別して明示する。機械の出力は結論ではなく候補リストとして扱う。最終的な判読は人間が行う。判読案を発表する際には、どのような機器でいかなる条件の下で撮影したかも併せて明らかにする。

政府レベルでの準備も進行中である。国家科学技術諮問会議は2026年4月16日に国家遺産庁の施行計画を議決した。2026年から2030年にかけて人工知能、ロボット、センサーを国家遺産の調査と復元に投入する計画である。本年度の投資規模は140億ウォンである。そのうち44億ウォンが知能型先端保存技術の開発に割り当てられた[4]。国立金海博物館は2026年6月23日に学術シンポジウムを開催した。テーマは「伽耶遺産研究の科学的視線」であった。非破壊分析、デジタル復元、3次元X-ray顕微鏡技術を扱う7件の発表が行われた[5]。肉眼では見ることでできなかった遺物の内部を覗き込む手法が、木簡の枠を超えて広がりを見せている。

参考資料

- [1] 奈良文化財研究所. 赤外線カメラを利用した木簡文字画像鮮明化システムの開発. 学術機関リポジトリデータベース. <https://irdb.nii.ac.jp/01144/0004944317> (2026年9月3日閲覧)
- [2] 경남일보. (2025, 8월 7일). 함안 성산산성서 새로운 목간 2점 명문 확인. 경남일보. <https://www.gnnews.co.kr/news/articleView.html?idxno=616028> (2026年9月3日閲覧)
- [3] 奈良文化財研究所, 東京大学史料編纂所. 木簡・くずし字解読システム MOJIZO. <https://aimojizo.nabunken.go.jp/> (2026年9月3日閲覧)
- [4] 뉴시스. (2026, 4월 16일). "AI, 블록체인으로 국가유산 보존관리"...과기자문회의, 유산청 시행계획 의결. 뉴시스. https://www.newsis.com/view/NISX20260416_0003594216 (2026年9月3日閲覧)
- [5] 경남일보. (2026, 6월 17일). 국립김해박물관, 23일 '가야학술제전' 심포지엄 개최. 경남일보. <https://www.gnnews.co.kr/news/articleView.html?idxno=639314> (2026年9月3日閲覧)

8.3. 科学捜査隊の追跡2：KuroNetとMiwo

続け字は切り離せない

日本には江戸時代までに記された古文書が数百万冊残されている。これらの文書は「くずし字」と呼ばれる崩し書きの筆文字で記されていた。くずし字は8世紀から千年以上使われ続けた筆記体である。1900年に日本が文字制度を改定した。それ以降、学校でこの文字が教えられることはなくなった。そのため現代の日本人の大半は150年前の本すら読むことができない[1]。

解読が難しい理由は文字の形状だけではない。筆を紙から離さずに複数の文字をつなげて書く書法がある。前の文字の最後の画と次の文字の最初の画が一本につながる。そうすると文字と文字の間に境界線が残らない。一行が丸ごと一本の長い線のようにも見える。

コンピューターによる文字認識は、長年にわたり同じ順序で作動してきた。まず文字を一つひとつ四角い枠で切り抜く。次に切り抜いた枠内の画像がどの文字であるかを判定する。この手法は境界線があって初めて機能する。続け書きされたくずし字の前では、最初の段階から破綻してしまった。前段階が間違っていれば後段階も間違えるため、精度を向上させる手立てがなかった。

文字を切り分けないKuroNet

人文学オープンデータ共同利用センターの研究陣が別の道を切り拓いた。ターリン・クラヌワット、アレックス・ラム、北本朝展が共同でモデルを構築した。その名はKuroNetである。2019年の文書分析認識国際会議で初めて発表された[1]。2020年には構造を改良し、その成果を学術誌に改めて掲載した[2]。

KuroNetは文字をあらかじめ切り出さない。古文書の1ページを丸ごと入力として受け取る。モデルの中心構造はU-Netである。U-Netは画像を少しずつ縮小して特徴を抽出する。その後、再び元の大きさへ拡大するニューラルネットワークである。縮小する側と拡大する側が対称なので、アルファベットのU字形に描かれる。縮小する間に筆画の太さと方向を学ぶ。拡大する間にその情報を元の位置へ戻す。

このニューラルネットワークは、ページ全体から二つのことを同時に予測する。一つは文字の中心がどこにあるかである。もう一つは、その位置の文字が何かである。結果は確率マップとして出る。文字がある可能性を画面の各所に示した画像である。確率が高い場所をたどると、文字の位置と名前が同時に決まる。切り出す段階がないので、続け書きの文字でも崩れない。

学習には日本の国文学研究資料館が作った資料を使った。日本古典籍くずし字データセットという名前である。100万字を超える文字標本が入っている[2]。同じ研究陣は、機械学習の研究者が使いやすいように縮小版も公開した。手書き数字資料をまねて作ったKuzushiji-MNISTがそれである。文字数を増やした版と、漢字だけを集めた版もあわせて公開した[3]。そのおかげで、日本古典を知らない研究者もこの問題に取り組めるようになった。

KuroNetは2022年10月に役目を引き継いだ。認識モデルの名前はRuriに変わった[4]。Ruriは物体検出技術をくずし字に合わせて整えたモデルである。文字の背後の色や模様が複雑な場合でも、以前よりよく読める。同じデータセットを補強して学習し、以前は読めなかった文字を読み取る[5]。

手の中に入ったMiwo

サーバーで動いていたモデルを携帯電話に入れることが次の課題だった。そうして生まれたアプリがMiwoである。2021年8月30日に初めて公開された[5]。日本語表記はMiwoで、英語表記はMiwoである[6]。

使い方は簡単である。古い文書を携帯電話のカメラで撮る。しばらくすると、画面に現代の文字へ移した結果が現れる。原本と結果をスライダーで重ねて比べることができる。変体仮名が何だったのかも表示される。変体仮名は、現在は使わない古いひらがなの形である。結果を文字ファイルとして書き出す機能もある。撮影した写真はサーバーに残さない。AndroidとiPhoneのどちらでも無料で使える[6]。

2022年10月26日に出た1.1版から、認識モデルがRuriに変わった[5]。同じ月、このアプリは2022年度グッドデザイン賞を受けた[7]。利用記録は着実に積み重なった。公開後に認識した画像は、2026年9月3日時点で3,523,373件である[6]。図書館、寺院、地域の文書館で人々が古い文書を撮っているという意味である。

市民がともに読む

機械が先に読み、人が直す方式はほかの場所でも使われている。みんなで翻刻というオンライン翻刻事業がある。翻刻とは、古い文字を現代の文字に移し書きする作業である。表記はMinna de Honkokuで、意味はみんなで移し書きすることである。2017年に京都大学古地震研究会で始まった。現在は国立歴史民俗博物館と東京大学地震研究所が共同で運営している[8]。

参加方法は開かれている。インターネットブラウザーさえあれば、誰でも古い文書を移し書きできる。画面にはAIが先に読んだ下書き結果が表示される。人は誤った文字だけを直せばよい。このサイトは人文学オープンデータ共同利用センターのくずし字認識サービスを取り込んで使っている。これまでに参加者約10,800人が4,800万字を超える文章を移し書きした[8]。

移し書きした文章は、ただ蓄積されるだけではない。人が直した正解が再び学習資料になる。学習資料が増えればモデルがよくなる。モデルがよくなれば下書き結果が正確になり、人の手が少なくてすむ。この循環が回り始めると速度がつく。

読み取った文書の使い道も広い。この事業は地震研究者たちが始めた。古い日記や官庁文書には、地震、津波、凶作の日付と被害が記されている。くずし字のために手をつけられなかった記録がデータに変われば、過去の災害周期を数えることができる。古い筆文字が今日の防災資料になる。

ここまでが日本の紙の古文書の話である。KuroNet、Ruri、Miwoはいずれも紙に書かれたくずし字を読むモデルである。泥の中の木簡をこのモデルが自動で読み取った事例は、まだ確認されていない。木簡は文字数が少なく、基盤が木であり、墨が多く消えている。学習に使う正解資料も、紙の古文書ほど蓄積されていない。韓国と日本の木簡データがさらに集まれば、状況は変わりうる。今は期待であって、成果ではない。

参考資料

- [1] Clanuwat, T., Lamb, A., & Kitamoto, A. (2019). KuroNet: Pre-modern Japanese Kuzushiji character recognition with deep learning. 2019 International Conference on Document Analysis and Recognition (ICDAR), 607-614. <https://doi.org/10.1109/ICDAR.2019.00103>

- [2] Lamb, A., Clanuwat, T., & Kitamoto, A. (2020). KuroNet: Regularized residual U-Nets for end-to-end Kuzushiji character recognition. SN Computer Science, 1, 177. <https://doi.org/10.1007/s42979-020-00186-z>
- [3] Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., & Ha, D. (2018). Deep learning for classical Japanese literature. arXiv. <https://arxiv.org/abs/1812.01718>
- [4] ROIS-DS 人文学オープンデータ共同利用センター. KuroNetくずし字認識サービス(AI OCR). CODH. <https://codh.rois.ac.jp/kuronet/> (2026年9月3日閲覧)
- [5] ROIS-DS 人文学オープンデータ共同利用センター. みを(miwo)とは? CODH. <https://codh.rois.ac.jp/miwo/about/> (2026年9月3日閲覧)
- [6] ROIS-DS 人文学オープンデータ共同利用センター. みを(miwo): AIくずし字認識アプリ. CODH. <https://codh.rois.ac.jp/miwo/> (2026年9月3日閲覧)
- [7] 国立情報学研究所. (2022, 10月 26日). 世界初のAIくずし字認識アプリ「みを(miwo)」が2022年度グッドデザイン賞を受賞. 国立情報学研究所. <https://www.nii.ac.jp/news/release/2022/1026.html> (2026年9月3日閲覧)
- [8] みんなで翻刻. みんなで翻刻について. <https://honkoku.org/> (2026年9月3日閲覧)

8.4. 明らかになった真実

月池木簡が知らせた王室の暮らし

慶州の東宮と月池は、統一新羅王室の離宮と池である。1975年の月池発掘で木簡が初めて出土した。その後も調査が続き、数量が増えた。国立慶州博物館は、ここから出た遺物を月池館で展示している[1]。

月池木簡は王室の一日を見せてくれる。ある木簡は壺に吊るした札である。札木簡は端の左右に溝が彫られている。その溝にひもを巻き、物に結びつけた。今日の宅配伝票と用途が似ている。月池から出た木簡には、nokgapとongという文字がある。nokgapは鹿で作った塩辛で、ongは壺である。新羅王室が塩辛の種類と漬けた時期を記して管理していたという意味である[2]。

このような記録は歴史書には出てこない。三国史記は王、戦争、制度を記した。台所に何が入ってきたかは記さなかった。木簡はその隙間を埋める。泥の中の木板一枚が宮殿の家計簿になる。

韓国木簡学会と国立慶州博物館は、2026年4月3日と4日に国際学術大会を開いた。主題は新羅王京慶州の空間と文字生活だった。慶州華白コンベンションセンターで二日間行われた。月城の濠、東宮、月池から出た資料を一緒に並べた。木簡、碑石、墨書によって新羅王京の姿を描き直そうとする試みだった。韓国と日本の研究者がともに発表した[3]。

城山山城木簡の処罰報告

咸安城山山城は、新羅が6世紀に築いた山城である。城壁の内側の低湿地から木簡が続けて出土した。ここから出た墨書木簡は、2025年8月時点で247点である。韓国で出土した古代木簡の半数を超える数量である[4]。

第18次発掘調査で新たに判読した二点は、マツ類で作られていた。一点は四つの面に文字を書いた多面木簡である。もう一点は両面木簡である[4]。多面木簡は、四つの面のうち三つの面に行政内容を収めていた。人に処罰を下した記録である[5]。作られた時期は6世紀半ばと見られる[4]。

この判読は、前に見たハイパースペクトル画像で得た結果である。目でも通常写真でも見分けられなかった文字がよみがえった[4]。千年以上泥の中にあった新羅下級官吏の報告書が、文章として戻ってきた。城山山城木簡には、穀物を送った村の名前も多く記されている。6世紀の新羅が地方をどのように分けて治めたかを示す資料である。税がどの道を通って城まで来たのかもたどることができる。

百済王宮の文書の山

扶余官北里遺跡は、百済泗泚期の王宮跡として知られる場所である。国立扶余文化遺産研究所は、2024年と2025年に第16次発掘調査を行った。ここで木簡類329点が出土した。文字が記された木簡が82点で、削屑が247点である[6]。国内の単一遺跡としては最大数量であり、泗泚期木簡の中では早い時期のものである[7]。

木簡には年代を記したものが混じっている。庚申年と癸亥年という干支が出る。それぞれ540年と543年と見られる。百済が熊津から泗沘へ都を移した直後の記録という意味である。官吏の人事記録、国の財政を記した帳簿、官職名があわせて確認された[7]。竹で作った百済の管楽器の実物も初めて出土した[8]。

2026年7月22日と23日には国際学術大会が開かれた。国立扶余文化遺産研究所、韓国木簡学会、百済学会が共同で用意した。主題は扶余官北里遺跡の楽器と木簡だった。釜山海雲台区で開かれた。朴淳發忠南大学名誉教授の基調講演で始まった。三つの分科で十二本が発表された。韓国、中国、日本の研究者が官北里木簡の意味をともに検討した[8]。

碑石の前では慎重になる

木簡で通じた方法が古い文字すべてに通じるわけではない。基盤が石に変わると事情が変わる。広開土大王碑がその事例である。414年に建てられたこの碑石には、高句麗の歴史が刻まれている。碑面は長く風雨にさらされ、文字がすり減った。ここでは墨ではなく、彫られた溝を読まなければならない。

そのため研究者たちは拓本を比べる。原石拓本は、碑面に石灰を塗る前に採った拓本である。もとの石に紙を当て、直接たたいて採った。これまで確認されたものは十数本である。19世紀後半には別のものが出た。文字をはっきり見せるため、碑面に石灰を塗って採った石灰拓本である。双鉤加墨本というものも複数作られた。碑文に紙を当て、文字の周囲に線を引いたものである。その余白を濃い墨で埋め、拓本のように見せた。これはたたいて採った拓本ではなく、写し描いた複製本である。作る間に文字が誤って移されることがある[9]。そのため同じ文字が拓本ごとに違って見えることが生じた。

論争の中心は辛卯年条の文章である。この文章の解釈が古代の日韓関係史と絡み、長く争われた。文字が人の手で変えられたかどうか争点だった[10]。本来の文字を知るには、早い時期の原石拓本と複数の石灰拓本を比べなければならない。各拓本を見分ける基準を立てる作業が先である[9]。

3次元計測とAIがこの論争に答えを与えたという話が出回っている。確認してみると、根拠として示された研究を見つけられない場合が多い。木簡で通じた方法が碑石でも通じるだろうという期待と、実際の研究結果は違う。前に見た技術は、見えなかったものを見えるようにする。見えるようにしたものの意味を定める作業は、その次である。

したがって結論は二重である。赤外線とハイパースペクトル画像は、新羅と百済の木簡で消えた墨をよみがえらせた。無分割認識モデルは、日本の紙の古文書にあった分けられないくずし字を読んだ。二つの技術が扱った対象は違うが、方向は同じである。人の目が見逃した筆画を機械が先に拾い出す。同時に、機械が描いた筆画と原本に残った筆画を区別する規律がなければ、同じ技術がなかった歴史を作る。筆文字を目覚めさせたのは機械であり、その文字に責任を持つのは人である。

参考資料

- [1] 국립경주박물관. 월지관. 국립경주박물관. <https://gyeongju.museum.go.kr/kor/html/sub02/0201.html> (2026年9月3日閲覧)
- [2] 국사편찬위원회. 목간. 우리역사넷. https://contents.history.go.kr/mobile/tt/view.do?levelId=tt_b36 (2026年9月3日閲覧)
- [3] 경북일보. (2026, 3월 31일). 1500년 전 신라의 목소리 깨어난다...경주서 국제 목간 학술대회. 경북일보. <https://www.kyongbuk.co.kr/news/articleView.html?idxno=4068614> (2026年9月3日閲覧)
- [4] 경남일보. (2025, 8월 7일). 함안 성산산성서 새로운 목간 2점 명문 확인. 경남일보. <https://www.gnnews.co.kr/news/articleView.html?idxno=616028> (2026年9月3日閲覧)
- [5] 경남도민일보. (2025, 8월 7일). 누군가를 처벌한 신라 행정 보고, 함안 성산산성 목간서 확인. 경남도민일보. <https://www.idomin.com/news/articleView.html?idxno=943461> (2026年9月3日閲覧)
- [6] 파이낸셜뉴스. (2026, 2월 5일). 부여 관북리 '백제 피리' 첫 확인..목간 329점 무더기 발견. 파이낸셜뉴스. <https://www.fnnews.com/news/202602051448564101> (2026年9月3日閲覧)
- [7] 경향신문. (2026, 2월 17일). 부여 관북리서 발견된 목간 329점, '백제 피리'만큼이나 중요한 이유. 경향신문. <https://www.khan.co.kr/article/202602171427001> (2026年9月3日閲覧)
- [8] 이데일리. (2026, 7월 21일). 백제 왕궁터 출토 목간 329점·횡적 발굴 성과, 국제학술대회서 조명. 이데일리. <https://edaily.co.kr/News/Read?mediaCodeNo=257&newsId=03276726645516816> (2026年9月3日閲覧)

[9] 국립중앙박물관. 광개토대왕비 탁본. 국립중앙박물관 큐레이터 추천 소장품.

<https://www.museum.go.kr/MUSEUM/contents/M0501000000.do?schM=view&relicRecommendId=254462>

(2026年9月3日閲覧)

[10] 국사편찬위원회. 광개토대왕릉비. 우리역사넷. https://contents.history.go.kr/mobile/kc/view.do?levelId=kc_r100300

(2026年9月3日閲覧)

第9章. エーゲ海の失われた謎：線文字Aと幾何学的格子

9.1. 事件ファイル

クレタ島の宮殿から出た三つの文字

エーゲ海南部にクレタ島がある。青銅器時代に、この島でミノア文明が育った。ミノア人は船を造り、地中海を行き来した。彼らはオリーブ油、ワイン、穀物を運んだ。宮殿はこれらの品物を集め、人々に分け与えた。

1900年ごろ、英国の考古学者アーサー・エヴァンズがクノッソス宮殿跡を発掘した。彼は地中から文字が刻まれた粘土板を見つけた。文字は一種類ではなかった。エヴァンズは見つけた文字を三つに分けた。クレタ象形文字、線文字A、線文字Bである[1]。

ミノア文明は千年以上続いた。その後、青銅器時代後期に大きな変化を経験した。北のサントリーニ島で火山が噴火したことが、その一つである。噴火した時期は紀元前17世紀から紀元前16世紀の間と見られている。正確な年代は今も議論中である[7]。火山灰がクレタ島まで飛来し、津波が海岸を襲ったと見られている。紀元前1450年ごろには、ミケーネ・ギリシア人が島を支配した[1]。火山とミケーネの支配の間には、百年以上の時間がある。二つの出来事をすぐに原因と結果として結びつけることは難しい。その後、ミノア人の言葉は記録から消えた。

三つの文字は使われた時期がそれぞれ異なる。線文字Aは紀元前1750年ごろから紀元前1450年ごろまで使われた。線文字Bはそれより遅く現れた。ミケーネ・ギリシア人がクレタ島を支配した後に使われた文字である。二つの文字は形の似た記号をいくつも共有している[1]。

粘土板が残ったのは偶然だった。ミノア人は粘土板を焼かなかった。湿った土に文字を刻み、乾かした後、一年ほど使って捨てた。ところが宮殿で火災が起き、土板が陶器のように焼かれた。火で焼かれた板だけが三千五百年を耐えた。今日残った資料は、火が残した資料である。

エヴァンズが分けた三つの文字のうち、クレタ象形文字が最初に現れた。人の頭や器の形がそのまま残っている文字である。線文字Aは、その絵を線で簡略化した姿に近い。二つの文字は二百年近く同じ時期に併用された。一方が消えた後に他方が生まれたのではない。線文字Aがクレタ象形文字をまねて作られたと見る見解がある。しかし三つの文字がまっすぐ一列につながると断定することはできない。記号を共有していることは確かだが、つながった道筋はまだ明らかになっていない[1]。

線文字Bは読まれた

線文字Bは1950年代に読まれた。英国の建築家マイケル・ヴェントリスが成し遂げた。彼は線文字Bが記した言葉がギリシア語の古い形であることを明らかにした。古典学者ジョン・チャドウィックが彼の研究を助けた[1]。

ヴェントリスは一つの表を描いて始めた。横列に子音を、縦列に母音を置いた表である[8]。このような表を音節格子と呼ぶ。音のまとまり一つ一つが入る枠をあらかじめ用意した図である。ヴェントリスは、どの記号がどの枠に入るかを統計で絞り込んでいった[1]。

彼は絞り込んだ枠にクノッソスやファイストスのような地名を当てはめた。地名が合うと、残りの記号も次々に解けた。学者たちは同じ方法を線文字Aにも当てはめてみた。線文字Bの音価を、似た記号にそのまま移してみたのである[3]。音は出た。しかしその音がどのような言葉なのか、誰にも理解できなかった[1]。

音価を移しても意味が出ない状況はなじみにくい。文字を読むことと言葉を理解することは別だからである。ハングルを知る人は、なじみのない外国語をハングルで書いても読むことができる。音は出るが、意味は分からない。線文字Aの前に立つ学者たちの置かれた立場もこれと同じである。

二つの言語で書かれた石がない

古い文字を読み解くとき、学者たちはたいてい対になる文書を探す。同じ内容を二つの言語で並べて書いた文書である。このような文書を二言語碑文と呼ぶ。エジプト象形文字はロゼッタ・ストーンのおかげで解かれた。ロゼッタ・ストーンには同じ法令が三つの文字で刻まれていた。そのうち一つが、学者たちがすでに読めた古代ギリシア語だった。

メソポタミアの楔形文字も似た道をたどった。イランのベヒストウンの崖に、三つの言語で刻まれた碑文が残っていた。この碑文は三つの言語すべてが、最初は読めなかった。そのうち古代ペルシア語が先に解かれた。記号が少なく、語と語の間が分かっていたからである。学者たちは先に解いた古代ペルシア語を足場にして、残りを読んだ。

クレタ島ではこのような文書は出ていない。線文字Aの横にギリシア語訳を書いた遺物はない。ミノア人が残した文字は、ミノア人の言葉だけで書かれている。比べる相手がないため、意味を確認する道がふさがる[1]。

手がかりになりそうな印も見えない。エジプト象形文字を読んだシャンポリオンにはカルトゥーシュがあった。王の名を囲む細長い枠である。この枠の中にある文字は人名だという事実が、あらかじめ定まっていた。シャンポリオンはここからいくつかの音価を拾い、残りを解いた。線文字Aの粘土板には、このような枠がない[1]。

問題はもう一つある。線文字Aが記した言葉の根を知らないという点である。学者たちはこの言葉をミノア語と呼ぶ。ミノア語がどの語族に属するかは明らかになっていない。インド・ヨーロッパ語族だという主張もあり、西アジア系統だという主張もある。その中で根拠が比較的良好とされるのは、セム語系統とリキア語系統である。リキア語は現在のトルコ南西部で使われていた古い言葉である[1]。どちらも学界の同意を得ていない。

線文字Aはクレタ島だけから出るわけではない。エーゲ海の多くの島々とギリシア本土でも発見された[1]。火山に埋もれたサントリーニ島の遺跡からも出た。ミノア人が行き来した航路に沿って、文字も広がったのである。青銅器時代の地中海がどれほど広くつながっていたかを示す資料である。

学者たちが線文字Aに取り組み続ける理由もここにある。線文字Aはヨーロッパの地で出た古い文字に入る。ミノア人が自分たちの言葉で残した記録はこれだけである。線文字Aが解ければ、彼らが何を信じ、何を売り買いたかが明らかになる。ギリシア神話に残るミノタウロスの物語の根もここにある。

1,400点の短い帳簿

残っている資料の量も十分ではない。銘文とは、遺物に刻んだり描き入れたりした文字をいう。これまで確認された線文字Aの銘文は約1,370点である。そこに書かれた記号をすべて合わせても7,400個前後である[1]。エジプトやメソポタミアの記録物とは比べられない量である。

文章の長さも短い。粘土板の大半は倉庫の帳簿である。何がいくつ入ってきて、誰にどれだけ出ていったかを記した。人名、物品名、数字が並んでいるだけである。文らしい文は少ない。文がなければ、文法を突き止める方法も減る。

記号の数を数えるときは、層を分けなければならない。音のまとまりを記す記号は97個である。物の名前をまると表す記号は50個ほどである。数を記す記号が4個、分数を記す記号が17個ある。二つの記号を重ねて作った組み合わせも約150個知られている[6]。これらをすべて集めた標準記号目録は300個に達する[2]。このうち頻繁に出る記号は多くない。残りは数回出て終わる記号である。コンピューターに統計を任せるには、同じ記号が何度も出てくる必要がある。数回しか出ない記号は、統計が捉えられない。

そこで研究者たちは資料そのものを作り直すことにした。手で模写した図を、コンピューターが扱える形に変える作業である[2]。2026年には、刻み跡を光で読み直す方法も出た[4]。目で見落としていた浅い溝まで浮かび上がらせる方法である。エーゲ海の沈黙を呼び覚ます仕事は、このように資料を整えることから再び始まった。

参考資料

- [1] Braović, M., Krstinić, D., Štula, M., & Ivanda, A. (2024). A systematic review of computational approaches to deciphering Bronze Age Aegean and Cypriot scripts. *Computational Linguistics*, 50(2), 725-779.
https://doi.org/10.1162/coli_a_00514
- [2] Salgarella, E., & Castellan, S. (2021). SigLA: The Signs of Linear A. A palaeographical database. In *Grapholinguistics in the 21st Century 2020: Proceedings (Grapholinguistics and Its Applications, Vol. 5, pp. 945-962)*. Fluxus Editions.
<https://sigla.phis.me/> (2026年9月3日閲覧)
- [3] Younger, J. G. (n.d.). Linear A texts and inscriptions in phonetic transcription. University of Kansas.
<http://people.ku.edu/~jyounger/LinearA/> (2026年9月3日閲覧)
- [4] Giorgi, L., & Lopez, S. (2026). Aegean Bronze Age writing systems through RTI: New potential and perspectives. *Digital Applications in Archaeology and Cultural Heritage*, 41, e00508.
<https://www.sciencedirect.com/science/article/abs/pii/S2212054826000159> (2026年9月3日閲覧)
- [5] GreekReporter. (2026, 3月 14日). New technology reveals hidden secrets of 3,500-year-old Bronze Age Aegean scripts. [greekreporter.com](https://greekreporter.com/2026/03/14/new-technology-reveals-hidden-secrets-aegean-bronze-age-scripts/).
<https://greekreporter.com/2026/03/14/new-technology-reveals-hidden-secrets-aegean-bronze-age-scripts/> (2026年9月3日閲覧)
- [6] Del Frio, M. (2022). Linear A. Mnamon: Ancient writing systems in the Mediterranean. *Scuola Normale Superiore*.
<https://doi.org/10.25429/sns.it/lettere/mnamon022> (2026年9月3日閲覧)
- [7] Pearson, C. L., Brewer, P. W., Brown, D., Heaton, T. J., Hodgins, G. W. L., Jull, A. J. T., Lange, T., & Salzer, M. W. (2018). Annual radiocarbon record indicates 16th century BCE date for the Thera eruption. *Science Advances*, 4(8), eaar8241.
<https://doi.org/10.1126/sciadv.aar8241>
- [8] Stavrakakis, E. (2024). Ventris' Grids. *Bulletin of the Institute of Classical Studies*, 67(1), 86-103.
<https://doi.org/10.1093/bics/qbae033>

9.2. 科学捜査隊の追跡：教師なし音韻格子合わせ

辞書なしで対応を探す機械

翻訳機を作るには、ふつつ対訳文が必要である。日本語の文と英語の文を並べて積み上げ、機械に見せる。このような資料を並列コーパスと呼ぶ。線文字Aには並列コーパスがない[3]。

そこで研究者たちは、正解表なしで学ぶ方法を選んだ。正解をあらかじめ与えず、統計だけで規則を探させる方式である。このような方式を教師なし学習と呼ぶ。機械は記号がどのような順序でつながって出るかを数える。そして別の文字体系の統計と比べ、対応を探してみる。

2019年に米国マサチューセッツ工科大学の研究陣がこの方法を発表した。ジアミン・ルオ、ユアン・チャオ、レジーナ・バルジライが共同で作ったモデルである。彼らは二つの文字体系をつなぐ仕事を、最小費用流問題に変えた。記号と記号の間に水路を置き、水が流れる値を付ける方式である。値が最も小さく済む水路が、正しい対応である確率が高い[1]。

研究陣はこのモデルをウガリット文字にかけてみた。従来の最高成績を5パーセントポイント上回った。線文字Bにもかけてみた。ギリシア語と根が同じ語のうち67.3パーセントを正しく移した[1]。ただし線文字Bには答えがあった。その文字がギリシア語を記したという事実を、研究陣はあらかじめ知っていた。

研究陣が線文字Bを試験台に選んだのには理由がある。答えを知っている問題でなければ、モデルの成績を測れない[1]。答えのない問題にすぐかけると、結果が正しいかどうかを知る方法がない。だからすでに読まれた文字で先に性能を測り、未解読文字へ移っていく[3]。

モデルが出すものは完成した翻訳文ではない。記号ごとに付きそうな音価の候補一覧である[1]。人がこの一覧を受け取り、遺物と比べながら選り分ける。コンピューターは人が見る分量を減らす仕事を担う[3]。

音の形を座標に変える

線文字Aは事情が違う。どの言葉と比べるべきかから明らかになっていない[3]。2021年に同じ研究室が、この条件を扱うモデルを発表した。ルオ、フレデリック・ハートマン、エンリコ・サントスが、バルジライ、チャオとともに作った[2]。

このモデルは二つの難しさを同時に扱う。第一の難しさは語境界がない点である。古い文字には分かち書きがない場合が多い。第二の難しさは親族言語が定まっていない点である[2]。

研究陣は音の形を数値座標に変えて入れた。国際音声記号をもとに、子音と母音の性質を座標に置いたのである。唇で出す音どうしは座標が近い。舌根で出す音とは座標が遠い。この座標があれば、機械が見当違いの対応を作りにくくなる[2]。

研究陣はこのモデルをゴート語とウガリット文字で試した。その次に、まだ意味が解けていないイベリア語にかけた。イベリア文字は記号の音価がかなり知られている。解けていないのは、その音が記している言葉である。研究陣はイベリア語がバスク語と親族であるという証拠を見つけられなかった。この結果は学界の主流見解とずれていなかった。どの言語が親族であり得るか、候補を比べる手順も合わせて示した[2]。同じ方法は線文字A研究にもそのまま使える[3]。

空欄が教える規則

機械が比べるのは記号の形だけではない。記号が隣とどのように並ぶかという傾向も計算に入る。ある記号は語頭にだけ出る。ある記号は末尾にだけ付く。二つの記号が並んで付くことがまったくない場合もある[3]。

このような統計を、共に現れる度合い、すなわち共起統計と呼ぶ。共起統計を集めると、表の空欄が見えてくる。同じ横列に置かれた記号は、子音が同じである確率が高い。同じ縦列に置かれた記号は、母音が同じである確率が高い[8]。ヴェントリスが手でしていた仕事を、コンピューターがはるかに速く繰り返すのである[3]。

コンピューターはこの計算を人よりはるかに速く終える。記号が300個なら、対応の数は数万通りになる。人が紙に書きながら数えるには手に余る量である。ヴェントリスはこの仕事を何年もかけて成し遂げた[3]。

ここまでは音の仕組みを描く仕事である。仕組みを描いても、その音が何を意味するかは出てこない。格子を埋めることと言葉を理解することは別である[3]。

文字の形をデータにする

機械が統計を出すには、資料が機械の読める形でなければならない。紙の本に載った図だけでは計算させることができない。この仕事を担ったのがSigLAという資料窓口である。ケンブリッジのエステル・サルガレッタとレンヌのシモン・カステランが共同で作った[4]。

SigLAには標準記号300個と、手で模写した銘文400点余りが収められている。銘文の中では、個別の記号3,000個以上を確認できる。資料と図版は誰でも使えるように公開されている。記号一つの線を切り出して比べることもできる[4]。

2026年3月に発表された研究は、資料の層をさらに一段広げた。ラヴィニア・ジョルジとサラ・ロペスが反射変換イメージングをエーゲ海の文字に適用した。同じ遺物に光をさまざまな方向から当て、写真を何枚も撮る方法である。これらの写真を合わせると、表面の細かな起伏が現れる。彫刻刀がどの線を先に刻んだかまで調べることができる[7]。

こうした資料が蓄積されると、ほかの文字にも役立つ。2022年にボローニャ大学の研究陣がキプロ・ミノア文字を見直した。ミケーレ・コラツツァ、ファビオ・タンブリーニ、ミゲル・ヴァレリオ、シルヴィア・フェラーラによる共同研究である。彼らは事前知識を入れないニューラルネットワークで記号の形を比べた。その結果、三つの群に分けてきた従来の分類を再検討すべきだという結論に至った。一つの文字体系と見るほうが統計に合うということである[5]。

コンピューターが出した結果は、すぐに論文になるわけではない。考古学者と文献学者が見直す過程を経る[3]。どの粘土板がどの部屋から出たのかを照合する。同じ書記が書いた字かどうか、線の癖を比べる[7]。この検討を通過して初めて、学界の資料になる[3]。

エーゲ海の三つの文字と一緒に扱う研究も続いている。クレタ聖刻文字、線文字A、キプロ・ミノア文字を同じ場に置く方法である。三つの文字を比べると、一つの文字だけを見ていたときには見えなかった規則が現れる。コンピューターは、こうした大量比較で力を発揮する[3]。

機械が代わりにできないこと

こうした方法は、線文字Aの意味まで明らかにしたわけではない。越えるべき壁が三つ残っている。

第一は、比べる相手である。教師なしアラインメントは、未解読文字と既知の文字との間に近縁性があるときに力を発揮する[1]。ミノア語には生き残った親族言語がない[3]。座標は描かれるが、その座標を固定する基準点がない。

第二は、資料の量である。銘文1,370点は、機械学習に使うには少ない[3]。資料が少ないと、大きなニューラルネットワークはもっともらしい幻を作り出す。こうした誤りを幻覚と呼ぶ。

第三は、検証である。2026年7月にThe Conversationに掲載された記事が、この問題を指摘した。ダブリンシティ大学のジェーン・アドキンズが書いた記事である。彼女は、統計でパターンを合わせる作業は、存在しない意味を作り出すことはできないと記した。既知の言語や対訳文書のような足場があってこそ、偶然の一致と本物の規則を分けられるという指摘である。機械は候補を減らすだけであり、正しいかどうかを見分ける仕事は人間の役割として残る[6]。

参考資料

- [1] Luo, J., Cao, Y., & Barzilay, R. (2019). Neural decipherment via minimum-cost flow: From Ugaritic to Linear B. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics. <https://aclanthology.org/P19-1303/> (2026年9月3日閲覧)
- [2] Luo, J., Hartmann, F., Santus, E., Barzilay, R., & Cao, Y. (2021). Deciphering undersegmented ancient scripts using phonetic prior. Transactions of the Association for Computational Linguistics, 9, 69-81. https://doi.org/10.1162/tacl_a_00354
- [3] Braović, M., Krstinić, D., Štula, M., & Ivanda, A. (2024). A systematic review of computational approaches to deciphering Bronze Age Aegean and Cypriot scripts. Computational Linguistics, 50(2), 725-779. https://doi.org/10.1162/coli_a_00514
- [4] Salgarella, E., & Castellan, S. (2021). SigLA: The Signs of Linear A. A palaeographical database. In Grapholinguistics in the 21st Century 2020: Proceedings (Grapholinguistics and Its Applications, Vol. 5, pp. 945-962). Fluxus Editions. <https://sigla.phis.me/> (2026年9月3日閲覧)
- [5] Corazza, M., Tamburini, F., Valério, M., & Ferrara, S. (2022). Unsupervised deep learning supports reclassification of Bronze age cypriot writing system. PLOS ONE, 17(7), e0269544. <https://doi.org/10.1371/journal.pone.0269544>
- [6] Adkins, J. (2026, 7月 27日). Cracking the code: can AI help us decipher ancient languages? The Conversation. <https://theconversation.com/cracking-the-code-can-ai-help-us-decipher-ancient-languages-288238> (2026年9月3日閲覧)
- [7] Giorgi, L., & Lopez, S. (2026). Aegean Bronze Age writing systems through RTI: New potential and perspectives. Digital Applications in Archaeology and Cultural Heritage, 41, e00508. <https://www.sciencedirect.com/science/article/abs/pii/S2212054826000159> (2026年9月3日閲覧)
- [8] Stavrakakis, E. (2024). Ventris' Grids. Bulletin of the Institute of Classical Studies, 67(1), 86-103. <https://doi.org/10.1093/bics/qbae033>

9.3. 明らかになった真実

数字と分数は解けた

線文字Aで先に解けたのは、数を数える方法である。ミノアの人々は、十を単位にまとめて数える十進法を使っていた。縦線一本が1で、横線一本が10である。丸は100で、光の筋が伸びた丸は1000である[3]。この規則は、図を見るだけでも推測できた。

難しかったのは分数記号だった。粘土板には十七個の分数記号が出てくる[8]。二つの記号を重ねて書いたものまで数えると、三十個近くになる[9]。どの記号がどれだけの値を意味するのかは、長い間解けなかった。

2021年にボローニャ大学の研究陣がこの問題を扱った。ミケーレ・コラッツァ、シルヴィア・フェラーラ、バルバラ・モンテッキ、ファビオ・タンブリーニ、ミゲル・ヴァレリオによる共同研究である。彼らは、記号の形を調べる古い方法とコンピューター計算を組み合わせた[1]。

研究陣はまず、記号が粘土板で守っている規則を整理した。どの記号同士が並んで出てきて、どの記号同士は並ばないのかを数えた。次に、値の組み合わせをコンピューターで調べた。候補は400万通り近くあった。統計検定と、ほかの文明の分数体系との比較によって、この候補を絞っていった[1]。

残った答えはこうだった。ミノアの分数体系で最小の値は60分の1だった。この体系で、60を分母とする値の大半を書けた。10分の1を意味する記号は、線文字Bにもそのまま受け継がれた。線文字Bでこの記号は、乾いた穀物を量る単位として使われた。二つの文字の記録の間には、約250年の時差がある。その時差を越えても、計算法は生き残った[1]。

計算法が解けると、帳簿の構造も見え始めた。粘土板は、品目記号と数量を組み合わせる書く方式を守っている。同じ書式が複数の遺跡で繰り返される。クレタ島の宮殿が同じ会計規則を共有していたという意味である[3]。意味が分からなくても、文書の性格はこのように現れる。

地名は読めるが、意味は解けなかった

音の面でも得られたものがある。米国カンザス大学のジョン・ヤンガーは、線文字Aの銘文を音に移した一覧を公開してきた。線文字Bで同じ形として使われる記号に、同じ音価を入れてみた転写である[2]。

この転写では、クレタ島の古い地名と見られる語がいくつも出てくる。クノッソスやファイストスのように、後の時代にも残った名前である[2]。地名が合うということは、音価仮説が完全に誤っているわけではないという意味である[3]。

音価仮説には限界も明らかである。二つの文字で形が同じに見えても、音まで同じだったという保証はない。線文字Bを作った人々は、他人の文字を借りて自分たちの言葉を書いた。借りて使う過程で、音が移った可能性がある[3]。そのため、転写文は暫定的な読みとしてだけ使う[2]。

しかし、そこまでである。地名を除くと、残りの語の意味はつながらない。音のとおりに読めても、何を言っているのかは分からない。音と意味の間に橋が架かっていない状態である[3]。

2026年の騒動

2026年夏に、線文字Aを読んだという発表が出た。AI技術者トム・ディ・ミノが出した主張である。彼は言語学を正式に学んだ人ではない[4]。

ディ・ミノは、Pythonスクリプトを書いて二つの資料窓口を調べたと明らかにした。手で模写した資料を集めたGORILAと、先に見たSigLAである。最初の発表で、彼は記号40個に音価を与え、語408個を一覧にしたと述べた[4]。8月には41ページの文書を出し、一覧を508項目に増やした。記号42個の読みと、銘文443件の翻訳も載せた。彼は、ミノア語がヘブライ語とアラム語が属するセム語族に属すると主張した[5]。

学界はこの主張を受け入れなかった。査読を経た論文が出ていないからである[4]。ラトガース大学とケンブリッジ大学で検討が進んでいると伝えられた。しかし、結論はまだ出ていない[5]。線文字Aを読んだという主張は、過去百年の間に何度も出ている。そのため研究者は、新しい主張に慎重に接する[6]。

同じ8月には、別の解釈も出た。ユクタス山頂の聖所から出た石の卓の銘文に関するものである。神に酒を注いで捧げた献酒卓である。この石に刻まれた二つの名前が、後のギリシア神とつながるという提案だった。この提案も仮説にとどまっている。線文字Aが確実に解けていないからである[7]。

この話は、AI研究全体にも残るものがある。機械は、規則があるところで規則をよく見つける。分数記号のように、前後が合わなければならぬ資料で成果が出た[1]。一方で、規則だけでは届かない場所がある。どの音がどの意味と結びつくかは、約束の問題である。その約束を知る人がすべて消えると、計算でよみがえらせる道は狭くなる[6]。

残った資料を開いておくこと

研究方法も変わった。以前は、銘文の図版は高価な図録にしか載っていなかった。数冊の本を手に入れた人だけが統計を出せた。今は資料がインターネットで公開されている。学生でもダウンロードして記号を数えられる[3]。

資料が開かれると、検証も速くなる。誰かが新しい読みを主張すれば、別の人が同じ資料で確認できる[6]。2026年夏の騒動でも、検証過程はそのまま動いた[4]。主張する側と検討する側が同じ資料を見ていた[5]。

資料が公開されたことで、研究の入り口も低くなった。大学に所属していない人でも、記号の統計を出して見ることができる。新しい目が増えると、見落としていた規則がつかまることもある。その代わり、検証されていない主張も一緒に増える。公開と検証は対で動かなければならない[6]。

まだ読めない文字

AIが線文字Aで成し遂げたことは四つある。機械は、散らばった資料を一か所に集めてくれた。記号の形と線を数字に変え、比べられるようにした[3]。分数記号の値は、コンピューター計算が人間の判読を助けて絞り込んだ。候補が数百万個ある問題を、人間が検討できる大きさに縮めてくれた[1]。

機械ができなかったことも明らかである。ミノア語がどのような言語なのかは、今も分かっていない。ロゼッタ・ストーンの役目を果たす文書は、まだ地中にあるか、存在しない[3]。音を写し取っても、その言葉が何を指すのかは出てこない[6]。

キプロ・ミノア文字の事例は、この違いをよく示している。ニューラルネットワークは、従来の分類が揺らぐという結果を出した[10]。だからといって、その文字を読み解いたわけではない。機械が整理してくれたのは文字の構造であり、言葉の意味ではなかった。

残された道は二つに見える。一つは新しい遺物が出ることである。二つの言語で書かれた文書一点があれば、状況は変わる。もう一つは、資料をさらに細かく整えることである。線一本まで記録しておけば、後の研究はその上から始められる[3]。

線文字Aは、今も読めない文字である。機械はその前で扉を開けてくれたのではなく、錠前の形を描いてくれた。扉を開ける仕事は、新しい遺物と人間の判断にかかっている[6]。エーゲ海の粘土板は、まだ自らの言葉を語っていない。

参考資料

- [1] Corazza, M., Ferrara, S., Montecchi, B., Tamburini, F., & Valério, M. (2021). The mathematical values of fraction signs in the Linear A script: A computational, statistical and typological approach. *Journal of Archaeological Science*, 125, 105214. <https://doi.org/10.1016/j.jas.2020.105214>
- [2] Younger, J. G. (n.d.). Linear A texts and inscriptions in phonetic transcription. University of Kansas. <http://people.ku.edu/~jyounger/LinearA/> (2026年9月3日閲覧)

- [3] Braović, M., Krstinić, D., Štula, M., & Ivanda, A. (2024). A systematic review of computational approaches to deciphering Bronze Age Aegean and Cypriot scripts. *Computational Linguistics*, 50(2), 725-779.
https://doi.org/10.1162/coli_a_00514
- [4] GreekReporter. (2026, 7월 30일). AI engineer claims to have deciphered Linear A. [greekreporter.com](https://greekreporter.com/2026/07/30/ai-engineer-claim-decipher-linear-a/).
<https://greekreporter.com/2026/07/30/ai-engineer-claim-decipher-linear-a/> (2026年9月3日閲覧)
- [5] GreekReporter. (2026, 8월 13일). AI engineer claims to have deciphered Linear A, identifies Minoan language as Semitic. [greekreporter.com](https://greekreporter.com/2026/08/13/ai-engineer-claim-linear-a-minoans-semitic-language/).
<https://greekreporter.com/2026/08/13/ai-engineer-claim-linear-a-minoans-semitic-language/> (2026年9月3日閲覧)
- [6] Adkins, J. (2026, 7월 27일). Cracking the code: can AI help us decipher ancient languages? *The Conversation*.
<https://theconversation.com/cracking-the-code-can-ai-help-us-decipher-ancient-languages-288238>
(2026年9月3日閲覧)
- [7] GreekReporter. (2026, 8월 14일). Linear A inscription points to Minoan origins of Greek gods Dionysos and Demeter. [greekreporter.com](https://greekreporter.com/2026/08/14/minoan-linear-a-inscription-greek-gods-dionysos-demeter/).
<https://greekreporter.com/2026/08/14/minoan-linear-a-inscription-greek-gods-dionysos-demeter/>
(2026年9月3日閲覧)
- [8] Del Frio, M. (2022). Linear A. *Mnamon: Ancient writing systems in the Mediterranean*. Scuola Normale Superiore.
<https://doi.org/10.25429/sns.it/lettere/mnamon022> (2026年9月3日閲覧)
- [9] Salgarella, E., & Castellan, S. (2021). SigLA: The Signs of Linear A. A palaeographical database. In *Grapholinguistics in the 21st Century 2020: Proceedings (Grapholinguistics and Its Applications, Vol. 5, pp. 945-962)*. Fluxus Editions.
<https://sigla.phis.me/> (2026年9月3日閲覧)
- [10] Corazza, M., Tamburini, F., Valerio, M., & Ferrara, S. (2022). Unsupervised deep learning supports reclassification of Bronze age cypriot writing system. *PLOS ONE*, 17(7), e0269544. <https://doi.org/10.1371/journal.pone.0269544>

第10章. 太陽の周期と動物印章の暗号をハックする： インダス文字とロンゴロンゴの音

10.1. 事件ファイル

印章に刻まれた四つの文字

インダス文明は紀元前2600年ごろから紀元前1900年ごろまで続いた[1]。基盤となった地域は、今日のパキスタンとインド北西部である。都市の大きさと数を見ると、青銅器時代でも有数の文明である。青銅器時代とは、鉄より青銅を多く使っていた古い時代だ。都市と村が広がった土地は100万平方キロメートルを超えると知られている[2]。

1920年代にハラッパーとモヘンジョダロで大規模な発掘が始まった[1]。この作業を統括した人物は、英国の考古学者ジョン・マーシャルである。彼はインド考古調査局の責任者だった。ハラッパーの現場は1921年にダヤ・ラム・サーニーが担当して掘った。モヘンジョダロの現場は1922年にラカルダス・バネルジが担当した。マーシャルはこれらの発掘をまとめ、新しい文明の存在を世に知らせた[3]。

地中からまっすぐな煉瓦の道と排水施設が出てきた。そして手のひらより小さい四角い印章が大量に出てきた[1]。

都市の姿だけでも分かることは多い。道は碁盤の目のようにまっすぐ伸び、各家には下水溝がつながっていた。煉瓦は大きさをそろえて焼かれた。重さを量る分銅にも規格があった。計画を立て、規則を守る社会だったという意味だ。ところが彼らが残した文字は読めない。

印章は滑石という柔らかい石で作られた。滑石は爪でも跡が付く石である。印章の下側には動物の絵が彫られている。一本角の獣、こぶのある牛、象のような絵である。絵の上の空いた場所には、記号が一列に並んでいる[1]。

印章は押して写す道具である。柔らかい粘土に押し当てると、絵と記号がそのまま移る。荷物を縛ったひもの上に粘土を付け、印章を押せば封印になる。インダスの都市では、印章を押した粘土の封印も一緒に出てきた[1]。

学者たちはこの記号の集まりをインダス文字、またはハラッパー文字と呼ぶ[1]。名前には文字と入っているが、まだ読めない。音も意味も分からない。発見されてから100年以上たってもそうである[2]。

読めないという言葉には二つの意味が重なっている。記号の音を知らないという意味が一つだ。記号の意味を知らないという意味がもう一つだ。今はその両方が分からない状態である。

問題は長さだ。一つの印章に刻まれた記号は、ふつつ四つか五つだけである[1]。一面に刻まれた長い記録でも十七文字を超えない。三面に分けて刻んだものをすべて数えてようやく二十六文字になる[4]。教室で使うノートの一行も埋められない量である。

印章や器に刻んだこのような文字を銘文と呼ぶ。これまで集められたインダス銘文は3千点前後である[5]。点数は少なくないが、一つ一つが短すぎる。短い文章だけを集めても、文の組み立てを知るのは難しい。主語と述語がどのように置かれるかを見る機会がないからだ。

海を越えた木板の記号

南太平洋の真ん中にラパ・ヌイという島がある。私たちにはイースター島という名前の方がなじみ深い。人が住む隣の土地まで2千キロメートルを超える。巨大な石像モアイで知られる場所でもある。この島からも、意味の分からない記号が記された木板が出てきた。この記録は今日ロンゴロンゴと呼ばれる[6]。

記号は筆で書いた文字ではない。鋭い黒曜石のかけらや骨で作った道具で木を彫って刻んだ。刻まれた形は人、鳥、魚に似ている[6]。手足を上げた人の形がよく出てくる。

読む方向もなじみがない。読む人は板の下側の左端から始める。一行を読み終えると、板を180度回す。上下が逆になった状態で次の行を読む。行ごとに板を裏返すこの方式を逆牛耕式と呼ぶ[6]。分かりやすく言えば、ひっくり返して読む方式である。古代ギリシャにも、行ごとに方向を変える文字があった。そちらは板を回さず、文字の向きだけを変える。

この方式は研究者をもう一度試す。どの行が最初の行なのかを決めてこそ順序が定まる[6]。順序を誤れば、記号が続く順番が丸ごと入れ替わる。コンピューターに資料を入れる前に、人がまず判断しなければならない部分である。

残った遺物は多くない。世界の博物館と個人収集家に散らばっている真品は26点である。すべてラパ・ヌイ島の外にある[6]。そこに刻まれた記号をすべて合わせても約1万5千文字だ[7]。本一冊にも満たない量である。

遺物がこれほど少ないのには理由がある。19世紀に島の人口が大きく減り、木板も多く消えた。残ったものはヨーロッパの博物館と修道院へ散らばった。そのうち4点がローマのある修道会に保管されている[8]。

この木板がいつ作られたのかも、長く争われてきた問題だった。ヨーロッパ人が島に入った後にまねて作ったのではないかという疑いがあった。2024年にシルヴィア・フェラーラの研究チームが、ローマの木板4点の年代を測った[8]。木に残った炭素の量を測って年代を知る方法を使った。この方法を放射性炭素年代測定と呼ぶ。

4点のうち3点は19世紀の木だった。残り1点ははるかに古い木だと分かった。ヨーロッパの船が島に着いた1720年代より200年以上前である[8]。ラパ・ヌイの人々が自ら文字を作った可能性が高まった。

研究チームは一つ条件を付けておいた。木を切った時期が記号を刻んだ時期と同じとは限らない。木の年代は、刻みがそれより後だという事実だけを示す。古い木を後で削って使った可能性があるからだ。早い年代が出た木板が1点だけだという点も残る[8]。

これは文字なのか

二つの記号体系をめぐる、学界は長く争った。争いは解釈から始まったのではない。これが言葉を書き取った文字で間違いないのかを先に問うた[1][9]。

2004年にスティーブ・ファーマー、リチャード・スプロート、マイケル・ヴィッツェルが共同で論文を発表した。彼らはインダス記号は文字ではないと主張した。記録が短すぎ、長い文書が一つもない点を根拠にした。つまり、この記号は家のしるしや物の所有表示、または護符に近いという意味である[9]。

反対側では、統計で答えを探そうとした。記号が出る頻度と前後に付く規則を数えてみようということだ[5]。人の言葉には口癖のような統計上の習慣がある。その習慣が記号にもあるかを測ってみようという方法である。

この方法の利点は、意味を知らなくても使える点にある。記号を番号に置き換え、順序だけを数えればよい。コンピューターが得意な仕事でもある。

解釈を阻む三つの壁

二つの記号体系はいずれもまだ解釈されていない。前をふさぐ壁は三つある。

照合する文章がない。エジプトのヒエログリフはロゼッタ・ストーンのおかげで解けた。ロゼッタ・ストーンには同じ内容がギリシャ語でも記されていた。フランスの学者シャンポリオンは、知っている文字と知らない文字を並べて対応させた。同じ内容を二つの言語で並べて記した文章を対訳本と呼ぶ。インダス遺跡とラパ・ヌイ島では、そのような記録は出ていない[2][6]。

どの言葉を記したものかも分からない。記号が書き取った実際の言葉を基層言語と呼ぶ。インダス文字の基層言語をめぐる、ドラヴィダ語系統説とインド・アーリア語系統説が対立する。知られているどの語族にも属さないという主張もある[2]。

残る壁は資料の量である。今日の人工知能は膨大な文章を食べて育つ。ところがここにあるのは、インダス印章数千点と木板26点だけである[1][6]。資料が足りなければ、人工知能はもっともらしい嘘を作り出す。この現象を幻覚と呼ぶ。

この三つの壁は互いを支えている。基層言語が分かれば、音価を合わせてみることができる。対訳本があれば、基層言語を突き止めることができる。資料が多ければ、統計で手がかりをつかめる。どれか一つが解けないので、残りも同じ場所にとどまっている。

2025年1月5日、インドのタミル・ナードゥ州が賞金をかけた。チェンナイで開かれたインダス文明発見100周年国際学術大会の開幕式で出た発表である[10]。インダス文字を考古学の専門家が認めるほど解き明かした人に100万ドルを与えるという内容である[10][11]。2026年9月までに知られている受賞者はいない。

参考資料

- [1] Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009). Entropic evidence for linguistic structure in the Indus script. *Science*, 324(5931), 1165. <https://doi.org/10.1126/science.1170391>
- [2] Khan, M. A., Khurshid, S. K., & Aslam, M. (2025). Modern algorithms for ancient scripts: A review of AI-based techniques in Indus civilization research. *Cultural and Historical Heritage: Preservation, Presentation, Digitalization*, 11(2), 11-21. <https://doi.org/10.55630/KINJ.2025.110201>
- [3] Grand Valley State University. (n.d.). Archaeologist of the month: John Marshall. gvsu.edu. <https://www.gvsu.edu/archaeology/module-spotlight-view.htm?entryId=A748C17B-A353-F70E-D7422D41F7889433> (2026年9月3日閲覧)
- [4] Farmer, S. (n.d.). Claims concerning the longest Indus "inscription". safarmer.com. <https://safarmer.com/indus-longestinscription/> (2026年9月3日閲覧)
- [5] Yadav, N., Joglekar, H., Rao, R. P. N., Vahia, M. N., Adhikari, R., & Mahadevan, I. (2010). Statistical analysis of the Indus script using n-grams. *PLoS ONE*, 5(3), e9506. <https://doi.org/10.1371/journal.pone.0009506>
- [6] Australian Museum. (2026, 5月 28日). Kohau Rongorongo: Talking tablet. australian.museum. <https://australian.museum/learn/cultures/pasifika-collections/rapa-nui-collections/kohau-rongorongo-talking-tablet/> (2026年9月3日閲覧)
- [7] Ross, L. (2026, 3月 31日). Reading Rongorongo (Version 2). Zenodo. <https://doi.org/10.5281/zenodo.19362491> (2026年9月3日閲覧)
- [8] Ferrara, S., Tassoni, L., Kromer, B., Wacker, L., Friedrich, M., Tonini, F., Lastilla, L., Ravanelli, R., & Talamo, S. (2024). The invention of writing on Rapa Nui (Easter Island): New radiocarbon dates on the Rongorongo script. *Scientific Reports*, 14, 2794. <https://doi.org/10.1038/s41598-024-53063-7>
- [9] Farmer, S., Sproat, R., & Witzel, M. (2004). The collapse of the Indus-script thesis: The myth of a literate Harappan civilization. *Electronic Journal of Vedic Studies*, 11(2), 19-57. <https://doi.org/10.11588/ejvs.2004.2.620>
- [10] The Tribune. (2025, 1月 6日). Stalin announces \$1 mn prize for deciphering Indus Valley script. [tribuneindia.com. https://www.tribuneindia.com/news/india/stalin-announces-1-mn-prize-for-deciphering-indus-valley-script/](https://www.tribuneindia.com/news/india/stalin-announces-1-mn-prize-for-deciphering-indus-valley-script/) (2026年9月3日閲覧)
- [11] Smithsonian Magazine. (2025, 1月 30日). Officials are offering \$1 million to anyone who can decode this ancient script. [smithsonianmag.com. https://www.smithsonianmag.com/smart-news/officials-are-offering-1-million-to-anyone-who-can-decode-this-ancient-script-180985922/](https://www.smithsonianmag.com/smart-news/officials-are-offering-1-million-to-anyone-who-can-decode-this-ancient-script-180985922/) (2026年9月3日閲覧)

10.2. 科学捜査隊の追跡: 情報エントロピー測定器

文字の指紋を測る者

インダス印章とラパ・ヌイの木板は互いに会ったことがない。ところが二つの場所の研究者が同じ道具を手にとった。意味の分からない記号列を前にして、できることが似ていたからだ。

読めない文字を前にして、研究者たちは別の問いを選んだ。何を意味するのかと問う代わりに、どのように並んでいるのかと問うた。意味が分からなくても、順序は数えられるからだ。

このとき使う道具が情報エントロピーである。エントロピーは、次に何が来るかをどれほど知りにくいかを測る値だ。値が大きければ、次を予測しにくいという意味である。値が小さければ、次がほぼ決まっているという意味である。

この概念は通信工学から生まれた。電話線で信号を送るとき、どれほどの情報が載るのかを測るために作られた物差しである。意味とは関係なく、順序だけを見て計算する。読めない文字に当てられる理由はここにある。

教室の座席を思い浮かべると理解しやすい。くじ引きで座れば、隣に誰が来るか分らない。このときエントロピーは高い。番号順に座れば、隣に来る人は決まる。このときエントロピーは低い。

人の言葉は二つの極端の間にある。文法が順序を縛るが、言う内容は自由に選ぶ。だから人が使う言葉のエントロピーは中間地帯に置かれる。この中間地帯を物差しにしようというのが研究の出発点である。

日本語でも確かめられる。「学校へ」の次に来る言葉はいろいろある。行く、行った、通ってきたのいずれも可能だ。しかし、どんな言葉でも来られるわけではない。選択肢はあるが絞られているこの状態が、言語の特徴である。

インダス記号を測ってみる

2009年にラジェシュ・ラオとニシャ・ヤダヴを含む研究チームが、この物差しをインダス記号に当てた。彼らは前の記号を知っているとき、次の記号の不確実性を計算した[1]。前の記号を知るとき、次の記号がどれほど予想されるかを測るこの値を条件付きエントロピーと呼ぶ。

比較対象も一緒に用意した。一方には英語、サンスクリット語、シュメール語のような人の言葉を置いた。もう一方には人の言葉ではないものを置いた。生物の遺伝子配列、コンピュータープログラムのコード、無作為に作った記号列である[1]。

結果は、インダス記号が人の言葉の側に近く出たというものだった。無作為の記号列より規則がはっきりしており、コンピューターコードより自由だった。研究チームはこの結果をScienceに掲載した[1]。

同じ研究チームは確率モデルも一緒に示した。前の記号が決まると、次の記号が出る確率を表にする方式である。このように前の記号だけを見て次を計算するモデルをマルコフモデルと呼ぶ。表をのぞくと、どの記号が文頭によく来るかが現れる。どの記号が文末によく来るかも現れる[2]。

同じ研究チームは翌年、記号の頻度も調べた。約3千点の銘文を集め、そのうち状態のよい1,548点を選んだ。少数の記号が使用の大半を占め、残りの記号はまれに出た。この形は人の言語で常に見られる分布である[3]。

この分布には名前がある。ジップの法則と呼ばれる。よく使う単語のいくつかが文章の大半を埋めるという規則である。日本語でもwa、ga、wo、niのような語が上位を占める。反対に、一度だけ出て終わる語はとても多い。インダス記号も同じ形を示した[3]。

置かれる順序にも規則が現れた。23個の記号が文章の終わりの80パーセントを占めた。反対に、文頭の80パーセントを満たすには82個の記号が必要だった[3]。終わりに来る記号がはるかに限られているという意味である。日本語で文末に助詞や語尾が集まる姿と似ている。

研究チームは前の記号から次の記号を当てるモデルも作った[2]。このモデルは摩耗して読みにくい記号を約75パーセントの確率で復元した[3]。規則があってこそ出せる成績である。

復元するという言葉を誤解してはいけない。モデルが埋めるのは記号の形だけである。その記号がどんな音だったのかは、今も分らない。欠けたパズルの一マスを埋めたのであって、絵を読んだのではない。

反論と反論への応答

このモデルには限界もある。前の一つか二つの記号だけを見て次を当てる方式だからである[2]。人の文は、はるかに離れた語同士が呼応する。資料が短ければ、その遠い呼応を学ぶことはできない。

この結果に対して、すぐに反論が出た。リチャード・スプロートが2010年に批判文を発表した。エントロピー値が似ているという事実だけでは、文字だと断定できないという指摘である。人の言葉ではない記号体系も、似た値を出すことができるということだ[4]。

ラオ研究チームは同じ年に同じ学術誌で答えを出した。自分たちはインダス記号が文字であることを証明したわけではないと明らかにした。文字である可能性を統計で裏づけたただけだという説明である[5]。

スプロートは2014年に批判を一步進めた。彼は人の言葉ではない記号資料をいくつも集めた。メソポタミアの神の標識、北アメリカのトーテムポール、納屋の壁に描く星模様がそれである。天気予報に使う絵も入れた。これらの資料とインダス記号を同じ物差しで比べた[6]。

彼が作った判定手順は、インダス記号を人の言葉ではない側に分類した。スプロートは統計だけでは足りないと書いた。解読に成功するか、その文化が文字を書いたという考古学的証拠が別に必要だという主張である[6]。この論争はまだ終わっていない。

2026年4月に出た新しい分析も同じ問いの前に立つ。アシン・ナイルはインダス銘文1,916点を対象に検査を設計し直した。彼は人の言葉ではない記号体系をまねた偽資料をコンピューターで作った。本物のインダス資料と偽資料の統計を並べて比べた[7]。

インダス資料はどの偽資料ともびたりとは合わなかった。二つの偽資料の中間に置かれた。実在するどの非言語記号体系も、インダス統計をそのまま再現できなかった。ナイルは、これが文字かどうかの判定は保留すべきだと書いた。この文章は学術誌の審査を経る前にインターネットに先に載せたプレプリントである[7]。

記号を見分ける目

統計の前にしなければならないことがある。石に刻まれたぼやけた跡を記号として写し取る仕事である。この作業は長い間、人の目と手に頼ってきた。

写し取りは思ったより難しい。印章は4千年の間にすり減った。同じ記号でも刻んだ人によって形が少しずつ違う。二つの記号と見るか、一つの記号と見るかで判断が分かれることもある。この判断が変わると、後に続く統計も変わる。

2017年にサティシュ・パラニアパンとロノジョイ・アディカリがこの仕事をコンピューターに任せた。印章写真を入れると記号一覧が出る処理の流れを作った[8]。

2025年にはヴァイシュナヴィ・ディクシット研究チームがこの流れを引き継いだ。印章写真から記号を認識するニューラルネットワークと、動物の絵を認識するニューラルネットワークを別々に作った。資料を五つに分け、交代で試験した。二つのニューラルネットワークはいずれも平均85パーセント前後の正確度を示した[9]。

この問題を減らすには、判断基準を公開し、互いに比べなければならない。2025年にムザファル・アリ・カーン研究チームがその作業を整理した。2009年から2025年までに出了インダス文字関連の人工知能研究を集め、種類別に分けた。よく繰り返される誤りもあわせて書いた[10]。

2025年にアトゥル・シャルマとシュバジット・ロイ・チョードゥリーが一つの道具を公開した。名前はAI-EPIGRAPHYである。利用者が記号の順序を入れると、確率の高い解釈候補を返す。頻度分析と連鎖確率計算を組み合わせた道具である[11]。

2026年5月には資料基盤そのものを作り直す試みが発表された。ベアタ・メジェシ研究チームは、希少で読めない文字を一か所に集めるデータベースを作った。高画質写真、写し取った記号、注釈、出典情報を一つの枠に収めた。散らばった資料を集めることが、解読より先の課題だという判断である[12]。

参考資料

[1] Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009). Entropic evidence for linguistic structure in the Indus script. *Science*, 324(5931), 1165. <https://doi.org/10.1126/science.1170391>

- [2] Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009). A Markov model of the Indus script. *Proceedings of the National Academy of Sciences*, 106(33), 13685-13690. <https://doi.org/10.1073/pnas.0906237106>
- [3] Yadav, N., Joglekar, H., Rao, R. P. N., Vahia, M. N., Adhikari, R., & Mahadevan, I. (2010). Statistical analysis of the Indus script using n-grams. *PLoS ONE*, 5(3), e9506. <https://doi.org/10.1371/journal.pone.0009506>
- [4] Sproat, R. (2010). Ancient symbols, computational linguistics, and the reviewing practices of the general science journals. *Computational Linguistics*, 36(3), 585-594. https://doi.org/10.1162/coli_a_00011
- [5] Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2010). Entropy, the Indus script, and language: A reply to R. Sproat. *Computational Linguistics*, 36(4), 795-805. https://doi.org/10.1162/coli_c_00030
- [6] Sproat, R. (2014). A statistical comparison of written language and nonlinguistic symbol systems. *Language*, 90(2), 457-481. <https://doi.org/10.1353/lan.2014.0031>
- [7] Nair, A. (2026, 4월 20일). How non-linguistic is the Indus sign system? A synthetic-baseline scorecard. arXiv:2604.17828. <https://arxiv.org/abs/2604.17828> (2026年9月3日閲覧)
- [8] Palaniappan, S., & Adhikari, R. (2017). Deep learning the Indus script. arXiv:1702.00523. <https://arxiv.org/abs/1702.00523>
- [9] Dixit, V., Hussain, N., Basak, S., Atturu, D., Mitra, D., & Bhattacharya, U. (2025). Deep learning in archiving Indus script and motif information. *Journal of Computer Applications in Archaeology*, 8(1), 156-169. <https://doi.org/10.5334/jcaa.175>
- [10] Khan, M. A., Khurshid, S. K., & Aslam, M. (2025). Modern algorithms for ancient scripts: A review of AI-based techniques in Indus civilization research. *Cultural and Historical Heritage: Preservation, Presentation, Digitalization*, 11(2), 11-21. <https://doi.org/10.55630/KINJ.2025.110201>
- [11] Sharma, A., & Roy Chowdhury, S. (2025). AI-EPIGRAPHY: An interactive tool for computational decipherment of the Indus Valley script. *Proceedings of IndiaHCI '25*. <https://doi.org/10.1145/3768633.3770145>
- [12] Megyesi, B., Rattenborg, R., Lang, B., Waldispuhl, M., & Heder, M. (2026). Building a corpus and database for rare and undeciphered scripts. *Proceedings of the Fourth Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2026) @ LREC 2026*, 184-196. ELRA Language Resources Association. <https://aclanthology.org/2026.lt4hala-1.17/> (2026年9月3日閲覧)

10.3. 明らかになった真実

月を数える木板

ラパヌイの木板の中に、ママリという名が付いたものがある。この木板には、ほかの場所とは形が違う区間がある。三日月形の記号が連続している[1]。

ドイツの学者トーマス・バルテルが20世紀半ばにこの区間に注目した。彼はこれを月の日付を書いた記録だと見た[1]。1990年にジャック・ギーが資料を整理し直し、同じ結論を示した[2]。

2011年にポール・ホーリーがこの解釈をさらに固めた。ママリ木板の暦は三十夜から成る。この数字は実際の月の周期と合う。ポリネシアの多くの島の暦とも合う[1]。

ホーリーはタヒチとニュージーランドの暦も並べて比べた。夜ごとに付く名前が島ごとに似ていた。ラパヌイの暦はほかの島々より三日ほど遅れていた。島が遠く離れている間に生じた違いと見られる[1]。

記号の配置も規則に従っていた。三日月記号の間に別の記号が入り、八つのまとまりに分ける。最後の二つの三日月は、月が見えない夜を指す[1]。

同じ配列は木板の外にも現れる。島の岩絵の中に、亀の模様と三日月と一緒に刻まれたものがある。そこでも最後の二つの三日月が亀の甲羅の中に置かれている。木板が月のない夜を別に分けて描いた方式と同じである[1]。木板と岩は同じ数え方を分け持っていたことになる。

この解釈はロンゴロンゴ研究で広く認められている。それでも読み取れたのは日付の順序である。ラパヌイの人々がその夜を何と呼んだかは分からない。

人工知能が出した答えとその重み

月を数えることと太陽を数えることはつながっている。日食は太陽と月が重なるときに起こる。だから暦を読み取ったという主張は、すぐに空全体を読み取ったという主張へ広がる。

2026年3月にローガン・ロスがプレプリント論文を出した。彼は個人用コンピューターで言語モデルを動かして、ロンゴロンゴを分析した。89個の記号から成る辞書を作り、これで残る記録の97パーセント近くを覆ったと主張した。ママリ木板が日食と月食を予測する表だという主張も示した[3]。

この主張はまだ学界の検証を経ていない。ロスの文章は学術誌の審査を通った論文ではない。インターネットの保存場所に先に載せたプレプリントである[3]。プレプリントは誰でも載せられる。内容が正しいかは、ほかの研究者が後から確かめなければならない。

このような主張を扱うときは、検査方法もあわせて見なければならない。残るロンゴロンゴ記録は26点、約1万5千字である[3]。資料がこれだけしかない場所では、偶然合う結果が出やすい。89個の記号に音価を自由に割り当てれば、どんな文でも作り出せる。だから学界は、ほかの人が同じ方法で同じ結果を得るかをまず問う。

本当の解読なら通らなければならない試験がある。解読案を作るときに使わなかった別の木板を持ってきて読んでみることである。そこでも意味の通る文が出なければならない。ミケーネ・ギリシア語を記した線文字Bがその試験を通った例である[4]。ロンゴロンゴとインダス文字は、まだその門を越えていない。

機械ができることとできないこと

これまで人工知能がこの二つの文字で成し遂げたことは明らかである。

機械はぼやけた跡から記号を見分ける[5]。人が目で照合すれば数か月かかる仕事を数時間で終える。摩耗して読みにくい記号を前後の文脈から復元することもある[6]。散らばった遺物の写真と記録を一か所に集め、検索できるようにする[7]。

資料を集める仕事は目立ちにくい。しかしこの仕事がないと、その後続く計算もない。写真の明るさと角度がばらばらなら、記号を比べられない。写し取りの規則が研究者ごとに違えば、統計もずれる。2026年に出たデータベースは、その基準を一つにそろえようとする試みである[7]。

統計の側でも成果がある。インダス記号が無作為配列ではない事実は、複数の研究が確認した[6][8]。配列に規則があるという結論は繰り返し再現された。だからといって、これが文字だという結論にはならない。規則があるということと、言葉を書き取ったということは違う。

しかし機械ができていないことは、さらに大きく残っている。インダス記号一つ一つの音価は今も分からない[9]。ロンゴロンゴ記号の音価も分からない[3]。二つの文字はいずれも、意味が確定した語がまだない。インダス印章に書かれたものが人名なのか物の名なのかも決まっていない[9]。

理由は資料の性質にある。対訳本があれば解読はずっとやさしくなる。インダス遺跡にもラパヌイ島にもそのような文章はない[9]。対訳本なしに解けた例がないわけではない。線文字Bはそうのように解かれた。その代わり、そのときは基盤となる言語を推測する手がかりがあり、資料も十分にあった[4]。インダス文字は基盤となる言語が何かも決まっていない[9]。ロンゴロンゴは残る文章が少なすぎる[3]。

2026年4月の分析は、この事情を数字で改めて示した。インダス銘文1,916点は、人の言葉の統計とも、人の言葉ではない記号の統計ともびたりとは合わなかった[10]。百年以上続いた問いが、まだ開かれているという意味である。

待つ側に立って

読めない文字を扱う研究は二つの態度を求める。一つは新しい道具を最後まで押し進める態度である。もう一つは出た結果を疑う態度である。

人工知能は前者をよく行う。数え、比べ、候補を並べる。後者は人の役目である。候補の中で何がその時代により得たかを判断しなければならない。その判断には、遺跡の位置、墓から出た物、残っている口承がともに入る。

ラパヌイの暦解釈がその例である。三日月記号を数えることは計算である。その三十夜をポリネシアの多くの島の暦と比べることは考証である[1]。二つの作業がかみ合ったとき、解釈は説得力を得た。

インダス印章の動物の絵もそのような資料である。角が一本ある獣が何を意味したのか、私たちは知らない。ママリ木板の三十夜も同じである。順序は分かったが、その夜に人々が何をしたのかは分からない。

分からないと言うことにも価値がある。どこまで明らかになり、どこからが推測なのか線を引かなければ、次の研究は始まらない。線をぼかすと、もっともらしい話が事実のように振る舞う。読めない文字を扱う研究では、この線引きは結果を出すことと同じくらい重要である。

この章で扱った二つの文字は、互いに出会ったことがない。時代も3千年以上離れ、暮らした場所も地球の反対側である。ところが研究者が直面する難しさはほとんど同じである。短い資料、存在しない対訳本、分からない基盤言語である。同じ道具が二つの場所でともに使われる理由はここにある。

タミル・ナードゥ州が掲げた100万ドルも、その事情を映している[11]。賞金がかけてから一年以上たって、受け取る人は現れていない。計算機は速くなったが、鍵はまだ出ていない。

参考資料

- [1] Horley, P. (2011). Lunar calendar in rongorongo texts and rock art of Easter Island. *Journal de la Societe des Oceanistes*, 132, 17-38. <https://doi.org/10.4000/jso.6314>
- [2] Guy, J. B. M. (1990). The lunar calendar of Tablet Mamari. *Journal de la Societe des Oceanistes*, 91(2), 135-149. <https://doi.org/10.3406/jso.1990.2882>
- [3] Ross, L. (2026, 3月 31日). Reading Rongorongo (Version 2). Zenodo. <https://doi.org/10.5281/zenodo.19362491> (2026年9月3日閲覧)
- [4] Chadwick, J. (1958). *The decipherment of Linear B*. Cambridge University Press.
- [5] Palaniappan, S., & Adhikari, R. (2017). Deep learning the Indus script. arXiv:1702.00523. <https://arxiv.org/abs/1702.00523>
- [6] Yadav, N., Joglekar, H., Rao, R. P. N., Vahia, M. N., Adhikari, R., & Mahadevan, I. (2010). Statistical analysis of the Indus script using n-grams. *PLoS ONE*, 5(3), e9506. <https://doi.org/10.1371/journal.pone.0009506>
- [7] Megyesi, B., Rattenborg, R., Lang, B., Waldispuhl, M., & Heder, M. (2026). Building a corpus and database for rare and undeciphered scripts. *Proceedings of the Fourth Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2026) @ LREC 2026*, 184-196. ELRA Language Resources Association. <https://aclanthology.org/2026.lt4hala-1.17/> (2026年9月3日閲覧)
- [8] Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R., & Mahadevan, I. (2009). Entropic evidence for linguistic structure in the Indus script. *Science*, 324(5931), 1165. <https://doi.org/10.1126/science.1170391>
- [9] Khan, M. A., Khurshid, S. K., & Aslam, M. (2025). Modern algorithms for ancient scripts: A review of AI-based techniques in Indus civilization research. *Cultural and Historical Heritage: Preservation, Presentation, Digitalization*, 11(2), 11-21. <https://doi.org/10.55630/KINJ.2025.110201>
- [10] Nair, A. (2026, 4月 20日). How non-linguistic is the Indus sign system? A synthetic-baseline scorecard. arXiv:2604.17828. <https://arxiv.org/abs/2604.17828> (2026年9月3日閲覧)
- [11] The Tribune. (2025, 1月 6日). Stalin announces \$1 mn prize for deciphering Indus Valley script. [tribuneindia.com. https://www.tribuneindia.com/news/india/stalin-announces-1-mn-prize-for-deciphering-indus-valley-script/](https://www.tribuneindia.com/news/india/stalin-announces-1-mn-prize-for-deciphering-indus-valley-script/) (2026年9月3日閲覧)

第11章. 人工知能、破損したギリシャ語碑文を読み直す： ファウンデーションモデルとIthaca

11.1. 事件ファイル

石に刻まれた国の記録

古代ギリシャの人々は重要な出来事を石に刻んだ。新しい法律を制定すると、その内容を石柱に刻んで広場に建てた。税金を徴収した帳簿も石に残した。神殿に奉納した品々の目録も石に残した。石や青銅に刻まれたこうした文章を碑文と呼ぶ。碑文を集めて解読する学問は金石学である。石や金属に刻まれた文字を研究する学問という意味である[4]。

碑文は大抵、その時代の人々が直接残した記録である。後世の人が書き写す過程で改変した痕跡が少ない。そのため、歴史学者にとっては貴重な資料である。ただし、常に最初の姿のままであるとは限らない。後から文字を削ったり彫り直したりした碑石もある。石を削り落として書き直した事例も残っている。アテネがどの都市という手を結んだのかも碑文が教えてくれる[4]。

碑文は多くの人が見るために作られた文章である。広場や神殿の前に建てておき、通りかかる人に読ませた。学校の廊下に貼られた案内文と用途が似ている。そのため、碑文にはその日の役人の名前と日付がはっきりと記される。歴史学者はこの名前と日付を手がかりにする[4]。

消えた文字、欠損部

問題は時間である。碑石は数千年の間、風雨にさらされて立ち続けていた。石の表面は少しずつ摩耗し、文字の溝は浅くなった。地中に埋もれた碑石は湿気と微生物によって侵食された。碑石が割れて破片の一部しか残らないこともあった[1]。

その結果、碑文には文字が消えた空欄が生じる。学者たちはこの空欄を欠損部と呼ぶ。ラテン語ではラクーナ (lacuna) である[1]。欠損部は文章の真ん中に穴を開けてしまう。人名が失われれば、誰が行ったことなのか分からない。数字が失われれば、いくら納めたのか分からない。都市名が失われれば、どの国と交わした約束なのか分からない。

学者たちは古くからこの空欄を埋めようとしてきた。同時代で似た内容の他の碑文を探して比較照合した。ギリシャ語碑文には定型の文章構造があった。法令を布告する文書は、ほぼ同じ順序で文章を並べていた。そのため、前後が残っていれば失われた単語を推測することができる[1]。

碑石は重いため、研究室へ運ぶことができない。そこで学者たちは湿らせた紙を碑石に当てて押し当てた。紙が乾くと、文字の溝がそのまま浮き上がった。この紙の写しを拓本という。学者たちは拓本を持ち帰り、机の上で文字を解読した。古い印刷図録には写真の代わりに書き写した文字だけが掲載された。そのため、後世の人は石の形状を見られないまま文字だけを読まなければならなかった[4]。

推測には規則がある。まず空欄が何文字であったかを数える。石に残る文字と文字の間隔を測れば、おおよそ見当がつく。次に文字数が合致する単語だけを候補に挙げる。このように補った文字は必ず角括弧の中に記す。形状が崩れて読みにくい文字の下には小さな点を打つ。この取り決めをライデン方式 (Leiden conventions) と呼ぶ。石に残った文字と人間の推測を区別するために作られた規則である[1, 4]。

シグマの3画をめぐる論争

碑文を解読するには、いつ刻まれた文章であるかも知る必要がある。昔からの方法は字形を見て時期を推定することである。字形は時代とともに少しずつ変化するからである[2]。

アテネの碑文でよく登場する例がシグマである。シグマはギリシャ文字の一文字である。初期の時代には3画のシグマが刻まれた。後には4画のシグマが広く使われた。長い間、学者たちは3画のシグマを基準にしてきた。そのような文字があれば、紀元前446年頃より前の碑文とみなした[2]。

しかし、この基準に疑問を抱く学者たちが現れた。書体が変化する速度は地域によって異なっていたからである。古い書体を長く維持した地域もあった。同一の都市内であっても、職人ごとに彫る癖が異なっていた[2]。

そのため、アテネが他都市と結んだ条約碑文の年代をめぐる論争が続いた。一方は紀元前446年頃よりも前だと主張した。もう一方は紀元前420年代に刻まれた文章だと主張した[2]。紀元前420年代はペロポネソス戦争の真っ最中であった。この戦争は紀元前431年に始まった。年代が数十年変われば、歴史の筋書きも変わってくる。アテネが力で同盟都市を押さえつけた時期が丸ごとずれてしまうからである。

空欄を埋める作業には危険が伴う。ある学者が補った単語が、次の世代には原文であるかのように読まれてしまうからである。角括弧を落としたひとつの印刷本が、推測を事実へと変えてしまうことさえある[1]。そのため、金石学者たちは自らの推測に常に目印を残した。

碑文が本来どこに立っていたのかも突き止めなければならない。ギリシャの各都市はそれぞれ異なる方言を使っていた。同じ単語であっても都市によって綴りが異なっていた[2]。学者たちはこの綴りの違いや官職名を見て出土地を推測した[4]。出土地とは、その碑石が地中から発見された場所のことである。

人の手ではすべてを読み切れない

解読すべき碑文は増え続けている。パッカード人文科学研究所が収集したギリシャ語碑文は178,551点にのぼる[2]。ローマ時代のラテン語碑文資料も176,000点を超える[3]。一方、碑文を解読できる学者は世界でもごくわずかしかない[4]。

データを整える作業からして容易ではない。この資料を機械が読み取れるように処理したところ、78,608点が残った。半分以上の資料が短すぎるか状態が悪いために除外された。機械に学習させるデータを選別するだけでも膨大な作業であった[2]。

空欄を埋める作業は人間にとっても難しい。訓練を受けた金石学者たちに文字が消えた碑文を渡し、空欄を埋めさせた実験がある。人間が最初に出した回答が正解だった割合は25%であった[2]。4回のうち3回は間違えたという意味である。それ以前の実験では、人間が補った文字のエラー率が57.3%と算出されていた[1]。

パピルスの事情はさらに深刻である。パピルスはエジプトの水辺に生える草の名である。昔の人々はその茎の内側を薄く削いだ。削いだものを縦横に重ね、叩いて圧着し乾燥させた。そうして作られた記録材料もパピルスと呼ぶ。エジプトの乾燥した砂の中から数十万枚が出土した。まだ誰も読んでいないギリシャ語パピルスは100万点に達すると推定されている[5]。この紙には納税の領収書や手紙、契約書が記されている。古代人の日常がそのまま詰まった記録である。

解読作業は箱を開けるだけでは終わらない。紙が破れ、インクが薄れて消えかかっている状態だからである。保存処理を施し、散らばった破片をつなぎ合わせる作業が先決である。その上でようやく文字を判読できる。これらすべての作業をこなせる人材が不足している[4]。

資料を読む方式にも問題があった。長い間、碑文資料は書物の中の文字目録として蓄積されるだけであった。石の写真や材質の情報は一緒に収録されていなかった[4]。機械に入力できる情報が文字だけに限られるという意味である。

発掘は現在も続いている。毎年、新たな碑石や新たなパピルスが地中から発見される。解読すべき資料は増える一方だが、解読できる専門家の数は変わらない[4]。この格差は年を追うごとに広がっていった。

そのため、多くの記録が解読されないまま博物館の箱の中に眠ることとなった。石や紙は存在していても、そこに記された文章を蘇らせる人材が足りなかった。研究者たちはこの行き詰まりを打開するため、機械を導入した[4]。

参考資料

- [1] Assael, Y., Sommerschield, T., & Prag, J. (2019). Restoring ancient text using deep learning: a case study on Greek epigraphy. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 6368-6375). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1668>
- [2] Assael, Y., Sommerschield, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. Nature, 603 (7900), 280-283. <https://doi.org/10.1038/s41586-022-04448-z>
- [3] Assael, Y., Sommerschield, T., Cooley, A., Pavlopoulos, J., Shillingford, B., Herms, B., Suresh, P., Maynard, B., Grayston, J., Wulgaert, R., Prag, J., Mullen, A., & Mohamed, S. (2025). Contextualizing ancient texts with generative neural networks. Nature, 645(8079), 141-147. <https://doi.org/10.1038/s41586-025-09292-5>
- [4] Sommerschield, T., Assael, Y., Pavlopoulos, J., Stefanak, V., Senior, A., Dyer, C., Bodel, J., Prag, J., Androutsopoulos, I., & de Freitas, N. (2023). Machine learning for ancient languages: A survey. Computational Linguistics, 49(3), 703-747. https://doi.org/10.1162/coli_a_00481
- [5] Reply. (2026, 3月 19日). Austrian Academy of Sciences is developing the Ancient Greek AI "Apollo" with Mistral AI and Reply. reply.com. <https://www.reply.com/en/newsroom/news/austrian-academy-of-sciences-is-developing-the-ancient-greek-ai-apollo-with-mistral-ai-and-reply> (2026年9月3日閲覧)

11.2. 科学捜査隊の追跡：古代テキスト自動補完装置IthacaとApollo

穴埋め問題で学ぶ機械

人工知能が古代の文章を読む方法は、国語の試験の穴埋め問題に似ている。機械に無傷の碑文を大量に見せる。その上でいくつかの文字を隠し、何が入るかを当てさせる。回答が間違っていれば、機械内部の数値を少しずつ修正する[4]。この過程を何百万回も繰り返すことで、機械はギリシャ語の文章の癖を習得する。

このように学習した機械は単語を暗記するのではない。ある単語の後にどのような単語が来やすいかを確率として記憶する[4]。そのため、初めて目にする碑文であっても、空欄にふさわしい候補を提示することができる。

PythiaからIthacaへ

2019年、研究者たちがPythiaというモデルを発表した。ヤニス・アサエル、テア・ゾンマーシルト、ジョナサン・プラグが開発した。Pythiaは、欠損したギリシャ語碑文の文字を予測するニューラルネットワークである[1]。ニューラルネットワークとは、データを観察して規則を自ら習得する計算構造である。人間の脳の神経細胞の結合を模して作られた。

Pythiaは文字単位と単語単位を同時に読み取る。単語の半分が壊れた資料を扱うには、両方の視点が必要だからである。人間が補った文字のエラー率が57.3%であったのに対し、Pythiaは30.1%を記録した。正解がPythiaの上位20個の候補に含まれていた割合は73.5%であった[1]。

2022年にはIthacaが登場した。Google DeepMindと複数の大学が共同で開発した。オックスフォード大学、ヴェネツィア・カ・フォスカリ大学、アテネ経済ビジネス大学が参加した。研究成果は2022年3月、学術誌『Nature』に掲載された[2]。名称は叙事詩『オデュッセイア』に登場するイタカ島に由来する。長い放浪の末に故郷へ帰還するという意味が込められている。

Ithacaはどのように判断するのか

Ithacaはトランスフォーマーと呼ばれる構造を採用している。トランスフォーマーは、文中の単語同士が互いにどれほど関連しているかを評価するニューラルネットワークである。文字が欠損した箇所は、空欄を表す記号に置き換えられて入力される。モデルはその記号を置き換える文字の確率を計算する[2]。

Ithacaが担う役割は3つある。欠損した文字を復元し、碑文が本来どこにあったかを特定し、いつ刻まれた文章かを推定することである。この3つを個別に学習するのではなく、同時に学習する。方言と綴りの癖は場所を特

定する手がかりとなる。単語の用法と文章構造は時期を特定する手がかりとなる。ひとつのニューラルネットワークがこれら複数の手がかりを同時に扱うことで、互いに補完し合う[2]。

Ithacaは石の写真を見ない。書き写された文字のみを読んで判断する[2]。画の形状から時期を判定していた従来の手法とはアプローチが異なる。

年代を提示する方式も異なる。Ithacaは特定の年をひとつだけ断定して示すことはしない。紀元前800年から西暦800年までを10年単位に分割する。そして各区間に確率を割り当ててグラフとして提示する[2]。学者はグラフのピークを見て、どの時期が有力であるかを判断する。

学習に用いたデータは、パッカード人文科学研究所の碑文コレクションである。前述した78,608点はそのデータである。10回以上出現する単語を数えるだけでも35,884語にのぼった[2]。

研究陣はIthacaを開発するにあたり、3つの目標を掲げた。学者と協働すること、判断を支援すること、そして判断プロセスを可視化することである。人間に取って代わる機械を目指したのではなかった。学者の研究プロセスに自然に組み込まれるツールを目指したのである[2]。

Ithacaはひとつの答えだけを提示するのではない。空欄に入る候補を確率とともに複数提示する。どの文字に着目してそのように判断したのかも画面上に表示する。学者は機械が何を根拠にして回答したのかを目で確認できる。最終的な判断は人間の手に委ねられる。研究陣はIthacaのコードとデータを一般に公開した[2]。

20個の候補が持つ意味

機械が提示する答えはひとつではない。確率が高い順に候補を並べる。最初の候補が正解だった割合をTop-1正解率と呼ぶ。正解が上位20個の候補に含まれていた割合はTop-20正解率である[1]。

この区別が重要である。学者に必要なのは単一の正解ではない。検討に値する候補のリストである。20個の中に正解が含まれていれば、学者はそこから選ぶことができる[1]。自分では思いつかなかった単語をそのリストから発見することもある[2]。

ラテン語へと渡ったAeneas

同じ研究陣はラテン語碑文へと対象を広げた。2025年7月、『Nature』誌にAeneasというモデルを発表した[3]。名称はローマ建国神話の主人公Aeneasに由来する[5]。

Aeneasは176,000点を超えるラテン語碑文データをもとに学習した。ローマ、ハイデルベルク、クラウス・スラビ（Clauss-Slaby）の3つの碑文データベースを統合した資料である。10文字以内が欠損した箇所において、正解が上位20個の候補に含まれていた割合は73%であった。何文字が欠損しているか不明な場合でも58%を記録した[3]。欠損した長さを把握できないまま答えることは、人間にとっても極めて難しい難題である。

碑文がローマのどの属州から出土したかも特定した。属州とはローマが統治していた地方領土のことである。62の選択肢の中から72%の確率で正解した[5]。刻まれた時期については、歴史学者が定めた推定範囲から平均13年以内にまで絞り込んだ[3, 5]。

Aeneasには新たな機能が備わっている。碑文の写真と共に入力して判断を下すことができる[3]。書き写された文字情報だけでなく、石の形状や外観まで併せて観察するという意味である。類似した文章構造を持つ他の碑文を検索し、並べて提示することもできる。学者が数日かけて探していた類似の事例を、わずか数秒で抽出してくれる[5]。

研究チームはこの道具をローマ皇帝アウグストゥスの記録に試してみた。Aeneasは、その文が刻まれた時期として二つの区間を同時に示した。一つは紀元前10年から紀元前1年までだった。もう一つは西暦10年から西暦20年までで、確率はこちらのほうが高かった[5]。学界で分かれていた二つの説が、確率グラフの上にそのまま現れた。

研究チームは歴史学者23人にこの機能を使わせた。学者たちは、碑文10点のうち9点で新しい研究アイデアを得たと答えた[3, 5]。このモデルは誰でも使えるよう公開された[5]。

ギリシア語をまるごと学ぶアポロ

2026年3月、オーストリア科学アカデミーが新しいモデルを発表した。名前はアポロである。AI企業Mistral AIと情報技術企業Sail Replyが共同で作っている[7]。Ithacaが碑文一点を扱う道具だとすれば、アポロはギリシア語全体を学ぶモデルを目標にしている[6]。

このようなモデルをファウンデーションモデルと呼ぶ。多くの用途に広く使うため、幅広く学習しておいた基本モデルという意味である。一つの作業だけを行うように作られた道具ではない[4]。碑文の復元にも、文献検索にも、文字判読にも同じ基盤を使う[6]。

アポロが学ぶ資料は約6億語分のギリシア語文献である。そこに、すでに出版された碑文とパピルス数万点が加わる[6, 7]。研究チームは、これがこれまで集められたギリシア語資料の中で大規模なものだと明らかにした[7]。

オーストリア国立図書館は多くのパピルスを所蔵している。この図書館だけで数万点が保管されている[7]。その中には、まだ読まれていない文書が混ざっている。

研究を率いる人物は、オーストリア考古学研究所のアンナ・ドルガノフである。彼女は、まだ誰も読んでいないギリシア語パピルスが世界に100万点あると述べた[7]。アポロは、こうした資料から文字を復元し、内容を探し出す役割を担う。手書き文字を読み取ることも目標に含まれている[6, 7]。研究チームは、この作業が数年ではなく数時間で終わり得ると明らかにした[7]。

参考資料

- [1] Assael, Y., Sommerschield, T., & Prag, J. (2019). Restoring ancient text using deep learning: a case study on Greek epigraphy. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 6368-6375). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1668>
- [2] Assael, Y., Sommerschield, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. Nature, 603 (7900), 280-283. <https://doi.org/10.1038/s41586-022-04448-z>
- [3] Assael, Y., Sommerschield, T., Cooley, A., Pavlopoulos, J., Shillingford, B., Herms, B., Suresh, P., Maynard, B., Grayston, J., Wulgaert, R., Prag, J., Mullen, A., & Mohamed, S. (2025). Contextualizing ancient texts with generative neural networks. Nature, 645(8079), 141-147. <https://doi.org/10.1038/s41586-025-09292-5>
- [4] Sommerschield, T., Assael, Y., Pavlopoulos, J., Stefanak, V., Senior, A., Dyer, C., Bodel, J., Prag, J., Androutsopoulos, I., & de Freitas, N. (2023). Machine learning for ancient languages: A survey. Computational Linguistics, 49(3), 703-747. https://doi.org/10.1162/coli_a_00481
- [5] Google DeepMind. (2025, 7月 23日). Aeneas transforms how historians connect the past. [deepmind.google](https://deepmind.google/blog/aeneas-transforms-how-historians-connect-the-past/). (2026年9月3日閲覧)
- [6] Mistral AI. (2026). Austrian Academy of Sciences. [mistral.ai](https://mistral.ai/customers/austrian-academy-of-sciences/). (2026年9月3日閲覧)
- [7] Reply. (2026, 3月 19日). Austrian Academy of Sciences is developing the Ancient Greek AI "Apollo" with Mistral AI and Reply. [reply.com](https://www.reply.com/en/newsroom/news/austrian-academy-of-sciences-is-developing-the-ancient-greek-ai-apollo-with-mistral-ai-and-reply). (2026年9月3日閲覧)

11.3. 明らかになった真実

62パーセントと72パーセント

研究チームはIthacaの能力を数字で確認した。文字を意図的に隠した碑文を与え、補わせる試験である。同じ試験を訓練を受けた金石学者にも出した[1]。

人が一人で解いたとき、最初の答えが正しかった割合は25パーセントだった。Ithacaが一人で解いたときは62パーセントだった。2倍を超える差である[1]。

この試験には三つ目の条件があった。人がIthacaの候補一覧を見ながら一緒に解く条件である。すると正確度は72パーセントに上がった[1]。人だけより高く、機械だけよりも高い値である。

この数字がこの研究の核心である。機械は78,608点をすべて読んだ記憶で確率を計算する。人はその碑文がどのような歴史の中にあるかを知っている。機械が候補を20個並べると、人が歴史知識で選別する。二つが合わさるとき、正確度が最も高かった[1, 2]。

場所と時期を当てる能力も確認した。Ithacaはギリシア世界を84地域に分けて学習した。出土地が分からない碑文の本来の故郷を71パーセントの正確度で当てた。刻まれた時期については、歴史学者が定めた範囲から平均29.3年離れた値を示した。人の目で字体を測っていた方法では、誤差が100年に達することもあった[1]。

紀元前421年という答え

Ithacaは古くからの論争にも答えを示した。先に見たアテネ条約碑文の年代問題である。三画のシグマのために、紀元前446年ごろより前のものと見られていた碑文群である[1]。

研究チームはこれらの碑文をIthacaに入れた。Ithacaは文字の形ではなく、文そのものを見て判断する。どの語がどの語と結びついて使われるかを調べる。どの官職名とどの表現が一緒に出てくるかも調べる[1]。

Ithacaが示した平均年代は紀元前421年だった。古い基準より一世代後である。この値は、新しく出た別の証拠と同じ方向を向いていた。近年の学者が改めて付けた年代とは平均5年の差だった。古い基準で付けた年代は27年も離れていた[1]。

機械が一人で歴史を書き換えたわけではない。長く分かれていた二つの説のうち一方に重みを与えた[1, 2]。学者たちは、字体一つで年代を決めつけていた習慣を見直すことになった。一つの道具が100年続いた論争の方向を変えた。

機械が作り出すもっともらしい文

このような成果があっても、残る問題がある。学習した資料が一方に偏っている点である。ギリシア語碑文資料にはアテネのものが多く、時代も古典期に集中している。資料が少ない周辺地域の碑文は、その分だけ正確度が下がる[2]。

機械がなぜそのように答えたのか分かりにくいという問題もある。ニューラルネットワークの中には数百万個の数字が入っている。どの文字により注目したかは示すが、その答えを選んだ理由を言葉で説明することはできない[2]。

空欄が長くなるときの問題が起きる。数文字を補うことは得意だが、複数の語がまるごと消えた場所は難しい。Aeneasの試験でも、消えた長さが分からないとき、成績は73パーセントから58パーセントに下がった[3]。このときモデルは、文法に合い、もっともらしいが、実際には存在しなかった文を作り出すことがある。このように作り出した答えを幻覚と呼ぶ。機械が事実のように作った誤った答えという意味である。幻覚は古い記録を扱うとき危険である。存在しなかった歴史が事実のように固まることもあるためである[2]。

2026年8月に公開されたある研究は、この限界を正面から扱った。複数のモデルと複数の条件で、損傷した古文書を復元してみた実験である。研究チームは、復元を完全に自動化して任せることはできないと結論づけた。文書のどの場所を補うかによって成績が大きく変わった。決まった文句が繰り返される文書はうまく補えた。人によって違って書かれた文は難しかった。消えた長さを知っているかどうかによっても結果は分かれた。研究チームは、機械を古文書学者を助ける道具として作るべきだと明らかにした[4]。

研究者たちはこの危険を減らすために規則を置いた。機械が補った文字にも角括弧を付ける。確率が低い候補は学者が選別する。古い金石学者たちが守っていた表記上の約束を、機械にもそのまま適用する[2]。

そのため、Ithacaが出す答えは歴史的事実ではない。検討してみる仮説の一覧である。学者が別の資料と照らし合わせて確認して初めて史料になる[2]。史料とは、歴史を明らかにする根拠資料である。

7年の間に起きたこと

この分野の変化の速さは数字で比べることができる。2019年、Pythiaは人の誤り率57.3パーセントを30.1パーセントに下げた[5]。2022年、Ithacaは最初の答えの正確度を62パーセントまで引き上げた[1]。2025年、Aeneasはラテン語に移り、年代誤差を平均13年まで縮めた[3]。2026年には、ギリシア語全体を学ぶアポロが始まった[6]。

7年の間に対象言語が増え、正確度が上がった。扱う資料も文字から写真へ広がった[3]。その間、変わらなかったものが一つある。最後の判断は人が行うという原則である[2]。

写真と文脈、次の一步

研究は広がり続けている。2026年に出たある論文は、広く使われる大規模言語モデルをギリシア語資料で再学習させた。Metaが公開したLlama 3.1モデルを使った実験である。パピルス復元で最初の答えが正しかった割合は73.5パーセントだった。碑文復元では63.7パーセントが出た。正解が候補20個の中に入った割合は、それぞれ86.0パーセントと83.0パーセントだった。碑文専用で作られたモデルでなくても、この作業を行えるという意味である[7]。

文字画像をよみがえらせる研究も出た。2026年4月に公開されたMesaは、碑文写真で消えた画を補う方法である。同じ石碑に良く残った文字を手本にして、摩耗した部分の模様をよみがえらせる[8]。文を補うモデルと写真を補うモデルが並んで育っている。

失われた悲劇も実際に戻ってきている。2022年、エジプトのフィラデルフィア遺跡でパピルス一枚が見つかった。西暦3世紀に書かれたと見られる紙である。2024年、学者たちがその内容を読み、発表した。エウリピデスを書いた悲劇『ポリュイドス』と『イノ』の台詞97行だった。二つの作品は、それまで短い引用文と筋書きだけで知られていた[9]。

この復元は人の手と目が成し遂げた。機械が行ったことではない。そのような紙が世界にさらに100万点残っている。人の目だけでは、この紙の山をすべて読むことはできない。アポロのようなモデルが狙う場所は、まさにここである[6]。

道具が公開され、研究方法も変わった。IthacaとAeneasはコードと資料が公開されている[1, 10]。大学研究室でなくても、誰でもウェブで使ってみることができる。碑文写真を一枚アップロードすると、候補一覧と地図と年代グラフが出る[10]。以前なら数年かかった比較作業である[3]。

残された仕事はなお多い。ギリシア語パピルスだけで100万点が読まれないまま残っている[6]。碑文資料は地域と時代が均等ではない[2]。機械がうまく扱えない長い欠損部もそのまま残っている[3]。これらの問題は、より良い資料とより多くの検証によって一つずつ減らしていかなければならない[4]。

機械が候補を出し、人が選ぶ。Ithacaを62パーセントから72パーセントに引き上げた方法が、その答えである[1]。沈黙していた石と紙が再び語るには、両者がそろっていなければならない。

参考資料

- [1] Assael, Y., Sommerschild, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. *Nature*, 603 (7900), 280-283. <https://doi.org/10.1038/s41586-022-04448-z>
- [2] Sommerschild, T., Assael, Y., Pavlopoulos, J., Stefanak, V., Senior, A., Dyer, C., Bodel, J., Prag, J., Androutsopoulos, I., & de Freitas, N. (2023). Machine learning for ancient languages: A survey. *Computational Linguistics*, 49(3), 703-747. https://doi.org/10.1162/coli_a_00481
- [3] Assael, Y., Sommerschild, T., Cooley, A., Pavlopoulos, J., Shillingford, B., Herms, B., Suresh, P., Maynard, B., Grayston, J., Wulgaert, R., Prag, J., Mullen, A., & Mohamed, S. (2025). Contextualizing ancient texts with generative neural networks. *Nature*, 645(8079), 141-147. <https://doi.org/10.1038/s41586-025-09292-5>
- [4] Zhang, S., Caraffa, E., Zuffrano, A., Modesti, M., & Colavizza, G. (2026, 8月 28日). Text restoration of ancient documents with language models (arXiv:2608.28170). arXiv. <https://arxiv.org/abs/2608.28170> (2026年9月3日閲覧)

- [5] Assael, Y., Sommerschield, T., & Prag, J. (2019). Restoring ancient text using deep learning: a case study on Greek epigraphy. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 6368-6375). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1668>
- [6] Reply. (2026, 3월 19일). Austrian Academy of Sciences is developing the Ancient Greek AI "Apollo" with Mistral AI and Reply. reply.com. <https://www.reply.com/en/newsroom/news/austrian-academy-of-sciences-is-developing-the-ancient-greek-ai-apollo-with-mistral-ai-and-reply> (2026年9月3日閲覧)
- [7] Cullhed, E. (2026). Instruction-tuning pretrained causal language models to restore ancient Greek papyri and inscriptions. Digital Scholarship in the Humanities, 41(1), 69-76. <https://doi.org/10.1093/llc/fqaf131>
- [8] Toulatzis, V., Theodoridou, S., & Fudos, I. (2026, 4월 19일). MESA: A training-free multi-exemplar deep framework for restoring ancient inscription textures (arXiv:2604.17390). arXiv. <https://arxiv.org/abs/2604.17390> (2026年9月3日閲覧)
- [9] O'Grady, E. (2024, 10월 11일). Unearthed papyrus contains lost scenes from Euripides' plays. The Harvard Gazette. <https://news.harvard.edu/gazette/story/2024/10/unearthed-papyrus-contains-lost-scenes-from-euripides-plays/> (2026年9月3日閲覧)
- [10] Google DeepMind. (2025, 7월 23일). Aeneas transforms how historians connect the past. deepmind.google. <https://deepmind.google/blog/aeneas-transforms-how-historians-connect-the-past/> (2026年9月3日閲覧)

第12章. 永遠に燃えない仮想図書館： 超高解像度マルチスペクトルアーカイブと未来

12.1. 事件ファイル

アレクサンドリア図書館が消えた過程

エジプトのアレクサンドリアには、古代世界最大の図書館があった。プトレマイオス朝が紀元前3世紀ごろに建てた。王たちは世界の本を一か所に集めようとした。港に入った船の本は取り上げて写させた。写本を持ち主に返し、原本は図書館が持った。古い記録は巻物が数十万本あったと伝えている [1].

多くの人は、この図書館が一度の大火で消えたと考える。しかし研究者たちは、ゆっくり崩れていったと見る。紀元前145年には、王が外国人学者を追い出した。学者が去ると、本を世話する手もともに去った。ローマ時代には建物が何度も壊れた。新しい本を買い入れる資金も途絶えた [1].

古代の文章は、もともと残りにくい。当時の書記たちは、新しい文章を作るより、ある本を写す仕事を多くした。多くの手を経たパピルス巻物は、百年ほどたつと古びて読みにくくなった。乾燥したエジプトの地中では、はるかに長く残ることもあった。そのため、誰かが時機を逃さず写し直さなければ文章は続かなかった。写す手が途絶えると、その本は世界から消えた。

私たちが今日読む古代文献は、その時代の文章の一部だ。名前だけ伝わり、本文は消えた本のほうがはるかに多い。古代ギリシアの悲劇作家たちが書いた作品の大半も、そうして失われた。残った本は、運よく各地に写本が散らばっていたものだ。

破壊は今も続く

記録を壊す力は、昔だけにあったものではない。戦争は今も図書館と記録保管所を崩す。ユネスコはウクライナで壊れた文化施設を一つずつ確認して数えている。2026年7月29日時点で確認された場所は545か所だ。その中には博物館43か所、図書館22か所、記録保管所5か所が含まれる。宗教施設と古い建物も数百か所がともに被害を受けた [2].

戦争は遺物を壊すだけではない。管理が止まった遺物保管倉庫から、遺物がひそかに持ち出されることもある。どこから出たのかという記録が消えた遺物は、研究価値を大きく失う。粘土板一枚がどの部屋でどの板と一緒にあったのかが、そのまま情報だからだ。

気候も記録を脅かす。海辺にある遺跡は波に削られ、少しずつ崩れる。地中の遺物は雨が増えたり地下水が上がったりすると、より速く傷む。発掘する速度より、消える速度が速い場所もある。

壊れた遺物がすべて消えるわけではない。形が残っていれば、再び立て直す道がある。そのため、事故が起きる前にあらかじめ測っておくことが重要だ。戦争が続く地域では、建物と遺物を3次元で測っておく作業が進んでいる。

素材ごとに寿命が違う

記録を載せる素材は、それぞれ別の形で壊れる。メソポタミアの粘土板は、泥を押して文字を刻んだ板だ。宮殿が燃えたときに生じた熱は、粘土板を陶器のように硬く焼いた。火に遭わなかった粘土板は、水が入ると再び泥に戻る。

粘土板には別の限界もある。書記たちは使い終えた板を水でぬらし、表面をならして再び使った。そのため古い板はたいてい再び泥に戻った。王国が崩れる直前の数年分だけが残る場合が多い理由だ。宮殿が燃えたとき、書庫にあった最後の板だけが硬く焼かれたからだ。

パピルスと皮と紙はさらに弱い。これらの素材は植物と動物に由来する有機物だ。暑く湿った場所では、数百年ももたずに腐る。ローマ兵が使っていた蠟板では、蠟がまるごと消えた。残ったのは、鉄筆が木板まで引っかいて通った浅い傷だけだ。

硬い素材だから安全というわけでもない。亀の腹甲と牛の肩甲骨に刻んだ甲骨文字も、地中で押されて割れた。今日残る甲骨は、ほとんどが手のひらより小さい破片だ。一つの文が三つの破片に散らばり、別々の倉庫に入った場合もある [3]。

火が救った図書館

記録が生き残る方法は、ときに意外だ。アッシリア王アッシュルバニパルはニネベに大きな書庫を置いた。彼は紀元前669年ごろから国を治めた。書庫には神話、占い、医学、行政文書が集められていた [4]。

紀元前612年ごろ、ニネベは焼けた。紙の本であれば、そのときすべて消えていただろう。ところが粘土板は火で焼かれ、かえって硬くなった。これらの板は1840年代から1930年代まで、何度も発掘された。現在、大英博物館にはその破片が3万点以上ある [4]。

だからといって、完全な状態で出てきたわけではない。都市が崩れるとき、板はわざと割られ、宮殿のあちこちに捨てられた。発掘者たちが出会ったのは、整理された書架ではなく、散らばった破片の山だった。一つの作品の前後が別々の部屋から出ることは珍しくなかった [4]。これらの破片を再び合わせることが、今日の計算考古学の大きな課題だ。計算考古学は、コンピューターで古い遺物を研究する分野だ。

すでによみがえった文字

デジタルに移す作業にも壁がある。世界の遺物は、人の手よりはるかに多い。先に見たニネベの破片の山がその例だ [4]。一点一点撮影し、読み取った文字まで付けるには長い時間がかかる。そのため、機械が人の速度を上げなければならない。

消えつつある遺物を救う方法は二つの方向に分かれる。一つは物そのものを長く守る保存だ。もう一つは、物に触れずに内部を読み、コンピューターに移す作業だ。後者の方法は、すでに何度も成果を上げている。

イスラエルのエン・ゲディから出た巻物は、火で焼けて炭の塊になっていた。手で広げようとすれば、そのまま崩れる状態だった。研究チームは、物を切らずに内部を見るCT撮影で中をのぞいた。コンピューターは巻かれた皮の層を数学的に追い、平面に開いた。その上で金属成分が混じったインクが現れ、レビ記の冒頭が読まれた [5]。

ヴェスヴィオ火山灰に埋もれたヘルクラネウム巻物にも道が開けた。2026年6月25日、研究陣は巻物一つを最初から最後まで読み取ったと発表した。この巻物の名前はPHerc. 1667だ。中にはギリシア語で22段分の倫理の話が含まれていた [6]。課題に指定された封印巻物13本には、新しい賞金がかけられた。締め切りは2027年6月25日だ [7]。

仮想図書館は、写真を大量に集めておいた倉庫ではない。遺物の形、インク成分、読み取った文章が一か所に結び付けられていなければならない。そうして初めて、後日に新しい方法が出たとき、古い資料を再計算できる。原本が消えた後も検証が続くには、このまとまりが必要だ。

撮影と判読には資金と人手がかかる。どの遺物を先に撮るかを決めることも容易ではない。危険な地域の遺物は、撮影そのものが難しい。そのため、どの国も順序を決めて作業する。

これらの事件は一つの問いを残す。物が最後には壊れても、その中の文章は残せるのか。その答えを探す場所が仮想図書館だ。

参考資料

- [1] Mark, J. J. (2023, 7월 25일). Library of Alexandria. World History Encyclopedia.
https://www.worldhistory.org/Library_of_Alexandria/ (2026年9月3日閲覧)

- [2] UNESCO. (2026). Damaged cultural sites in Ukraine verified by UNESCO. UNESCO.
<https://www.unesco.org/en/ukraine-war/damaged-cultural-sites> (2026年9月3日閲覧)
- [3] Ma, Y. R. (2026). Towards Computational Chinese Paleography. arXiv:2601.06753. <https://arxiv.org/abs/2601.06753>
(2026年9月3日閲覧)
- [4] The British Museum. What was Ashurbanipal's Library? The British Museum.
<https://www.britishmuseum.org/research/projects/what-was-ashurbanipals-library> (2026年9月3日閲覧)
- [5] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science Advances, 2(9), e1601247.
<https://doi.org/10.1126/sciadv.1601247>
- [6] Vesuvius Challenge. (2026, 6月 25日). An entire Herculeaneum scroll has been read for the first time. scrollprize.org.
<https://scrollprize.org/firstscroll> (2026年9月3日閲覧)
- [7] Vesuvius Challenge. (2026). Open Prizes. scrollprize.org. <https://scrollprize.org/prizes> (2026年9月3日閲覧)

12.2. 科学捜査隊の追跡: 知識グラフと光のスペクトル分析

遺物をつなぐ大きな地図

昔の研究室では、遺物の形と文字を別々に扱った。形は保存科学者が測り、文字は文献学者が読んだ。二つの結果は別々の引き出しに入った。そのため、同じ遺物を二度研究しながらも、答えを合わせることができなかった。

計算考古学は、この二つを一つの地図の上に載せる。その地図の名前が知識グラフだ。知識グラフは、情報を点と線で描いた図だ。点は物であり、線は物同士の関係だ。

この地図では、粘土板一枚が点になる。その板を書いた書記も点になり、板が出た都市も点になる。インクの成分と3次元形状情報も、それぞれ点になる。板に書かれた人名と物の名前も点として取り出しておく。コンピューターは、これらの点をつなぐ線を探して学習する。

地図が作られると、質問の形が変わる。以前は板一枚を置いて、何の文字かを尋ねた。今は、このインクを使った別の文書がどこにあるかを尋ねることができる。同じ書記が書いた板がどの都市から出たのかも、一度に探せる。

2026年2月、ポーランドのヤギェウォ大学研究チームがこのような地図を公開した。この地図には、人、事件、物の関係が約320万行入っている。研究チームは、機械が関係を推薦し、その理由まで説明するようにした。専門家たちはその説明を読み、使えるかどうかを確認した [1]。

形を測る層

地図の第一層は遺物の外形だ。古い石碑と粘土板は表面が摩耗し、文字の溝が浅くなる。目では溝が残っているか分かりにくい。そのため研究員たちは3次元スキャナーで表面をなぞる。

スキャナーは表面を数百万個の点に変えて保存する。コンピューターはその点をつないで三角形の網にする。この網の上で、仮想の光をさまざまな方向から当てることができる。光の方向を変えると、浅い溝にも影ができ、文字が現れる。

機械は溝の断面の形も学ぶ。人が道具で彫った跡は、断面が整い、方向が一定だ。自然にできた亀裂はそうではない。この違いを学んだモデルは、文字と傷を分ける。

同じ方法で、割れた断面も測る。二つの破片の断面の起伏が互いにかみ合うかを、機械が照合する。この計算は、人が破片を直接持って合わせてみる作業に代わる。手の届かない別の国の博物館にある破片とも合わせてみることができる。

巻かれた巻物には、別の計算も使われる。エン・ゲディ巻物で、研究チームは皮の層をコンピューター内で一層ずつ分けた。そして各層を平面に開き、文字を立ち上げた [2]。

光でインクを読む層

地図の第二層はインクだ。インクは素材ごとに光を異なる形で受け止める。マルチスペクトル撮影は、目に見える光だけでなく、赤外線と紫外線まで分けて撮る方法だ。ある波長では、下地が暗く、インクが明るく浮かぶ。

この違いのおかげで、消された文字もよみがえる。昔の人々は高価な羊皮紙を節約するため、文字を消して再び書いた。消した表面にはインク成分が少し残る。複数の波長で撮って重ねると、その痕跡が下から浮かび上がる。このように二度書かれた文書をパリンプセストと呼ぶ。

エジプトのシナイ山にある聖カタリナ修道院には、このような文書が多い。アメリカとギリシアの研究陣が、ここで74冊をマルチスペクトルで撮影した。撮影した面は6,800ページに達する。その下から、消された文章305編が再び現れた。それらの文章は、ギリシア語とシリア語をはじめとする10の言語で書かれていた [3]。

成分そのものを測る方法もある。ラマン分光法は、インクに光を当て、跳ね返ってくる光の模様を読む。模様は物質ごとに異なるため、指紋の役割をする。AIはこの模様を見て、煤で作った墨なのか、鉄成分の入ったインクなのかを分ける。インクの配合は工房ごとに少しずつ異なり、出所を推測する手がかりになる。

ヘルクラネウム巻物には、さらに強い光が使われた。研究陣は、きわめて強いX線を作る放射光加速器で巻物の内部を撮影した。焼けたパピルスと炭素墨は密度が似ており、そのままでは区別できない。ディープラーニングモデルが微細な筋目の違いを学び、文字があった場所を見つけ出した [4]。

文章と文章をつなぐ層

地図の第三層は、文書同士の関係だ。古代の本は写されながら広がった。写すたびに小さな誤りが一つずつ入り込んだ。学者たちはこの誤りを追い、写本の系譜を描く。

コンピューターは、この作業に生物学で使われていた計算法を借りてくる。遺伝子配列を比較し、生物の系統樹を描く方法に似ている。文書の差を数え、どの写本がどの写本から出たのかを並べる。このように描いた系譜は、失われた原本の姿を推測させてくれる。

筆跡も関係をつなぐ手がかりだ。書記ごとに筆を押す力と線の角度が違う。AIはこの癖を数値に変えて保存する。その数値が近い文書同士をまとめると、書記の関係網が現れる。

すでに動いているアーカイブ

こうした層を集め、実際に動いているアーカイブがいくつもある。トリスメギストスは、古代世界の文書に番号を付ける事業だ。ここは100を超える言語の資料およそ96万件を一つの体系にまとめた。同じ遺物を国ごとに違う名前と呼んでいた問題は、この番号で解けた [5]。

ペラギオス・ネットワークは、散らばった資料を互いにつなぐ共同体だ。彼らは場所、人、時間をつなぐ共通の方式を定める。古い地名が時代ごとに変わっても、一つの座標で結ばれるようにする [6]。ミュンヘンで運営されている電子バビロニア図書館がある。ここは世界各地の楔形文字の破片を一つのデータベースに集める。学者たちはここで破片を検索し、失われた文を取り戻す [7]。

共通番号は退屈に見えるが、力が強い。番号が同じでなければ、異なる機関の資料は一つの行としてつながらない [5]。写真は博物館に、読み取られた文は大学に、発掘記録は研究所にある場合が多い。番号がこの三つを結びつけられれば、散らばった情報が一つの遺物の物語として集まる。

資料を公開する方法も変わった。以前は写真と読み取られた文を、論文が出た後でなければ見ることができなかった。近年の研究チームは、スキャン資料とプログラムをともに公開する。そうすれば、装置を持たない国の研究者も同じ資料で再計算できる [8]。

まだ残る課題

ヴェスヴィオ・チャレンジは2026年7月10日に残る問題を整理して公開した。押しつぶされた場所では、巻物の層が互いに分離しない。コンピューターが層を追っている途中で、隣の層へ誤って飛び移ることもある。インクが表面からどれほど深く染み込んだのかも、まだ確実ではない。一つの巻物で学んだモデルが、別の巻物では力を失う。資料の規模が大きいため、普通のコンピューターでは扱いにくい。研究陣が定めた次の目標時点は2027年6月である[8]。

古い文字を扱う別の分野も同じ壁にぶつかる。中国古文字研究を整理した2026年の論文は、資料不足を大きな難題として挙げた。そして人間の判断に代わるよりも、それを助ける方向を勧めた[9]。

参考資料

- [1] Kutt, K., Sroka, E., Ishchuk, O., & do Valle Miranda, L. (2026). A three-stage neuro-symbolic recommendation pipeline for cultural heritage knowledge graphs. arXiv:2602.19711. <https://arxiv.org/abs/2602.19711> (2026年9月3日閲覧)
- [2] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science Advances, 2(9), e1601247. <https://doi.org/10.1126/sciadv.1601247>
- [3] Early Manuscripts Electronic Library. Sinai Palimpsests Project. EMEL. <https://www.emelibrary.org/sinai-palimpsests-project/> (2026年9月3日閲覧)
- [4] Vesuvius Challenge. (2026, 6月 25日). An entire Herculaneum scroll has been read for the first time. scrollprize.org. <https://scrollprize.org/firstscroll> (2026年9月3日閲覧)
- [5] KU Leuven. Trismegistos portal. Trismegistos+. <https://tplus.kuleuven.be/trismegistos-portal> (2026年9月3日閲覧)
- [6] Pelagios Network. Pelagios Network. <https://pelagios.org/> (2026年9月3日閲覧)
- [7] Ludwig-Maximilians-Universität München & Bayerische Akademie der Wissenschaften. electronic Babylonian Library. <https://www.ebl.badw.de/> (2026年9月3日閲覧)
- [8] Vesuvius Challenge. (2026, 7月 10日). Open problems: Why reading every Herculaneum scroll is still a challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026年9月3日閲覧)
- [9] Ma, Y. R. (2026). Towards Computational Chinese Paleography. arXiv:2601.06753. <https://arxiv.org/abs/2601.06753> (2026年9月3日閲覧)

12.3. エピローグ

十一の事件を振り返る

本書は十一の事件を追ってきた。始まりは火山灰に埋もれたヘルクラネウムの巻物だった。触れれば崩れる炭の塊を、コンピューターが仮想的に広げた。焼けたパピルスと炭素インクは、X線で見るとほとんど同じに見える。機械は表面に残った微細な筋の違いを学び、文字を見つけ出した[1]。

エン・ゲディのレビ記巻物も、仮想展開で最初の句を取り戻した。こちらでは金属成分を含むインクがX線にはっきり写った[2]。イザヤ書巻物では筆跡が手がかりだった。人工知能は筆画の微細な癖を測り、書記が一人ではないことを示した。人間の目には同じ文字にしか見えなかった部分だった[3]。

四つ目の事件からは、断片を合わせる作業が続いた。散らばった粘土板と羊皮紙の破片が、断面の形によって再びつながった。偽造された遺物を見分ける作業にも、同じ計算が使われた。『ギルガメシュ叙事詩』の欠けた行は、破片をつなぎ合わせて埋められた。

メソポタミアの領収書は、機械翻訳によって言葉を取り戻した。数千年前の穀物帳簿と手紙が、今日の言葉に移された。亀の甲羅の甲骨文字は、3次元スキャンと専用モデルでつなぎ合わされた[4]。新羅の木簡も文字認識モデルが読み取った。日本の草書体文書も同じ方法で解かれた[5]。

まだ開かれていない扉も見た。線文字A、インダス文字、ロンゴロンゴには、今も答えがない。機械は文字配列の規則性を測り、それが言語がどうかを検討した。意味を読む仕事は、次の世代の役割として残った。

最後の事件は古代碑文の自動補完だった。Ithacaはギリシア語碑文の欠けた文字を補うモデルである。このモデルは文字を補いながら、碑文が作られた場所と時期もあわせて指摘する[6]。続いて登場したAeneasはラテン語碑文を扱う。Aeneasは転写された文字と碑文写真をともに見る[7]。

十一の事件には共通点がある。どれも機械だけで解いた問題ではなかった。物理学者が撮影を担当し、計算機科学者が計算を担当した。古代語を読む学者が結果を確認し、意味を付けた。三つの分野が同じ遺物をめぐって出会ってこそ、一行がよみがえった。

機械がまだできないこと

ここまで来る間に、機械の弱点もともに明らかになった。第一の壁は資料不足である。線文字Aは残った記号をすべて合わせても7,400個ほどである。この程度の量では、機械が規則を学ぶのは難しい[8]。今日の言語モデルは数十億語を読んで育つ。消えた古語の資料は、その規模にはるかに及ばない。

資料が少ない場所で無理に答えを出すと、存在しない語を作り出す。空欄が長くなるほど、答えの信頼度は下がる。文脈にはもっともらしいが、歴史には存在しなかった文が出ることもある。そのような文が一度論文に載れば、次の研究がそれを事実として受け取る。

学習資料が一方に偏っても、答えは傾く。ある地域の碑文が多ければ、別の地域の碑文までその地域の様式で読んでしまう。機械はなぜそのように答えたのか、自分で説明できないことが多い。そのため近年の研究は、根拠もともに示す方向へ移っている[9]。

巻物読解の研究チームも、残る問題を一覧にして公開した。押しつぶされた場所では層が互いに分離しない。一つの巻物で学んだモデルが、別の巻物では力を失うこともある。何ができないのかを明らかにすることが、次の歩みを作る[10]。

機械が示した文字を扱う規則も必要である。確実に読んだ文字と推測した文字は、表示を変えなければならない。どのモデルがどの資料によってそのように読んだのかも、あわせて記録しなければならない。そうしてこそ、後に別の研究者が同じ計算を再現できる。

原本は原本として残る

デジタル復元がどれほど精巧でも、原本に取って代わることはできない。学界はスキャンした資料を原本ではなく、デジタル上の分身と見る。撮影には人間の選択が入る。どの光を使い、どの角度で撮るかを人間が決めるからである。その選択が変われば、結果の画像も変わる。

実物にはスキャナーが収められない情報が残っている。粘土板表面の土粒の構造がそうである。亀の甲羅の微細な穴の配列も、実物でしか見ることができない。インクがパピルスの中へどれほど染み込んだのかも、実物でしか測れない[10]。

新しい測定技術は今後も登場する。いま捨てた遺物には、その技術を使うことができない。だから遺物を切ったり、無理に広げたりすることは控えなければならない。仮想図書館の目的は原本を捨てることではない。原本にあまり触れなくても読めるようにすることに目的がある[2]。遺物に触れる回数が減れば、原本の寿命は延びる。

ともに読む未来

人工知能は歴史家を押しつける道具ではない。Ithacaの研究はこの点を数字で示した。Ithacaが単独で欠けた文字を復元したとき、正確度は62パーセントだった。歴史家が単独で行ったときは25パーセントだった。両者が協力すると72パーセントに上がった[6]。

この数字は二つのことを同時に語っている。この実験では、機械だけでも人間よりよかった。そして人間とともに行ったとき、成績はさらに上がった。どちらか一方だけでは、ここまで到達できなかった。

機械は候補を速く大量に出す。人間はその候補がその時代にあり得るかを確かめる。このように人間と機械が交互に確認する方法を、学界は人間が輪の中にいる協業と呼ぶ。この輪が回るには、古い文を読める人が存在し続けなければならない。読む人がいなければ、機械が取り出した文字も、また一つの謎になる。

教えることも研究と同じくらい重要である。古代語を読む方法は、人から人へと受け継がれてきた。その鎖が切れれば、機械が取り出した文字を検証する人がいない。新しい世代は、古い文字とコンピューターをともに学ばなければならない。

この輪は専門家だけのものではない。日本では草書体を読んでもくれる無料アプリ「みを」が登場している。写真を撮ると、画面に現代の文字が表示される。古い文書を初めて見る人も、最初の一行を読むことができる[5]。博物館のガラスケースの前でスマートフォンを構えることが、研究の始まりになることもある。

これからすべきことはまだ多い。ヘルクラネウムには封印された巻物がまだ残っている[11]。ウクライナでは今も図書館と文書館が破壊されている[12]。世界の古い文は、今なお消える速度のほうが速い。

本書を読んだ人にできることもある。公開されたスキャン資料は、誰でもダウンロードして見ることができる[10]。断片を合わせるコンテストには学生も参加する[11]。古い文書を転写する作業に市民が加わる場も多い。一人が一行ずつ引き受けても、資料は蓄積される。

記録を守ることは昔話だけではない。今日私たちが作る資料も、いつか古い資料になる。ファイル形式が変われば、現在の文書も読めなくなる。そのため図書館は資料を新しい形式へ移す作業を続ける。仮想図書館を築く作業には、管理する手が常につきまとう。

それでも方向は定まった。2000年前に火山灰に埋もれた巻物が、再び読まれた[1]。一つの巻物から得た方法は、次の巻物研究の出発点になる[11]。資料とコードを公開しておけば、他国の研究者も同じ道を進む[10]。遺物はいずれ土に戻るとしても、その中の文は残る。火に烧けない図書館は、そうして一行ずつ築かれている。

参考資料

- [1] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. <https://scrollprize.org/firstscroll> (2026年9月3日閲覧)
- [2] Seales, W. B., Parker, C. S., Segal, M., Tov, E., Shor, P., & Porath, Y. (2016). From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. *Science Advances*, 2(9), e1601247. <https://doi.org/10.1126/sciadv.1601247>
- [3] Popović, M., Dhali, M. A., & Schomaker, L. (2021). Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsaa). *PLOS ONE*, 16(4), e0249769. <https://doi.org/10.1371/journal.pone.0249769>
- [4] Ma, Y. R. (2026). Towards Computational Chinese Paleography. arXiv:2601.06753. <https://arxiv.org/abs/2601.06753> (2026年9月3日閲覧)
- [5] ROIS-DS Center for Open Data in the Humanities. *みを(miwo): AIくずし字認識アプリ*. CODH. <https://codh.rois.ac.jp/miwo/> (2026年9月3日閲覧)
- [6] Assael, Y., Sommerschild, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. *Nature*, 603 (7900), 280-283. <https://doi.org/10.1038/s41586-022-04448-z>
- [7] Assael, Y., Sommerschild, T., Cooley, A. E., Shillingford, B., Pavlopoulos, J., Suresh, P., Herms, B., Grayston, J., Maynard, B., Dietrich, N., Wulgaert, R., Prag, J., Mullen, A., & Mohamed, S. (2025). Contextualizing ancient texts with generative neural networks. *Nature*. <https://doi.org/10.1038/s41586-025-09292-5>
- [8] Braović, M., Krstinić, D., Štula, M., & Ivanda, A. (2024). A systematic review of computational approaches to deciphering Bronze Age Aegean and Cypriot scripts. *Computational Linguistics*, 50(2), 725-779. https://doi.org/10.1162/coli_a_00514
- [9] Kutt, K., Sroka, E., Ishchuk, O., & do Valle Miranda, L. (2026). A three-stage neuro-symbolic recommendation pipeline for cultural heritage knowledge graphs. arXiv:2602.19711. <https://arxiv.org/abs/2602.19711> (2026年9月3日閲覧)
- [10] Vesuvius Challenge. (2026, 7월 10일). Open problems: Why reading every Herculaneum scroll is still a challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026年9月3日閲覧)
- [11] Vesuvius Challenge. (2026). Open Prizes. scrollprize.org. <https://scrollprize.org/prizes> (2026年9月3日閲覧)

[12] UNESCO. (2026). Damaged cultural sites in Ukraine verified by UNESCO. UNESCO.
<https://www.unesco.org/en/ukraine-war/damaged-cultural-sites> (2026年9月3日閲覧)

付録. 私だけのAI考古学旅行記： 一般向けグローバル探索ポータル

A.1. Vesuvius Challenge公式データアーカイブ

何を見ることができるか

Vesuvius Challengeは国際大会である。火で焼けた巻物をコンピューターで読む作業を競う。巻物が出た場所はヘルクラネウムである。ヴェスヴィオ火山のそばにあった古代ローマ都市である。巻物はパピルスで作られた。パピルスは草を押し固めて作った昔の紙である。公式アドレスはscrollprize.orgである[1]。このサイトは大会で使う資料を誰にでも公開している[2]。会員登録も不要で、費用もかからない[2][5]。

最初の画面の上部にはメニューが並んでいる。Prizes, Problems, Tutorials & In-depth, Dataが前に来る。その後にMilestones, Community, Press, Donateが続く。2026年9月現在、巻物とパピルス断片45点がスキャンされている。そのうち1点は最初から最後まで読まれた[1]。

2026年6月25日に大会チームはその1点の知らせを発表した。PHerc. 1667と呼ぶ巻物である[3]。巻物を開かずには最後まで読み取った初めての例である[3][4]。展開されたパピルスの長さは約1.4メートルである。古代の本は文字を縦の欄に分けて書いた。この欄一つを列と呼ぶ。この巻物ではギリシャ文字で書かれた列が二十二個ほど出た。内容はストア派の倫理を扱う文章である。人間の本性と衝動、道徳的成長を語る。紀元前2世紀ごろの文章と見られる[3]。

巻物全体が完全に残ったわけではない。今残っているのは巻物の内側の芯である。外側の層は19世紀と1980年代に手で開こうとして壊れた。表面がはがれ落ちた場所では文字も一緒に消えた。そのため読み取った文章にはところどころ空白がある[3]。

スキャンはフランスのグルノーブルで行った。欧州シンクロトロン放射光施設ESRFである。とても明るいX-rayを作る大きな研究装置である。このうちBM18ビームラインを使った[4]。ビームラインはX-rayを引き出して物体に当てる実験通路である。

家で試す手順

まずscrollprize.orgに入り、Get Startedボタンを押す。この画面は初めて来る人に三つの段階を案内する[5]。

第一段階はインク探しである。目に見えるインクを画面上で塗って示す。するとモデルがその印を学び、残りを予測する[5]。

第二段階は仮想展開である。コンピューターで巻物を開いてみる作業である。CT断面写真からパピルスの一層をたどる。CT断面写真とは、内部を薄く切って撮るように撮影した写真である。その層を平らに広げて結果を確認する[5]。

第三段階は全体の過程を身につけることである。スキャン、展開、読み取りを小さな規模で続けて試す[5]。

体験の順序と実際の復元の順序は違うことを知っておく必要がある。体験はやさしいインク探しから始まる。実際の復元ではスキャンが先である。その次に巻物を仮想的に開く。インクを読む作業は最後である[5]。

資料を直接ダウンロードするにはscrollprize.org/dataへ行く。資料は公開保存先s3://vesuvius-challenge-open-data/にある。ウェブブラウザでも開いて見ることができる。使い方を示すサンプルファイルはGitHubにある[2]。

挑戦者には賞金もかかっている。2026年9月時点で残っている賞金は214万ドルである。2027グランドプライズは100万ドル規模である。1位は80万ドルを受け取る。締切は2027年6月25日である。まだ文字が見つからない巻物にも賞金がかかっている。面積4平方センチメートルの中で文字10個を読み取ればよい。このとき

巻物1点ごとに5万ドルが与えられる。毎月与えられるProgress Prizeもある。その月に選ばれた最優秀提出物には2万ドルが行く[6]。

注意する点

資料はCC BY-NC 4.0条件で公開される。収益を得る目的では使えない。使うときは出典を明らかにしなければならない。古い資料には引用すべき論文が別にある。巻物1番から4番、断片1番から6番がそうである[2]。このときはParsonsらの2023年論文を記す[2][7]。

ファイルサイズが大きいことも前もって知っておく必要がある。3次元スキャン資料は全体で数テラバイトに及ぶ[2]。ディープラーニングを動かすには性能のよいグラフィックカードが必要である。家のコンピューターが弱ければ、資料を目で見て回ることから始めればよい。

賞金に挑戦するときは条件が付く。大会のDiscord部屋に参加しなければならない。コードはMITのような自由な条件で公開しなければならない。受賞発表までは結果を外に知らせない[6]。

参考資料

- [1] Vesuvius Challenge. (2026). Vesuvius Challenge: Reading the Herculaneum Scrolls with machine learning. scrollprize.org. <https://scrollprize.org/> (2026年9月3日閲覧)
- [2] Vesuvius Challenge. (2026). Data. scrollprize.org. <https://scrollprize.org/data> (2026年9月3日閲覧)
- [3] Vesuvius Challenge. (2026, 6月 25日). An entire Herculaneum scroll has been read for the first time. scrollprize.org. <https://scrollprize.org/firstscroll> (2026年9月3日閲覧)
- [4] Angelotti, G., Parsons, S., Nicolardi, F., et al. (2026). Complete virtual unwrapping and reading of a rolled Herculaneum papyrus. arXiv:2606.29085. <https://arxiv.org/abs/2606.29085> (2026年9月3日閲覧)
- [5] Vesuvius Challenge. (2026). Get Started. scrollprize.org. https://scrollprize.org/get_started (2026年9月3日閲覧)
- [6] Vesuvius Challenge. (2026). Open Prizes. scrollprize.org. <https://scrollprize.org/prizes> (2026年9月3日閲覧)
- [7] Parsons, S., Parker, C. S., Chapman, C., Hayashida, M., & Seales, W. B. (2023). EduceLab-Scrolls: Verifiable recovery of text from Herculaneum papyri using X-ray CT. arXiv:2304.02084. <https://doi.org/10.48550/arXiv.2304.02084>

A.2. eBL電子バビロニア文学ポータル

ポータルの骨組み

eBLはelectronic Babylonian Libraryの略である。日本語では電子バビロニア図書館である。古代メソポタミアの人々は楔形文字で文章を書いた。楔形文字は湿った粘土板に押し付けて刻んだ釘の形の文字である。このポータルはその粘土板断片を一か所に集める。集める対象はバビロニア文学を書いた断片である。世界のすべての楔形文字資料を収める場所ではない[1]。

ポータルはドイツのミュンヘン・ルートヴィヒ・マクシミリアン大学にある。エンリケ・ヒメネス教授が率いる。事業は2018年4月に始まった。アレクサンダー・フォン・フンボルト財団のソフィア・コワレフスカヤ賞助成金で運営している[2]。

アドレスはwww.ebl.lmu.deである。ログインなしで入ることができ、利用料もない。画面左側には八つの項目がある。eBL Project, Library, Corpus, Signsが前に来る。その後にAkkadian DictionaryとBibliographyが来る。最後にはNewsとArchaeologyがある[1]。

Libraryは断片一つ一つを集めておいた場所である。以前はFragmentariumと呼ばれていた。eBL研究員たちはここで断片を接合した事例を約1,200件見つけた[1]。Corpusは作品単位で本文を見せる。ギルガメシュ叙事詩は約3,150行に整理されている。創造神話エヌマ・エリシュは1,096行である[3]。2024年に出たデータ論文が規模を示している。このポータルは粘土板約25,000点を収める。転写文は35万行を超える[4]。転写文とは楔形文字をラテン文字に移して書いた文章である。

画面で何を試せるか

Libraryを押すと検索ボックスが出る。名前、年、題名で探すことができる。単語単位検索と高度検索もある。何を探すか分からなければI'm feeling luckyボタンを押す。すると任意の断片が一つ開く[5]。

断片一つを開くと画面がいくつかの欄に分かれる。上部には所蔵機関、寸法、出土地が書かれる。字体と文書タイプも一緒に出る。たとえばK.5743は大英博物館のクウンジク・コレクションである。クウンジクは古代都市ニネヴェがあった丘である。その丘から出た遺物をまとめて呼ぶ名前である。寸法は3.81センチメートル×2.54センチメートルである。字体は新アッシリア式である[6]。新アッシリア帝国時代に使われた文字の形である。

中央には転写文が行番号とともに出る。下部には粘土板の写真が付く。どの断片と接合したかも一緒に書かれる[6]。

文章を引用するときに使う方式もサイトが知らせてくれる。断片番号とeBL editionを書く。その後にアドレスとアクセス日を付ける[1]。

新着情報と注意する点

News項目でニュースレターを見ることができる。2026年5月28日付の22号が最新である。この号でポータルは新しい画面構成を出した。所蔵機関もいくつか増えた。仁川にある国立世界文字博物館もその一つである。大英博物館の写真3,100枚以上が新しく上がった。イェール・ピーボディ博物館の写真も1,000枚以上上がった。辞書に新しい単語を直接登録する機能もできた[7]。

使う前に知っておく点がある。Libraryの転写文は完成版ではなく暫定版である。研究陣が今も直し続けている。Corpusに整理された作品本文も同じである。これまで集めた断片で組み合わせた中間結果である。新しい断片が付くと行数と内容が変わる。したがって学校の課題や文章に使うときはアクセス日も一緒に書く[1]。

転写文はCC BY-NC-SA 4.0条件で公開される。写真は条件が違う。写真を再利用するには別途許可を受けなければならない[1]。粘土板を持つ博物館が許可する側である[1][4]。画面下の写真説明に付いた案内を先に読まなければならない[1]。

参考資料

- [1] electronic Babylonian Library. (2026). About: Library. ebl.lmu.de. <https://www.ebl.lmu.de/about/library> (2026年9月3日閲覧)
- [2] Jiménez, E. (2023, 5月 25日). AI is helping us read ancient Mesopotamian literature. The Conversation. <https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091> (2026年9月3日閲覧)
- [3] electronic Babylonian Library. (2026). Corpus. ebl.lmu.de. <https://www.ebl.lmu.de/corpus> (2026年9月3日閲覧)
- [4] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Härtinen, A., Heinrich, A., Mitto, T., Rozzi, G., Sáenz, L., & Jiménez, E. (2024). Transliterated cuneiform tablets of the electronic Babylonian Library platform. Journal of Open Humanities Data, 10(1), 19. <https://doi.org/10.5334/johd.148>
- [5] electronic Babylonian Library. (2026). Library. ebl.lmu.de. <https://www.ebl.lmu.de/library> (2026年9月3日閲覧)
- [6] electronic Babylonian Library. (2026). K.5743, eBL edition. ebl.lmu.de. <https://www.ebl.lmu.de/library/K.5743> (2026年9月3日閲覧)
- [7] electronic Babylonian Library. (2026, 5月 28日). eBL Newsletter 22. ebl.lmu.de. <https://www.ebl.lmu.de/news> (2026年9月3日閲覧)

A.3. Miwo使用案内

どのようなアプリか

日本の古い本と文書には崩し字が多く使われた。この文字をくずし字と呼ぶ。江戸時代の版本と古文書でよく見られる[3]。今日の日本人の多くはこの文字を読めない。

miwoはこの崩し字を写真で撮り、現代の文字に変えてくれるスマートフォンアプリである。日本の人文学オープンデータ共同利用センターCODHが作った[1]。料金は無料である[2]。アプリ名は『源氏物語』第十四帖「滯標」から取った。滯標は船の航路を知らせる標識である。古文書の海で道案内になるという意味である[3]。

アプリは2021年8月30日に初めて出た。2022年10月26日に1.1版へ上がった。このとき「瑠璃」という新しい認識モデルが入った。瑠璃は文字の背景に模様や色があってもうまく見つけ出す[1]。学習には日本古典籍くずし字データセットを使った。機械学習のために作った古い文字写真の集まりである。文字100万字を超える資料である。このアプリは2022年グッドデザイン賞を受けた[3]。2026年9月時点で、このアプリが認識した写真は350万枚を超える[1]。

インストールと使用手順

まずアプリストアを開く。iPhoneとiPadはApp Storeを使う。Android機器はGoogle Playを使う。検索窓にみまたはmiwoを入れる[1]。App Store基準でアプリのサイズは約52メガバイトである。iPhoneとiPadはiOS 11.0以上でよい[2]。

アプリを起動すると写真を撮る画面が出る。古い文書を画面に収めて撮影ボタンを押す。すでに撮っておいた写真を読み込んでよい。その次に認識ボタンを押すと、人工知能が文字を読む[1]。

結果画面にはいくつかの機能がある。画面には調整バーが一つ置かれている。このバーを左右に動かすと二つの画面が重なって見える。元の写真と読み取った文字を並べて比べることができる。日本語の仮名文字を押す機能もある。その文字が元はどの漢字から出たのかを知らせてくれる。読み取った文章はコピーボタンで他のアプリに移すことができる[1]。結果を保存しておき、後で再び取り出して見ることもできる[1][3]。

インターネットが有効になっていなければならない点を覚えておく必要がある。写真は認識のためにCODHサーバーへ送られる。送った写真はサーバーに残さない[1]。

よく読める資料と読みにくい資料

アプリを使う前に限界を知っておくと失望が少ない。人工知能は江戸時代の版本を中心に学習した。その時期の本は木版印刷本が多い。木の板に文字を彫り、紙に刷り出した本である。そのため、その時期の版本はよく読める。他の時代の資料や手書き文書は精度が落ちる[3]。

紙の状態も結果を変える。染みが付いた場所や虫に食われた場所は読みづらい。紙の模様、照明、影も妨げになる[3]。明るく均一な光の下で撮るほうがよい。

読めない対象もある。石碑や看板に刻んだ崩し字は現在対応していない[3]。

開発陣は認識結果に誤りがある可能性があるとは明らかにしている。人工知能による崩し字認識は完全にはならないと明言している。そのため重要な資料は人が再確認しなければならない[3]。

古い文字を現代の文字に移して書く作業を翻刻という。この作業を多くの人で分担する場所が別にある。「みんなで翻刻」という市民参加型翻刻サイトである。日本語に移せば「みんなの翻刻」である。miwoを作った所とは別の機関が運営している。アプリが読んでくれた文章を人が整える作業と性格が同じである。アドレスはhonkoku.orgである。これまで1万800人を超える人が参加した。移して書いた文字は4,800万字を超えた。2025年8月に二度目の刷新を行った。新しい版は試験運用中である。国立歴史民俗博物館が参加している。東京大学地震研究所と京都大学の研究陣も一緒に運営している[4]。

参考資料

[1] ROIS-DS 人文学オープンデータ共同利用センター. (2026). AIくずし字認識アプリ「みを (miwo)」. CODH.
<https://codh.rois.ac.jp/miwo/> (2026年9月3日閲覧)

[2] Clanuwat, T. (2022, 10월 27일). miwo (Version 1.1) [Mobile app]. App Store.
<https://apps.apple.com/jp/app/miwo/id1581794085> (2026年9月3日閲覧)

[3] ROIS-DS 人文学オープンデータ共同利用センター. (2026). みを (miwo) とは?. CODH.
<https://codh.rois.ac.jp/miwo/about/> (2026年9月3日閲覧)

[4] みんなで翻刻. (2026). みんなで翻刻 Minna de Honkoku. honkoku.org. <https://honkoku.org/> (2026年9月3日閲覧)

このAI書庫の本を支援する

本書が役に立ったなら、今後の翻訳や公開版の制作、開かれた人工知能の知識づくりを
PayPal で支援できる。

PayPal

<https://paypal.me/kimkjkj>



QRコードを読み取るか、上のリンクを開いてください。

金京鎮弁護士