AI 침묵의 글자를 깨우다

인공지능이 되살린 고대의 기록

저자 : 김경진 변호사

목차

제1부 화염이 남긴 수수께끼: 비침습 고고학과 가상 전개 기술

1장 화산재가 삼킨 철학자의 서재: 헤르쿨라네움의 탄화 두루마리
베수비오 화산에 구워진 두루마리를 열지 않고 읽어 낸 과정

1. 사건 파일
2. 과학 수사대의 추적 1: 가상 양파 껍질 벗기기
3. 과학 수사대의 추적 2: 잉크의 미세 지문 찾기
4. 밝혀진 진실

2장 타버린 성궤 속에서 피어난 말씀: 엔게디의 봉인된 레위기
금속이 섞인 먹물의 흔적으로 되살린 엔게디의 성경 본문

1. 사건 파일
2. 과학 수사대의 추적: 엑스레이 디지털 정화
3. 밝혀진 진실

제2부 사해 동굴의 비밀 코드와 서기들의 필적 지문

3장 이사야 두루마리에 숨겨진 두 명의 그림자: 서기 감정
필적을 재는 기계가 찾아낸 이사야 두루마리의 두 서기

1. 사건 파일
2. 과학 수사대의 추적: 글씨체에 남겨진 생체 지문
3. 밝혀진 진실

4장 수만 조각의 점토·양피지 퍼즐을 맞추는 디지털 접착제
조각의 모서리와 필체를 맞춰 문서를 되살리는 디지털 접착제

1. 사건 파일
2. 과학 수사대의 추적 1: 기하학적 엣지 친화도와 물리적 지문
3. 과학 수사대의 추적 2: 위작 탐정 AI
4. 밝혀진 진실

제3부 진흙 점토판의 대화를 읽어내다: 전산 설형문자학

5장 점토판의 부러진 단면을 맞추다: 길가메시 서사시의 전산학적 재건
부러진 점토판 단면을 이어 붙여 되찾은 길가메시 서사시

1. 사건 파일
2. 과학 수사대의 추적: 클레이 엣지 디지털 융합과 Fragmentarium
3. 밝혀진 진실

6장 고대 메소포타미아 영수증과 상인들의 속삭임: AI 번역
고대 메소포타미아의 아카드어를 기계가 번역하기까지

1. 사건 파일

2. 과학 수사대의 추적: 아카드어 AI 번역
3. 밝혀진 진실

제4부 동아시아 고고학의 비밀: 진흙 속 목간과 뼈 위의 붉은 문자

7장 **거북 배딱지 속에 잠들었던 제국의 미래: 갑골문과 3D 가상 편철**
갑골의 글자를 읽고 깨진 뼈를 화면에서 다시 맞추다

1. 사건 파일
2. 과학 수사대의 추적: 갑골 전용 멀티모달 모델과 3D 가상 편철
3. 밝혀진 진실

8장 **진흙 구덩이에서 살아 돌아온 붓글씨: 신라 목간과 흘림체 판독 인공지능**
신라 목간의 흘림체 붓글씨를 읽어 내는 판독 인공지능

1. 사건 파일
2. 과학 수사대의 추적 1: 디지털 먹물 돋보기
3. 과학 수사대의 추적 2: KuroNet과 Miwo
4. 밝혀진 진실

제5부 침묵하는 문자들의 최후 반격: 미해독 문자의 전산 언어학

9장 **에게해의 잃어버린 수수께끼: 선형문자 A와 기하학적 격자**
선형문자 A를 소리의 격자에 얹어 풀어 보려는 시도

1. 사건 파일
2. 과학 수사대의 추적: 비지도 음운 격자 맞추기
3. 밝혀진 진실

10장 **태양의 주기와 동물 인장의 암호를 해킹하다: 인더스 문자와 롱고롱고의 소리**
인더스 문자와 롱고롱고에 정보 엔트로피를 들이대다

1. 사건 파일
2. 과학 수사대의 추적: 정보 엔트로피 측정기
3. 밝혀진 진실

제6부 디지털 알렉산드리아 도서관과 미래의 고고학

11장 **인공지능, 깨진 그리스 비문을 다시 읽다: 파운데이션 모델과 이타카**
깨진 그리스 비문의 빈칸을 채우는 이타카와 아폴로

1. 사건 파일
2. 과학 수사대의 추적: 고대 텍스트 자동완성 장치 Ithaca와 Apollo
3. 밝혀진 진실

12장 **영원히 타지 않는 가상 도서관: 초고해상도 다분광 아카이브와 미래**
다분광 촬영과 지식 그래프가 짓는 불타지 않는 도서관

1. 사건 파일
2. 과학 수사대의 추적: 지식 그래프와 빛의 스펙트럼 분석

3. 에필로그

부록 **나만의 AI 고고학 여행기: 대중을 위한 글로벌 탐사 포털**

누구나 집에서 따라 해 보는 고대 문자 탐사 안내

1. Vesuvius Challenge 공식 데이터 아카이브
2. eBL 전자 바빌로니아 문학 포털
3. Miwo 사용 안내

제 1 부

화염이 남긴 수수께끼: 비침습 고고학과 가상 전개 기술

01 CHAPTER

제1장. 화산재가 삼킨 철학자의 서재: 헤르쿨라네움의 탄화 두루마리

베수비오 화산에 구워진 두루마리를 열지 않고 읽어 낸 과정

1. 사건 파일

화산이 서재를 덮은 날

서기 79년에 이탈리아 남부의 베수비오 화산이 터졌다. 화산은 뜨거운 가스와 화산재를 한꺼번에 쏟아냈다. 산 아래 바닷가에는 헤르쿨라네움이라는 도시가 있었다. 로마의 부자들이 여름을 보내던 휴양 도시였다. 이 도시도 하루 만에 땅속에 묻혔다[1].

도시 바깥에는 큰 저택이 한 채 서 있었다. 오늘날 학자들은 이곳을 파피루스 빌라라고 부른다. 집주인은 루키우스 칼푸르니우스 피소 카이소니누스로 보인다. 그는 율리우스 카이사르의 장인이었다. 이 집에는 그리스 철학자 필로데모스가 드나들었다. 필로데모스는 에피쿠로스 학파의 학자였다. 저택 한쪽에는 두루마리를 모아 둔 서재가 있었다. 그의 글이 유난히 많이 나와 그와 이어진 서고로 본다[1].

파피루스는 고대 지중해 세계의 종이다. 이집트 나일강 가에 자라는 풀로 만든다. 줄기 속살을 얇게 저며 가로로 죽 늘어놓는다. 그 위에 다시 세로 방향으로 한 겹을 얹는다. 물을 뿌려 누르면 풀의 진액이 접착제 노릇을 한다. 말리면 한 장의 종이가 된다. 그래서 파피루스 표면에는 가로세로로 엇갈린 결이 남는다. 이 결은 나중에 중요한 단서가 된다[7].

그날 서재로 뜨거운 가스가 밀려들었다. 두루마리들은 불길 없이 그대로 구워졌다. 종이 역할을 하던 파피루스가 새까만 숯으로 변했다. 이렇게 타서 숯이 되는 것을 탄화라고 부른다. 두루마리는 쪼그라들고 표면이 부풀어 올랐다[2]. 그 위로 화산 퇴적물이 20미터 넘게 쌓였다. 화산 퇴적물은 화산재와 돌가루가 쌓여 굳은 층이다[1].

땅속 깊이 묻히자 공기가 닿지 않았다. 공기가 닿지 않으면 식물 섬유는 잘 썩지 않는다. 재앙이 유물을 지키는 뚜껑 노릇을 했다. 두루마리들은 그렇게 2,000년을 버텼다[2].

땅속 터널에서 나온 검은 막대

1738년에 나폴리 왕실이 이 일대를 파기 시작했다. 1750년에는 우물을 파던 인부가 색색의 모자이크 바닥을 만났다. 저택의 위치가 그렇게 드러났다. 발굴 책임은 카를 베버가 맡았다. 그는 스위스 출신의 군사 공학자였다[1].

베버는 굳은 화산 퇴적물에 아래로 곧게 굴을 팠다. 이런 굴을 수직 갱도라고 부른다. 그리고 좁은 터널을 미로처럼 파 들어갔다. 터널 안은 어둡고 공기가 나빴다. 벽이 무너질 위험도 늘 있었다. 그런 조건에서도 베버는 저택의 평면도를 꼼꼼히 그렸다. 그 지도 덕분에 어느 방에서 무엇이 나왔는지 알 수 있다[1].

1752년부터 인부들은 검은 막대 같은 물건을 꺼내기 시작했다. 처음에는 그것이 숯덩이인 줄 알았다. 알고 보니 파피루스 두루마리였다. 이 저택에서 나온 두루마리는 약 1,100개로 알려져 있다. 고대 세계의 서고가 통째로 남은 유일한 사례로 알려져 있다[1].

이 서고에는 필로데모스의 글이 유난히 많았다. 완성본과 사본, 초고가 함께 놓여 있었다. 학자들은 그가 이 집에서 글을 쓰고 정리했다고 본다[1]. 힘들게 펼친 두루마리에는 철학과 수사학, 윤리에 관한 글이 있었다. 수사학은 말과 글로 남을 설득하는 방법을 다루는 학문이다. 펼치지 못한 두루마리에 무엇이 담겼는지는 아무도 몰랐다[2].

억지로 펴다가 부순 두루마리

학자들은 그 안의 글을 읽고 싶어 했다. 문제는 두루마리가 만지면 바스러진다는 것이었다[2]. 1753년에 안토니오 피아조 신부가 기계를 하나 만들었다. 실과 추로 두루마리를 조금씩 당겨 펴는 장치였다[3]. 이 기계는 몇 권을 살렸지만 안쪽 겹을 찢기도 했다[2].

19세기에는 영국 화학자 험프리 데이비가 나섰다. 그는 염소 기체와 요오드 증기, 산을 두루마리에 썼다. 약품으로 겹을 떼어 내려는 방법이었다. 결과는 좋지 않았다. 약품은 파피루스를 무르게 만들어 부서뜨렸다[3].

물리적으로 펴려는 시도는 대부분 유물을 망가뜨렸다. 글자가 적힌 부스러기가 바닥에 흩어져 사라지기도 했다. 결국 학계는 개봉 시도를 멈추었다. 지금도 열리지 않은 두루마리가 수백 권 남아 있다[2]. 이 유물들은 나폴리와 파리, 옥스퍼드의 소장 기관에 나뉘어 보관된다[4].

엑스선을 막아선 탄소의 벽

과학자들은 열지 않고 읽는 길을 찾았다. 먼저 떠오른 도구는 컴퓨터 단층촬영이었다. 흔히 CT라고 부르는 기술이다. CT는 물건을 자르지 않고 속의 밀도 차이를 재는 장치다. 밀도가 높은 물질은 엑스선을 많이 막는다. 그래서 단면 사진에 밝게 나온다[5].

금속 성분이 섞인 잉크는 CT 사진에서 하얗게 빛난다. 중세 유럽에서는 철 성분이 든 잉크가 널리 쓰였다. 이스라엘 엔게디에서 나온 탄 두루마리도 비슷했다. 이 유물은 고대 이스라엘의 히브리어 사본이다. 잉크에 금속이 섞인 것으로 보여 사진에 자국이 남았다. 그래서 2016년에 열지 않고 판독되었다[5].

헤르쿨라네움의 필경사들은 사정이 달랐다. 필경사는 글을 손으로 베껴 쓰던 사람이다. 그들은 나무를 태운 그을음에 물과 수지를 섞어 먹을 만들었다. 먹의 주성분은 탄소다. 그런데 화산 열기에 구워진 파피루스도 탄소가 되었다. 종이와 글씨가 같은 물질이 된 것이다[6].

밀도가 같으면 엑스선은 두 곳을 똑같이 지나간다. 단면 사진에는 경계선이 나타나지 않는다. 학자들은 이 벽을 탄소 잉크 문제라고 불렀다. 2015년에 이탈리아와 프랑스 연구진이 다른 방법을 시도했다. 위상차 촬영이라는 기법이다. 엑스선이 물질의 경계를 지날 때 진행 방향이 살짝 꺾이는 성질을 쓴다. 밀도가 같아도 경계면에서는 신호가 달라진다. 연구진은 이 방법으로 말린 두루마리 안에서 글자 몇 개의 윤곽을 잡아냈다. 하지만 문장을 읽을 정도는 아니었다[6]. 오랫동안 사람들은 이 벽을 넘지 못한다고 믿었다[2].

참고 자료

[1] Herculaneum Society. The Villa of the Papyri. University of Oxford. https://www.herculaneum.ox.ac.uk/index.php/the-villa-of-the-papyri/ (2026년 9월 3일 접속)

[2] Vesuvius Challenge. FAQ. scrollprize.org. https://scrollprize.org/faq (2026년 9월 3일 접속)

[3] Villa of the Papyri. Wikipedia. https://en.wikipedia.org/wiki/Villa_of_the_Papyri (2026년 9월 3일 접속)

[4] 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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

2. 과학 수사대의 추적 1: 가상 양파 껍질 벗기기

열지 않고 읽는다는 생각

미국 켄터키 대학교의 브렌트 실즈 교수는 다른 길을 택했다. 유물을 여는 대신 컴퓨터 안에서 펴자는 생각이었다. 실물은 그대로 두고 3차원 사진만 얻는다. 그다음 화면 위에서 겹을 한 장씩 떼어 낸다. 양파 껍질을 벗기듯 안쪽으로 들어가는 방식이다[1].

연구팀은 이 방법을 가상 전개라고 이름 붙였다. 영어로는 버추얼 언래핑이라고 부른다. 2016년에 이 방법이 처음 성공했다. 대상은 앞 절에서 본 엔게디 두루마리였다. 이 유물은 1970년에 옛 회당 터에서 나왔다. 겉모습은 새까맣게 탄 덩어리였다. 연구팀이 화면에서 펼쳐 보니 히브리어 글씨가 드러났다. 구약 레위기의 첫머리였다[1].

가상 전개는 다섯 단계로 이루어진다. 3차원 사진 얻기, 겹 나누기, 잉크 값 입히기, 평평하게 펴기, 이어 붙이기다. 엔게디 사본에서는 이 차례가 그대로 통했다. 잉크가 사진에 이미 보였기 때문이다[1].

헤르쿨라네움 두루마리는 엔게디 사본보다 조건이 나빴다. 화산 열기에 눌려 단면이 찌그러졌다. 둥근 모양이 무너져 겹과 겹이 서로 들러붙었다. 게다가 잉크가 사진에 보이지 않았다. 다섯 단계 가운데 둘째와 셋째가 함께 발목을 잡았다[2].

그래서 헤르쿨라네움에서는 차례를 바꾸었다. 먼저 겹을 나누고 평평하게 편다. 잉크 찾기는 그다음에 따로 한다. 1.3절이 그 잉크 찾기 이야기다[2].

입자가속기가 찍은 3차원 사진

첫 단계는 두루마리의 속을 통째로 찍는 일이다[1]. 병원 CT는 보통 1밀리미터 안팎까지 구별한다. 파피루스 겹은 그보다 훨씬 얇다. 겹 하나의 두께는 머리카락 굵기에도 못 미친다. 병원 장비로는 겹과 겹이 뭉개져 보인다[3].

연구팀은 입자가속기를 찾아갔다. 영국의 다이아몬드 라이트 소스를 처음 찾아갔다[3]. 뒤에는 프랑스 그르노블의 유럽 싱크로트론 방사광 시설도 함께 썼다[4]. 이 시설들은 아주 밝고 곧게 나아가는 엑스선을 만든다. 밝은 빛일수록 짧은 시간에 또렷한 사진을 얻는다[3].

연구팀은 두루마리를 통에 세워 단단히 고정했다. 촬영 중에 조금만 흔들려도 사진이 흐려지기 때문이다. 이렇게 얻은 자료의 해상도는 복셀 하나에 약 8마이크로미터였다. 복셀은 깊이 값까지 가진 입체 화소를 말한다. 그림판의 점 하나를 작은 정육면체로 바꾼 것이라고 보면 된다[3].

자료의 양은 어마어마했다. 두루마리 한 권을 찍으면 약 300테라바이트가 쌓인다. 스마트폰 수천 대의 저장 공간을 채우는 크기다. 이 자료를 다루는 일 자체가 새로운 과제가 되었다[4].

프랑스 그르노블의 시설은 4세대 광원으로 불린다. 2020년에 새 장비가 가동에 들어갔다. 유럽 21개 나라가 1억 5천만 유로를 들여 성능을 끌어올렸다. 이 시설은 원래 물질과 생명 현상을 살피려고 지은 곳이다. 고대 두루마리를 읽는 데 쓰일 줄은 처음에 아무도 몰랐다[4].

겹을 갈라내는 체적 분할

3차원 사진을 얻으면 다음은 겹을 나누는 단계다. 이 작업을 체적 분할이라고 부른다. 입체 사진에서 겹을 하나씩 갈라내는 일이다. 컴퓨터는 단면 사진마다 종이 표면이 지나간 길을 따라간다. 그 선들을 이어 붙이면 종이 한 장의 입체 모형이 생긴다. 이 모형을 메시라고 부른다[1].

처음에는 사람이 마우스로 경계를 하나씩 찍었다[2]. 두루마리 한 권에 담긴 종이 길이는 1미터가 넘는다[4]. 손으로 따라가기에는 너무 길고 촘촘했다. 게다가 겹이 서로 눌러붙어 있어 경계가 흐릿했다[2].

대회 참가자들이 자동화 도구를 내놓으면서 속도가 빨라졌다. 스위스의 율리안 쉴리거가 만든 프로그램이 대표적이다. 이 프로그램은 이웃한 단면 사진의 무늬를 좇아 겹을 이어 간다. 사람이 며칠 걸리던 일을 몇 시간으로 줄였다[2].

이 일은 한 연구실이 혼자 할 수 있는 규모가 아니었다. 주최 측은 촬영 자료를 인터넷에 그대로 열어 두었다. 참가자들은 자기가 만든 프로그램의 소스 코드를 공개했다. 남이 만든 도구 위에 다음 사람이 기능을 얹었다. 판독문과 코드는 지금도 누구나 내려받을 수 있다[5].

평평하게 펴서 이어 붙이기

겹을 나누면 종이 모형은 여전히 둥글게 말려 있다. 이대로는 글자가 휘어 보인다. 그래서 모형을 평평하게 펴는 계산을 한다. 이 단계를 평탄화라고 부른다[1].

컴퓨터는 모형의 꼭짓점을 작은 구슬로 바꾼다. 구슬과 구슬 사이는 용수철로 이어져 있다고 가정한다. 그다음 이 그물을 바닥에 눕히는 계산을 반복한다. 용수철이 늘어나거나 줄어드는 정도를 최소로 맞춘다. 그러면 글자의 모양이 크게 일그러지지 않는다[1].

마지막은 이어 붙이기다. 두루마리는 한 번에 다 펴지지 않는다. 컴퓨터는 여러 구역으로 나누어 따로 편다. 펴 놓은 구역들을 다시 한 장으로 맞춰야 한다. 이때 파피루스 특유의 격자 섬유 무늬를 쓴다. 두 구역의 섬유 무늬가 맞물리면 원래 이웃이었다는 뜻이다. 이 대조 작업까지 끝나면 가상 전개가 완성된다[1].

다섯 단계는 한 번에 끝나지 않는다. 어느 단계에서 실수가 나면 뒤 단계가 모두 어긋난다. 겹을 잘못 나누면 다른 장의 글자가 섞여 들어온다. 그래서 연구진은 같은 두루마리를 여러 번 되풀이해 처리한다. 결과를 눈으로 확인하고 설정을 바꾸어 다시 돌린다[5].

참고 자료

[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, 2월 5일). Vesuvius Challenge 2023 Grand Prize awarded: we can read the scrolls! scrollprize.org. https://scrollprize.org/grandprize (2026년 9월 3일 접속)

[3] Vesuvius Challenge. FAQ. scrollprize.org. https://scrollprize.org/faq (2026년 9월 3일 접속)

[4] 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 (2026년 9월 3일 접속)

[5] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

3. 과학 수사대의 추적 2: 잉크의 미세 지문 찾기

탄소 위에 쓴 탄소

가상 전개가 끝나도 화면은 아직 텅 비어 있었다. 종이는 펴졌지만 글자가 보이지 않았다[1]. 원인은 앞서 본 탄소 문제였다. 파피루스도 탄소이고 먹도 탄소였다. 엑스선은 둘을 구별하지 못한다[2].

실즈 교수 연구팀은 가설을 하나 세웠다. 먹이 아예 흔적을 남기지 않았을 리는 없다는 것이었다[1]. 먹은 액체 상태로 종이에 닿는다. 액체는 섬유 사이로 스며든다. 그리고 마르면서 표면을 조금 잡아당긴다[3].

그러면 종이 표면의 모양이 미세하게 달라진다. 밀도는 같아도 굴곡은 달라진다. 연구팀은 이 굴곡을 찾기로 했다[3]. 눈으로 볼 수 없는 크기라서 기계에 맡겨야 했다[1].

여기서 쓰인 도구가 기계학습이다. 사람이 규칙을 일일이 적어 주는 방식이 아니다. 정답을 붙인 예시를 잔뜩 보여 주고 스스로 규칙을 찾게 한다. 고양이 사진 수천 장을 보여 주면 고양이를 알아보는 것과 같은 이치다. 잉크 찾기에서는 먹이 있는 곳과 없는 곳을 표시한 표본을 보여 준다[1].

크랙클이라는 단서

2023년 3월에 베수비오 챌린지라는 대회가 열렸다[1]. 실즈 교수와 나트 프리드먼, 대니얼 그로스가 함께 시작했다[4]. 주최 측은 촬영 자료를 인터넷에 무료로 풀었다. 상금을 걸고 전 세계 프로그래머를 불러 모았다[1].

처음 몇 달 동안은 아무도 잉크를 찾지 못했다. 그해 여름에 케이시 핸드머라는 참가자가 실마리를 잡았다. 그는 펴진 표면 자료를 그냥 눈으로 오래 들여다보았다. 그러다 특이한 무늬를 발견했다. 잔금이 자잘하게 갈라진 자국이었다[1].

가뭄이 든 논바닥을 떠올리면 된다. 마른 진흙은 그물처럼 갈라진다. 먹이 묻은 곳에서도 비슷한 잔금이 생겼다. 참가자들은 이 무늬를 크랙클이라고 불렀다. 핸드머는 이 발견으로 첫 잉크 상을 받았다[1].

핸드머는 발견 즉시 온라인 대화방에 알렸다. 이 공유가 대회의 흐름을 바꾸었다. 다른 참가자들이 곧바로 이 무늬를 학습 재료로 삼았다[1].

작은 창으로만 보게 하기

미국 대학생 루크 패리터가 먼저 움직였다. 그는 크랙클 무늬를 기계학습 모형에 가르쳤다. 모형은 어디에 먹이 있는지 확률로 답한다. 학습이 진행될수록 답이 정확해졌다[1].

문제는 학습 자료가 적다는 것이었다. 자료가 적으면 기계는 답을 외워 버린다. 이것을 과적합이라고 부른다. 시험 문제를 이해하지 않고 답만 외운 학생과 같다[1].

참가자들은 자료를 돌리고 뒤집어 양을 늘렸다. 그림을 돌려도 잔금 무늬의 성질은 그대로다. 기계는 그 덕분에 무늬 자체의 특징만 익혔다[1].

입력 창의 크기도 손을 봤다. 처음에는 한 번에 넓은 영역을 보여 주었다. 넓게 보면 글자 모양이 통째로 눈에 들어온다. 그러자 기계가 글자를 지어내기 시작했다. 먹이 없는 곳에 그럴듯한 획을 그려 넣었다. 이것을 인공지능의 환각이라고 부른다[1].

연구자들은 창을 64 곱하기 64 화소로 줄였다. 그 정도 넓이에서는 글자 모양이 드러나지 않는다. 기계는 잔금의 질감만 보고 판단해야 한다. 지어내기가 크게 줄었다. 대신 실제 붓 자국이 화면에 그대로 나타났다[1].

이 계산을 맡은 모형은 두 종류를 이어 붙여 만들었다. 하나는 3차원 합성곱 신경망이다. 작은 정육면체 안의 무늬를 훑어 특징을 뽑아내는 구조다. 다른 하나는 트랜스포머다. 떨어져 있는 신호끼리 서로 얼마나 관련이 있는지 따지는 구조다. 두 구조를 함께 쓰면 얇은 겹 속의 흐릿한 무늬를 놓치지 않는다[1].

기계가 내놓은 화면은 곧바로 정답으로 받아들여지지 않는다. 고문서 전문가들이 그 화면을 따로 읽는다. 그리스어 낱말의 배열이 문법에 맞는지 살핀다. 이어지는 문장이 뜻이 통하는지도 확인한다. 기계와 사람의 판독이 어긋나면 그 대상은 다시 계산한다[1].

표면의 굴곡만으로

이후에도 잉크를 찾는 방법은 계속 나아졌다. 참가자 유세프 나데르는 미리 익히기 방법을 썼다. 정답표가 없는 두루마리 자료를 먼저 대량으로 보여 주는 방식이다. 기계는 파피루스 섬유의 생김새를 스스로 익힌다. 그다음 적은 정답표로 잉크 찾기만 가르친다[1].

2026년 3월에는 굴곡 자체를 다룬 연구가 나왔다. 조르조 안젤로티와 페데리카 니콜라르디, 폴 헨더슨, 브렌트 실즈가 함께 썼다. 연구진은 이미 펼쳐진 파피루스 표본의 표면을 광학 장비로 쟀다. 높낮이만 담긴 자료를 기계에 학습시켰다. 밝기나 밀도 정보는 넣지 않았다[3].

그런데도 기계는 먹이 있는 곳을 가려냈다. 표면의 높낮이만으로 잉크를 찾을 수 있다는 뜻이다. 연구진은 어느 정도로 촘촘히 재야 하는지도 함께 계산했다. 이 수치는 다음 촬영의 해상도 기준이 된다. 이 논문은 같은 해 학술지 사이언티픽 리포츠에 실렸다[3].

그래도 잉크 찾기는 아직 완성된 기술이 아니다. 먹이 옅게 남은 구간은 여전히 흐리게 나온다. 겹이 눌러붙은 곳에서는 촬영 자료 자체가 뿌옇다. 서로 다른 종이 두 장이 한 장으로 붙어 버리기도 한다. 한 두루마리에서 잘 되던 모형이 다른 두루마리에서 헛도는 일도 있다. 그래서 사람이 중간마다 눈으로 확인하고 손을 본다. 연구는 지금도 한 걸음씩 나아가고 있다[5].

참고 자료

[1] Vesuvius Challenge. (2024, 2월 5일). Vesuvius Challenge 2023 Grand Prize awarded: we can read the scrolls! scrollprize.org. https://scrollprize.org/grandprize (2026년 9월 3일 접속)

[2] 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

[3] Angelotti, G., Nicolardi, F., Henderson, P., & Seales, W. B. (2026). Ink detection from surface topography of the Herculaneum papyri. Scientific Reports. https://doi.org/10.1038/s41598-026-58467-1

[4] 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 (2026년 9월 3일 접속)

[5] Vesuvius Challenge. (2026, 7월 10일). Open problems: why reading every Herculaneum scroll is still a challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026년 9월 3일 접속)

4. 밝혀진 진실

2023년에 되찾은 첫 낱말

첫 성과는 낱말 하나였다. 2023년 10월 12일에 발표가 나왔다. 주인공은 미국의 대학생 루크 패리터였다. 그때 나이는 스물한 살이었다. 그가 읽어 낸 낱말은 포르퓌라스였다. 보라색을 뜻하는 그리스어다. 고대 글에서 흔히 쓰이는 말은 아니다. 패리터는 첫 글자 상으로 4만 달러를 받았다. 2위인 유세프 나데르는 1만 달러를 받았다. 열지 않은 두루마리에서 낱말을 읽은 첫 사례였다[7].

2024년 2월 5일에 대회 결과가 발표되었다. 대상은 세 사람이 함께 받았다. 이집트의 유세프 나데르, 미국의 루크 패리터, 스위스의 율리안 쉴리거였다. 상금은 70만 달러였다[1].

이들이 읽어 낸 유물은 파리 학술원이 가진 두루마리다. 이름은 PHerc. Paris. 4다. 세 사람은 이 두루마리에서 2,000자 넘는 그리스어를 되살렸다. 열다섯 개 단의 글이었다. 단은 두루마리에 세로로 나눠 쓴 글의 칸이다. 두루마리 전체의 약 5퍼센트에 해당한다[1].

심사 기준은 까다로웠다. 140자짜리 구절 네 곳을 제출해야 했다. 각 구절에서 글자의 85퍼센트 이상을 읽어야 통과였다. 고문서 전문가들이 이름을 가린 채 검토했다[1].

내용은 즐거움에 관한 논의였다. 드문 것이 흔한 것보다 반드시 더 즐거운지 따진다. 본문에는 음식을 예로 든 구절이 있다. 드물다고 해서 그것이 곧 더 즐겁다고 믿지는 않는다는 취지다. 에피쿠로스 학파의 사고가 그대로 드러난 문장이다[1].

제목을 되찾은 두루마리

다음 성과는 제목이었다. 2025년 5월 5일에 첫 제목 상이 나왔다. 수상자는 마르셀 로트와 미하 노바크였다. 상금은 6만 달러였다[2].

대상 유물은 옥스퍼드 보들리언 도서관의 PHerc. 172였다. 화면에 드러난 글자는 필로데모스의 '악덕에 대하여'였다. 제1권으로 읽혔다. 나폴리 페데리코 2세 대학교의 페데리카 니콜라르디 연구진이 확인했다. 말려 있는 두루마리의 제목을 열지 않고 읽어 낸 첫 사례다[2].

이 두루마리에서는 뒤이어 70개가 넘는 단이 읽혔다. 탐욕과 오만, 아첨 같은 성격의 흠을 다룬 글이다. 이제 논의의 흐름을 처음부터 따라갈 수 있다[3].

처음으로 끝까지 읽은 한 권

2026년 6월 25일에 더 큰 소식이 나왔다. 두루마리 한 권을 처음부터 끝까지 읽었다는 발표였다. 유물은 나폴리 국립도서관의 PHerc. 1667이다. 대회 참가자들은 이를 4번 두루마리라고 부른다[4].

촬영은 프랑스의 유럽 싱크로트론 방사광 시설에서 이루어졌다. 그 뒤 가상 전개와 잉크 찾기를 차례로 거쳤다. 되살아난 글의 길이는 약 1.4미터였다. 단으로는 스물두 개쯤이다. 헤르쿨라네움 두루마리를 통째로 읽은 것은 이때가 처음이다[4].

내용은 윤리와 사람의 성품을 다룬 그리스어 논고다[3]. 본문에는 스토아 철학자 크리시포스의 조카가 언급된다. 그 이름은 아리스토크레온이다[4]. 글이 쓰인 시기는 기원전 3세기 말에서 2세기 무렵으로 본다. 저자가 누구인지는 아직 확정되지 않았다[3].

연구 결과는 논문으로도 공개되었다. 조르조 안젤로티와 스티븐 파슨스, 페데리카 니콜라르디를 포함해 27명이 이름을 올렸다[5]. 자료와 코드, 판독문도 함께 공개되었다[4].

연구진은 검증 절차도 함께 밝혔다. 2023년에 읽은 PHerc. Paris. 4를 더 높은 해상도로 다시 찍었다. 새 자료로 얻은 판독문은 예전 결과와 그대로 맞아떨어졌다. 기계가 지어낸 글이 아니라는 증거다. 같은 방법을 다른 두루마리에 적용할 수 있다는 뜻이기도 하다[4].

아직 남은 600권

같은 발표에는 다른 소식도 담겼다. PHerc. 139라는 두루마리의 제목이 읽혔다. 필로데모스의 '신들에 대하여' 제8권이었다. 그전까지 이 책은 제1권만 알려져 있었다. 이 저작이 여덟 권 넘게 이어졌다는 사실이 처음 드러났다[3].

읽지 못한 두루마리는 아직 600권이 넘는다[3]. 대회는 2027년 그랑프리를 걸어 두었다. 상금은 모두 100만 달러다. 1위에게 80만 달러가 돌아간다. 대상은 지정된 열세 권의 두루마리다. 각 단에서 남아 있는 글자의 70퍼센트 이상을 읽어야 한다. 마감은 2027년 6월 25일이다[6].

이 이야기에는 아직 빈칸이 많다. 글자를 읽어도 그 뜻을 새기는 일은 따로 남는다. 판독문 한 단을 학술적으로 다듬는 데도 오랜 시간이 든다. 저자가 밝혀지지 않은 두루마리도 남아 있다[3].

그래도 방향은 분명해졌다. 2,000년 동안 아무도 열지 못한 책이 화면 위에서 펼쳐졌다. 불에 타 사라진 줄 알았던 서재가 조금씩 돌아오고 있다[4].

이 일을 한 사람은 한 명이 아니다. 촬영을 맡은 물리학자와 겹을 나눈 프로그래머가 있었다. 고대 그리스어를 읽어 준 고문서 학자들도 있었다. 대회에 참가한 학생과 회사원도 손을 보탰다. 2,000년 만에 책이 열린 것은 이 사람들이 함께 만든 결과다. 남은 두루마리도 같은 방식으로 열릴 수 있다[4].

참고 자료

[1] Vesuvius Challenge. (2024, 2월 5일). Vesuvius Challenge 2023 Grand Prize awarded: we can read the scrolls! scrollprize.org. https://scrollprize.org/grandprize (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[4] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[6] Vesuvius Challenge. Open Prizes. scrollprize.org. https://scrollprize.org/prizes (2026년 9월 3일 접속)

[7] Vesuvius Challenge. (2023, 10월 12일). First word discovered in unopened Herculaneum scroll by 21yo computer science student. scrollprize.org. https://scrollprize.org/firstletters (2026년 9월 3일 접속)

02 CHAPTER

제2장. 타버린 성궤 속에서 피어난 말씀: 엔게디의 봉인된 레위기

금속이 섞인 먹물의 흔적으로 되살린 엔게디의 성경 본문

1. 사건 파일

사해 서쪽의 오아시스

에인게디는 사해의 서쪽 기슭에 있는 오아시스다. 엔게디라고도 적는다. 둘레는 메마른 사막이지만 이곳에는 샘물이 솟는다. 물이 있으니 사람이 모였다. 사람들은 수천 년 동안 이 물가에 마을을 이루고 살았다. 구약 성경 사무엘상에도 이 땅의 이름이 나온다. 다윗이 사울 왕을 피해 에인게디 광야에 숨었다는 이야기다.

유적 조사에 따르면 이 언덕에는 기원전 8세기 후반부터 사람이 살았다. 그리고 서기 600년 무렵의 큰불로 마을이 무너졌다[1]. 천삼백 년이 넘는 세월이 흙 속에 층층이 쌓여 있는 땅이다.

사해 둘레는 비가 적고 공기가 건조하다. 이런 기후는 옛 물건을 오래 남기는 데 도움이 된다. 나무와 가죽과 천이 다른 곳보다 덜 썩는다. 사해문서가 이 지역에서 나온 까닭도 여기에 있다.

1970년에 이곳에서 큰 발굴이 시작되었다. 발굴은 단 바라그와 에후드 네처가 이끌었다. 두 사람은 예루살렘 히브리 대학교 고고학 연구소에 있었다. 요세프 포라트는 1970년 5월 5일 이 두루마리를 파냈다[1]. 발굴단이 찾아낸 것은 옛 유대교 회당의 터였다. 회당은 유대인 공동체가 모여 기도하고 성경을 읽던 건물이다. 바닥에는 색이 다른 작은 돌을 맞춰 넣은 모자이크가 깔려 있었다.

회당 안에는 성궤를 두는 공간이 있다. 성궤는 거룩한 성경 두루마리를 넣어 두는 상자다. 유대교 공동체는 이 상자를 회당에서 거룩한 곳으로 여긴다. 발굴단은 그 성궤에서 불에 탄 잔해를 걷어냈다. 잔해 속에서 검은 덩어리가 나왔다[1].

성궤에서 나왔다는 사실 자체가 중요한 실마리다. 회당의 성궤에는 아무 문서나 넣지 않는다. 공동체가 읽고 지키던 성경 두루마리를 넣는다. 검은 덩어리의 정체를 짐작할 근거가 여기에 있었다.

옛 성경 사본은 그 자체로 귀한 자료다. 인쇄기가 없던 시절에는 사람이 손으로 옮겨 적었다. 옮겨 적을 때마다 글자가 바뀔 여지가 생긴다. 오래된 사본이 나오면 본문이 어떻게 전해졌는지 확인할 수 있다. 성궤에서 나온 검은 덩어리에 학자들이 매달린 이유다.

불이 남긴 검은 덩어리

불은 기록을 없애는 힘이다. 종이도 가죽도 불 속에서는 재가 된다. 그런데 에인게디에서는 사정이 조금 달랐다. 잔해에 눌린 두루마리는 산소가 모자란 상태에서 천천히 구워졌다. 재로 흩어지지 못하고 숯처럼 굳었다. 이렇게 숯처럼 굳는 변화를 탄화라고 부른다. 공기가 모자란 불 속에서 타면 이런 일이 생긴다.

숯가마에서 나무가 숯이 되는 과정과 원리가 같다. 나무를 그냥 태우면 재만 남는다. 흙으로 덮어 공기를 막고 구우면 검은 숯이 남는다. 탄화한 두루마리도 그렇게 만들어졌다.

탄화한 덩어리는 무너진 건물의 재와 흙 밑에 묻혔다. 공기가 닿지 않으니 썩는 일도 멈췄다. 가죽은 원래 잘 썩는 재료다. 그대로 땅 위에 놓였다면 오래전에 사라졌을 것이다. 유물을 부순 불이 유물을 남긴 힘이기도 했다.

발굴단이 꺼낸 덩어리는 얼핏 숯덩이로 보였다. 그러나 끊어진 옆면에는 둥글게 말린 겹이 남아 있었다. 고고학자들은 이것을 말아 놓은 가죽 두루마리로 판단했다[1]. 유물에는 발견된 땅의 이름을 붙였다. 에인게디 두루마리라는 이름이 그렇게 생겼다.

이 두루마리는 동물 가죽을 다듬어 만든 양피지다[1]. 이탈리아 헤르쿨라네움에서 나온 두루마리는 파피루스로 만들었다. 파피루스는 풀줄기를 눌러 붙인 식물성 종이다. 두 유물은 바탕 재료부터 다르다. 이 차이는 뒷날 읽는 방법을 가르는 갈림길이 된다.

양피지는 만드는 데 손이 많이 가는 재료다. 동물 가죽에서 털과 기름을 걷어내고 씻는다. 틀에 걸어 팽팽하게 당긴 채로 말린다. 그렇게 다듬은 얇은 판에 글을 적는다. 종이보다 질기고 오래 가지만 값이 비쌌다.

발견 상태도 흔치 않았다. 오래된 두루마리는 대개 잘게 부서진 조각으로 나온다. 1947년부터 쿰란 동굴에서 나온 사해문서도 조각이 많다. 에인게디의 유물은 말린 모양을 그대로 지킨 채 나왔다. 안쪽에 몇 겹이 더 남아 있다는 뜻이었다.

손댈 수 없는 유물

글을 읽으려면 두루마리를 펴야 한다. 그런데 이 유물은 펼 수가 없었다. 열기에 가죽이 오그라들면서 겹끼리 달라붙었기 때문이다. 겉은 돌처럼 딱딱하고 속은 부스러지기 쉬웠다. 손끝으로 건드리면 가루가 될 상태였다[1].

방법이 없지는 않았다. 칼로 겹을 갈라 보는 길이 있다. 약품을 발라 가죽을 무르게 하는 길도 있다. 두 방법 모두 유물을 망가뜨린다. 한 번 부서진 두루마리는 되돌릴 수 없다.

겉을 보는 것만으로는 글자가 있는지도 알 수 없었다. 탄화한 표면은 온통 검고 무늬가 없다. 먹물도 검고 가죽도 검으니 눈으로는 갈리지 않는다. 안쪽에 글이 있다는 것도 겹의 모양에서 짐작한 추측이었다.

문화재를 다루는 사람들은 이 점을 늘 앞에 둔다. 유물은 세상에 하나뿐이고 실패를 한 번도 감당할 수 없다. 그래서 보존 과학에는 손대지 않고 조사한다는 원칙이 있다. 이런 조사 방법을 비침습 조사라고 부른다. 침습은 안으로 파고든다는 뜻이고 비침습은 그 반대다. 에인게디 두루마리는 비침습 조사만 허락되는 유물이었다.

45년의 침묵

두루마리는 이스라엘의 유물 수장고로 들어갔다. 뒷날 이스라엘 유물청이 이 유물을 맡아 지켰다[1]. 상자 안에서 기다리는 일이 그때의 유일한 보존법이었다. 수장고는 온도와 습도를 일정하게 지키는 공간이다. 공기와 빛과 벌레를 막아 유물이 더 상하지 않게 한다. 읽지 못하는 유물이라도 상태를 지키는 일은 계속된다.

사해문서 보존을 맡은 프니나 쇼어가 관리를 이어받았다. 그는 뒷날 이 두루마리를 읽어 낸 연구에도 이름을 올린다[1].

기다림은 1970년부터 2015년까지 이어졌다[1][2]. 45년 동안 아무도 그 안의 글자를 보지 못했다. 겉면이 까맣기만 해서 글자가 있는지조차 알기 어려웠다. 학자들 사이에서는 영영 읽지 못할 유물이라는 말도 나왔다.

기다림에는 이유가 있었다. 1970년의 기술로는 이 덩어리를 다룰 수 없었다. 물체 속을 잘게 나누어 찍는 촬영 장치도 흔하지 않았다. 찍는다 해도 그 많은 자료를 계산할 컴퓨터가 없었다. 유물은 그대로 두고 기술이 자라기를 기다렸다.

그동안 세계 곳곳에서 비슷한 고민이 이어졌다. 이탈리아의 헤르쿨라네움 두루마리도 같은 벽에 막혀 있었다. 물리적으로 펴려던 옛 시도는 유물을 부수기만 했다. 두 유물은 사는 곳도 재료도 달랐지만 문제는 하나였다. 손대지 않고 안쪽을 읽는 방법이 필요했다.

침묵을 깬 것은 새로운 손기술이 아니었다. 물건을 자르지 않고 속을 들여다보는 촬영과 컴퓨터 계산이었다. 실즈 연구진은 손대지 않고 옛 글을 읽는 방법을 오래 연구해 왔다. 이 연구는 미국 국립과학재단과 구글의 지원을 받았다[2]. 오래 쌓은 방법을 시험할 유물이 필요했고 에인게디 두루마리가 그 대상이 되었다.

2015년에 미국 켄터키 대학교의 브렌트 실즈 연구진이 이 일을 맡았다. 연구 결과는 2016년 9월 21일 학술지 사이언스 어드밴시스에 실렸다[1][2].

이제 남은 물음은 하나였다. 손대지 않고 어떻게 안쪽 글자까지 읽어 냈을까. 답은 촬영 장치와 계산의 순서에 있다.

참고 자료

[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, 9월 22일). Virtually Unwrapping the En-Gedi Scroll. research.uky.edu. https://research.uky.edu/news/virtually-unwrapping-en-gedi-scroll (2026년 9월 3일 접속)

2. 과학 수사대의 추적: 엑스레이 디지털 정화

자르지 않고 속을 보는 촬영

연구진이 고른 첫 도구는 마이크로 CT다. CT는 컴퓨터 단층 촬영을 뜻하는 말이다. 병원에서 몸속을 들여다볼 때 쓰는 그 촬영과 원리가 같다. 엑스선을 여러 각도에서 쏘아 물체 속의 밀도를 잰다. 컴퓨터는 그 값을 모아 입체 지도를 만든다. 마이크로 CT는 이것을 아주 작은 눈금으로 하는 장치다.

에인게디 두루마리의 촬영에는 브루커 스카이스캔 1176이라는 장치가 쓰였다. 촬영을 맡은 곳은 이스라엘의 메르켈 테크놀로지스다. 이 회사는 촬영 자료를 이스라엘 유물청에 기증했다. 촬영 눈금은 한 칸이 18마이크로미터였다[1]. 1마이크로미터는 1밀리미터의 1000분의 1이다. 머리카락 굵기의 몇 분의 일에 해당하는 크기다.

이 눈금 크기에는 까닭이 있다. 층을 갈라 보려면 가죽 한 겹보다 훨씬 잘게 재야 한다. 눈금이 겹의 두께보다 크면 두 층이 한 덩어리로 뭉개진다. 촬영 눈금은 읽을 수 있는 층의 한계를 정하는 조건이다.

촬영이 만든 자료는 사진 한 장이 아니다. 아주 작은 정육면체 알갱이가 빽빽하게 쌓인 입체 덩어리다. 이 알갱이를 복셀이라고 부른다. 사진의 화소가 평면 알갱이라면 복셀은 입체 알갱이다. 알갱이 하나에는 그곳의 밀도를 나타내는 숫자가 들어 있다.

이 단계에서 유물은 손대지 않고도 컴퓨터 안으로 옮겨졌다. 이제부터 다룰 대상은 가죽이 아니라 숫자다. 숫자는 아무리 만져도 부서지지 않는다. 계산이 틀리면 처음부터 다시 하면 된다. 비침습 조사의 힘이 여기에 있다.

금속이 섞인 먹물이 남긴 밝은 점

입체 자료를 얻었다고 글자가 보이지는 않는다. 글자가 보이려면 먹물과 바탕의 밀도가 달라야 한다. 이 조건이 헤르쿨라네움 두루마리에서는 무너졌다. 그곳의 먹물은 숯가루로 만들었고 탄 파피루스도 탄소 덩어리다. 둘의 밀도가 비슷해서 엑스선으로는 잘 갈리지 않는다.

에인게디 두루마리는 조건이 좋았다. 연구진은 스캔 자료에서 글자가 놓인 곳의 밀도가 둘레보다 높다는 사실을 확인했다. 그래서 이 먹물에 철이나 납 같은 금속이 섞여 있을 것으로 보았다. 다만 겉면이 드러난 곳이 없어서 성분을 직접 분석하지는 못했다. 연구진은 이 한계를 논문에 그대로 적었다[1].

금속은 엑스선을 많이 흡수한다. 흡수가 많은 곳은 화면에서 밝게 나온다. 탄 가죽은 엑스선을 덜 흡수해 어둡게 나온다. 글자와 바탕의 밝기 차이가 이렇게 생긴다. 이 차이가 컴퓨터가 글자를 찾아낼 수 있는 첫 실마리가 되었다. 에인게디 두루마리가 헤르쿨라네움보다 먼저 읽힌 까닭도 여기에 있다.

다만 이 먹물의 성분은 아직 확정되지 않았다. 연구진이 내놓은 것은 밀도가 높다는 관찰과 그에 따른 추정뿐이다[1]. 옛 서기가 무엇을 섞었는지는 시료를 얻어야 알 수 있다. 확정되지 않은 것을 확정된 사실처럼 다루지 않는 것이 이 분야의 원칙이다.

가상 전개의 네 단계와 얼룩 걷어내기

밝은 점을 찾았다고 글이 읽히지는 않는다. 글자는 둥글게 말린 겹 위에 흩어져 있다. 이 겹을 컴퓨터 안에서 펴야 한다. 이 작업을 가상 전개라고 부른다. 실즈 연구진은 가상 전개를 네 단계로 나누었다[1].

첫 단계는 겹의 표면을 찾아내는 일이다. 컴퓨터는 입체 자료를 얇게 잘라 가며 가죽 층의 좌표를 짚는다. 이렇게 모은 점들을 이어 얇은 판 모양의 면을 만든다. 겹이 서로 붙어 있으면 이 단계에서 길을 잃는다. 두 층을 한 층으로 붙여 읽으면 글자도 뒤섞인다.

두 번째 단계는 그 면에 밝기 값을 입히는 일이다. 면이 지나가는 좌표의 촬영 값을 읽어 면 위에 옮겨 그린다. 밀도가 높던 먹물이 이때 밝은 획으로 나타난다.

세 번째 단계는 구겨진 면을 평평하게 펴는 일이다. 입체를 평면으로 누르면 모양이 늘어나거나 찌그러진다. 지구본을 종이 지도로 펴면 극지방이 넓어지는 것과 같은 문제다. 계산은 이 뒤틀림을 되도록 줄이는 쪽으로 면을 편다.

네 번째 단계는 따로 펴낸 여러 면을 이어 붙이는 일이다. 면과 면의 경계를 맞춰 한 장의 큰 그림으로 합친다[1].

네 단계 가운데 첫 단계가 까다롭다. 층을 한 번 잘못 짚으면 뒤의 세 단계가 모두 어긋난다. 잘못 펴낸 그림에는 없던 글자가 생기기도 한다. 연구자들이 이 단계를 손으로 확인하는 까닭이다.

이 네 단계를 지나면 검은 원통이 한 장의 평면이 된다. 연구진은 가죽 다섯 겹을 이 방식으로 펴냈다. 펼친 면의 넓이는 약 94제곱센티미터로 추산되었다[1]. 공책 한 쪽보다 조금 작은 넓이다.

펼친 그림에도 문제는 남는다. 가죽의 두께가 곳곳마다 다르고 주름도 남아 있다. 촬영 과정에서 생긴 잡음도 섞인다. 그래서 어떤 데는 지나치게 밝고 어떤 데는 어둡다. 글자와 얼룩이 섞이면 읽기 어렵다.

연구진은 이 얼룩을 계산으로 걷어냈다. 먼저 둘레 화소의 평균 밝기를 재서 바탕의 밝기 지도를 만든다. 이 지도는 넓게 퍼진 밝기 흐름만 담는다. 그다음 원래 그림을 이 바탕 지도로 나눈다. 나누기를 거치면 넓게 퍼진 밝기 차이가 사라진다. 좁은 폭에 몰려 있는 먹물 신호만 또렷하게 남는다. 이 과정을 거친 그림에서 히브리 문자가 드러났다[1].

2026년에도 이어지는 숙제

가상 전개는 그 뒤로 다른 유물에도 쓰였다. 2026년 6월 25일 베수비오 챌린지가 소식을 알렸다. 헤르쿨라네움 두루마리 한 권을 처음부터 끝까지 읽었다는 내용이다. 유물 번호는 PHerc. 1667이다. 프랑스 그르노블의 유럽 싱크로트론 방사광 시설에서 촬영이 이루어졌다. 펼쳐 낸 그리스어 본문은 약 1.4미터 길이의 스물두 단이다. 기원전 2세기의 윤리학 저술로 확인되었다[2].

싱크로트론은 빠르게 달리는 전자를 자기장으로 휘게 하는 장치다. 전자가 휠 때 아주 강한 엑스선이 나온다. 이 빛으로 찍으면 작은 밀도 차이도 드러난다. 실험실 촬영기로는 얻기 어려운 선명함이다. 에인게디에서 쓴 장치보다 한 단계 위의 도구다.

이 성과에는 촬영 방법의 개선이 있었다. 새 촬영법에서는 탄소 먹물이 자료 안에서 직접 보였다. 밀도 차이가 작아 보이지 않던 먹물을 잡아낸 것이다. 에인게디에서 저절로 주어졌던 밝기 차이를 촬영 기술로 만들어 낸 것이다. 남은 두루마리 600여 권에도 같은 방법을 쓸 수 있다고 알려졌다[3].

숙제는 아직 남아 있다. 2026년 7월 10일에 공개된 정리는 남은 문제를 여섯 가지로 꼽는다. 겹이 눌린 구간에서는 엑스선이 흩어져 화면이 뿌예진다. 층을 찾는 계산은 구멍을 내거나 두 층을 붙여 버린다. 복셀 자료에는 어느 층이 어느 층에 이어지는지가 담겨 있지 않다. 사람이 만든 학습용 표시는 알갱이 단위로 정확하지 않다. 한 두루마리에서 배운 모형이 다른 두루마리에서는 잘 듣지 않는다. 먹물이 층의 어느 깊이에 있는지도 명확하지 않다[4].

해결 방법도 함께 제시되었다. 정답 표시가 모자라면 자료 자체에서 규칙을 배우게 하는 길이 있다. 스스로 배우는 방식의 학습 모형이 그 후보로 꼽힌다. 층을 찾는 계산에도 새 신경망 구조가 쓰이고 있다[4].

새로운 실마리도 나왔다. 2026년 3월 29일 공개된 연구는 먹물이 표면의 모양을 바꾼다는 점에 기댔다. 연구진은 이미 펼쳐진 파피루스의 표면 굴곡을 정밀하게 재어 기계

학습에 가르쳤다. 그 결과 굴곡만으로도 글자가 놓인 곳을 가릴 수 있었다. 이 연구는 굴곡을 읽으려면 어느 정도의 촬영 눈금이 필요한지도 따져 보았다[5]. 에인게디에서 밀도로 얻었던 실마리를 다른 성질에서 다시 찾은 연구다.

참고 자료

[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, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

[3] ASNT. (2026, 8월 13일). 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 (2026년 9월 3일 접속)

[4] Vesuvius Challenge. (2026, 7월 10일). Open Problems: Why Reading Every Herculaneum Scroll Is Still a Challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026년 9월 3일 접속)

[5] Angelotti, G., Nicolardi, F., Henderson, P., & Seales, W. B. (2026, 3월 29일). Ink Detection from Surface Topography of the Herculaneum Papyri. arXiv:2603.27698. https://arxiv.org/abs/2603.27698 (2026년 9월 3일 접속)

3. 밝혀진 진실

화면에 떠오른 레위기

컴퓨터 화면에 떠오른 글자는 히브리 문자였다. 연구진은 가죽 다섯 겹을 펴냈고 그 안에서 두 단의 글이 나왔다. 본문은 구약 성경의 세 번째 책인 레위기였다. 읽어 낸 곳은 레위기 1장과 2장이다[1].

합쳐진 그림에는 두루마리 한 장의 글 서른다섯 줄이 담겼다. 이 가운데 열여덟 줄은 글자가 남아 있었다. 나머지 열일곱 줄은 학자들이 복원한 것이다. 복원은 컴퓨터가 글을 지어내는 일이 아니다. 남은 획과 알려진 본문을 맞춰 빠진 곳을 채우는 작업이다. 한 줄에 든 글자와 사이 공백은 서른셋에서 서른넷이었다[1].

논문이 남은 줄과 복원한 줄을 나누어 밝힌 점이 중요하다. 어디까지가 확인된 사실인지 독자가 구분할 수 있기 때문이다. 남은 것과 채워 넣은 것을 섞어 발표하면 뒷사람이 검증할 수 없다. 덕분에 다른 학자들이 같은 자료로 다시 따져 볼 수 있었다.

레위기 1장은 번제의 규칙을 다룬다. 번제는 제단 위에서 제물을 온전히 태워 바치는 제사다. 2장은 곡식으로 드리는 제사의 규칙을 다룬다. 고운 가루에 기름을 섞고 향을 얹는 방법이 여기에 나온다. 불에 탄 성궤에서 불에 태워 바치는 제사의 규칙이 나온 것이다.

복원된 글에는 레위기의 첫 구절도 들어 있다[1]. 여호와가 회막에서 모세를 불러 말했다는 문장이다. 책의 첫머리가 남았다는 것은 이 두루마리가 레위기 한 권의 시작이었음을 뜻한다.

두 단이라는 말도 뜻이 있다. 옛 두루마리는 글을 세로로 긴 칸에 나누어 적는다. 한 칸을 다 채우면 옆 칸으로 넘어간다. 두 단이 나왔다는 것은 그런 칸 두 개가 이어져 나왔다는 뜻이다.

읽어 낸 것은 글자만이 아니다. 서기가 붓을 눌러 쓴 획의 굵기까지 화면에 남았다. 사람이 손으로 옮겨 적은 사본이라는 증거다.

천 년의 빈칸을 메우다

연구진은 두루마리의 나이도 쟀다. 방사성 탄소 연대 측정으로 서기 3세기에서 4세기라는 값이 나왔다[1]. 이 방법은 유물에 남은 탄소의 변화를 재서 나이를 알아낸다. 이 값은 옛 성경 사본 연구의 오랜 빈칸을 메운다.

빈칸이 무엇인지부터 보자. 쿰란 동굴에서 나온 사해문서 성경 사본들은 기원전 3세기부터 기원후 1세기 사이의 것이다. 반면 온전한 형태로 남은 표준 히브리어 성경 사본은 서기 1000년 무렵의 것이다. 두 무리 사이에는 천 년 가까운 빈칸이 있다. 이 기간에 본문이 어떻게 옮겨졌는지 알려 주는 사본이 드물었다.

빈칸이 크면 의심도 커진다. 중세 사본이 옛 본문을 그대로 옮겼는지 확인할 길이 없기 때문이다. 학자들 사이에서는 중세 필사자들이 본문을 다듬었을 것이라는 견해도 있었다.

에인게디 두루마리는 이 빈칸의 한복판에서 나왔다. 그리고 읽어 낸 본문은 중세 히브리어 성경의 자음 본문과 완전히 같았다[1]. 히브리어는 닿소리 글자만 적고 홀소리는 적지 않는다. 그 닿소리만 적힌 본문을 자음 본문이라고 부른다. 이 중세 본문을 맛소라 본문이라고 부른다. 유대교 필사자들이 규칙을 정해 지켜 온 표준 본문이다. 문단을 나

눈 위치까지 같았다[2]. 이 결과는 오랜 세월 동안 본문이 그대로 옮겨졌음을 보여 준다.

유대교 필사자들은 성경을 옮겨 적을 때 규칙을 엄격하게 지켰다고 알려져 있다. 글자를 세어 확인하는 방법도 그런 규칙 가운데 하나다. 에인게디 두루마리는 그 규칙이 실제로 지켜졌음을 보여 주는 물증이 되었다. 사본 한 점이 오랜 논쟁에 자료 하나를 보탠 것이다.

이 사본에는 기록 하나가 더 붙었다. 지금까지 성궤 안에서 나온 모세오경 사본 가운데 연대가 앞선다[1]. 모세오경은 구약 성경의 첫 다섯 권을 함께 이르는 말이다.

확정된 것과 남은 것

밝혀진 것과 아직 논의 중인 것은 나누어 보아야 한다.

밝혀진 것은 이렇다. 읽어 낸 본문은 레위기 1장과 2장이다. 탄소 연대 측정값은 서기 3세기에서 4세기를 가리킨다. 자음 본문은 중세 표준 본문과 같다. 먹물의 밀도가 가죽보다 높아 금속 성분이 섞였을 것으로 보인다[1].

남은 문제도 분명하다. 먹물의 성분은 직접 분석하지 못했다[1]. 겉으로 드러난 면이 없어 시료를 얻을 수 없었기 때문이다. 글씨체를 근거로 연대를 다르게 보는 견해도 있다. 고문서학자 아다 야르데니는 서기 1세기 후반이나 2세기 초로 보았다. 고문서학자는 옛 글씨체를 견주어 연대를 재는 학자다. 이에 대해 드루 롱에이커는 탄소 연대 측정이 더 믿을 만하다고 반박했다[2]. 두 견해는 아직 정리되지 않았다.

보관 기간도 풀리지 않은 문제다. 두루마리는 서기 3세기에서 4세기에 만들어졌다. 회당은 서기 600년 무렵에 불탔다[1]. 그렇다면 이 사본은 이백 년이 넘도록 성궤에 놓여 있었다. 공동체가 왜 오래된 사본을 계속 두었는지는 밝혀지지 않았다. 물려받은 사본을 소중히 여겼으리라는 추정만 있다.

읽어 낸 범위도 두루마리의 한 조각이다. 펼쳐 낸 것은 바깥쪽 다섯 겹이다[1]. 안쪽 깊은 곳은 겹이 더 심하게 눌려 있다. 눌린 구간에서 층을 갈라내지 못하는 문제는 2026년에도 남아 있다. 헤르쿨라네움 두루마리를 다루는 연구자들이 지금도 붙들고 있는 숙제다[3].

안쪽을 읽으려면 두 가지가 함께 좋아져야 한다. 층을 더 잘게 나누어 보는 촬영 눈금이 필요하다. 붙어 버린 층을 갈라내는 계산도 좋아져야 한다. 두 조건이 갖춰지면 에인게

디 두루마리는 다시 촬영대에 오를 수 있다. 유물이 그대로 남아 있기 때문에 가능한 일이다.

연대 측정 자체를 돕는 방법도 늘고 있다. 2025년에는 히브리어 옛 사본의 글씨체를 기계 학습으로 익혀 연대를 추정하는 모형이 발표되었다. 이 모형은 탄소 연대 측정값과 글씨체 자료를 함께 배웠다. 사람의 눈으로는 잡기 어려운 획의 굽이까지 재는 방식이다[4]. 이런 도구가 쌓이면 에인게디 두루마리의 연대 논쟁도 다시 검토될 수 있다.

가상 전개가 열어 준 길은 성경 사본에만 머물지 않는다. 접힌 채 봉인된 옛 편지에도 같은 방법이 쓰였다. 부서지기 쉬운 책과 문서에도 촬영과 계산이 들어간다. 손대지 않고 읽는다는 원칙이 자료를 다루는 기본이 되어 가고 있다.

컴퓨터가 내놓은 그림은 그 자체로 최종 답이 아니다. 2016년 연구에서 컴퓨터가 한 일은 층을 펴서 그림을 만드는 데까지다. 그 그림에서 글자를 읽고 뜻을 붙인 것은 사람이다. 그래서 고문서학자가 한 글자씩 다시 살펴야 한다. 이 사건이 남긴 것은 본문 한 편만이 아니다. 손대면 부서지는 유물을 다루는 방법이 하나 생겼다. 촬영으로 숫자를 얻고 계산으로 펴고 사람이 검토하는 순서다. 이 순서는 뒤에 나올 여러 유물에도 그대로 쓰인다.

에인게디에서 읽어 낸 레위기도 그 검토를 거쳐 학계의 기록이 되었다. 45년의 침묵을 깬 것은 기계와 사람이 함께 한 일이다. 촬영 장치는 밀도를 재고 계산은 층을 폈다. 그리고 사람이 글자를 읽고 뜻을 붙였다. 다음 장에서는 글자 모양 자체를 살피는 이야기가 이어진다.

참고 자료

[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, 9월 3일). Wikipedia. https://en.wikipedia.org/wiki/En-Gedi_Scroll (2026년 9월 3일 접속)

[3] Vesuvius Challenge. (2026, 7월 10일). Open Problems: Why Reading Every Herculaneum Scroll Is Still a Challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026년 9월 3일 접속)

[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

제 2 부

사해 동굴의 비밀 코드와 서기들의 필적 지문

03 CHAPTER

제3장. 이사야 두루마리에 숨겨진 두 명의 그림자: 서기 감정

필적을 재는 기계가 찾아낸 이사야 두루마리의 두 서기

1. 사건 파일

동굴에서 나온 가죽 두루마리

1947년에 사해 서쪽 쿰란 지역의 동굴에서 오래된 가죽 두루마리가 발견되었다. 염소를 치던 베두인 사람들이 절벽 동굴 안에서 항아리와 두루마리를 찾아냈다. 이 유물들은 사해문서라는 이름으로 불린다. 이후 여러 동굴에서 발굴이 이어졌다. 지금까지 모인 사해문서는 약 930점의 사본으로 정리되어 있다. 손으로 베껴 쓴 책을 사본이라고 부른다. 낱장으로 세면 가죽과 파피루스 파편이 수천 장을 넘는다[1]. 기록된 시기는 기원전 3세기부터 기원후 1세기 사이로 본다[2].

사해는 소금기가 짙은 호수이고 둘레는 메마른 사막이다. 비가 거의 오지 않고 공기가 건조하다. 동굴 안은 온도도 크게 흔들리지 않는다. 가죽과 파피루스가 썩지 않고 버틴 까닭이 여기에 있다. 이만한 규모의 옛 문서 묶음이 남은 곳은 드물다.

이 유물 가운데 학자들의 눈길을 붙든 두루마리가 하나 있다. 이사야 대두루마리다. 학계에서 쓰는 기호는 1QIsaa다. 쿰란 제1동굴에서 나온 이사야서 첫 번째 사본이라는 뜻이다. 기원전 2세기에 짐승 가죽 위에 히브리어로 쓴 사본이다. 길이는 약 7.3미터다. 구약 성경 이사야서의 내용을 거의 통째로 담고 있다. 글은 세로로 긴 54개의 단으로 나뉘어 있다[2]. 이 단을 학술 용어로 칼럼이라고 부른다. 사해 동굴에서 나온 성경 사본 중에 보존 상태가 제일 좋다.

두루마리는 여러 장의 가죽을 실로 이어 붙여 만들었다. 한 장에 몇 개의 단을 쓰고 다음 장으로 넘어간다. 다 쓴 두루마리는 둘둘 말아 보관한다. 읽을 때는 한쪽을 풀고 다른 쪽을 감아 가며 본다. 두루마리라는 이름이 여기서 나왔다.

이 사본의 값어치는 나이에서 온다. 사해문서에는 지금 남아 있는 히브리어 성경 사본 중 제일 오래된 것들이 들어 있다[3]. 이사야 대두루마리도 그중 하나다. 성경 본문이 오랜 세월 동안 어떻게 전해졌는지를 견주어 볼 자료가 생긴 것이다. 그런데 두루마리를 만든 사람에 대해서는 알려진 바가 없었다.

누가 썼는지를 두고 갈린 학계

두루마리를 읽는 일과 누가 썼는지 밝히는 일은 서로 다르다. 글자는 다 읽히는데 이 사본에는 베껴 쓴 사람의 이름이 없다. 글을 베껴 쓰는 일을 맡은 사람을 서기라고 부른다. 이 두루마리에는 서기의 서명도 없고 날짜도 없다[2]. 그래서 학자들은 글씨 모양만 보고 사람 수를 헤아려야 했다.

이 문제를 두고 학계는 두 갈림길에 섰다. 한쪽은 서기 한 사람이 처음부터 끝까지 혼자 썼다고 보았다. 두루마리 전체의 글씨 모양이 고르다는 것이 근거였다. 다른 쪽은 두 사람 이상이 나누어 썼다고 보았다. 중간을 지나면서 획의 생김새가 조금씩 달라진다는 것이 근거였다.

두 주장 모두 어려운 문제 하나를 안고 있었다. 한 사람이 쓴 글씨도 처음과 끝이 똑같지는 않다. 손이 피곤해지면 획이 느슨해지고 펜을 갈면 굵기가 바뀐다. 이 흔들림의 크기를 모르면 사람이 바뀐 것인지 판단할 수 없다. 두 주장 모두 눈으로 본 인상에 기대고 있었다. 어느 쪽이 맞는지 가릴 저울추가 없었다.

쿰란에는 두루마리를 만들고 지키던 공동체가 있었던 것으로 본다. 사람들은 두루마리를 항아리에 넣어 동굴에 감추었다. 전쟁이 다가오자 문서를 지키려 한 것으로 보인다. 감춘 사람들의 이름도 남아 있지 않다. 두루마리를 만든 손을 찾는 일은 그 사람들에게 다가서는 길이기도 하다.

이 다툼은 글씨 이야기에 그치지 않는다. 사본을 여러 사람이 나누어 만들었다면 서기들이 함께 일하는 조직이 있었다는 뜻이다. 성경 본문이 어떤 사람들의 손을 거쳐 오늘까지 왔는지를 밝히는 일과 이어진다[2]. 사해문서 연구자들이 이 물음을 오래 붙들고 있었던 까닭이 여기에 있다.

눈으로 하는 감정의 한계

옛 글씨를 연구하는 학문을 고문서학이라고 한다. 학자가 확대경을 들고 글자 모양을 살피는 방식이다. 글자의 높이와 기울기, 획의 굵기를 재고 그림으로 옮겨 적는다. 여러 사본의 글씨를 시대순으로 늘어놓아 변화의 흐름을 잡기도 한다. 오랫동안 이 방법이 유일한 감정 도구였다.

이 방법에는 약점이 있다. 학자의 눈과 경험에 기대는 폭이 넓다. 같은 글자를 보고도 학자마다 다른 평가를 내놓는 일이 잦았다[2]. 사람의 눈은 아주 작은 수치 차이를 재지 못한다. 갈대펜을 누른 힘의 세기나 잉크가 번진 두께는 눈으로 가늠하기 어렵다. 기억에도 한계가 있다. 몇 미터 떨어진 두 글자를 머릿속에 나란히 놓고 견주기란 쉽지 않다. 결국 의견 대립만 되풀이되었다.

또 하나의 벽은 그 시대의 교육 방식이었다. 서기가 되려는 사람은 스승의 글씨를 그대로 베끼는 훈련을 받았다. 자기만의 개성을 드러내지 않도록 배웠다. 그 결과 한 학교에서 나온 서기들의 글씨는 겉보기에 거의 똑같았다. 인간 학자에게는 이것이 짙은 안개였다. 두루마리 앞뒤의 미세한 차이가 한 사람의 컨디션 변화인지 사람이 바뀐 것인지 가릴 수 없었다[2].

감정이 어긋나면 뒤따르는 판단도 어긋난다. 서기가 한 사람이라면 사본 하나에 한 사람의 시간이 담긴다. 두 사람이라면 일을 나누어 맡는 조직이 있었다는 말이 된다. 사본을 만든 곳의 규모와 성격이 달라진다. 글씨 감정 하나에 여러 결론이 매달려 있었다.

교실에서 같은 글씨본을 놓고 받아쓰기를 한 아이들을 떠올려 보면 사정이 비슷하다. 다 같이 반듯하게 쓰려고 애쓴 공책들은 겉보기에 구별이 잘 되지 않는다. 그래도 획을 꺾는 습관은 아이마다 조금씩 다르다. 문제는 그 차이를 재는 자가 없었다는 것이다.

사정이 바뀐 것은 사본 사진이 디지털로 공개되면서부터다. 실물 두루마리는 빛과 습기에 약해 자주 펼칠 수 없다. 사진이 있으면 유물을 건드리지 않고도 몇 번이든 살펴볼 수 있다. 이스라엘 문화재청은 구글과 함께 사해문서 디지털 자료집을 만들었다. 메가비전이라는 장비로 여러 파장의 빛을 쬐어 유물을 촬영했다. 사람 눈에 보이지 않는 빛인 근적외선까지 담아 흐려진 글자를 되살렸다. 실물을 직접 들여다보는 것과 비슷한 화질을 목표로 삼았다[1].

컴퓨터가 읽을 수 있는 고해상도 사진이 쌓이자 새 도구가 들어올 길이 열렸다. 사진은 숫자의 모임이다. 화면을 이루는 작은 점 하나를 화소라고 부른다. 화소마다 밝기 값이

붙어 있다. 획의 굵기와 각도를 숫자로 잴 수 있다는 뜻이다. 눈으로 하던 감정을 계산으로 바꿀 준비가 끝난 것이다. 여기서부터 인공지능 탐정이 나선다.

참고 자료

[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 (2026년 9월 3일 접속)

[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, 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 (2026년 9월 3일 접속)

2. 과학 수사대의 추적: 글씨체에 남겨진 생체 지문

수사팀이 꾸려지다

네덜란드 흐로닝언 대학교 연구팀이 이 사건을 맡았다. 고문서학자 믈라덴 포포비치가 팀을 이끌었다. 인공지능 연구자 람베르트 스호마커와 마루프 다리가 함께했다[1]. 이들은 2017년에 사해문서 필적을 컴퓨터로 재는 방법을 먼저 발표했다[2]. 그 방법을 다듬어 이사야 대두루마리에 겨눈 결과가 2021년에 나왔다[1, 5].

연구팀이 세운 목표는 분명했다. 사람의 인상 평가를 숫자로 바꾸는 것이었다. 같은 사람이 쓴 글씨 안에서 생기는 흔들림의 크기를 먼저 잰다. 그다음 두루마리 앞뒤의 차이가 그 흔들림보다 큰지를 따진다. 차이가 더 크면 다른 사람이 썼다는 뜻이 된다.

2017년에 만든 특징 계산법은 사본끼리 견주는 데 쓰였다[2]. 연구팀은 그 계산법을 이사야 대두루마리 하나에 집중해 다시 적용했다. 한 사본 안에서 서기가 바뀐 곳을 찾는 일은 사본끼리 견주는 일보다 어렵다. 글씨의 차이가 훨씬 작기 때문이다.

이 방식에는 좋은 점이 하나 있다. 계산 절차를 적어 두면 누구든 같은 사진으로 같은 답을 얻는다. 학자의 안목에 기대지 않아도 된다. 결과가 틀렸다면 어느 단계에서 틀렸는지 되짚을 수 있다. 감정 결과를 놓고 다시 다투더라도 다툴 근거가 숫자로 남는다.

얼룩을 지우는 첫 단계

첫 작업은 사진에서 잉크만 남기는 일이다. 이 작업을 이진화라고 부른다. 모든 화소를 잉크와 배경 둘 중 하나로 갈라 놓는 처리다. 사진을 검은 글씨와 밝은 배경으로 나누는 일이다.

이 단계가 필요한 까닭이 있다. 두루마리는 약 2천 년 동안 동굴에 있었다. 가죽 표면에는 얼룩과 곰팡이 자국이 생겼고 잔금도 갔다. 컴퓨터가 보기에 이런 자국과 잉크 획은 둘 다 어두운 화소다. 배경을 그대로 두면 기계가 글씨가 아닌 가죽의 결을 학습한다. 그러면 서기 대신 가죽 종류를 구별하게 되어 감정이 어긋난다[1].

연구팀은 이 일을 맡을 신경망을 따로 만들었다. 이름은 바이넷이다. 신경망은 사람 뇌의 신경 세포 연결을 본떠 만든 계산 구조다. 바이넷은 손으로 잉크와 배경을 표시한 견본을 보고 그 판단 기준을 배운다. 학습을 마친 바이넷은 새 사진에서 얼룩을 걷어 내고 획만 남긴다.

여기서 지켜야 할 규칙이 있다. 원래 획의 굵기와 가장자리 모양을 건드리지 않아야 한다. 서기의 버릇이 담긴 곳이 바로 그 가장자리이기 때문이다. 잡티를 지우려다 획까지 매끈하게 다듬으면 증거가 사라진다. 바이넷은 잉크 흔적을 그대로 두는 쪽으로 설계되었다[1]. 이 기술은 2019년에 공개 초고로 먼저 나왔다[3]. 설계와 성능을 다시 정리한 논문은 2026년 국제 문서 인식 학술 논문집에 실렸다[4].

글씨 전체의 버릇을 재는 자

깨끗해진 글자에서 두 종류의 특징을 뽑는다. 첫 번째는 결에 관한 특징이고 두 번째는 모양에 관한 특징이다.

결에 관한 특징을 힌지라고 부른다. 힌지는 문의 경첩을 뜻하는 말이다. 획이 꺾이는 곳마다 두 방향의 각도를 잰다. 그 각도 짝이 얼마나 자주 나오는지 표로 만든다. 이 표는 글씨를 쓸 때 손목이 도는 버릇을 숫자로 담는다. 글자를 하나하나 알아보지 않아도 계산할 수 있다[1].

모양에 관한 특징은 프래글릿이라고 부른다. 획을 짧게 자른 토막을 뜻하는 말이다. 획을 잘라 모아 목록으로 만든다. 연구팀은 여러 사본에서 뽑은 토막들로 표준 목록을 만들었다. 새 글씨가 들어오면 그 토막이 목록의 어느 항목과 닮았는지를 세어 준다. 사람이 글자를 배울 때 익힌 획의 기본형이 여기에 드러난다[1].

이 두 특징에는 공통점이 있다. 글자의 뜻을 몰라도 계산할 수 있다는 것이다. 히브리어를 읽지 못하는 컴퓨터도 획의 각도와 굽은 모양은 잴 수 있다. 훼손된 글자가 섞여 있어도 남은 획에서 값을 뽑는다. 옛 문서를 다룰 때 쓸모가 큰 성질이다.

두 특징은 서로 다른 것을 본다. 힌지는 손의 움직임을 보고 프래글릿은 배운 글꼴을 본다. 연구팀은 둘을 각각 쓰고 합쳐서도 썼다. 서로 다른 잣대가 같은 답을 내는지 확인하기 위해서다[1].

알레프 한 글자에 남은 흔적

앞의 두 특징은 글씨 전체를 뭉뚱그려 잰다. 이번에는 글자 하나를 크게 확대해 들여다볼 차례다.

연구팀은 히브리어 자모 하나를 따로 골라 정밀 검사를 했다. 알레프다. 알레프는 히브리어의 첫 글자이고 두루마리에 제일 자주 나온다. 대각선 획 하나에 짧은 획 둘이 붙은 모양이다.

알레프를 고른 데는 까닭이 있다. 표본이 많을수록 통계가 든든해진다. 자주 나오는 글자라야 앞뒤 구역에서 골고루 모을 수 있다. 두루마리 전체에 알레프는 5,011개 있다. 연구팀은 그중 758개를 컴퓨터로 골라내어 분석에 썼다[1]. 획이 끊기지 않고 온전하게 남은 것만 추렸다.

고른 알레프는 두 무리로 나누었다. 두루마리 앞쪽에서 나온 것과 뒤쪽에서 나온 것이다. 각 무리의 이미지를 같은 위치에 겹쳐 평균 모양을 만들었다. 여러 장의 반투명 종이를 포개어 놓고 위에서 내려다보는 방식과 같다. 자주 겹치는 획은 진해지고 사람마다 다른 획은 흐려진다. 두 평균 그림을 서로 빼면 남는 것은 차이뿐이다. 그 차이가 어느 획에 몰려 있는지를 색으로 볼 수 있다[1].

두 무리로 갈라진 점들

숫자로 바뀐 특징은 값이 수백 개씩 늘어선 긴 목록이다. 사람은 이런 목록을 눈으로 견주지 못한다. 그래서 연구팀은 목록을 두 개의 값으로 줄여 평면에 점으로 찍었다. 값이 닮은 글씨는 가까이 모이고 다른 글씨는 멀어지도록 줄인다. 점의 위치를 조금씩 옮기는 계산을 되풀이해 만든 그림이다.

54개의 단 가운데 사진이 온전한 것을 모두 이 평면에 올렸다. 16번과 46번 단은 쓸 만한 사진이 없어 빠졌다. 점들이 찍히자 두 무리가 나타났다. 앞쪽 단들이 한 무리를 이루고 뒤쪽 단들이 다른 무리를 이루었다[1]. 두 무리 사이에는 눈에 보이는 틈이 있었다.

연구팀은 이 갈라짐이 우연인지 검증했다. 카이제곱 검정과 티 검정 같은 통계 방법을 썼다. 카이제곱 검정은 차이가 우연인지 따지는 계산이다. 티 검정은 두 무리의 평균 차이를 따지는 계산이다. 두 방법 모두 우연히 이런 모양이 나올 확률을 구한다. 그 확률이 아주 작으면 우연으로 보지 않는다. 한 사람의 글씨 안에서 생기는 흔들림만으로는 이만한 차이가 나오지 않았다[1].

여러 잣대가 같은 답을 낸 것도 중요하다. 힌지로 재도 프래글릿으로 재도 같은 곳에서 경계가 잡혔다. 알레프 평균 그림에서도 같은 곳이 갈렸다. 경계는 27번 단과 28번 단 사이에 놓였다[1]. 서로 다른 방법이 한곳을 가리켰다면 그 경계는 사진의 잡티 때문에 생긴 것이 아니다. 사람의 눈이 놓친 선을 기계가 그어 준 것이다.

연구팀은 반대 가능성도 따져 보았다. 사진의 밝기나 가죽의 상태가 앞뒤에서 달라 생긴 결과일 수 있다. 그러나 이진화로 배경을 걷어 낸 뒤에도 경계는 그대로 남았다[1]. 원인이 가죽이 아니라 글씨 자체에 있다는 뜻이다.

참고 자료

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

3. 밝혀진 진실

27번 단에서 바뀐 손

분석 결과는 복수 서기설의 손을 들어 주었다. 이사야 대두루마리는 두 사람이 나누어 썼다. 첫 번째 서기가 1번 단부터 27번 단까지 썼다. 두 번째 서기가 28번 단부터 54번 단까지 이어 썼다[1].

이 경계는 다른 증거와도 맞물린다. 27번 단의 아래쪽은 세 줄 높이만큼 비어 있다. 28번 단은 새로 이어 붙인 가죽 장에서 시작한다. 그곳은 이사야서의 새 장이 시작하는 곳이기도 하다[1]. 일을 넘겨받기에 알맞은 매듭이었다. 컴퓨터가 찾아낸 통계의 경계와 두루마리의 물리적 경계가 같은 곳에 있었다.

27번 단 뒤에서 서기가 바뀌었으리라는 짐작은 전에도 있었다. 가죽 장이 이어지는 곳과 본문이 나뉘는 곳을 보고 나온 견해였다. 다만 학계의 합의로는 이어지지 못했다. 인공지능 분석이 이 짐작에 수치를 붙여 주었다[2].

이 발견의 뜻은 크다. 성경 사본 하나를 여러 사람이 이어서 만들었다는 근거가 수치로 나왔다[1]. 두 서기는 이름도 남기지 않았고 서명도 남기지 않았다. 남긴 것은 손이 움직인 버릇뿐이었다. 2천 년 뒤의 기계가 그 버릇을 읽어 냈다. 포포비치는 이름은 끝내 알 수 없겠지만 필체를 통해 그들과 악수하는 느낌이라고 말했다[2].

같은 학교에서 배운 두 사람

두 서기의 글씨는 여전히 서로 닮아 있다. 기계가 잰 뒤에도 차이는 미세했다. 연구팀은 이 닮음 자체를 또 하나의 증거로 읽었다. 두 사람의 글씨 방식이 서로 가깝다는 뜻이다[1]. 연구팀은 두 서기가 같은 곳에서 훈련받았거나 같은 계통에서 나왔다고 보았다[2].

여기서 고대 서기 학교의 모습이 그려진다. 훈련생은 공인된 글꼴을 획 단위로 익혔다. 자기 개성을 눌러야 좋은 서기가 되었다. 그 규율 덕분에 한 두루마리 안에서 글씨가 고르게 유지되었다.

그래도 몸의 버릇까지 지울 수는 없었다. 갈대펜을 쥐는 각도는 사람마다 다르다. 손을 움직이는 속도와 누르는 힘도 다르다. 이런 차이는 획이 꺾이는 각도와 획의 가장자리 모양에 남는다. 사람의 눈은 이 미세한 차이를 모아서 판단하지 못한다. 컴퓨터는 수천

개의 값을 모아 평균과 흩어짐을 계산한다. 규율이 덮어 놓은 개인의 흔적이 그렇게 드러났다[1].

이 결과는 사본을 만드는 일이 공동 작업이었음을 보여 준다. 긴 두루마리 하나를 만들려면 여러 달이 걸린다. 한 사람이 끝까지 붙들고 있을 이유는 없다. 일을 나누어 맡고 넘겨주었을 가능성이 여기서 나온다[1]. 성경 본문은 이런 손과 손을 거쳐 오늘까지 전해졌다.

기계가 아직 못 하는 일

이 성과가 인공지능의 만능을 뜻하지는 않는다. 같은 연구팀이 만든 다음 도구가 그 한계를 보여 준다.

연구팀은 2025년에 이노크라는 연대 추정 모델을 내놓았다. 이노크는 글씨 모양을 보고 사본이 쓰인 시기를 맞히는 모델이다. 방사성 탄소 연대는 오래된 물건에 남은 탄소로 나이를 재는 방법이다. 이노크는 그 방법으로 연대를 잰 표본 24개를 보고 배웠다. 연대를 모르는 사본 135점에 적용한 결과도 함께 발표했다. 예측 오차는 평균 27.9년에서 30.7년 사이였다. 고문서학자들이 이 예측을 검토한 결과 79퍼센트가 타당하다는 평가를 받았다[3].

나머지 21퍼센트에서 논쟁이 벌어졌다. 고문서학자 크리스토퍼 롤스턴은 한 사례를 들어 반박했다. 이노크는 다니엘서를 담은 4Q114 사본을 기원전 230년에서 160년 사이로 보았다. 롤스턴은 다니엘서 7장부터 12장이 한 사건을 가리킨다고 보았다. 기원전 167년에 안티오코스 4세가 예루살렘 성전을 더럽힌 일이다. 그 본문이 사건보다 먼저 쓰였을 수는 없다. 그런데 이노크가 내놓은 범위는 대부분 167년보다 앞선다. 롤스턴은 이노크의 연대가 너무 이르다고 지적했다. 그는 이노크가 도구 상자에 든 여러 도구 중 하나여야 한다고 말했다[4].

이 논쟁이 알려 주는 바가 있다. 기계는 글자의 모양만 본다. 글에 적힌 역사적 사건이나 문장의 뜻은 읽지 못한다. 그래서 기계의 답은 확정이 아니라 후보다. 후보를 좁힌 뒤 역사학자가 문맥을 대조해 판정해야 한다. 어떤 글자를 재야 하는지 정한 사람도 고문서학자였다. 알레프를 고른 판단은 옛 글씨를 오래 본 경험에서 나왔다.

학습 자료의 수도 걸림돌이다. 이노크는 연대를 아는 표본 24개로 배웠다[3]. 옛 유물은 표본을 늘리기 어렵다. 연대를 재려면 유물을 조금 떼어 내 태워야 하기 때문이다. 적

은 표본으로 배운 모델은 배운 범위를 벗어나면 흔들린다. 서기 감정도 사정이 비슷하다. 자주 나오는 글자는 표본이 넉넉하지만 드문 글자는 몇 개뿐이다. 표본이 적은 곳에서 기계의 판단은 힘을 잃는다.

그래서 연구팀도 인공지능을 유일한 심판으로 두지 않는다. 가죽이 이어진 곳과 본문이 나뉜 곳, 학자들이 쌓아 온 견해가 함께 놓인다. 여러 증거가 한곳을 가리킬 때 결론이 선다. 이사야 대두루마리에서 답이 나온 것도 그런 방식이었다. 기계는 사람이 세운 물음에 숫자로 답했을 뿐이다.

다음 수사의 표적

이사야 대두루마리 연구는 더 큰 계획으로 이어졌다. 유럽연구위원회는 2026년 6월 23일에 포포비치에게 어드밴스드 그랜트를 주기로 했다. 연구비는 250만 유로이고 기간은 5년이다[5]. 새 계획의 이름은 트레이싱 스크라이브스 앤드 스크롤스다[6]. 서기와 두루마리의 자취를 따라간다는 뜻이다.

이스라엘 문화재청은 2026년 6월 30일에 이 계획에 참여한다고 밝혔다. 피사 대학교와 나폴리 페데리코 2세 대학교, 남덴마크 대학교, 뢰번 대학교도 함께한다. 베를린과 토리노의 이집트 박물관도 이름을 올렸다[6].

이번 질문은 누가 썼는가에서 어디서 만들었는가로 옮겨 간다. 연구팀은 사해문서 표본 약 250점을 조사한다. 가죽과 파피루스와 잉크의 화학 성분을 잰다. 그 성분을 고대 이집트 파피루스의 성분과 맞대어 본다. 재료가 어디서 왔는지 알면 사본을 만든 장소를 좁힐 수 있다. 인공지능은 이 화학 자료를 정리하고 글씨에 숨은 규칙을 찾는 일을 맡는다[7]. 포포비치는 실험실 분석과 옛 글씨 연구를 합치면 전에는 닿지 못한 질문에 다가설 수 있다고 말했다[6].

이 방법은 이사야 대두루마리에만 쓰이지 않는다. 사해문서에는 아직 필사자를 모르는 사본이 많이 남아 있다. 같은 손이 쓴 사본을 묶으면 서기 한 사람의 작업 목록이 만들어진다. 그런 목록이 쌓이면 서기들이 서로 어떻게 이어져 있었는지 그려 볼 수 있다[1]. 두 사람을 찾아낸 이 연구가 그 출발선이 되었다.

이 연구가 남긴 또 하나의 성과는 방법 자체다. 사진에서 잡티를 걷어 내고, 획을 숫자로 바꾸고, 통계로 검정하는 순서가 정리되었다. 다른 언어의 옛 문서에도 같은 순서를 적용할 수 있다. 이 책의 뒷장에서 만나게 될 점토판과 목간 연구도 비슷한 길을 걷는다.

이사야 대두루마리에서 시작된 수사는 아직 끝나지 않았다. 두 사람의 그림자를 찾아낸 도구가 이제 두루마리를 만든 장소를 찾아 나선다.

옛 문서를 읽는 일은 글자를 알아보는 데서 끝나지 않는다. 그 글자를 누가 어디서 썼는지까지 물어야 한다. 이사야 대두루마리는 몇 사람이 썼는가라는 물음에 숫자로 답한 첫 사례다. 두루마리는 2천 년 동안 아무 말도 하지 않았다. 사람은 이제 글씨의 떨림에서 그 대답을 읽는다.

참고 자료

[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, 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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

04 CHAPTER

제4장. 수만 조각의 점토·양피지 퍼즐을 맞추는 디지털 접착제

조각의 모서리와 필체를 맞춰 문서를 되살리는 디지털 접착제

1. 사건 파일

염소를 찾다 발견한 항아리

1947년 봄에 사해 서쪽 절벽에서 일이 벌어졌다. 베두인 소년들이 달아난 염소를 찾고 있었다. 소년들은 바위 사이에서 좁은 동굴 입구를 보았다. 동굴 안에는 키 큰 점토 항아리들이 서 있었다. 항아리 속에는 가죽에 글씨를 쓴 두루마리가 들어 있었다[1].

이 두루마리들이 사해문서다. 사해문서는 기원전 3세기 무렵부터 기원후 1세기 사이에 쓰인 유대 문헌이다. 그 뒤 아홉 해 동안 가까운 동굴 열 곳이 더 나왔다. 두루마리가 나온 동굴은 모두 열한 곳이다. 여기서 나온 글은 900편이 넘는다. 히브리어와 아람어와 그리스어로 쓰였다[1]. 대부분은 양이나 염소 가죽을 얇게 편 양피지에 쓰였다. 파피루스라는 식물 줄기 종이에 쓴 사본도 있다[7].

두루마리에는 히브리어 성경 사본 가운데 연대가 오래된 것들이 들어 있었다. 공동체의 규율을 적은 글과 달력을 적은 글도 나왔다[1]. 메마른 광야의 동굴이 이 글자를 2,000년 동안 지켰다.

동굴 근처에는 쿰란이라는 옛 마을 터가 있다. 학자 다수는 이곳에 살던 유대인 공동체가 두루마리를 남겼다고 본다. 이들은 성경을 베껴 쓰고 자기들의 규칙을 적어 두었다. 로마 군대가 다가오자 두루마리를 동굴에 숨긴 것으로 짐작한다.

제4동굴의 파편 더미

동굴마다 유물의 상태는 크게 달랐다. 제1동굴의 두루마리는 항아리 안에서 제법 온전하게 남았다. 제4동굴은 사정이 달랐다. 이 동굴의 두루마리는 항아리에 담기지 않은 채 바닥에 쌓여 있었다. 이곳 한 곳에서만 두루마리 약 500편의 파편 수천 점이 나왔다

[1].

동굴 안의 습기가 가죽을 서서히 상하게 했다. 벌레와 짐승이 가죽을 갉아 먹었다. 오랜 세월이 지나면서 두루마리는 잘게 부서졌다. 남은 것은 손톱만 한 파편 더미였다.

이스라엘 유물청이 보존하는 파편은 25,000점이 넘는다[2]. 파편 대부분에는 어느 두루마리에서 떨어져 나왔는지 적혀 있지 않다. 파편 하나에 글자가 두세 자만 남은 것도 흔하다. 학자들은 이것들을 큰 유리판 위에 늘어놓고 눈으로 맞추었다.

맞출 단서는 몇 가지뿐이었다. 글자 모양이 닮았는지 보았다. 가죽 색이 비슷한지 보았다. 부러진 테두리가 맞물리는지 대 보았다. 이 방법으로는 한 번에 두세 점밖에 견주지 못한다. 파편이 만 단위로 늘어나면 견줄 짝의 수는 걷잡을 수 없이 커진다.

수작업 퍼즐 맞추기는 오래 걸렸다. 사람의 눈은 파편 수만 개의 테두리를 한꺼번에 견주지 못한다. 게다가 파편의 테두리는 이미 닳고 오그라들어 있었다. 몇십 년 동안 학자들이 매달렸어도 제자리를 찾지 못한 파편이 많이 남았다.

문서를 처음 읽을 권한이 소수 학자에게만 주어진 것도 문제였다. 바깥 연구자들은 사진조차 보기 어려웠다[1]. 1991년 9월에 미국 헌팅턴 도서관이 이 빗장을 풀었다. 도서관은 소장하던 사진 필름을 조건 없이 열었다[3]. 사진이 열리자 컴퓨터로 파편을 맞추려는 시도가 시작되었다.

유물이 여러 나라로 흩어진 것도 걸림돌이었다. 발견 초기에 두루마리 일부는 베들레헴의 골동품상을 거쳐 팔려 나갔다[1]. 발굴 비용을 댄 나라의 기관이 파편을 나누어 받기도 했다. 그래서 한 두루마리에서 떨어진 파편이 여러 도시에 따로 놓이게 되었다. 두 파편이 붙는지 보려면 사람이 비행기를 타야 했다.

시장에 나타난 가짜 파편

사해문서 파편의 값이 오르면서 다른 문제가 생겼다. 2002년 무렵부터 골동품 시장에 사해문서 파편이라는 물건이 나왔다. 성경 구절이 적힌 조각이 70점가량 나타났다[4]. 학자들은 이 물건들을 "포스트 2002 파편"이라고 불렀다. 이 물건들에는 어느 동굴에서 언제 나왔는지를 알려 주는 기록이 없었다[6].

미국 워싱턴의 성서박물관도 이런 파편 16점을 사들여 전시했다. 여러 학자가 글자 모양이 어색하다고 의심했다. 박물관은 미술품 위조 조사 회사에 정밀 감정을 맡겼다. 조

사를 이끈 사람은 콜레트 롤 조사관이다[4].

2020년 3월에 결과가 나왔다. 16점 모두가 현대에 만든 가짜였다[4][5]. 조사팀은 3차원 현미경으로 표면을 들여다보았다. 재료는 진짜 사해문서의 얇은 양피지가 아니었다. 적어도 15점은 두껍고 결이 거친 가죽이었다. 잉크는 이미 갈라진 틈 속에 고여 있었다[4]. 가죽이 갈라진 다음에 글씨를 썼다는 증거였다.

이 사건은 성서 고고학계 전체를 흔들었다. 조지워싱턴대학교의 크리스토퍼 롤스턴 교수는 가짜가 지금도 계속 시장에 나온다고 말한다. 그는 자기 물건이 진짜라고 믿는 사람을 자주 만나지만 거의 모두 가짜라고 밝혔다. 발굴 기록이 남아 있는 유물과 시장에서 산 유물의 차이가 여기서 드러났다[6].

부서진 파편을 맞추는 일과 가짜를 걸러 내는 일은 뿌리가 같다. 둘 다 사람의 눈으로는 잡히지 않는 미세한 물리 흔적을 읽어야 한다. 인공지능이 이 일에 들어왔다. 2026년 6월에 유럽연구위원회가 새 연구비를 배정했다. 사해문서의 출신지를 밝히는 5년짜리 연구에 250만 유로를 주었다. 연구 이름은 "서기와 두루마리 추적"이다. 이 연구는 재료의 성분 분석과 인공지능과 옛 글씨 연구를 함께 쓴다[7].

이 장에서 따라갈 사건은 두 가지다. 하나는 흩어진 파편을 다시 잇는 일이다. 다른 하나는 가짜를 걸러 내는 일이다. 두 일은 같은 자료를 놓고 서로 다른 질문을 던진다. 앞의 질문은 이 둘이 원래 하나였는가다. 뒤의 질문은 이것이 원래 옛것인가다.

참고 자료

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[3] Deseret News. (1991, 9월 22일). 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/ (2026년 9월 3일 접속)

[4] Greshko, M. (2020, 3월 13일). "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 (2026년 9월 3일 접속)

[5] Solly, M. (2020, 3월 16일). 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/ (2026년 9월 3일 접속)

[6] Govier, G. (2026, 4월 8일). 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/ (2026년 9월 3일 접속)

[7] The Jerusalem Post. (2026, 6월 30일). 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 (2026년 9월 3일 접속)

2. 과학 수사대의 추적 1: 기하학적 엣지 친화도와 물리적 지문

찢어진 테두리를 숫자로 바꾸다

부서진 파편을 맞추는 첫 단서는 테두리 모양이다. 종이를 손으로 찢으면 톱니처럼 들쭉날쭉한 선이 생긴다. 이웃한 두 파편의 선은 서로 맞물린다. 사람은 이 선을 눈대중으로 견준다. 컴퓨터는 이 선을 숫자로 바꾼다.

컴퓨터는 테두리를 따라가며 굽은 정도를 잰다. 굽은 정도를 순서대로 늘어놓으면 파도 모양의 수열이 된다. 두 파편의 수열이 서로 뒤집어 겹칠 때 잘 맞으면 결합 후보가 된다. 이 방법을 엣지 친화도 계산이라고 부른다. 우리말로는 테두리 맞물림 점수다.

이 점수를 우리말로는 테두리 맞물림 점수라고 부를 수 있다. 두 파편의 테두리가 얼마나 잘 맞는지를 나타내는 수치다.

테두리만 보는 방법에는 한계가 뚜렷하다. 옛 파편의 테두리는 이미 닳아 있다. 마르면서 오그라들어 원래 모양을 잃은 곳도 많다. 그래서 연구자들은 테두리 밖의 단서를 함께 본다.

재료 자체가 남긴 지문

테두리만으로는 부족하다. 파편의 가장자리는 이미 닳아 있는 경우가 많다. 그래서 연구자들은 재료 자체를 본다.

파피루스는 식물 줄기를 가늘게 갈라 가로세로로 겹쳐 만든다. 그래서 표면에 줄기 무늬가 결처럼 남는다. 원래 한 장이었던 두 파편은 이 결이 이어진다[1].

이스라엘 하이파대학교의 로이 아비트볼 연구팀이 이 결을 읽었다. 연구팀은 파편 테두리 부근을 작은 네모로 잘랐다. 신경망이 네모 두 개를 보고 결이 이어지는지 점수를 냈다. 그 점수를 모아 파편 전체의 짝을 판정했다[1].

시험 대상은 상태가 나쁜 사해문서 파편 묶음이었다. 정답을 모르는 묶음이라 틀린 후보도 많이 나왔다. 그래도 맞을 수 없는 짝의 98퍼센트를 지웠다[1]. 사람이 볼 분량이 50분의 1로 줄었다.

이 결과의 뜻은 분명하다. 컴퓨터가 짝을 확정한 것이 아니다. 컴퓨터는 볼 필요가 없는 짝을 치웠다. 파편 맞추기에서 인공지능이 맡는 몫은 대개 이런 걸러 내기다.

프랑스 보르도의 앙투안 피론 연구팀은 여기에 쌍둥이 신경망을 썼다. 쌍둥이 신경망은 두 그림을 나란히 넣고 같은 출신인지 판정한다. 연구팀은 이 모델의 이름을 파피에스넷이라고 붙였다. 모델은 파피루스 파편 약 500점으로 배웠다. 글자가 아니라 종이 결만 보고 짝을 골랐다. 실제 자료에서 짝의 79퍼센트를 맞혔다[2].

양피지에도 비슷한 지문이 있다. 가죽에는 털이 났던 자리인 털구멍 자국이 남는다. 같은 짐승의 가죽에서 나온 파편은 이 자국의 배열이 이어진다. 벌레가 두루마리를 뚫고 지나간 구멍도 단서다. 말려 있던 두루마리를 펴면 벌레 구멍이 줄지어 되풀이된다. 두루마리는 안쪽 겹이 짧고 바깥 겹이 길다. 그래서 구멍 사이 간격이 바깥으로 갈수록 조금씩 넓어진다. 이 간격을 재면 파편이 몇 번째 겹이었는지 짐작할 수 있다.

이런 물리 지문을 다루려면 사진부터 정리해야 한다. 이스라엘 유물청의 사해문서 사진에는 자와 색상표와 번호판이 함께 찍혀 있다. 배경은 검은색이라 잉크와 헷갈린다. 브론슨 브라운더보스트 연구팀은 이 방해물을 먼저 지우는 네 단계 처리 방법을 만들었다. 연구팀은 정답이 붙은 자료 묶음도 함께 공개했다[3].

정답이 붙은 자료 묶음을 공개하는 일은 뜻이 크다. 다른 연구자가 같은 자료로 자기 방법을 시험할 수 있다. 두 방법 가운데 어느 쪽이 나은지도 숫자로 견줄 수 있다. 고문서 연구가 이렇게 서로 검증하는 분야로 바뀌는 중이다.

점토판과 뼈와 종이로 번진 기술

같은 생각이 다른 유물에도 옮겨 갔다. 독일 뮌헨 루트비히막시밀리안대학교에 엔리케 히메네스 교수가 있다. 그는 전자 바빌로니아 도서관(eBL) 사업을 이끈다. 이 사업의 파편 맞추기 도구가 프래그멘타리움이다[4].

메소포타미아 사람들은 젖은 점토에 갈대 펜으로 쐐기문자를 눌러 썼다. 이 점토판은 깨지기 쉽다. 연구팀은 대영박물관과 이라크박물관에서 점토판 사진을 찍었다. 2018

년부터 파편 22,000점가량을 자료로 다루었다[5]. 컴퓨터는 파편에 남은 글자를 옮겨 적은 문자열을 서로 견준다. 이때 생물학에서 유전자 배열을 견줄 때 쓰는 방법을 빌려 왔다. 옛 서기들이 지역마다 철자를 다르게 썼기 때문이다[4].

성과는 숫자로 나타났다. 이 사업은 5년 만에 새 결합 약 1,500건을 찾았다. 앞선 150년 동안 학자들이 찾은 결합은 약 5,000건이었다[4].

점토판에는 양피지에 없는 단서가 하나 더 있다. 부러진 면의 입체 모양이다. 점토판은 대개 햇볕에 말렸고 일부만 불에 구웠다. 굳는 동안 속이 고르지 않게 단단해진다. 그래서 깨진 면마다 다른 굴곡이 생긴다. 이 면을 입체로 재어 두면 맞물리는 짝을 찾을 수 있다.

중국에서는 갑골 파편에 같은 기술을 쓴다. 갑골은 상나라 사람들이 점을 치면서 거북 배딱지와 소뼈에 새긴 기록이다. 허난대학교의 장충성 연구팀은 갑골 결합 자료 묶음을 만들었다. 연구팀의 결합 알고리즘은 상위 10위 안 정확도 98.39퍼센트를 얻었다[6]. 장잔 연구팀도 갑골 파편 사진으로 딥 리조이닝 모델을 훈련했다. 이 연구팀 역시 자료 묶음을 함께 공개했다[7].

상위 10위 안 정확도라는 말은 풀어서 볼 필요가 있다. 기계는 정답 하나를 내놓지 않는다. 가능성이 높은 후보 열 개를 순서대로 내놓는다. 그 열 개 안에 정답이 들어 있을 확률이 98.39퍼센트다[6]. 마지막 판단은 사람이 한다.

자료를 한곳에 모으는 일도 함께 이루어졌다. 안양사범대학교는 은계문연이라는 갑골 자료 플랫폼을 운영한다. 2023년 4월 기준으로 갑골 사진 23만여 장이 올라 있다. 갑골문 관련 저작 152부와 논문 3,000여 편도 함께 모았다[8].

돈황 막고굴에서 나온 불경 파편도 나라마다 흩어져 있다. 천밍쿤 연구팀은 2026년에 새 방법을 내놓았다. 이 방법은 사람이 정답을 달아 주지 않아도 스스로 배운다. 이런 학습 방식을 자기지도 학습이라고 부른다. 연구팀은 온전한 사본과 짝이 밝혀진 파편을 함께 모았다. 짝을 모르는 외톨이 파편도 자료에 넣었다. 컴퓨터는 이 자료에서 스스로 연습 문제를 만들어 냈다. 연구팀은 테두리 방향을 함께 담는 새 부호 방식을 설계했다. 그 뒤 쌍둥이 신경망으로 두 파편의 닮은 정도를 점수로 매겼다[9].

사해문서 쪽에서는 유럽과 이스라엘 연구자들이 함께 온라인 작업대를 만들었다. 이름은 스크립타 쿰라니카 엘렉트로니카다. 이 작업대에서 연구자는 파편 사진을 화면 위에서 옮기고 돌려 가며 배치한다. 배치한 결과는 다른 연구자와 나눌 수 있다[10]. 파편 맞추기가 한 사람의 머릿속 작업에서 여럿이 함께 검토하는 작업으로 바뀌었다.

참고 자료

[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, 5월 25일). AI is helping us read ancient Mesopotamian literature. The Conversation. https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091 (2026년 9월 3일 접속)

[5] Phys.org. (2023, 2월 2일). 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 (2026년 9월 3일 접속)

[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, 4월 8일). "殷契文渊"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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

3. 과학 수사대의 추적 2: 위작 탐정 AI

가짜 16점을 잡아낸 방법

성서박물관이 사들인 사해문서 파편 16점은 모두 가짜였다[1][2]. 조사팀은 눈으로 글자만 본 것이 아니다. 재료와 잉크를 물리와 화학의 방법으로 뜯어보았다. 이 감정에 인공지능이 쓰이지는 않았다. 사람이 현미경과 성분 분석으로 잡아냈다[1].

성서박물관의 감정은 한 번에 끝나지 않았다. 2018년 10월에 독일 연방재료시험연구소가 다섯 점을 먼저 조사했다. 이 다섯 점은 고대 유물의 특징과 맞지 않아 전시에서 빠졌다[5]. 그 뒤 남은 것까지 모두 조사한 결과가 2020년에 나왔다[1][2].

첫 번째 증거는 바탕 재료였다. 진짜 사해문서는 얇게 편 양피지에 쓰였다. 성서박물관 파편 가운데 적어도 15점은 가죽이었다. 이 가죽은 양피지보다 두껍고 결이 거칠다[1].

두 번째 증거는 잉크가 놓인 자리였다. 조사팀은 3차원 현미경으로 표면을 들여다보았다. 갈라진 틈 속에 잉크가 고여 있었다. 찢어진 가장자리에서는 잉크가 아래로 흘러내린 자국이 보였다[1].

이 자국이 왜 결정적인지는 순서를 생각하면 알 수 있다. 진짜 사본은 평평한 양피지에 먼저 글씨를 썼다. 그 뒤 오랜 세월 동안 가죽이 마르면서 갈라졌다. 그래서 진짜는 갈라진 틈에서 글자 획이 끊긴다. 가짜는 순서가 거꾸로다. 위조범은 이미 갈라진 옛 가죽을 구해서 그 위에 잉크를 얹었다. 잉크는 벌어진 틈 속으로 흘러 들어갔다[1].

세 번째 증거는 표면에 발린 물질이었다. 조사팀은 파편이 호박빛 액체에 담겼던 흔적을 찾았다. 이 액체는 짐승 가죽으로 만든 아교로 보인다. 위조범은 진짜 사해문서 표면의 반들거리는 느낌을 흉내 내려 한 것이다[1].

세 가지 증거는 서로 다른 방향에서 같은 답을 가리켰다. 하나만으로는 우연일 수 있다. 셋이 겹치면 우연이 아니다. 유물 감정에서 서로 다른 종류의 증거를 함께 모으는 이유가 여기에 있다.

손이 남기는 흔적

가짜를 만드는 사람은 진짜 사진을 보고 글자를 따라 그린다. 이 따라 그리기가 흔적을 남긴다.

사람이 익숙한 글씨를 쓸 때는 손이 멈추지 않고 흐른다. 낯선 글자 모양을 베낄 때는 다르다. 획을 그리다가 멈추고 다음 획의 위치를 확인한다. 붓끝의 속도가 떨어지고 선이 미세하게 떨린다. 이 차이는 확대해서 보면 획의 바깥선 모양으로 나타난다.

컴퓨터는 이런 차이를 수치로 잡아낸다. 2026년 6월에 영국 브래드퍼드대학교의 하산 우가일 연구팀이 관련 연구를 발표했다. 연구팀은 옛 소묘 작품이 진짜인지 판정하는 인공지능을 만들었다. 이 모델은 그림에서 뽑은 숫자 다섯 개만 본다. 무늬가 되풀이되는 정도와 밝기 값이 흩어진 정도를 잰다. 밝은 곳과 어두운 곳의 차이도 잰다. 질감이 고른 정도와 선이 잔가지처럼 갈라진 정도를 더한다[3].

연구팀은 화가 열 명의 진품 소묘로 모델을 가르쳤다. 작품의 연대는 1510년 무렵부터 1917년까지 걸쳐 있다. 자료는 공개 자료를 내놓는 미술관 여섯 곳에서 모았다. 메트로폴리탄 미술관과 애슈몰린 박물관이 여기에 든다[3].

판정 900건으로 시험한 결과가 논문에 실렸다. 균형 정확도는 87.6퍼센트였다. 진품을 진품으로 받아들인 비율은 77.8퍼센트였다. 남의 작품을 그 화가의 것으로 잘못 받아들인 비율은 2.6퍼센트였다[3].

이 연구에는 눈여겨볼 결과가 하나 더 있다. 널리 쓰이는 큰 딥러닝 모델은 이 일에서 오히려 성적이 나빴다. 큰 모델은 진짜 작품을 가짜로 밀어내는 일이 잦았다. 자료가 적을 때는 문제에 맞게 설계한 작은 모델이 나았다[3].

이 결과는 고문서 연구에도 참고가 된다. 남아 있는 고대 사본은 수가 적다. 큰 모델을 배불리 먹일 만한 자료가 없다. 무엇을 볼지 미리 정해 주는 설계가 오히려 낫다는 뜻이다.

글씨체를 재는 기계

같은 기술은 진짜 유물을 연구할 때도 쓰인다. 네덜란드 흐로닝언대학교의 믈라덴 포포비치 연구팀은 사해문서 글씨체를 인공지능으로 분석했다. 연구팀은 신경망으로 잉크가 지나간 자국만 뽑아냈다. 그 뒤 글자의 굽은 정도와 기울기를 수치로 바꾸었다[4].

연구팀은 탄소 연대 측정을 마친 사본 24점으로 모델을 훈련했다. 모델의 이름은 에녹이다. 에녹은 연대를 모르는 사본 135점의 나이를 추정했다. 평균 오차는 약 28년에서 31년이었다. 여러 사본이 지금까지 알려진 것보다 오래된 것으로 나왔다[4].

이 기술은 진위 판정에도 도움이 된다. 진짜 서기의 글씨는 시대마다 일정한 통계 특징을 지닌다. 가짜는 이 분포에서 벗어난다. 사람의 눈은 두 글씨가 닮았다고 느끼지만 기계는 차이를 숫자로 내놓는다.

다만 이 도움은 참고 자료를 하나 더 얻는 수준이다. 에녹 연구에는 조심할 점이 있다. 이 모델은 연대를 맞히도록 훈련되었다. 진짜와 가짜를 가리도록 훈련된 것은 아니다. 연대 추정이 이상하게 나온다고 해서 곧바로 가짜라고 말할 수는 없다. 기계의 결과는 그 기계가 배운 일의 범위 안에서만 뜻을 지닌다.

기계가 대신하지 못하는 것

기계의 판정에도 한계가 있다. 성서박물관 파편의 진위를 최종 판정한 것은 인공지능 하나가 아니었다. 현미경 관찰과 화학 시험과 문헌학 지식이 함께 쓰였다[1][2].

출처 기록의 무게는 그대로 남는다. 출처 기록은 유물이 어디서 나왔는지 적은 기록이다. 롤스턴 교수는 발굴 기록이 있는 유물과 시장에서 산 유물을 나누어 보라고 말한다. 발굴로 나온 두루마리에는 어느 동굴에서 언제 나왔는지가 남아 있다. 시장에 나온 파편에는 그 기록이 없다[5].

기계는 사람이 놓치는 미세한 신호를 읽어 준다. 사람은 그 신호가 무엇을 뜻하는지를 판단한다. 위작을 가려내는 일은 이 두 가지가 함께 있어야 끝난다.

참고 자료

[1] Greshko, M. (2020, 3월 13일). "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 (2026년 9월 3일 접속)

[2] Solly, M. (2020, 3월 16일). 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/ (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

4. 밝혀진 진실

확정된 결합과 후보 결합을 가르는 선

컴퓨터가 두 파편이 붙는다고 말해도 그것으로 끝이 아니다. 기계는 확률을 내놓을 뿐이다. 연구자들은 확정된 결합과 후보 결합을 엄격히 나눈다.

확정된 결합은 여러 종류의 증거가 함께 맞을 때만 인정한다. 먼저 부러진 면의 모양이 빈틈없이 맞물려야 한다. 다음으로 재료의 결이 이어져야 한다. 파피루스는 줄기 무늬가 이어져야 한다[1]. 양피지는 털구멍 자국이 이어져야 한다. 그다음으로 경계를 넘어가는 글자 획이 자연스럽게 이어져야 한다. 마지막으로 이어 붙인 글이 그 언어의 문법과 맞아야 한다.

후보 결합은 이 조건 가운데 일부만 맞는 경우다. 벌레 구멍의 간격으로 파편의 위치를 짐작한 결과가 여기에 든다. 잉크의 성분이 같아 한 두루마리 무리로 묶은 결과도 후보에 든다. 이런 결과는 연구의 출발점이 되지만 결론이 되지는 않는다. 컴퓨터가 걸러 준 후보를 사람이 확인하는 순서다[1].

이 구분을 지키지 않으면 위험하다. 확률이 높다는 이유로 두 파편을 붙이면 없던 문장이 생긴다. 그 문장이 인용되면 잘못은 다음 연구로 번진다. 그래서 자료 기록에는 확정과 후보를 따로 적어 둔다.

기계가 내놓는 숫자도 이 구분을 전제로 한다. 걸러 내기 비율이 높다고 결합이 확정되지는 않는다[1]. 상위 후보 안에 정답이 들어갈 확률도 마찬가지다[2]. 어느 쪽이든 마지막 판정은 사람이 한다.

되살아난 글

이 방식으로 실제 문헌이 되살아났다. 프래그멘타리움은 2018년부터 점토판 파편 22,000점가량을 다루었다. 이 프로그램은 새 결합과 새 사본을 여럿 찾아냈다. 연구팀은 2023년 2월에 프래그멘타리움을 일반에 열었다. 길가메시 서사시의 디지털 판본도 함께 내놓았다[3].

되살아난 글 가운데 길가메시 서사시가 있다. 길가메시 서사시는 메소포타미아의 영웅 이야기다. 2022년 11월에 프로그램이 파편 하나를 알아보았다. 이 파편은 기원전 130

년 무렵에 베껴 쓴 판본에서 나왔다. 지금까지 확인된 길가메시 사본 가운데 시기가 늦다. 그 시대까지 사람들이 이 이야기를 베껴 썼다는 뜻이다[3].

쐐기문자는 지역마다 철자가 달랐다. 서기들이 자기 사투리에 맞추어 글자를 바꾸어 썼기 때문이다. 사람이 눈으로 찾을 때는 이 차이가 큰 벽이었다. 연구팀은 유전자 배열을 견주는 방법을 빌려 이 벽을 넘었다. 조금 다른 문자열끼리도 견줄 수 있게 되었다[4].

사해문서 쪽에서도 자료 기반이 넓어졌다. 이스라엘 유물청은 파편 25,000점이 넘는 유물을 고해상도 사진으로 찍어 공개했다[5]. 연구자들은 이 사진 위에서 파편을 옮겨 가며 배치한다[6]. 자와 색상표 같은 방해물을 자동으로 지우는 처리 방법도 공개되었다[7].

2026년에 열린 다음 문

2026년 6월에 시작한 "서기와 두루마리 추적" 연구가 다음 문을 연다. 이 연구가 던지는 물음은 파편 맞추기보다 한 걸음 앞이다. 사해문서를 어디에서 누가 썼는가를 묻는다[8].

연구팀은 양피지와 파피루스와 잉크 시료 약 250점을 조사한다. 시료란 검사에 쓰는 작은 표본을 말한다. 여기에 인공지능 글씨체 분석과 옛 글씨 연구를 겹친다. 이스라엘 유물청과 피사대학교와 나폴리대학교가 함께한다. 남덴마크대학교와 벨기에 뢰번대학교도 참여한다. 베를린과 토리노의 박물관이 소장한 이집트 파피루스와도 견준다[8].

시료의 성분이 같으면 같은 공방에서 나왔을 가능성이 커진다. 이 정보는 파편 맞추기에도 쓰인다. 재료의 출신이 같은 것끼리 먼저 묶으면 견줄 짝의 수가 줄어든다.

남은 파편과 남은 일

성과가 쌓였어도 아직 제자리를 찾지 못한 파편이 많다. 유물청의 보관 상자에는 어느 두루마리에서 나왔는지 모르는 파편이 남아 있다[8]. 점토판 쪽에도 아직 찾지 못한 결합이 많다[4].

장비의 차이도 문제로 남아 있다. 입체 스캐너로 유물을 재려면 유물을 장비 앞으로 옮겨야 한다. 유물을 밖으로 내보내기 어려운 박물관도 많다. 그래서 어떤 유물은 사진 자료만 있고 입체 자료가 없다. 이런 유물은 부러진 면을 견주는 방법을 쓸 수 없다.

기술이 남긴 숙제도 있다. 파편 사진은 나라마다 다른 장비로 찍혔다. 조명과 배경이 다르면 같은 알고리즘도 다른 답을 낸다. 그래서 연구자들은 정답이 붙은 공개 자료 묶음을 만들었다. 서로의 결과를 같은 잣대로 견주기 위해서다[7]. 돈황 파편 연구도 사람이 정답을 달지 않아도 배우는 쪽으로 방향을 잡았다[9].

이 기술이 사람을 밀어내는 것은 아니다. 기계는 견줄 짝을 줄이고 사람은 뜻을 판단한다. 되살아난 문장이 그 시대의 말투와 맞는지는 사람이 읽어야 안다. 두 역할이 겹치지 않기 때문에 함께 일할 수 있다.

공개도 달라졌다. 예전에는 소수의 학자만 원본 사진을 보았다. 지금은 점토판 자료와 사해문서 사진이 인터넷에 열려 있다[4][5]. 학생도 화면에서 파편을 옮겨 볼 수 있다[6].

디지털 접착제는 파편을 물리적으로 붙이지 않는다. 유물은 상자 안에 그대로 있다. 컴퓨터 화면 위에서만 두루마리가 다시 이어진다. 그래서 판정이 틀리면 언제든 되돌릴 수 있다. 2,000년 동안 흩어져 있던 글이 이렇게 조금씩 문장으로 돌아오고 있다.

참고 자료

[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, 2월 2일). 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 (2026년 9월 3일 접속)

[4] 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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[6] Scripta Qumranica Electronica. (n.d.). Scripta Qumranica Electronica. University of Haifa. https://megillot.haifa.ac.il/index.php/en/research/scripta-qumranica-electronica (2026년 9월 3일 접속)

[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, 6월 30일). 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

(2026년 9월 3일 접속)

[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

제 3 부

진흙 점토판의 대화를 읽어내다: 전산 설형문자학

05 CHAPTER

제5장. 점토판의 부러진 단면을 맞추다: 길가메시 서사시의 전산학적 재건

부러진 점토판 단면을 이어 붙여 되찾은 길가메시 서사시

1. 사건 파일

4천 년을 건너온 이야기

길가메시 서사시는 지금 남아 있는 이야기 문학 가운데 나이가 많은 축에 든다. 이보다 이른 수메르 노래들도 전해진다. 주인공은 메소포타미아의 도시 우루크를 다스린 왕 길가메시다[1]. 그는 친구 엔키두와 함께 먼 삼나무 숲으로 간다. 두 사람은 그 숲을 지키는 괴물 훔바바와 싸운다[2]. 뒤에 엔키두가 죽자 길가메시는 죽지 않는 길을 찾아 떠난다[1].

이 이야기를 적은 바탕은 종이가 아닌 진흙판이다. 고대의 서기들은 갈대를 비스듬히 깎아 뾰족한 막대를 만들었다. 이 막대를 첨필이라고 부른다. 첨필로 젖은 진흙을 눌러 찍으면 쐐기 모양의 자국이 남는다. 선을 긋는 것이 아니라 찍어서 만든 글자다. 진흙이 마르면 그 자국은 그대로 굳는다. 그래서 이 글자를 쐐기문자라고 부른다. 서기가 되려는 학생들은 글씨 연습으로 이 서사시를 베껴 썼다. 그 덕분에 사본 파편이 중동 여러 유적에서 나온다. 이집트의 아마르나 유적에서도 이 서사시의 파편이 나왔다. 메소포타미아에서 멀리 떨어진 곳까지 이야기가 퍼져 있었다는 뜻이다[1].

전해지는 판본은 크게 두 가지다. 하나는 기원전 2천년기 초에 적힌 고바빌로니아 판본이다. 다른 하나는 뒤에 정리된 표준 바빌로니아 판본이다. 표준 판본은 열두 장의 점토판으로 짜여 있다. 그 가운데 열한 장이 본편이다. 열두 번째 판은 뒤에 덧붙은 부록으로 본다. 오늘 우리가 읽는 번역본은 이 표준 판본을 바탕으로 삼는다. 빈 곳은 고바빌로니아 판본과 다른 파편으로 메운다. 표준 판본을 다듬은 사람은 신-레키-운닌니라는 서기로 알려져 있다. 그는 흩어져 있던 옛 노래들을 하나의 긴 작품으로 엮었다. 판마다 담을 내용을 정하고 줄 수를 맞추었다. 그런데 이 표준 판본에도 빈 곳이 아주 많다. 처음부

터 끝까지 온전하게 남은 사본은 하나도 없다[1].

이야기의 내용은 지금 읽어도 낯설지 않다. 길가메시는 힘이 세지만 백성을 괴롭히는 왕이었다. 신들은 그를 상대할 사람으로 엔키두를 만들어 보낸다. 둘은 싸우다가 친구가 된다. 함께 훔바바를 물리치고 하늘의 황소도 쓰러뜨린다. 그 대가로 엔키두가 병들어 죽는다. 친구의 죽음을 본 길가메시는 자기도 죽는다는 사실을 깨닫는다. 제11판에는 큰 홍수에서 살아남은 사람을 찾아가 이야기를 듣는 장면이 들어 있다[1].

이 이야기는 2천 년 넘게 잊혀 있었다. 쐐기문자를 읽을 사람이 사라졌기 때문이다. 19세기에 학자들이 쐐기문자를 다시 해독하면서 서사시가 세상에 돌아왔다. 니네베에서 실어 온 점토판 상자를 뒤지던 연구자가 홍수 이야기를 알아본 것이 시작이었다. 그 뒤로 150년 동안 학자들은 흩어진 사본을 모아 본문을 조금씩 채워 왔다[1].

니네베와 기원전 612년의 불

표준 판본의 사본은 대부분 한 고대 도서관에서 나왔다. 기원전 7세기 아시리아의 왕 앗슈르바니팔이 니네베 왕궁에 세운 장서고다. 이 도서관은 점술서, 의학서, 사전, 문학 작품을 두루 모았다. 왕은 제국 곳곳에서 점토판을 거두어들이게 했다. 신하들에게 편지를 보내 없는 책을 찾아 오라고 시켰다. 앗슈르바니팔은 쐐기문자를 직접 읽을 줄 아는 왕이었다[3].

니네베는 기원전 612년 무렵 큰 불에 휩싸였다. 왕궁이 무너지고 도서관도 함께 불탔다. 이 불은 유물에 두 가지 일을 한꺼번에 했다. 그늘에서 말리기만 했던 진흙판이 불길 속에서 단단하게 구워졌다[4]. 구워진 점토판은 가마에서 나온 벽돌처럼 오래 견디는 몸을 얻었다. 같은 불길 속에 파피루스나 가죽이 있었다면 재가 되어 사라졌을 것이다.

그러나 무너지는 건물의 무게는 그 단단한 판을 산산조각 냈다. 지금 대영박물관이 보관한 니네베 출토 점토판과 파편은 3만 점을 넘는다[3][4]. 19세기 중반 발굴단이 흙더미에서 거두어 배에 실어 온 유물이다[3]. 하나의 이야기가 수십 개의 파편으로 흩어진 상태로 우리에게 왔다.

파편의 크기는 제각각이다. 손바닥만 한 것도 있고 엄지손톱만 한 것도 있다. 한 파편에 남은 글은 몇 줄에 지나지 않는 경우가 많다. 그 몇 줄이 서사시 열두 판 가운데 어디에 들어가는지 아무도 알려 주지 않는다. 글을 읽어 내는 일과 위치를 찾아 주는 일은 서로 다른 작업이다.

파편 하나가 제자리를 찾으면 뜻이 달라지기도 한다. 앞뒤가 이어져야 누가 한 말인지, 어느 장면인지 알 수 있다. 떨어져 있을 때는 뜻 모를 낱말의 나열이던 글이 문장이 된다. 접합 한 건이 이야기 한 장면을 통째로 되살리기도 한다.

35만 줄을 손으로 맞출 수는 없다

깨진 파편을 맞추는 일은 오랫동안 사람의 손과 기억에 맡겨졌다. 학자는 수장고에서 파편 하나를 집어 든다. 그리고 다른 파편들의 사진과 손 모사도를 떠올리며 짝을 찾는다. 두 파편이 같은 판에서 나왔다고 판단하면 그것을 접합이라고 부른다.

이 방식은 사람의 기억 용량에서 막힌다. 아시리아학이 시작된 뒤 150년 동안 이렇게 이어 붙인 접합은 약 5,000건이다[5]. 남은 파편의 수는 그보다 훨씬 많다. 대영박물관은 니네베에서 나온 점토판만 3만 점 넘게 가지고 있다[3]. 그 밖의 소장품까지 더하면 수는 훨씬 늘어난다. 그 가운데 많은 수는 고대 이후로 읽힌 적이 없다[5]. 세계 박물관에 남은 쐐기문자 점토판을 모두 세면 50만 점쯤으로 본다[6].

일이 어려운 까닭이 하나 더 있다. 파편들이 여러 나라에 흩어져 있다. 대영박물관, 베를린, 예일 대학교, 필라델피아의 펜 박물관이 저마다 소장품을 가진다[7]. 두 파편을 직접 맞춰 보려면 두 대륙을 오가야 한다.

쐐기문자 자체도 손이 많이 간다. 같은 낱말을 서기마다 조금씩 다르게 적었다. 지역과 시대에 따라 철자가 흔들린다[5]. 사람이 눈으로 두 파편의 글을 비교해도 같은 작품인지 알아채기 어렵다.

게다가 쐐기 기호 하나가 여러 소리로 읽힌다. 같은 자국이 문맥에 따라 다른 값을 가진다. 이런 글자를 수십만 줄 놓고 짝을 찾는 일은 사람의 눈으로 끝낼 수 없다.

그래서 독일 뮌헨의 루트비히 막시밀리안 대학교에 있는 엔리케 히메네스 교수가 다른 길을 골랐다. 2018년에 그는 바빌로니아 문헌을 되도록 온전하게 되살리는 사업을 시작했다. 세계에 흩어진 점토판 판독문을 한곳에 모으는 일이다[5]. 이 사업의 이름이 전자 바빌로니아 문헌 도서관, 줄여서 eBL이다. eBL이 만든 파편 맞추기 작업실이 프래그멘타리움이다. 2024년까지 이 작업실에는 약 2만 5천 점의 점토판 판독문이 쌓였다. 줄 수로 세면 35만 줄이 넘는다[7].

이 일이 급한 까닭은 시간에 있다. 점토판은 불에 구워졌어도 영원하지 않다. 소금기와 습기가 표면을 부풀리고 글자를 지운다. 박물관 수장고에서 읽히기를 기다리는 점토판이 지금도 수십만 점이다[5]. 그 가운데 얼마나 많은 이야기가 들어 있는지는 아무도 모른다.

쐐기문자를 읽는 사람의 수도 문제다. 세계에 아시리아학 연구자는 많지 않다. 남은 점토판 가운데 번역된 것은 일부에 그친다[6]. 한 사람이 평생 읽을 수 있는 점토판에는 한계가 있다. 기계의 도움을 찾은 까닭이 여기에 있다.

참고 자료

[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 Gilgameš. 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 (2026년 9월 3일 접속)

[4] British Museum. (n.d.). What was Ashurbanipal's Library? britishmuseum.org. https://www.britishmuseum.org/research/projects/what-was-ashurbanipals-library (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[7] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Hätinen, 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. 과학 수사대의 추적: 클레이 엣지 디지털 융합과 Fragmentarium

글자를 알파벳으로 옮겨 적는 첫 단계

프래그멘타리움의 출발점은 글자다. 연구자는 점토판 사진을 보고 쐐기문자를 로마자로 옮겨 적는다[1]. 이 작업을 음역이라고 한다. 우리말로 치면 한자를 한글 소리로 옮겨 적는 일과 비슷하다. 쐐기 기호 하나하나를 소리값에 맞춰 알파벳으로 바꾸는 일이

다. 깨져서 읽을 수 없는 곳은 빈칸 기호로 표시해 둔다.

이렇게 만든 음역문은 컴퓨터가 다룰 수 있는 문자열이 된다. eBL은 2018년 5월부터 2023년 8월까지 이 음역문을 쌓았다. 그 결과 약 2만 5천 점의 점토판, 35만 줄이 넘는 글이 검색 가능한 자료가 되었다. eBL이 관리하는 점토판 기록 목록은 26만 건이 넘는다. 대영박물관, 예일 바빌로니아 컬렉션, 펜 박물관의 소장품이 여기에 들어 있다[1].

음역은 사람이 한다. 기계가 사진에서 글자를 완벽하게 뽑아내지는 못하기 때문이다. eBL 연구진은 다섯 해 넘게 이 음역 작업을 이어 갔다. 그 결과물은 누구나 내려받을 수 있는 공개 자료로 나왔다[1]. 기계에게 일을 시키려면 먼저 사람이 자료를 만들어 주어야 한다.

유전자 검사를 닮은 문자열 맞추기

핵심 알고리즘은 뜻밖의 분야에서 왔다. 생물학자들이 유전자 서열을 비교할 때 쓰는 문자열 정렬 방법이다[2]. 유전자는 A, T, G, C 네 글자가 길게 이어진 사슬이다. 두 생물의 유전자를 비교할 때는 글자가 조금 달라도 같은 위치로 맞춰 놓고 본다.

쐐기문자에서도 같은 문제가 생긴다. 서기마다 같은 낱말을 다르게 적었기 때문이다. 히메네스 교수는 이 흔들림이 유전자의 돌연변이와 닮았다고 보았다. 그래서 생물학의 정렬 알고리즘을 쐐기문자에 맞게 고쳐 썼다[2].

컴퓨터가 실제로 세는 것은 짧은 글자 토막이다. 음역문을 몇 글자씩 잘라 낸 토막을 엔그램이라고 부른다. 시스템은 두 점토판이 공유하는 엔그램의 개수를 센다. 긴 토막을 함께 가진 짝일수록 더 높은 점수를 받는다. 점수가 높은 짝이 사람 앞에 후보로 올라온다.

프래그멘타리움은 웹 화면으로 열린다. 세계 어느 나라의 연구자든 접속해서 같은 자료를 본다. 런던의 파편과 뮌헨의 파편을 한 화면에 나란히 놓고 비교할 수 있다[8]. 짐을 싸서 비행기를 탈 필요가 없어졌다.

이 방식은 철자가 흔들려도 견딘다. 낱말 하나가 다르게 적혀 있어도 주변 토막이 겹치면 점수가 남는다. eBL은 이 짝짓기 프로그램을 공개 저장소에 올려 두었다. 다른 연구자도 같은 도구를 내려받아 쓸 수 있다[6].

컴퓨터는 글자 읽기 자체도 돕는다. 쐐기 기호 하나가 여러 소리로 읽히는 일이 잦다. eBL이 학습시킨 모델은 이 여러 뜻 가운데 맞는 것을 고르는 문제에서 98퍼센트의 정확도를 냈다[2].

이 일이 왜 필요한지는 예를 들면 쉽다. 같은 쐐기 자국을 어떤 문장에서는 하나의 소리로 읽는다. 다른 문장에서는 낱말 전체를 뜻하는 기호로 읽는다. 잘못 읽으면 짝짓기 점수가 흐트러진다. 읽기를 바로잡는 일이 곧 접합의 정확도를 높이는 일이다.

성과는 숫자로 나타났다. 아시리아학이 150년 동안 손으로 이어 붙인 접합은 약 5,000건이다. eBL은 다섯 해 만에 약 1,500건의 접합을 더 찾았다[2]. 음역문 자료집을 만드는 과정에서만 약 1,250건의 접합이 드러났다[1]. 직접 맞닿지는 않지만 같은 작품에 속한 파편도 수천 점 확인되었다[2].

깨진 곳을 채우는 언어 모델

파편을 이어 붙여도 글은 여전히 군데군데 비어 있다. 이 빈 곳을 라쿠나라고 부른다. 여기서부터는 언어 모델이 나선다.

2023년에 텔아비브 대학교와 아리엘 대학교 연구진이 아카드어 번역 모델을 내놓았다. 이 모델은 쐐기문자 부호를 넣어도, 알파벳 음역문을 넣어도 영어로 옮긴다. 두 가지 입력을 모두 다룬 것이다. 영어를 아카드어로 되돌리는 일은 다루지 않았다. 학습에 쓴 자료는 이미 출판된 점토판 판독문이다[3].

2026년에는 반대 방향까지 다루는 도구가 나왔다. 태블릿크래프트라는 이름의 공개 소프트웨어다. 이 도구는 아카드어에서 영어로, 영어에서 아카드어로 양쪽 번역을 한다. 학습 자료는 11만 6천 쌍의 문장이다. 쐐기 기호 변환표는 1만 4천 개가 넘는 대응을 담았다. 개발자는 남은 50만 점의 점토판을 읽을 도구가 필요하다고 밝혔다[4].

이런 모델은 빈칸의 앞뒤 문맥을 보고 들어갈 만한 낱말을 확률로 내놓는다. 바빌로니아 시에는 같은 구조의 두 줄을 나란히 놓는 규칙이 있다. 모델은 이런 형식도 단서로 삼는다. 모델이 내놓는 것은 확정된 답이 아닌 후보 목록이다.

글씨체까지 읽는 2026년의 도구

2026년 5월 18일, 독일 뷔르츠부르크 대학교가 새 도구를 공개했다. 마인츠 학술원과 도르트문트 공과대학교가 함께 만들었다. 이름은 팔레오그라피쿰이다. 이 도구는 점토

판 사진에서 쐐기 기호 하나하나의 생김새를 배운다. 그리고 같은 방식으로 눌러 쓴 기호를 전체 소장품에서 찾아 준다[5].

이 도구가 다루는 사진은 7만 장이고, 담긴 기호는 500만 개가 넘는다. 다섯 파편의 글씨체를 비교하는 일에 예전에는 사흘이 걸렸다. 이 도구를 쓰면 5분이면 끝난다. 사람마다 글씨체가 다르듯 서기마다 쐐기를 누르는 각도와 간격이 다르다. 그 차이를 실마리로 같은 판에서 나온 파편을 골라낼 수 있다[5].

이 도구가 다루는 히타이트 쐐기문자는 기호가 375개다. 지금까지 알려진 히타이트 점토판 파편은 3만 점에 이른다. 연구를 이끈 다니엘 슈베머 교수는 이 도구로 수천 시간을 아낄 수 있다고 밝혔다[5].

연구진의 다음 목표는 서기 개인을 알아보는 일이다. 같은 사람이 쓴 점토판을 모두 찾아내겠다는 뜻이다. 그 일이 되면 서기 한 사람이 어디서 일했고 어디로 옮겨 갔는지 그릴 수 있다. 3,500년 전 사람들의 직업 생활이 자료로 드러나는 것이다[5].

부러진 진흙 모서리를 맞추는 다른 길

글자가 아닌 모양으로 파편을 맞추려는 연구도 오래 이어져 왔다. 깨진 점토판에는 부러진 면이 남는다. 이 면을 파단면이라고 한다. 두 파편의 파단면이 서로 맞물리면 같은 판에서 나온 것이다.

이 길을 걸어온 대표 사업이 가상 쐐기문자 점토판 복원 사업이다. 영국 킬 대학교의 산드라 울리가 이 사업을 이끈다. 팀 콜린스와 에를렌트 겔켄이 함께한다. 연구진은 파편을 3차원 스캐너로 훑는다. 그리고 두 파편의 부러진 면이 맞물리는지 컴퓨터가 판단하게 한다. 오래된 파편은 모서리가 닳아 있다. 이 닳음까지 계산에 넣는다[9].

이 방법의 성과 하나가 대홍수 이야기에서 나왔다. 아트라하시스라는 홍수 서사시의 제3판 파편이 그것이다. 한 파편은 런던에, 다른 파편은 제네바에 있었다. 두 곳은 1,000킬로미터쯤 떨어져 있다. 실물을 맞대 볼 수 없어 접합을 확인하지 못하던 짝이었다. 3차원 자료를 겹쳐 보자 두 파편이 이어졌다[9]. 이 방법은 2014년에 논문으로 나왔다[10].

3차원 자료를 다루는 연구는 지금도 이어진다. 2026년 3월에 한 논문이 공개되었다. 점토판의 3차원 점군 자료로 유물의 정보를 분류하는 신경망을 다룬다. 점군은 물체 표

면을 수많은 점의 좌표로 기록한 자료다. 3차원 스캐너로 점토판을 훑으면 이런 자료가 나온다. 표면의 굴곡과 두께가 숫자로 남는다. 이 연구는 앞선 트랜스포머 기반 모델보다 나은 성적을 냈다[7].

다만 지금 프래그멘타리움이 실제로 쓰는 주된 무기는 3차원 형상이 아니다. 사진과 음역문, 곧 글자와 그림이다[1][2]. 파단면 맞추기는 파편을 하나하나 스캔해야 해서 품이 많이 든다. 두 길은 서로를 대신하지 않고 함께 간다.

사진에서 기호를 찾아내는 일도 빠르게 나아갔다. 2026년 6월에 eBL 연구진이 새 논문을 냈다. 점토판 사진에서 쐐기 기호의 자리를 자동으로 찾아 주는 방법이다. 대상이 된 파편 사진은 8만 7천 장이 넘는다. 찾아낸 기호는 290만 개에 이른다. 앞선 방법보다 성적이 크게 올랐다[11].

참고 자료

[1] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Hätinen, 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, 5월 25일). AI is helping us read ancient Mesopotamian literature. The Conversation. https://theconversation.com/ai-is-helping-us-read-ancient-mesopotamian-literature-205091 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[5] 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 (2026년 9월 3일 접속)

[6] Electronic Babylonian Literature. (n.d.). ngram-matcher [소프트웨어]. GitHub. https://github.com/ElectronicBabylonianLiterature/ngram-matcher (2026년 9월 3일 접속)

[7] Hagelskjær, F. (2026). A novel network for classification of cuneiform tablet metadata. arXiv:2603.03892. https://arxiv.org/abs/2603.03892 (2026년 9월 3일 접속)

[8] electronic Babylonian Library. (n.d.). Fragmentarium. ebl.lmu.de. https://www.ebl.lmu.de/fragmentarium (2026년 9월 3일 접속)

[9] International Association for Assyriology. (2022, 10월 22일). In the spotlight: The "Virtual Cuneiform Tablet Reconstruction Project". iaassyriology.com. https://iaassyriology.com/in-the-spotlight-the-vctr-project/ (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

3. 밝혀진 진실

길가메시 서사시에 돌아온 줄들

프래그멘타리움은 길가메시 서사시의 파편 20점을 새로 찾아 주었다. 그 덕분에 100줄이 넘는 본문이 제자리로 돌아왔다. 사라졌던 장면도 몇 개 되살아났다. 엔키두가 길가메시에게 훔바바를 죽이지 말라고 말리는 장면이 그 하나다. 두 사람이 훔바바를 죽인 뒤 니푸르로 가는 장면도 새로 이어졌다. 니푸르는 신 엔릴을 모시던 종교의 중심 도시다[1].

이 방법이 얼마나 빠른지 보여 준 사례가 2025년에 나왔다. 바그다드 대학교의 안마르 파딜과 히메네스 교수가 바빌론을 찬양하는 긴 시를 되살렸다. 이 시는 1,000년 넘게 잊혀 있었다. 연구진은 eBL 플랫폼으로 같은 시에 속한 사본 30점을 찾아냈다. 손으로 했다면 수십 년이 걸렸을 일이다. 흩어진 사본을 모으자 250줄에 이르는 시의 대부분이 다시 이어졌다[3].

이 시는 바빌론이라는 도시의 하루를 보여 준다. 강물과 성문과 사람들의 일이 나온다. 여사제들이 맡았던 일에 관한 새 기록도 들어 있다. 사본이 30점이나 나왔다는 사실은 이 시가 널리 읽혔다는 뜻이다. 학교에서 학생들이 베껴 쓰던 글이었을 가능성이 크다[3].

이 작업에 걸린 시간이 방법의 힘을 보여 준다. 사본 30점을 손으로 찾으려면 세계 여러 박물관의 목록을 하나씩 뒤져야 한다. eBL 플랫폼에서는 같은 일이 검색 한 번으로 끝난다[3].

이 방법은 길가메시 서사시에만 쓰이지 않는다. 바빌로니아의 점술서, 의학서, 사전, 기도문에도 같은 방법이 들어간다. eBL이 찾은 접합은 여러 종류의 글에 걸쳐 있다[5].

사람의 손이 먼저 푼 삼나무 숲

기계가 모든 것을 푼 것은 아니다. 길가메시 서사시에서 널리 알려진 새 발견은 사람의 손에서 나왔다.

2011년에 이라크 술라이마니야 박물관이 점토판 한 점을 사들였다. 파루크 알라위(Al-Rawi)와 앤드루 조지가 이 점토판을 읽었다. 두 사람은 2014년에 결과를 발표했다. 판은 표준 판본 제5판의 앞부분과 끝부분을 담고 있었다. 여기서 약 20줄의 새 본문이 나왔다[4].

새 본문은 삼나무 숲을 소리로 그린다. 원숭이들이 시끄럽게 울고 매미가 합창한다. 온갖 새가 지저귄다. 이 소리들이 숲을 지키는 훔바바를 날마다 즐겁게 했다. 숲을 처음 본 두 사람이 놀라 멈춰 서는 장면도 나온다. 훔바바를 죽인 뒤 길가메시가 나무를 벤 일을 후회하는 구절도 있다[4]. 메소포타미아의 평원에는 원숭이가 사는 숲이 없다. 그래서 이 묘사가 어디서 왔는지가 학자들의 물음으로 남았다.

이 점토판이 나온 경로는 깨끗하지 않았다. 박물관은 유물 상인에게서 이것을 사들였다. 전쟁 뒤 이라크에서 도굴된 유물이 흩어지는 것을 막으려는 조치였다. 어느 유적에서 나왔는지는 기록이 없다[4]. 이런 유물은 파편의 짝을 찾을 단서가 더 적다.

이 발견은 인공지능이 없어도 나왔다. 그러나 앞으로 같은 종류의 발견을 얼마나 자주 할 수 있느냐는 도구에 달렸다. 술라이마니야 판이 세상에 알려지기까지는 우연과 사람의 눈이 필요했다. eBL은 그 우연을 검색으로 바꾸려 한다[5].

둥근 방주 이야기

대홍수 이야기도 사정이 비슷하다. 길가메시 서사시 제5판이 삼나무 숲을 담았다면, 제11판은 큰 홍수를 담았다[6]. 아래에서 볼 점토판은 길가메시 서사시의 사본이 아니다. 같은 홍수 전승을 담은 더 오래된 판이다[8].

대영박물관의 어빙 핑클 박사는 2009년에 한 점토판을 살펴보았다. 개인 소장자가 박물관으로 가져온 판이다. 기원전 1750년 무렵에 적힌 것으로 그는 보았다. 크기는 가로 11.5센티미터, 세로 6센티미터쯤이다. 여기에 60줄이 적혀 있다[8]. 내용은 신이 한 사람에게 배를 만들라고 지시하는 이야기다. 짐승을 둘씩 태우라는 말도 나온다[7].

눈에 띄는 것은 배의 모양이다. 이 배는 길쭉한 상자가 아니라 둥근 배다. 강에서 쓰던 코라클이라는 둥근 바구니 배와 같은 꼴이다[7]. 점토판에는 밧줄과 나무 틀이 적혀 있다. 물이 새지 않게 바를 역청의 양도 나온다. 역청은 검고 끈적한 천연 기름이다. 핑클 박사는 이 판독을 2014년에 책으로 펴냈다[8].

이 점토판은 인공지능이 읽은 것이 아니다. 사람이 한 줄씩 읽어 냈다. 인공지능의 몫은 다른 곳에 있다. 세계 곳곳에 흩어진 홍수 이야기 사본을 한곳에 모으고, 서로 겹치는 줄을 찾아 주는 일이다[5].

기계가 하는 일과 사람이 하는 일

이 분야는 결과를 세 층으로 나누어 다룬다.

첫 층은 물리적 접합이다. 두 파편의 부러진 면이 실제로 맞물리는 경우다. 이때 두 파편은 하나의 점토판에서 나왔다고 확정된다.

둘째 층은 문헌적 접합이다. 두 파편이 직접 맞닿지는 않지만 같은 작품의 서로 다른 부분임이 밝혀지는 경우다. 프래그멘타리움이 찾아 주는 대부분이 여기에 든다[5]. 같은 작품에 속한 파편이 이렇게 수천 점 확인되었다[2].

이 두 층은 성격이 다르다. 물리적 접합은 눈으로 확인되면 그대로 사실이 된다. 문헌적 접합은 확률에 기대므로 새 파편이 나오면 바뀔 수 있다. eBL은 두 층을 구분해서 기록한다[5].

셋째 층은 해석이다. 되살아난 본문이 무엇을 뜻하는지 따지는 일이다. 삼나무 숲의 원숭이 소리가 어디서 온 기억인지 기계는 알지 못한다. 둥근 배가 어떻게 성서의 네모난 방주가 되었는지도 기계는 답하지 못한다. 이 물음은 옛 글을 연구하는 학자와 역사학자의 몫이다.

기계가 내놓은 후보는 사람의 검토를 거쳐야 정식 기록이 된다. 학자는 후보 목록을 열어 두 파편의 사진을 나란히 본다. 글자의 모양과 줄 간격, 판의 두께까지 살핀 뒤에 접합을 확정한다[5]. 기계가 하는 일은 수많은 짝 가운데 볼 만한 몇 개를 골라 주는 데까지다[2].

2026년 6월에 나온 한 정리 글도 같은 선을 긋는다. 지금의 인공지능은 쐐기문자를 스스로 해독하지 못한다. 사람의 연구를 돕는 도구다[9]. 뷔르츠부르크 대학교의 팔레오

그라피쿰도 마찬가지다. 사흘 걸리던 글씨체 비교를 5분으로 줄여 줄 뿐이다. 무엇을비교할지 정하고 결과를 판단하는 것은 사람이다[10].

바뀐 것은 속도다. 150년 동안 5,000건이던 접합이 다섯 해에 1,500건 늘었다[2]. 길가메시가 찾아 헤맨 영원한 생명은 결국 그의 이야기가 얻었다. 흙 속에서 부서진 채로 2,600년을 기다린 문장들이 지금 하나씩 제자리로 돌아오고 있다.

남은 일은 아직 많다. 세계의 점토판 50만 점 가운데 읽힌 것은 일부다[11]. eBL의 음역 자료에 들어온 점토판은 2만 5천 점이다[5]. 히타이트 파편 3만 점은 이제 막 사진으로 정리되었다[10]. 부러진 면을 맞추는 3차원 작업은 이제 시작 단계다. 다음 세대의 학생이 이 목록을 이어받을 것이다. 그때 길가메시 서사시의 빈칸은 지금보다 줄어 있을 것이다. 니네베의 불이 남긴 파편은 아직 다 읽히지 않았다. 읽어 낼 도구는 해마다 늘고 있다.

참고 자료

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 Gilgameš. 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., Hätinen, 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 (2026년 9월 3일 접속)

[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/mesopota

mia/2026-06-14.html (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

06 CHAPTER

제6장. 고대 메소포타미아 영수증과 상인들의 속삭임: AI 번역

고대 메소포타미아의 아카드어를 기계가 번역하기까지

1. 사건 파일

진흙에 새긴 영수증

고대 메소포타미아 사람들은 젖은 진흙판에 글을 썼다. 갈대 끝을 각지게 깎아 철필을 만들었다. 철필은 잉크 없이 자국만 내는 필기도구다. 이 철필을 진흙에 눌렀다가 곧바로 뗐다[1]. 눌린 자국은 쐐기를 닮은 세모꼴이 되었다. 이 글자의 이름은 쐐기문자다. 학계에서는 설형문자라는 이름도 함께 쓴다. 설형은 쐐기 모양이라는 뜻이다.

진흙은 마르면 돌처럼 굳는다. 종이나 가죽과 달리 세월이 지나도 썩지 않는다. 불이 나도 재가 되지 않는다. 오히려 뜨거운 열이 진흙을 벽돌처럼 단단하게 구워 준다. 전쟁으로 도시가 불탈 때 점토판은 더 튼튼해졌다[3]. 그 덕분에 4천 년 전 장부가 오늘까지 남았다.

점토판은 손바닥만 한 것이 많다. 주머니에 넣고 다닐 만한 크기다. 서기는 젖은 진흙을 반죽해 판을 빚었다. 서기는 글을 대신 쓰고 읽어 주던 사람이다. 글을 다 쓰면 볕에 말렸다. 일부러 가마에 굽는 일은 드물었다. 지금 박물관에서 굽는 것은 보존을 위한 조치다[2]. 고칠 일이 생기면 표면을 문질러 지웠다.

지금까지 발굴된 점토판은 약 50만 장으로 추정된다[4]. 땅속에는 아직 캐내지 않은 점토판이 더 묻혀 있다. 여기 적힌 글은 왕의 자랑만이 아니다. 곡식을 몇 자루 받았다는 영수증이 있다. 은을 빌려주고 이자를 정한 계약서가 있다. 멀리 떠난 남편에게 아내가 보낸 편지도 있다[6].

영수증 한 장은 짧다. 물건 이름과 수량이 먼저 나온다. 그다음 주는 사람과 받는 사람의 이름이 온다. 끝에는 날짜와 증인의 이름을 적는다. 오늘 가게에서 받는 영수증과 순서

가 비슷하다. 이런 문서 한 장은 알맹이가 작다. 그러나 수만 장이 모이면 이야기가 달라진다. 어느 해에 곡식값이 올랐는지 드러난다. 어느 도시가 무엇을 사고팔았는지도 보인다.

번역되지 않은 대부분

문제는 이 글이 거의 읽히지 않는다는 데 있다. 점토판 가운데 현대어로 옮겨진 것은 일부뿐이다[5]. 나머지는 박물관 상자 안에서 잠자고 있다.

글을 쓰는 사람은 아무나 되지 못했다. 서기가 되려면 학교에서 오래 배웠다. 학생들은 기호를 진흙판에 되풀이해 썼다. 잘못 쓴 판은 물에 개어 다시 빚었다. 남아 있는 연습판도 유물로 발굴된다. 연습판은 그 시대의 교실을 보여 준다.

글자를 아는 사람은 적었다. 대부분은 서기에게 대신 쓰게 했다. 편지를 받으면 서기가 읽어 주었다. 편지의 문장이 정중한 틀을 갖춘 까닭이 여기에 있다. 말하는 사람과 적는 사람이 달랐기 때문이다.

점토판 한 장을 읽는 데도 여러 단계를 거친다. 먼저 기호 하나하나를 알아본다. 다음에 기호가 어떤 소리인지 정한다. 그다음 낱말을 끊고 뜻을 찾는다. 끝으로 문장을 엮어 번역한다. 익숙한 서식이면 빨리 끝난다. 낯선 문서는 오래 붙잡아야 한다.

이유는 읽을 사람이 모자라기 때문이다. 쐐기문자를 다루는 학문을 아시리아학이라고 한다. 학자 한 명을 길러내는 데 수십 년이 걸린다. 옛 언어와 옛 역사를 함께 익혀야 하기 때문이다. 그래서 이 글자를 읽는 사람은 세계에 아주 적다[5].

새로 나오는 점토판의 수는 학자가 읽는 속도를 넘어선다. 사람이 늘지 않으니 상자만 쌓인다. 2026년 6월 영국 요크 대학교가 새 연구 사업을 시작했다. 사업 이름은 "쐐기문자 유산에 대한 보편적 접근을 향하여"다. 쐐기문자 자료를 아랍어로도 옮겨 누구나 보게 하려는 계획이다. 이 사업은 2029년까지 3년 동안 이어진다[7].

자료를 한곳에 모으는 일도 함께 진행된다. 쐐기문자 디지털 도서관 계획이 그런 사업이다. 세계에 흩어진 점토판 정보를 모아 인터넷에 공개한다. 사진과 음역과 번역을 나란히 붙여 둔다. 음역은 쐐기 기호의 소리를 알파벳으로 적은 글이다. 지금까지 40만 점 넘는 유물이 목록에 올랐다[4]. 자료가 모여야 기계도 배울 거리를 얻는다.

점토판은 지금 여러 나라에 흩어져 있다. 이라크와 튀르키예의 박물관에 많이 있다. 런던과 파리와 베를린의 박물관도 큰 몫을 갖고 있다. 미국과 일본의 대학에도 나뉘어 있다. 한 문서의 앞뒤가 서로 다른 나라에 있으면 맞추기 어렵다. 사진을 인터넷에 함께 올려야 비로소 맞춰 볼 수 있다.

왕의 기록에 밀린 장부

일상 문서가 뒤로 밀린 데는 오래된 사정이 있다. 19세기와 20세기 초의 발굴단은 왕궁과 신전을 먼저 팠다. 커다란 석비와 금붙이가 사람들의 눈길을 끌었다. 학자들도 왕실 비문과 법전을 먼저 읽었다. 신화와 서사시를 옮기는 데 시간을 먼저 썼다.

그사이 상인의 영수증과 편지는 순서에서 밀렸다. 곡식 배급 목록은 잔심부름 기록으로 여겨졌다. 한 장씩 떼어 놓고 보면 알맹이가 없어 보였다. 학자들은 이 지루한 문서를 손으로 옮기는 일을 뒤로 미뤘다. 그렇게 보통 사람들의 목소리가 오래 묻혔다.

이 치우침 때문에 잃은 것이 있다. 값이 어떻게 오르내렸는지 알기 어려워졌다. 여자들이 무슨 일을 했는지 가려졌다. 아이와 하인의 이름은 목록에만 남았다. 왕의 비문은 자랑을 적는다. 영수증은 자랑하지 않는다. 거래한 사실만 적는다. 그래서 영수증이 더 정직한 자료가 되기도 한다.

하나의 쐐기가 여러 뜻을 갖는다

기계로 읽으려 해도 쐐기문자는 만만하지 않다. 기호 하나가 여러 역할을 겹쳐 맡기 때문이다[8]. 어떤 자리에서는 소리를 나타내는 음절 기호가 된다. 다른 자리에서는 낱말의 뜻을 통째로 나타낸다. 또 어떤 자리에서는 소리를 내지 않는다. 뒤에 오는 낱말이 나무인지 신인지 알려주는 표지로만 쓰인다. 이런 표지를 한정사라고 한다.

쐐기문자는 한 언어만 적은 글자가 아니다. 수메르어와 아카드어가 같은 글자를 함께 썼다. 뒷날 히타이트어와 엘람어도 이 글자를 빌려 썼다. 같은 기호가 언어마다 다른 소리를 낸다. 읽는 사람은 문서의 언어부터 정해야 한다.

나무를 뜻하는 수메르어 기호 기쉬가 좋은 예다. 가구 이름 앞에 붙으면 재료가 나무라는 표지다. 홀로 놓이면 나무라는 낱말 자체다. 다른 자리에서는 기쉬라는 소리만 나타낸다. 읽는 사람은 앞뒤 문장을 보고 역할을 골라야 한다.

숫자를 적는 방식도 지금과 다르다. 메소포타미아 사람들은 60을 한 묶음으로 삼았다. 시계의 60분과 60초가 여기서 왔다. 영수증의 수량도 이 방식으로 적혔다. 곡식과 은은 서로 다른 단위를 썼다. 같은 기호라도 곡식일 때와 은일 때 값이 달라진다. 기계는 이 단위 차이까지 함께 배워야 한다.

낱말을 붙여 쓴 점도 걸림돌이다. 옛 서기들은 낱말 사이를 띄우지 않았다. 컴퓨터는 어디서 낱말이 끊기는지 알아내기 어렵다[8]. 점토판이 깨지고 닳은 것도 문제다. 쐐기 자국이 얕아지면 사람 눈으로도 구분이 힘들다. 아카드어의 낱말 모양이 자주 바뀌는 것도 어렵다. 하나의 어근에 여러 어미가 붙는다. 주어인지 목적어인지에 따라 끝소리가 달라진다. 시제와 사람에 따라 다시 바뀐다. 사전에서 찾을 기본형을 먼저 알아내야 한다.

점토판이 깨지는 이유도 여럿이다. 땅속에서 눌리고 물에 젖는다. 발굴하다가 모서리가 떨어져 나가기도 한다. 옛날 유물 거래로 깨진 점토판이 여러 나라에 흩어졌다. 땅에서 꺼낸 뒤에는 소금기가 배어 나와 글자를 덮기도 한다[2]. 글자가 사라진 빈자리를 학자들은 결손부라고 부른다.

이 여러 겹의 벽 때문에 수십만 장의 장부가 침묵해 왔다.

참고 자료

[1] Truman State University. (2026). How it works. Truman State University Cuneiforms. https://exhibits.truman.edu/cuneiform/how-it-works/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[3] Bouscaren, D. (2023). Assyrian women of letters. Archaeology, 76(6). https://archaeology.org/issues/november-december-2023/features/assyrian-women-of-letters/ (2026년 9월 3일 접속)

[4] Cuneiform Digital Library Initiative. (2026). About the CDLI. cdli.earth. https://cdli.earth/about (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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

2. 과학 수사대의 추적: 아카드어 AI 번역

번역 통로를 둘로 내다

2023년 텔아비브 대학교 연구진이 먼저 길을 냈다. 가이 구테르츠와 동료들은 아카드어를 영어로 옮기는 번역기를 만들었다[1]. 아카드어는 메소포타미아에서 오래 쓰인 말이다. 쐐기문자로 적혔고 지금은 쓰는 사람이 없다.

연구진은 번역 통로를 둘로 냈다. 첫째 통로의 이름은 C2E다. 컴퓨터 글자로 옮겨 적은 쐐기 기호를 넣어 영어를 받는 길이다. 점토판 사진을 바로 읽는 방식은 아니다. 학자의 음역을 거치지 않으니 단계가 짧다. 둘째 통로의 이름은 T2E다. 학자가 쐐기 기호의 소리를 알파벳으로 옮겨 적은 글을 넣는 길이다. 이 알파벳 옮김을 음역이라고 한다. 음역 단계에서 기호의 뜻이 한 번 정리된다. 그래서 T2E 쪽 번역이 더 정확하다[1].

음역을 만드는 일도 쉽지 않다. 학자는 기호를 보고 소리를 정한다. 소리가 여럿이면 문맥을 보고 고른다. 이 고르기가 곧 판단이다. 판단이 한 번 끝난 글은 기계가 다루기 쉽다. T2E의 점수가 높은 까닭이 여기에 있다.

번역기의 속은 둘로 나뉜다. 앞쪽은 원문을 숫자 덩어리로 바꾼다. 뒤쪽은 그 숫자를 보고 영어 낱말을 하나씩 고른다. 앞쪽을 인코더, 뒤쪽을 디코더라고 부른다. 인코더는 낱말의 뜻과 순서를 함께 담는다. 디코더는 앞서 고른 낱말을 보면서 다음 낱말을 정한다.

사진에서 기호를 바로 읽는 연구는 따로 이어진다. 2026년 6월 전자 바빌로니아 도서관 연구진이 결과를 내놓았다. 깨진 점토판 8만 7천여 점의 사진을 기계가 훑었다. 찾아낸 기호 자국은 290만 개에 가깝다. 앞선 방법보다 검출 성적이 최대 37퍼센트 올랐다. 많이 깨진 점토판과 흐트러진 줄 배열에는 아직 약하다[6].

8천 편의 문서로 배운 기계

기계는 예문을 보고 번역을 배운다. 연구진은 오라크라는 공개 자료집에서 문서를 모았다. 오라크는 쐐기문자 문헌을 정리해 공개하는 인터넷 자료집이다. 여기서 모은 문서는 8,056편이었다. 문장 수로는 5만 개가 넘었다[1].

모아 온 문서는 종류가 여러 가지였다. 왕의 업적을 새긴 비문이 약 3천 편이었다. 행정 편지가 약 2천 편이었다. 재판 기록과 별점 보고서도 들어 있었다[1]. 문서마다 옆에 학자가 만든 영어 번역이 붙어 있었다. 기계는 이 짝을 수만 번 들여다보며 규칙을 스스로 찾았다.

기계가 배우는 과정은 시험공부와 비슷하다. 먼저 문제와 답을 함께 본다. 답을 가리고 스스로 풀어 본다. 틀리면 어디서 틀렸는지 되짚는다. 이 되짚기를 수백만 번 되풀이한다. 사람과 다른 점은 속도와 끈기다.

기계는 낱말을 통째로 다루지 않는다. 낱말을 더 작은 토막으로 쪼개어 다룬다. 이 토막을 토큰이라고 부른다. 드문 낱말도 토막으로 쪼개면 다룰 수 있다. 어미가 많이 붙는 아카드어에 이 방법이 맞는다.

번역 품질은 BLEU라는 점수로 잰다. 기계의 번역과 사람의 번역이 얼마나 겹치는지 세는 방법이다. 겹치는 낱말 묶음의 비율로 점수를 낸다. 0점부터 100점까지로 나타낸다. 30점을 넘으면 뜻이 통한다고 보는 일이 많다. 그러나 절대 기준은 아니다. 문서 종류와 문장 길이에 따라 해석이 달라진다.

연구진은 합성곱 신경망이라는 구조를 골랐다. 가까이 있는 정보끼리 묶어 배우는 구조다. 자료가 적을 때 학습이 빠르고 안정적이기 때문이다. 검증 결과 C2E는 36.52점을 받았다. T2E는 37.47점으로 더 높았다. 예전에 쓰던 자동 검색 방식은 27.09점과 23.51점에 그쳤다. 문장이 짧고 서식이 정해진 문서에서 성적이 좋았다[1].

자료를 나누는 방법도 중요하다. 배울 몫과 시험 볼 몫을 따로 떼어 둔다. 시험 볼 문서를 미리 보여 주면 점수가 헛되다. 사람 시험에서 문제를 미리 아는 것과 같다. 연구진은 이 구분을 지켜 점수를 냈다.

같은 원문도 사람마다 번역이 다르다. 그래서 점수만 보고 좋고 나쁨을 정하지 않는다. 점수 말고 사람의 평가도 함께 쓴다. 번역문을 학자에게 보여 주고 등급을 매기게 한다. 뜻이 통하는지 따로 묻는다. 틀린 곳이 있는지도 따로 묻는다. 학자가 쓸 만하다고 본 번역은 T2E에서 약 44퍼센트였다. C2E는 약 38퍼센트였다[1]. 점수와 사람의 평가가 어긋날 때가 있다. 그럴 때는 사람의 판단을 앞세운다.

아카드어 번역은 저자원 과제에 속한다. 저자원이란 배울 예문이 적다는 뜻이다. 영어와 프랑스어 번역기는 수천만 문장을 배운다. 아카드어가 가진 짝 문장은 그 천분의 일에도 못 미친다. 자료가 적으면 큰 모델이 도리어 헤맨다.

2025년에는 여러 연구팀이 겨루는 공동 과제도 열렸다. 과제는 낱말의 기본형을 찾고 빠진 글자를 맞히는 일이었다. 자료는 전자 바빌로니아 도서관과 아르시밥에서 가져왔다. 같은 자료로 성적을 재니 어느 방법이 나은지 바로 드러났다[2].

기본형 찾기는 번역의 밑작업이다. 이 작업을 학계에서는 레마 찾기라고 한다. 어미가 붙어 늘어난 낱말을 사전 형태로 되돌리는 일이다. 밑작업이 끝나야 낱말별 검색과 통계가 가능해진다.

그럴듯한 거짓말을 잡아내기

기계 번역에는 위험한 버릇이 하나 있다. 근거가 없는데도 그럴듯한 문장을 지어내는 것이다. 이 버릇을 환각이라고 부른다. 연구진은 환각을 두 가지로 나누어 살폈다. 하나는 원문을 보고도 뜻을 잘못 엮는 경우다. 다른 하나는 원문을 무시하고 딴 이야기를 만드는 경우다[1].

깨진 점토판에서 환각이 잘 생긴다. 글자가 사라진 빈자리를 기계가 상상으로 메우기 때문이다. 그래서 번역문은 반드시 학자가 다시 봐야 한다. 기계가 내놓은 문장은 초벌이지 최종 판정이 아니다.

환각을 줄이려는 방법은 여러 가지다. 깨진 곳을 빈칸 표시로 분명히 알려 준다. 기계가 그 대목을 모른다고 답하게 둔다. 같은 문장을 여러 번 번역해 보기도 한다. 답이 매번 달라지면 믿지 않는다. 학자는 이런 표시를 보고 어디를 다시 볼지 정한다.

쐐기 기호를 알파벳으로 옮기고 낱말을 끊는 일이 번역보다 먼저다. 같은 연구 흐름에서 이 작업도 기계에 맡겨 보았다. 양방향 LSTM이라는 신경망의 성적이 앞섰다. 정확도는 최고 97퍼센트에 이르렀다[3].

두 방향으로 열린 문

지금까지의 번역기는 한쪽으로만 흘렀다. 옛 글을 현대 영어로 옮기기만 했다. 2026년 5월 미국 서던캘리포니아 대학교의 왕자오후이가 반대 방향을 열었다. 그가 만든 도구의 이름은 태블릿크래프트다[4].

태블릿크래프트는 영어를 아카드어로도 옮긴다. 옮긴 결과를 쐐기 기호로 그려 점토판 그림까지 만들어 준다. 학습에는 양방향 문장 11만 6천 쌍을 썼다. 기호 대응표는 1만 4천여 개를 갖췄다. 시험 성적은 아카드어에서 영어로 49.1점이었다. 영어에서 아카드어로는 48.5점이 나왔다[4].

이 도구는 박물관 체험 기기와 대학 수업에서 쓰였다. 만든 사람이 분명히 밝힌 한계도 있다. 기계가 지어낸 아카드어를 진짜 옛 기록처럼 인용하면 안 된다. 도구가 잘 다루는 글은 신아시리아 시대의 왕실 비문이다[5].

두 방향 번역에는 쓸모가 있다. 학생이 자기 문장을 쐐기 기호로 바꿔 볼 수 있다. 기호를 직접 그려 보면 글자의 짜임이 눈에 들어온다. 박물관에서는 관람객이 자기 이름을 새겨 본다. 옛 글자가 구경거리에서 다루는 도구로 바뀐다.

앞의 연구들은 자료와 프로그램을 함께 공개했다[1][5]. 공개하면 다른 팀이 같은 시험을 되풀이해 볼 수 있다. 틀린 곳이 있으면 빨리 드러난다. 고대 언어 연구는 참여하는 사람이 적다. 공개가 발전 속도를 크게 좌우한다.

번역기를 쓰는 방법도 정리되고 있다. 학자는 기계의 초벌을 화면에 띄운다. 원문과 번역을 나란히 놓고 견준다. 미심쩍은 낱말에 표시를 남긴다. 표시가 쌓이면 다음 학습 자료가 된다.

참고 자료

[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, 5월 14일). TabletCraft: Bridging a 4,000-year cultural gap with bidirectional Akkadian NMT and cuneiform rendering. arXiv:2608.02609 (2026년 5월 14일 제출, 2026년 8월 공개). https://arxiv.org/abs/2608.02609 (2026년 9월 3일 접속)

[5] Buyukyildirim, K. (2026, 8월 11일). TabletCraft opens a two-way AI translation channel for archaeologists. Arkeonews. https://arkeonews.net/tabletcraft-opens-a-two-way-ai-translation-channel-for-archaeologists/ (2026년 9월 3일 접속)

[6] Che, W., Garces Arias, E., Niaz, A., Bender, A., & Jimenez, E. (2026, 6월 21일). 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 (2026년 9월 3일 접속)

3. 밝혀진 진실

카네시 상인들의 편지

튀르키예 중부 카이세리 근처에 퀼테페라는 언덕이 있다. 4천 년 전 이곳의 이름은 카네시였다. 아시리아 상인들이 여기에 무역 거점을 두었다. 상인들의 거점 공동체를 카룸이라고 불렀다. 카룸은 본디 강가의 부두를 뜻하는 말이다. 물건을 부리던 곳이 상인들의 구역이 되었다.

카네시의 카룸은 아시리아 상인들의 자치 조직이었다. 그들은 자기들끼리 규칙을 만들고 다툼을 가렸다. 현지 왕에게 세금을 내고 안전을 보장받았다. 물건은 당나귀 등에 실려 옮겨졌다. 아수르에서 카네시까지는 여러 주가 걸리는 먼 길이다.

1948년부터 이어진 발굴에서 점토판이 쏟아졌다. 지금까지 나온 문서는 약 2만 3천 5백 장이다. 발굴을 이끄는 피크리 쿨라코울루가 밝힌 숫자다[1]. 내용은 왕의 자랑이 아니다. 차용증과 화물 송장이 있다. 화물 송장은 실어 보낸 짐의 목록표다. 재판 기록과 집안 편지가 있다.

편지에는 재촉과 불평이 그대로 남아 있다. 돈을 언제 보낼지 묻는다. 물건이 늦는다고 나무란다. 몸조심하라는 인사를 붙인다. 아들에게 공부하라고 이르기도 한다. 4천 년 전 사람의 걱정이 문장에 그대로 있다.

이 문서에서 옛 경제의 얼개가 드러났다. 상인들은 아수르에서 직물과 주석을 실어 왔다. 아나톨리아에서 그것을 은과 금으로 바꾸었다. 세금과 통행료를 어떻게 낼지도 적혀 있었다. 다툼이 생기면 어떻게 판결했는지도 남았다.

주석은 청동을 만드는 데 꼭 필요한 금속이다. 청동은 구리와 주석을 섞어 만든다. 아나톨리아에서는 주석을 구하기 어려웠다. 주석은 먼 동쪽에서 실어 와야 했다[8]. 이 모자람이 긴 교역로를 만든 힘이었다.

아수르에 남은 여자들의 몫도 밝혀졌다. 상인의 아내와 딸들은 집에서 직물을 짜서 보냈다[2][8]. 그들은 생산량을 조절하고 값을 흥정했다. 남편에게 보낸 편지에서 시세를

따지기도 했다. 세실 미셸은 이런 문서 300여 편을 모아 영어로 옮겼다. 미셸은 프랑스 국립과학연구원과 함부르크 대학교에서 일한다[2].

봉투를 뜯지 않고 읽다

옛 상인들은 중요한 점토판을 진흙 봉투에 넣어 봉했다. 안을 읽으려면 봉투를 깨야 했다. 그런데 봉투에도 도장 자국이 찍혀 있다. 봉투를 깨면 그 정보가 사라진다.

봉투는 위조를 막는 장치였다. 안쪽 점토판에 계약 내용을 적는다. 그 위에 진흙을 덧씌워 봉투를 만든다. 봉투 겉면에 같은 내용을 다시 적는다. 증인들이 도장을 굴려 찍는다. 겉과 속이 다르면 고쳐 쓴 사실이 드러난다.

봉투를 살린 채 속을 읽는 방법이 필요했다. 2026년 6월 카이세리 고고학 박물관이 해답을 보여 주었다. 봉투에 든 점토판 하나를 뜯지 않은 채 읽어 낸 것이다. 프랑스 국립과학연구원과 함부르크 대학교가 함께한 작업이다. 물건을 자르지 않고 속을 들여다보는 CT 촬영과 같은 방법을 썼다. 안에 적힌 내용은 밀과 보리 거래 약속이었다. 쿨라코울루는 이 촬영이 보리 낟알 하나까지 보여 준다고 말했다[1]. 이렇게 얻은 글자 사진은 번역기의 입력 자료가 된다. 읽어 낼 문서가 늘어야 기계도 배울 거리를 얻는다.

5만 달러가 걸린 번역 대회

번역 대회도 열렸다. 딥 패스트 이니셔티브라는 단체가 대회를 준비했다[3]. 대회는 인공지능 경진 사이트 캐글에서 열렸다[4]. 과제는 카네시의 상업 문서를 영어로 옮기는 일이었다[3]. 대회는 2025년 12월 16일에 시작했다. 참가 마감은 2026년 3월 23일이었다. 상금은 5만 달러였다. 운영 비용은 한 금융 회사가 댔다[7].

대회 자료는 누구나 내려받을 수 있게 공개되었다[4]. 공개 자료는 대회가 끝난 뒤에도 남는다. 다음 연구자가 그 자료 위에서 다시 시작한다. 대회 하나가 자료 하나를 남기는 구조다.

퀼테페 문서 가운데도 아직 옮기지 못한 것이 많다[3]. 발굴은 지금도 해마다 이어진다[1]. 새 점토판이 나오면 목록에 더해진다. 읽을 사람이 늘지 않으면 목록만 길어진다. 기계의 도움이 필요한 까닭이 여기에 있다.

같은 상황은 한 유적에서 끝나지 않는다. 세계 여러 곳의 박물관에 비슷한 문서가 쌓여 있다. 아직 읽지 못한 상자가 그곳에도 있다. 서식이 정해진 문서는 어디에나 많다. 곡

식 장부와 세금 대장이 그렇다. 카네시에서 통한 방법은 다른 문서고에도 쓸 수 있다.

대회 안내문은 문서에 나오는 실제 인물들을 앞세웠다. 아슈르나다와 아슈르이디 같은 상인의 이름이다. 이슈타르라마시와 엔남아슈르의 이름도 함께 걸렸다[3]. 4천 년 전 한 상인 가족의 이름이 대회 주제가 된 것이다.

대회 방식에는 좋은 점이 있다. 몇몇 연구실이 아니라 세계의 개발자가 같은 문제에 달려든다. 같은 자로 성적을 재니 어느 방법이 나은지 바로 드러난다. 상금은 사람들을 모으는 힘이 된다.

이 문서군은 기계 번역에 잘 맞는 자료다. 서식이 정해져 있고 문장이 짧다. 같은 표현이 여러 문서에 되풀이된다. 기계는 되풀이되는 규칙을 잘 찾아낸다. 학자가 이미 옮겨 둔 문서도 적지 않다. 그 번역문이 기계의 교재가 된다.

기계가 하지 못하는 일

성과가 쌓였지만 남은 벽도 뚜렷하다. 첫 번째 벽은 자료의 부족이다. 요즘 언어의 번역기는 수억 개의 문장을 먹고 자란다. 아카드어는 짝을 이룬 문장이 수만 개뿐이다[5]. 자료가 적으면 드문 지명과 인명에서 틀리기 쉽다.

두 번째 벽은 기호의 여러 뜻이다. 같은 기호가 소리인지 낱말인지 표지인지 헷갈린다[6]. 앞뒤 문맥이 짧으면 기계도 헤맨다.

세 번째 벽은 기계의 판단 방식이다. 기계는 어떤 낱말이 올 확률이 높은지 셈한다. 그 문장이 역사적으로 맞는지는 판단하지 못한다. 그래서 학자의 검증이 마지막 단계에 꼭 필요하다[5]. 기계와 사람이 짝을 이룰 때 번역이 믿을 만해진다.

일을 나누어 맡는 방식이 뿌리를 내리고 있다. 기계는 수만 장을 훑어 대강의 뜻을 낸다. 학자는 그중에서 중요한 문서를 골라낸다. 고른 문서만 손으로 꼼꼼히 다시 읽는다. 이렇게 하면 읽는 문서의 수가 크게 늘어난다.

번역문에는 꼬리표가 필요하다. 어느 대목을 기계가 만들었는지 적어 둔다. 사람이 고친 곳도 표시한다. 그래야 뒷사람이 근거를 되짚을 수 있다. 역사 연구는 근거를 되짚는 일에서 시작한다.

번역에서 이름은 까다로운 대목이다. 사람 이름과 땅 이름은 뜻을 옮길 수 없다. 소리를 그대로 옮겨 적어야 한다. 그런데 같은 이름이 문서마다 다르게 적히기도 한다. 기계는 처음 보는 이름을 엉뚱한 낱말로 바꾸기 쉽다. 그래서 이름 목록을 따로 만들어 기계에 알려 준다.

그래도 방향은 분명하다. 예전에는 학자 한 명이 평생 수백 장을 읽었다. 이제는 기계가 초벌 번역을 먼저 만든다. 학자는 그 초벌을 고치는 데 시간을 쓴다. 상자 속에서 잠자던 4천 년 전 영수증이 차례를 기다린다. 왕이 아닌 보통 사람들의 목소리가 들리기 시작한다.

카네시의 상인들은 자기 편지가 오래 남으리라 생각하지 않았다. 그들은 그날의 셈을 적었을 뿐이다. 그 셈이 지금 한 시대의 경제를 보여 준다. 남는 것은 자랑보다 기록이다.

참고 자료

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[4] Kaggle. (2026). Deep Past Challenge: Translate Akkadian to English. kaggle.com. https://www.kaggle.com/competitions/deep-past-initiative-machine-translation (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[8] Bouscaren, D. (2023). Assyrian women of letters. Archaeology, 76(6). https://archaeology.org/issues/november-december-2023/features/assyrian-women-of-letters/ (2026년 9월 3일 접속)

제 4 부

동아시아 고고학의 비밀: 진흙 속 목간과 뼈 위의 붉은 문자

07 CHAPTER

제7장. 거북 배딱지 속에 잠들었던 제국의 미래: 갑골문과 3D 가상 편철

갑골의 글자를 읽고 깨진 뼈를 화면에서 다시 맞추다

1. 사건 파일

약방에서 발견된 3천 년 전 글자

1899년 청나라의 학자 왕의영이 옛 글자 하나를 알아보았다. 이 일에는 오래 전해 오는 이야기가 붙어 있다. 왕의영이 병을 앓아 용골이라는 약재를 받았다는 이야기다. 용골은 오래된 짐승 뼈를 갈아 만든 가루약이었다. 옛사람들은 이 뼈를 용의 뼈로 믿었다. 그가 받은 뼈는 아직 갈리지 않은 상태였다고 한다. 그 표면에서 낯선 무늬를 보았다는 것이다[1].

이 이야기는 사실로 굳어지지 않았다. 중국 학계는 이를 뒷받침할 기록이 없다고 본다. 고문자학자 이학근은 2007년 글에서 이 이야기를 부정했다. 약재로 팔린 용골은 이미 부수어져 글자를 볼 수 없었다는 것이다. 실제로는 골동상이 안양에서 나온 뼈를 사들여 옮겼다. 그 뼈가 금석학자 왕의영에게 흘러가 감정을 받았다고 본다[2].

왕의영은 이듬해 세상을 떠나 뼈의 출처를 찾지 못했다. 친구인 유악이 1903년에 이 글자를 모은 첫 책을 냈다. 1908년에는 학자 나진옥이 뼈의 출처를 밝혀냈다[1]. 뼈는 허난성 안양의 밭에서 나오고 있었다.

안양의 그 땅이 바로 은허다. 은허는 상나라의 마지막 도읍터였다. 상나라는 기원전 13세기부터 기원전 11세기까지 이곳을 다스렸다[3]. 정식 발굴은 1928년에 시작되었다. 그 전까지 농민들이 파낸 뼈는 약재로 팔려 나갔다. 글자가 새겨진 면을 긁어내고 파는 일도 있었다[1]. 3천 년을 버틴 기록이 약재로 사라질 뻔했다.

왕이 신에게 묻던 나라

은허의 뼈에 새겨진 글자를 갑골문이라 부른다. 갑골문보다 이른 문자 기록은 동아시아에서 알려진 것이 없다[3].

상나라 사람들은 큰일을 앞두면 신에게 물었다. 비가 올지 물었다. 농사가 잘될지 물었다. 전쟁에 나가도 되는지 물었다. 왕비가 아이를 낳는 날도 물었다. 이렇게 신의 뜻을 묻는 의식을 점복이라 한다. 점복은 왕실이 직접 관리하는 나라 사업이었다[3].

상나라의 왕비 부호는 군대를 이끌던 사령관이기도 했다. 부호의 이름은 갑골문 기록에 여러 번 나온다. 1976년 은허에서 부호의 무덤이 온전한 상태로 발굴되었다[4]. 무덤 속 청동기와 옥기는 그의 지위를 보여 주었다. 갑골문은 이렇게 무덤과 함께 옛사람의 삶을 알려 준다.

갑골문에 적힌 물음은 나라의 일에만 머물지 않았다. 왕의 치통을 걱정하는 물음도 있었다. 사냥을 나가도 되는지 묻는 물음도 있었다. 그래서 이 뼈들은 상나라 사람들의 하루를 보여 주는 기록이 된다[3].

불로 지진 뼈, 칼로 새긴 글자

상나라 사람들은 점복에 두 가지 재료를 썼다. 하나는 거북의 납작한 배딱지다. 다른 하나는 소의 넓은 어깨뼈다. 두 재료를 합쳐 갑골이라 부른다[3].

준비 과정은 꼼꼼했다. 서기들은 뼈의 겉면을 매끄럽게 다듬었다. 그다음 뒷면에 작은 홈을 줄지어 팠다. 파낸 홈에 뜨겁게 달군 막대를 댔다. 열을 받은 뼈는 겉면이 쩍 갈라졌다. 왕과 점쟁이는 갈라진 금의 방향과 각도를 살폈다. 그리고 그 금을 신의 대답으로 읽었다[3].

읽기가 끝나면 서기들이 칼을 들었다. 그들은 점복의 과정을 뼈 표면에 한 획씩 새겼다. 이 새긴 글이 바로 갑골문이다.

기록은 정해진 순서를 지켰다. 먼저 점을 친 날짜와 담당자 이름을 적었다. 다음으로 신에게 던진 질문을 적었다. 그다음 왕이 내린 길흉 판단을 적었다. 마지막으로 실제로 일어난 일을 적었다. 네 칸을 차례로 채우는 방식이다. 다만 네 칸이 다 남아 있는 기록은 드물다. 뼈가 깨져 한두 칸이 없어지기도 했다. 서기가 처음부터 빈칸으로 둔 경우도 있다[3]. 그래도 적는 순서 자체는 일정했다.

이 반듯함은 나중에 큰 도움이 된다. 서식이 정해져 있으면 빈칸을 좁힐 수 있기 때문이다. 앞 칸에 날짜가 있으면 뒤 칸에는 질문이 온다. 컴퓨터도 이 순서를 배워 빠진 글자를 짐작한다[3].

날짜를 적는 방법도 정해져 있었다. 상나라는 열 개와 열두 개의 기호를 짝지어 날을 세었다. 이 짝은 예순 날마다 한 바퀴를 돌았다. 뼈에 날짜 기호가 두 자만 남아 있어도 순서를 따져 볼 수 있다. 잘린 기록의 시기를 좁히는 실마리가 여기서 나온다[3].

15만 편으로 부서진 기록

상나라가 무너진 뒤 은허는 흙 속에 묻혔다. 그 오랜 세월 동안 뼈는 온전할 수 없었다. 흙의 무게가 얇은 배딱지를 눌러 쪼갰다. 물과 미생물이 뼈의 결을 갉아 먹었다. 원래 한 장이던 큰 배딱지가 수십 개로 갈라졌다.

지금까지 세상에 알려진 갑골 파편은 약 15만 편이다[5]. 이 유물들은 한곳에 모여 있지 않다. 19세기 말에 발견된 뒤 상당수가 골동품 시장으로 흘러갔다[1]. 그래서 한 뼈에서 떨어져 나온 파편들이 서로 다른 나라 박물관에 놓이게 되었다[5]. 중국 안양과 베이징에 있는 유물도 있다. 유럽과 한국의 기관이 가진 유물도 있다[6].

흩어진 파편을 원래 한 장으로 되돌리는 일을 편철이라 한다. 편철은 오래도록 사람의 눈에만 기댔다. 학자들은 흑백 탁본 수만 장을 확대경으로 들여다보았다. 탁본은 뼈 위에 종이를 대고 먹을 두드려 찍어 낸 사본이다. 동전 위에 종이를 대고 연필로 문질러 무늬를 뜨는 것과 비슷하다. 학자들은 뼈의 결과 깨진 선의 굴곡을 견주었다. 잘린 획의 끝이 이어지는지도 하나하나 살폈다[5].

이 방식은 느리고 힘들었다. 한 쌍을 찾는 데 몇 해가 걸리기도 했다. 짝을 찾지 못한 파편은 상자 속에 그대로 남았다. 짝이 없으면 문장도 잘린 채로 남는다. 질문만 남고 결과가 사라진 기록이 쌓였다.

일의 크기를 헤아려 보면 어려움이 더 뚜렷해진다. 파편이 15만 편이면 두 개를 고르는 경우의 수가 100억 쌍을 넘는다. 사람이 하루에 백 쌍을 견주어도 평생 다 볼 수 없다. 게다가 파편은 여러 나라에 흩어져 있다[5]. 두 파편을 나란히 놓고 볼 기회조차 오지 않는 경우가 많았다.

3천 자가 아직 말이 없다

파편을 다 붙여도 문제가 남는다. 글자 자체를 읽지 못하기 때문이다. 갑골문 전체에서 확인된 서로 다른 글자는 약 4,500자다. 그런데 학계가 뜻에 합의한 글자는 약 1,500자에 그친다. 나머지 3,000자쯤은 아직 뜻을 모른다[7].

읽기가 어려운 까닭은 여러 가지다. 칼자국이 닳아 윤곽이 흐려졌다. 뼈가 깨지면서 글자의 반쪽이 사라졌다. 서기마다 칼을 쥐는 습관이 달라 같은 글자를 다르게 새겼다. 그래서 뜻은 같은데 모양이 다른 글자가 여럿 생겼다. 이런 글자를 이체자라 한다[3]. 현대 한자 사전은 이 3천 년 전의 모양을 설명해 주지 못한다.

이 침묵은 상나라 역사에도 구멍을 남긴다. 제사의 이름을 읽지 못하면 제사의 종류를 셀 수 없다. 땅 이름을 읽지 못하면 나라의 경계를 그릴 수 없다. 사람 이름을 읽지 못하면 누가 무엇을 맡았는지 알 수 없다. 글자 하나의 침묵이 역사 한 장면의 침묵으로 이어진다.

뜻을 모르는 글자 3천 자와 짝을 잃은 파편 수만 점이 남았다. 이 침묵을 깨러 인공지능이 은허로 들어왔다.

참고 자료

[1] Mark, E. (2016, 2월 26일). Oracle bones. World History Encyclopedia. https://www.worldhistory.org/Oracle_Bones/ (2026년 9월 3일 접속)

[2] 姚小鸥. (2020, 7월 16일). 谁是甲骨文的最早发现者. 澎湃新闻. https://www.thepaper.cn/newsDetail_forward_8297008 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[4] Smarthistory. War and sacrifice: The tomb of Fu Hao. Smarthistory. https://smarthistory.org/tomb-of-fu-hao/ (2026년 9월 3일 접속)

[5] Zhang, C., Wang, B., Chen, K., Zong, R., Mo, B., Men, Y., Almpanidis, 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, 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 (2026년 9월 3일 접속)

[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

2. 과학 수사대의 추적: 갑골 전용 멀티모달 모델과 3D 가상 편철

흩어진 자료를 한 화면에 모으다

수사는 자료를 모으는 일에서 시작한다. 갑골 연구도 같았다. 안양사범대학과 중국사회과학원 연구진은 2019년에 은계문연이라는 자료 플랫폼을 열었다. 은계문연은 갑골 탁본과 글자 정보와 논문을 한곳에 모아 무료로 공개한다[1].

이 플랫폼은 사진만 올려 두지 않는다. 갑골 파편을 고해상도로 찍고 3차원 모형으로 만들어 올린다. 3차원 모형은 물건을 여러 방향에서 스캔해 컴퓨터 안에 입체로 되살린 자료다. 화면을 돌리면 뼈의 앞면과 뒷면과 깨진 옆면을 볼 수 있다. 연구진은 밀리미터 단위의 정밀도로 이 모형을 만든다. 얕게 파인 칼자국의 깊이도 이 자료에 담긴다[1].

해외로 나간 갑골을 되찾는 작업도 이 플랫폼과 이어졌다. 연구진은 실물 대신 3차원 자료를 들여오는 방식을 골랐다. 독일과 한국과 프랑스 등의 소장처와 협력해 900여 점의 자료를 모았다[1]. 실물은 원래 소장처에 두고 모습만 안양으로 돌아왔다.

2025년 10월에는 이 플랫폼 위에 은계행지라는 도구가 붙었다. 이 도구는 글자를 알아보고 탁본을 견주고 관련 문헌을 찾아 준다[1]. 자료를 쌓아 두던 창고가 질문에 답하는 조수로 바뀌었다.

자료가 얼마나 쌓였는지는 숫자로도 확인된다. 2026년에 나온 종합 정리 논문은 갑골 관련 공개 자료집 36종을 소개한다. 글자 인식용 자료집도 있고 편철 연구용 자료집도 있다[2].

여러 종류의 자료를 함께 쓰는 방식을 멀티모달이라 한다. 갑골 연구에서 이 말은 세 가지 자료를 같이 본다는 뜻이다. 하나는 먹으로 찍어 낸 흑백 탁본이다. 하나는 색과 질감이 살아 있는 컬러 사진이다. 나머지 하나는 깊이 정보가 담긴 3차원 스캔 자료다. 탁본은 획이 또렷하지만 깊이를 잃는다. 3차원 자료는 깊이를 담지만 획이 흐려 보인다. 셋을 겹쳐 보면 서로의 빈틈을 메울 수 있다[2].

사람이 글자를 읽는 순서를 따라 하다

2026년에 나온 알파오라클은 이 자료 위에서 움직이는 인공지능이다. 이 모델은 화중과학기술대학과 안양사범대학 등이 함께 만들었다. 설계 원칙은 사람 학자의 작업 순서

를 그대로 따르는 것이다[3].

사람 학자는 네 단계를 거친다. 먼저 탁본을 살펴 획을 정리한다. 다음으로 글자의 모양을 뜯어본다. 그다음 같은 글자가 나오는 다른 문장을 찾아본다. 마지막으로 옛 문헌 지식으로 뜻이 맞는지 따진다. 알파오라클은 이 네 단계를 각각 담당하는 부품으로 나뉘어 있다[3].

이 구조에는 이유가 있다. 답만 툭 내놓는 기계는 학자가 믿기 어렵다. 단계를 나누면 어느 단계에서 답이 갈렸는지 보인다. 학자는 그 단계로 되돌아가 기계와 의견을 맞댈 수 있다[3].

이 모델이 읽은 자료의 양은 사람이 평생 볼 수 있는 양을 넘는다. 탁본 73,883점을 다루었다. 스캔 자료 45,364점과 연구 논문 23,755편도 함께 다루었다. 연구진은 전문가 86명에게 이 도구를 써 보게 했다. 그 결과 한 글자를 분석하는 시간이 64퍼센트 줄었다. 참가자의 79퍼센트가 이 도구를 쓸모 있다고 답했다[3].

글자를 부품으로 쪼개서 보기

다른 연구진은 글자를 통째로 외우지 않는 방법을 택했다. 갑골문 글자는 되풀이되는 그림 부품으로 이루어져 있다[4]. 손 모양, 발 모양, 그릇 모양 같은 부품이 여러 글자에 나눠 들어간다. 부품의 뜻을 알면 처음 보는 글자도 짐작할 수 있다.

이 생각으로 2026년 4월에 발표된 것이 OB-Radix 자료집이다. OB-Radix에는 서로 다른 글자 934자가 담겼다. 그 글자를 이루는 부품 그림은 1,853점이다. 부품의 종류는 478가지다. 각 부품에는 전문가가 확인한 설명이 붙어 있다[4].

인공지능은 낯선 글자를 만나면 먼저 부품으로 쪼갠다. 그다음 부품들이 어떤 관계로 놓였는지 따진다. 두 부품이 위아래로 놓였는지 나란히 놓였는지가 뜻을 가른다. 모델은 이 관계까지 살펴 뜻을 제안한다[4].

같은 해 6월에는 오라클애널라이저라는 모델도 나왔다. 이 모델은 30억 개 규모의 작은 언어 모델을 갑골문 자료로 다시 훈련시킨 것이다. 연구진은 갑골문 추론 자료와 선호도 자료를 새로 만들어 붙였다. 선호도 자료는 사람이 더 좋다고 고른 답을 모은 자료다. 그 결과 작은 모델이 훨씬 큰 모델보다 좋은 성적을 냈다[5]. 자료를 잘 고르면 크기로 밀어붙이지 않아도 된다는 뜻이다.

흐려진 획을 되살려 그려 주는 모델

읽기 전에 먼저 보이게 만드는 연구도 있다. 2024년에 발표된 OBSD가 그런 도구다. 이 도구는 그림을 새로 그려 내는 확산 모델을 쓴다[6]. 확산 모델은 흐릿한 얼룩에서 점점 또렷한 그림을 만들어 내는 방식이다.

OBSD는 닳고 흐려진 갑골 글자 그림을 입력으로 받는다. 그리고 그 글자가 어떤 현대 한자로 이어지는지 그림으로 제안한다. 이 연구는 2024년 국제 학회에서 최우수 논문으로 뽑혔다[6]. 글자 읽기를 그림 그리기 문제로 바꾼 발상이 인정을 받았다.

깨진 파편의 단면을 컴퓨터가 맞춘다

글자 읽기와 나란히 파편 맞추기 연구도 진행되었다. 2022년에 공개된 OB-Rejoin은 편철 연구용 자료집이다. 실제 갑골 탁본 998장이 들어 있다. 같은 연구진은 S3-Net이라는 이어 붙이기 신경망도 함께 내놓았다[7].

이 신경망은 파편 테두리의 굴곡이 서로 얼마나 닮았는지를 계산한다. 두 테두리가 맞물리면 점수가 높아진다. 기계는 점수가 높은 순서로 후보를 늘어놓는다. 같은 연구진은 2020년에 앞선 연구도 발표했다. 그때는 후보 열 개 안에 정답을 넣는 비율이 98.39퍼센트였다[11].

퍼즐을 맞출 때 사람은 그림보다 파편의 모양을 먼저 본다. 기계도 같은 방법을 쓴다. 다만 기계는 하루에 수십만 번을 견줄 수 있다.

기계가 내놓는 것은 정답이 아니라 후보 목록이다. 목록의 맨 위가 틀리는 경우도 있다. 그래서 연구진은 열 개 안에 정답이 들었는지를 성능 기준으로 삼는다[7]. 사람이 열 개만 확인하면 되는 상황과 십만 개를 확인해야 하는 상황은 다르다.

수도사범대학 갑골문연구센터는 허난대학과 함께 편철 프로그램을 만들었다. 이름은 철다다이고 2020년부터 쓰였다. 이 프로그램은 탁본 18만 장을 한꺼번에 견주어 짝을 찾는다. 연구센터는 이 방식으로 갑골 파편 약 1,500건의 기록을 정리했다. 예전에는 한 파편의 짝을 찾는 데 여덟 해에서 열 해가 걸리기도 했다. 지금은 며칠이면 후보가 나온다[8].

문장 단위로 넘어가는 시험

글자 하나를 읽는 것과 문장을 이해하는 것은 다른 일이다. 2026년 6월 말에 공개된 S-OBI는 이 차이를 재는 시험 문제집이다. 연구진은 전문가가 확인한 원본 탁본 95점에서 문제를 뽑았다. 문제는 모두 695개이고 문장 단위로 답하게 되어 있다. 그리고 여러 대형 인공지능에게 같은 문제를 풀게 했다. 시험 결과 모델마다 성적 차이가 뚜렷했다. 문장 전체의 흐름을 잡는 일이 아직 어렵다는 사실도 드러났다[9].

이 시험이 알려 준 사실이 하나 더 있다. 갑골문 문제는 일반 인공지능이 잘 풀지 못한다는 사실이다[9]. 세상의 글을 두루 배운 모델도 3천 년 전 문장 앞에서는 흔들린다. 배운 적이 없는 글자와 어법이기 때문이다. 그래서 갑골 전용으로 다시 훈련하는 연구가 이어진다[5].

깨진 단면을 넘어 문맥을 잇는 연구도 나왔다. 한 연구진은 두 문장이 원래 같은 뼈에 있었는지 판정하는 과제를 만들었다. 이들은 OBID-ACR이라는 자료집을 공개했다. 그리고 글자 모양 정보와 문장 정보를 함께 쓰는 방법을 제안했다[10]. 파편을 실제로 붙이기 전에도 문맥을 이어 볼 길이 열린 것이다.

참고 자료

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[7] Zhang, C., Wang, B., Chen, K., Zong, R., Mo, B., Men, Y., Almpanidis, 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, 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 (2026년 9월 3일 접속)

[9] Li, Z., Chen, Z., Chen, T., & Zhai, G. (2026, 6월 30일). Beyond single character: Evaluating MLLMs for sentence-level oracle bone inscription understanding. arXiv:2606.31169. https://arxiv.org/abs/2606.31169 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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

3. 밝혀진 진실

글자 하나가 바뀌자 지도가 달라졌다

알파오라클이 내놓은 첫 성과는 글자 하나였다. 연구진은 오래 다투어 온 글자 하나를 이 시스템에 걸었다. 뒷날 한자 로(勞)로 이어지는 글자다[1]. 시스템은 획의 모양과 부품 구성을 정리했다. 그리고 그 글자가 나오는 다른 문장들을 찾아 나열했다. 이어서 옛 문헌 지식과 맞대어 뜻을 좁혔다.

시스템이 제안한 답은 이 글자가 땅 이름이거나 씨족 이름이라는 것이었다. 씨족은 한 조상에서 갈라져 나온 친척 집안을 뜻한다. 이 글자를 문장을 꾸미는 말로 읽던 해석과 다른 결과였다. 연구진은 이 해석이 고문자 학자들의 검토를 거쳤다고 밝혔다[1].

글자 하나의 뜻이 바뀌면 그 글자가 들어간 문장이 모두 바뀐다. 사람 이름이면 그 문장은 한 개인의 이야기다. 땅 이름이면 같은 문장이 지방 행정의 기록이 된다. 상나라가 그 땅을 어떻게 다스렸는지도 다시 그려야 한다. 연구진은 이 결과가 상나라의 행정과 사회를 보는 시각을 손보게 한다고 보고했다[1].

이 연구에서 눈여겨볼 것은 답 자체가 아니다. 시스템이 근거를 함께 내놓았다는 점이다. 알파오라클은 어떤 문장과 어떤 부품을 보고 그 답을 골랐는지 보여 준다[1]. 학자

는 그 근거를 따라가며 동의할지 말지 정할 수 있다. 사람이 검토할 수 있는 형태로 답을 내는 일이 이 연구의 핵심이었다.

몇 해 걸리던 짝짓기가 며칠로 줄었다

편철 쪽에서도 결과가 나왔다. 앞에서 본 철다다 프로그램이 낸 성과다. 이 프로그램은 학자들이 오래 놓쳐 온 짝을 300건 넘게 찾아냈다[2]. 백 년 넘게 손으로 찾아 온 짝 위에 더해진 수다.

사람이 하는 일은 그대로 남았다. 학자는 기계가 준 목록을 하나씩 확인한다. 확인은 실물 대조로 마무리한다.

안양사범대학 연구실도 자체 성과를 내놓았다. 2026년 5월까지 새로 이어 붙인 갑골이 155건이다. 같은 연구실이 겹친 기록을 가려낸 사례는 7,000건 가까이 된다[3]. 한 뼈가 여러 도록에 따로 실리는 일이 있었다. 그러면 한 뼈가 서로 다른 두 유물처럼 세어진다. 새로 붙인 수보다 겹침을 가려낸 수가 훨씬 많다. 기계가 하는 일이 새 발견만이 아니라 옛 기록의 점검이기도 하다는 뜻이다.

파편이 붙으면 문장이 길어진다. 한쪽에는 질문만 남고 결과는 다른 쪽에 있는 경우가 많다. 두 파편이 만나야 점을 친 날과 실제로 벌어진 일이 한 줄로 이어진다. 짧은 기록 여러 개가 아니라 하나의 사건 기록이 되는 것이다.

이 차이는 역사 서술에 곧바로 영향을 준다. 잘린 기록만 보면 사건이 흩어져 보인다. 이어진 기록을 보면 앞뒤 순서가 드러난다. 왕이 언제 물었고 언제 결과를 받았는지가 잡힌다.

겹침을 가려낸 기록 7,000건도 같은 무게를 지닌다[3]. 같은 뼈를 둘로 세면 자료의 수부터 틀어진다. 같은 문장을 서로 다른 두 기록으로 알고 인용하기도 한다. 기계가 탁본을 대량으로 다시 견주면서 이런 겹침을 걸러 냈다. 눈에 띄는 발견은 아니지만 기초 자료를 바로 세우는 일이다.

바다 건너 흩어진 뼈가 화면에서 만난다

편철을 가로막던 벽은 거리였다. 한 뼈에서 떨어진 파편들이 서로 다른 나라에 있으면 나란히 놓고 볼 수 없었다. 사진을 주고받아도 두께와 굴곡은 알 수 없었다. 3차원 자료는 이 벽을 낮추었다.

앞에서 본 해외 자료 900여 점이 이 벽을 넘은 사례다[3]. 실물은 원래 있던 곳에 그대로 있다. 옮겨 온 것은 표면의 형태 정보다.

안양의 연구자는 화면에서 그 파편을 돌려 가며 볼 수 있다. 베이징의 파편과 유럽의 파편을 같은 화면에 놓고 단면을 맞춰 볼 수도 있다.

공개의 방식도 달라졌다. 예전에는 소장 기관을 찾아가야 유물을 볼 수 있었다. 허가를 받는 데만 여러 달이 걸리기도 했다. 지금은 학생도 집에서 갑골 자료를 열어 볼 수 있다[3]. 자료를 다루는 사람의 수가 늘면 확인하는 눈도 늘어난다. 눈이 늘면 잘못 붙인 짝을 걸러 낼 기회도 늘어난다.

기계가 아직 못 하는 일

성과와 함께 한계도 분명해졌다. 2026년 6월 말에 공개된 S-OBI 시험이 그 한계를 숫자로 보여 주었다. 여러 대형 인공지능이 같은 갑골 문장 문제를 풀었다. 글자 하나를 알아보는 문제보다 문장 전체를 이해하는 문제에서 성적이 떨어졌다. 원인도 드러났다. 모델은 멀쩡히 남아 있는 글자부터 잘못 보았다. 그 오류를 그대로 안고 추론을 이어 가니 답이 틀어졌다[4]. 3천 년 전 문장의 흐름을 잡는 일은 아직 사람의 몫이 크다.

더 조심해야 할 문제는 지어내기다. 인공지능은 확률로 답을 고른다. 갑골문처럼 학습 자료가 적은 분야에서는 그럴듯하지만 실제로 없던 글자를 만들어 낼 수 있다. 상나라 말은 비교할 만한 이웃 언어 자료도 넉넉하지 않다[5]. 사전을 뒤져 답을 검산하기 어려운 상황이다.

자료의 치우침도 남아 있다. 지금까지 모인 갑골 자료는 특정 시기와 특정 발굴지에 쏠려 있다. 쏠린 자료로 배운 기계는 낯선 지역의 갑골을 만나면 실수하기 쉽다. 2026년에 나온 종합 정리 논문도 자료의 품질과 균형을 앞으로의 과제로 꼽았다[5].

숫자로 보면 갈 길이 아직 멀다. 뜻을 모르는 글자는 여전히 3,000자쯤 남아 있다. 알파오라클이 근거를 갖춰 풀어낸 글자는 그중 하나였다[1]. 한 글자를 푸는 데 든 품을 생각하면 남은 일이 크다. 그래도 방법이 정리되면 다음 글자는 조금 빨라진다.

사람과 기계가 나눠 맡는 일

그래서 지금의 연구 방식은 사람과 기계를 나란히 두는 쪽으로 굳어졌다. 기계는 후보를 넓게 내놓고 근거를 정리한다. 사람은 그 후보를 문헌과 실물로 확인한다.

앞에서 본 분석 시간 단축도 이 분업의 결과다. 기계가 판단을 대신한 것이 아니라 사람이 볼 자료를 미리 추려 주었다. 전문가들이 이 도구를 쓸모 있다고 답한 이유도 여기에 있다[1].

편철에서도 같은 방식이 굳어졌다. 예전에 여러 해가 걸리던 대조를 프로그램이 며칠 만에 끝낸다[2]. 학자는 그 후보를 들고 수장고로 간다. 실제 뼈를 맞대어 단면이 붙는지 확인한다. 마지막 판단은 사람의 손끝에서 이루어진다.

이 분업에는 한 가지 조건이 붙는다. 기계가 낸 답의 근거가 사람에게 보여야 한다는 조건이다. 근거가 보이지 않으면 학자는 확인할 방법이 없다. 확인할 수 없는 답은 역사 서술에 넣을 수 없다. 알파오라클이 단계를 나눈 이유도 여기에 있다[1].

3천 년 전 상나라 왕은 뼈의 갈라진 금을 보며 미래를 물었다. 오늘의 연구자는 같은 뼈의 갈라진 단면을 보며 과거를 묻는다. 물음의 방향은 반대지만 들여다보는 자국은 여전히 같다.

참고 자료

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

08 CHAPTER

제8장. 진흙 구덩이에서 살아 돌아온 붓글씨: 신라 목간과 흘림체 판독 인공지능

신라 목간의 흘림체 붓글씨를 읽어 내는 판독 인공지능

1. 사건 파일

나무에 글을 쓰던 시대

종이가 귀하던 시절이 있었다. 고대 동아시아 사람들은 나무를 얇게 다듬어 그 위에 글을 적었다. 글자를 적은 이 나무판을 목간이라고 부른다. 목간은 대개 손가락 두세 개를 붙인 너비다. 길이는 한 뼘 남짓한 것이 많다. 나무를 기둥처럼 다듬은 것도 있다. 이런 목간은 네 면 모두에 글을 쓸 수 있다. 여러 면에 글을 쓴 나무판이라는 뜻으로 다면 목간이라고 부른다.

목간의 쓰임은 여러 가지였다. 하나는 물건에 매다는 꼬리표다. 지방에서 왕경으로 곡식이나 젓갈을 보낼 때 목간을 끈으로 묶어 달았다. 거기에는 보낸 고을 이름과 물건 이름과 수량을 적었다. 다른 하나는 관청의 문서다. 관리의 이름과 등급, 받은 물건과 날짜를 적은 목간이 나온다[1]. 글씨를 익히려고 같은 한자를 여러 번 써 놓은 연습용 목간도 있다.

나무는 종이보다 단단해서 다시 쓸 수 있었다. 쓸모가 끝난 목간은 겉면을 칼로 얇게 깎아 냈다. 새 면이 드러나면 그 위에 다시 글을 썼다. 이때 깎여 나온 얇은 나뭇밥도 유적에서 함께 나온다. 이 부스러기를 삭설이라고 부른다. 글씨를 지우려고 깎아 낸 나무 부스러기라는 뜻이다. 부여 관북리 유적에서 나온 목간류 329점을 보면 알 수 있다. 그 가운데 247점이 삭설이었다. 문서를 만들고 고쳐 쓰던 시설이 가까이 있었다는 뜻이다[1].

역사학자들이 목간을 귀하게 여기는 까닭이 있다. 역사책은 대개 사건이 끝난 뒤에 쓰인다. 쓰는 사람의 판단과 왕조의 사정이 들어간다. 목간은 그날의 업무를 그날 적은 기록에 가깝다. 후대에 보여 주려고 꾸민 글이 아니다. 곡식 몇 섬을 누가 언제 보냈는지가

그대로 남았다. 사건이 벌어진 때에 만들어진 이런 기록을 일차 사료라고 부른다. 목간에도 잘못 적은 글자와 고쳐 쓴 자리는 있다. 그래서 연구자는 여러 점을 나란히 놓고 견준다.

목간은 한반도와 일본 열도에서 함께 나온다. 두 곳 모두 한자로 문서를 만들던 사회였다. 그러면서도 제 나라 말을 적으려는 궁리를 했다. 신라 사람들은 한자의 뜻과 소리를 섞어 우리말 어순을 적었다. 이 표기를 이두와 향찰이라고 부른다. 일본 사람들도 한자의 소리를 빌려 자기 말을 적었다. 목간은 이런 시도가 언제 시작되었는지 알려 주는 실물 증거다. 종이 문서는 거의 남지 않았기 때문이다. 두 나라 목간을 나란히 놓고 견주는 연구가 이어지는 까닭이 여기에 있다.

진흙이 지켜 준 천 년

목간이 나오는 곳은 대개 정해져 있다. 궁궐 곁의 연못 바닥이다. 성벽 안쪽에 물이 고인 저습지다. 도랑을 메운 진흙층이다. 사람들이 쓰고 버린 나무가 물속에 가라앉아 그대로 묻혔다.

물속 진흙에는 공기가 거의 들어가지 못한다. 나무를 썩게 만드는 것은 대개 공기를 마시며 사는 미생물이다. 공기가 막히면 이 미생물이 힘을 쓰지 못한다. 공기 없이 사는 미생물이 남아 있기는 하다. 그래도 나무가 삭는 속도는 크게 느려진다. 그래서 천 년이 넘도록 나무가 형태를 지켰다. 가죽과 종이가 흔적도 없이 사라지는 땅에서도 목간은 남았다. 무덤이 아닌 쓰레기 구덩이가 기록을 지켜 주었다.

대신 다른 변화가 일어났다. 나무 속으로 물이 스며들어 세포 벽을 채웠다. 흙 속의 철분과 여러 광물도 함께 들어갔다. 나무는 물을 잔뜩 머금은 상태가 되었다. 이런 나무를 물먹은 나무, 곧 포수 목재라고 부른다.

포수 목재는 꺼내서 그냥 말리면 안 된다. 물이 빠지면서 급하게 오그라들고 갈라진다. 발굴 현장에서 목간을 곧바로 물통에 담그는 이유가 여기에 있다. 보존 처리실에서는 나무 속의 물을 약품으로 천천히 바꿔 준다. 이 과정에만 몇 달에서 몇 해가 걸린다. 판독은 그 전에 끝내 두어야 한다. 처리를 거치면서 표면 색이 달라질 수 있기 때문이다.

검은 나무 위의 검은 글씨

광물이 스며든 나무는 색이 짙어진다. 갈색을 지나 새까맣게 변하는 것도 많다. 나무가 검게 변하는 이 현상을 흑화라고 한다. 문제는 글씨의 재료다. 옛사람들은 소나무를 태운 그을음을 아교에 개어 먹을 만들었다. 먹의 주성분은 탄소이고 색은 검다.

검게 변한 나무 위에 검은 먹으로 쓴 글씨가 있다. 밝은 조명을 아무리 비춰도 둘이 구별되지 않는다. 게다가 물속에서 오랜 세월을 보내는 동안 먹 알갱이가 상당수 씻겨 나갔다. 나뭇결의 어두운 세로줄이 붓 자국과 뒤섞인다. 흙과 부식물이 붓이 지나간 홈을 메운다. 사람 눈으로는 글자가 있는지조차 알기 어렵다.

판독이 어려우면 학설도 갈린다. 같은 목간을 두고 학자마다 다른 한자로 읽는 일이 생긴다. 한 글자가 바뀌면 문장 뜻이 바뀌고 역사 해석도 바뀐다. 그래서 목간 연구자들은 오래전부터 새로운 눈을 찾아 왔다. 사람 눈이 놓치는 것을 잡아 줄 도구가 필요했다.

읽어야 할 목간은 계속 늘고 있다. 경남 함안 성산산성에서는 신라 목간이 꾸준히 나왔다. 2025년 8월에도 명문 목간 두 점이 새로 확인되었다[2]. 명문은 유물에 적혀 있는 글자를 뜻한다. 충남 부여 관북리 유적에서는 백제 사비기 목간류가 무더기로 나왔다[1]. 경기 양주 대모산성에서는 5세기로 추정되는 목간이 나왔다. 사실이라면 지금까지 알려진 백제 문자 자료 가운데 이른 시기의 것이 된다[3].

일본에서도 사정은 같다. 나라문화재연구소는 각지에서 나온 목간 정보를 모아 공개한다. 이 데이터베이스의 이름은 목간고다[4]. 여기서는 목간에 적힌 글자로만 찾지 않는다. 용도와 연대와 재질과 출토 지역으로도 찾을 수 있다[5]. 읽어야 할 나무가 이렇게 쌓여 있다.

참고 자료

[1] 파이낸셜뉴스. (2026, 2월 5일). 부여 관북리 '백제 피리' 첫 확인..목간 329점 무더기 발견. 파이낸셜뉴스. https://www.fnnews.com/news/202602051448564101 (2026년 9월 3일 접속)

[2] 경남일보. (2025, 8월 7일). 함안 성산산성서 새로운 목간 2점 명문 확인. 경남일보. https://www.gnnews.co.kr/news/articleView.html?idxno=616028 (2026년 9월 3일 접속)

[3] 뉴시스. (2025, 11월 20일). 1500년 전 '백제 문자' 깨어났다…양주 대모산성서 5세기 목간 출토. 뉴시스. https://www.newsis.com/view/NISX20251120_0003410107 (2026년 9월 3일 접속)

[4] 奈良文化財研究所. 木簡庫(MOKKAN-KO). https://mokkanko.nabunken.go.jp/ja/ (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

2. 과학 수사대의 추적 1: 디지털 먹물 돋보기

눈에 보이지 않는 빛을 쓴다

사람 눈은 좁은 범위의 빛만 본다. 무지개의 빨강 바깥에도 빛이 있다. 이 빛을 적외선이라고 부른다. 파장이 길어서 눈에는 잡히지 않지만 카메라로는 찍을 수 있다. 리모컨이 텔레비전에 보내는 신호도 적외선이다.

목간 판독에 적외선을 쓰는 이유는 재료의 성질 때문이다. 먹의 탄소 알갱이는 적외선을 잘 흡수한다. 흑화된 나뭇결은 적외선을 많이 되돌려 보낸다. 사람 눈에는 똑같이 검은 두 물질이 적외선 카메라 앞에서는 갈라진다. 흡수한 쪽은 어둡게 찍히고 반사한 쪽은 밝게 찍힌다. 사라진 줄 알았던 붓 자국이 사진 위에 떠오른다.

주로 쓰는 빛은 파장 800나노미터 안팎의 근적외선이다. 나노미터는 1미터를 10억으로 나눈 길이다. 사람이 보는 빨간빛은 700나노미터 부근에서 끝난다. 그 바로 바깥의 빛을 쓴다는 뜻이다. 촬영은 바깥 빛을 막은 방에서 한다. 빛은 한쪽으로 쏠리지 않게 고르게 퍼뜨려 비춘다. 한쪽에서만 비추면 나뭇결의 골마다 그림자가 길게 생기기 때문이다.

적외선 촬영은 목간 연구의 기본 절차가 되었다. 발굴 보고서에 실리는 목간 사진 옆에는 대개 적외선 사진이 나란히 붙는다. 나라문화재연구소는 적외선 카메라로 목간 문자 이미지를 선명하게 만드는 장치를 일찍부터 개발해 왔다[1].

적외선에서 초분광으로

적외선 사진 한 장으로 모든 글자가 살아나지는 않는다. 얼룩의 위치와 나뭇결의 상태에 따라 잘 보이는 파장이 다르다. 그래서 등장한 방법이 초분광 영상이다.

보통 사진은 빨강과 초록과 파랑 세 값만 저장한다. 초분광 영상은 다르다. 한 장면을 수십에서 수백 개의 좁은 파장 구간으로 쪼개어 찍는다. 빛을 무지개색으로 잘게 갈라 따로 찍는다고 생각하면 된다. 화면의 한 점마다 파장별 밝기 값이 줄줄이 기록된다. 그래서 사진 한 장 대신 두꺼운 사진 묶음이 만들어진다.

이 값들을 견주면 먹이 남은 곳과 흙 얼룩이 앉은 곳을 구별할 수 있다. 먹과 흙은 파장마다 밝기가 달라지는 방식이 서로 다르기 때문이다. 어떤 글자는 900나노미터 부근에

서 잘 보인다. 어떤 글자는 다른 구간에서 살아난다. 연구자는 파장 구간을 골라 가며 제일 잘 읽히는 조합을 찾는다. 컴퓨터가 여러 구간을 자동으로 견주어 후보를 추리기도 한다. 사람이 한 장씩 눈으로 확인하던 일을 계산으로 줄인 것이다.

국립가야문화유산연구소는 2025년에 이 방법을 처음 썼다. 함안 성산산성 목간 판독에 초분광 영상을 넣은 것이다. 연구소는 이 방법이 기존 적외선 촬영보다 정밀하다고 밝혔다. 눈이나 일반 촬영으로 알아보기 힘든 문자까지 되살릴 수 있다는 설명이다[2].

글자 모양으로 글자를 찾는다

사진이 선명해져도 다음 문제가 남는다. 흐릿한 붓 자국이 어떤 한자인지 정해야 한다. 초서체로 흘려 쓴 글자는 모양이 크게 달라진다. 획이 몇 개 남지 않은 글자도 많다. 한자는 종류가 수만 자여서 사람이 사전을 뒤지는 데도 한계가 있다.

나라문화재연구소와 도쿄대학 사료편찬소는 이 일을 돕는 시스템을 만들었다. 이름은 모지조다. 사용자가 읽고 싶은 글자 그림을 잘라 올린다. 시스템은 목간과 고문서에서 모아 둔 문자 이미지를 뒤진다. 그 가운데 모양이 닮은 것을 찾아 보여 준다[3]. 사람이 후보를 보고 판단한다. 답을 정해 주는 기계가 아니라 비슷한 사례를 꺼내 주는 검색 도구다.

이런 도구가 바꾼 것은 작업 순서다. 예전에는 학자가 기억과 자전에 기대어 후보를 떠올렸다. 지금은 기계가 수만 건 가운데 닮은 글자를 먼저 골라 준다. 사람은 좁혀진 후보 안에서 문맥을 따져 고른다. 판독에 드는 시간이 크게 줄었다.

기계가 그린 획을 가려낸다

요즘의 생성형 인공지능은 흐린 그림을 또렷하게 만들어 준다. 끊어진 선을 이어 주고 빈 곳을 채워 준다. 목간 사진에 이 기술을 쓰면 결과가 보기 좋아진다. 여기에 함정이 있다.

기계가 채운 획은 나무 위에 실제로 남은 먹이 아니다. 학습한 글씨 수만 점의 평균에서 뽑아낸 추측이다. 나뭇결 무늬가 둥글게 남아 있으면 기계는 그것을 둥근 획으로 읽을 수 있다. 그러면 원래 없던 글자가 그럴듯하게 생겨난다. 사료가 오염되고 역사 해석이 틀어진다.

그래서 연구 현장에는 규칙이 있다. 원본 사진과 보정 결과를 반드시 나란히 싣는다. 실제로 확인되는 먹 자국과 기계가 만든 화소를 구분해 표시한다. 기계의 답은 결론이 아니라 후보 목록으로 다룬다. 최종 판독은 사람이 한다. 판독안을 발표할 때는 어떤 장비로 어떤 조건에서 찍었는지도 함께 밝힌다.

정부 차원의 준비도 진행 중이다. 국가과학기술자문회의는 2026년 4월 16일에 국가유산청의 시행계획을 의결했다. 2026년부터 2030년까지 인공지능과 로봇과 센서를 국가유산 조사와 복원에 넣는 계획이다. 올해 투자 규모는 140억 원이다. 그중 44억 원이 지능형 첨단 보존 기술 개발에 배정되었다[4]. 국립김해박물관은 2026년 6월 23일에 학술 심포지엄을 열었다. 주제는 가야유산 연구의 과학적 시선이었다. 비파괴 분석과 디지털 복원과 3차원 엑스레이 현미경 기술을 다루는 발표 일곱 건이 이어졌다[5]. 맨눈으로 볼 수 없던 유물 속을 들여다보는 방법이 목간 밖으로도 넓어지고 있다.

참고 자료

[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 (2026년 9월 3일 접속)

[2] 경남일보. (2025, 8월 7일). 함안 성산산성서 새로운 목간 2점 명문 확인. 경남일보. https://www.gnnews.co.kr/news/articleView.html?idxno=616028 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[4] 뉴시스. (2026, 4월 16일). "AI, 블록체인으로 국가유산 보존관리"…과기자문회의, 유산청 시행계획 의결. 뉴시스. https://www.newsis.com/view/NISX20260416_0003594216 (2026년 9월 3일 접속)

[5] 경남일보. (2026, 6월 17일). 국립김해박물관, 23일 '가야학술제전' 심포지엄 개최. 경남일보. https://www.gnnews.co.kr/news/articleView.html?idxno=639314 (2026년 9월 3일 접속)

3. 과학 수사대의 추적 2: KuroNet과 Miwo

이어 쓴 글씨는 나눌 수 없다

일본에는 에도 시대까지 쓰인 옛 문서가 수백만 권 남아 있다. 이 문서들은 쿠즈시지라는 흘림 글씨로 적혔다. 쿠즈시지는 8세기부터 천 년 넘게 쓰인 필기체다. 1900년에 일본이 문자 제도를 바꾸었다. 그 뒤로 학교에서 이 글씨를 가르치지 않게 되었다. 그래서 지금의 일본인 대다수는 150년 전 책도 읽지 못한다[1].

읽기 어려운 이유는 글자 모양만이 아니다. 붓을 종이에서 떼지 않고 여러 글자를 이어 쓰는 서법이 있다. 앞 글자의 끝 획과 뒷 글자의 첫 획이 하나로 이어진다. 그러면 글자

와 글자 사이에 경계선이 남지 않는다. 한 줄이 통째로 긴 선 하나처럼 보인다.

컴퓨터 문자 인식은 오랫동안 같은 순서로 작동했다. 먼저 글자 하나하나를 네모 상자로 잘라 낸다. 그다음 잘린 상자 안의 그림이 어떤 글자인지 맞힌다. 이 방식은 경계선이 있어야 돌아간다. 이어 쓴 흘림 글씨 앞에서 첫 단계부터 무너졌다. 앞이 틀리면 뒤도 틀리므로 정확도를 올릴 방법이 없었다.

글자를 나누지 않는 쿠로넷

일본 인문학 오픈데이터 공동이용센터의 연구진이 다른 길을 냈다. 타린 크라누왓과 알렉스 램과 기타모토 아사노부가 함께 모델을 만들었다. 이름은 쿠로넷이다. 2019년 문서 분석 인식 국제학회에서 처음 발표했다[1]. 2020년에는 구조를 다듬고 결과를 학술지에 다시 실었다[2].

쿠로넷은 글자를 미리 자르지 않는다. 고문서 한 쪽을 통째로 입력받는다. 모델의 중심 구조는 유넷이다. 유넷은 그림을 점점 줄여 특징을 뽑는다. 그다음 다시 원래 크기로 키우는 신경망이다. 줄이는 쪽과 키우는 쪽이 대칭이어서 알파벳 U 모양으로 그려진다. 줄이는 동안 붓 획의 굵기와 방향을 익힌다. 키우는 동안 그 정보를 원래 위치에 되돌린다.

이 신경망은 쪽 전체에서 두 가지를 한꺼번에 예측한다. 하나는 글자의 중심이 어디에 있는지다. 다른 하나는 그 위치의 글자가 무엇인지다. 결과는 확률 지도로 나온다. 글자가 있을 가능성을 화면 곳곳에 표시한 그림이다. 확률이 높은 곳을 따라가면 글자의 위치와 이름이 동시에 정해진다. 자르는 단계가 없으니 이어 쓴 글씨에서도 무너지지 않는다.

학습에는 일본 국문학연구자료관이 만든 자료를 썼다. 일본 고전적 쿠즈시지 데이터셋이라는 이름이다. 100만 자가 넘는 글자 표본이 들어 있다[2]. 같은 연구진은 기계학습 연구자들이 쉽게 쓰도록 축소판도 공개했다. 손글씨 숫자 자료를 본떠 만든 쿠즈시지 엠니스트가 그것이다. 글자 수를 늘린 판과 한자만 모은 판도 함께 내놓았다[3]. 덕분에 일본 고전을 모르는 연구자도 이 문제에 뛰어들 수 있었다.

쿠로넷은 2022년 10월에 임무를 넘겼다. 인식 모델의 이름은 루리로 바뀌었다[4]. 루리는 물체 검출 기술을 흘림 글씨에 맞게 다듬은 모델이다. 글자 뒤의 색이나 무늬가 복잡한 경우에도 이전보다 잘 읽는다. 같은 데이터셋을 보강해 학습했고 예전에 못 읽던

글자를 읽어 낸다[5].

손안에 들어온 미오

서버에서 돌던 모델을 손전화에 넣는 일이 다음 과제였다. 그렇게 나온 앱이 미오다. 2021년 8월 30일에 처음 공개했다[5]. 일본어 표기는 みを이고 영문 표기는 Miwo다[6].

쓰는 법은 간단하다. 옛 문서를 손전화 카메라로 찍는다. 잠시 뒤 화면에 요즘 글자로 옮긴 결과가 나타난다. 원본과 결과를 밀대로 겹쳐 견줄 수 있다. 변체가나가 무엇이었는지도 표시된다. 변체가나는 지금은 쓰지 않는 옛 히라가나 모양이다. 결과를 글자 파일로 내보내는 기능도 있다. 촬영한 사진은 서버에 남기지 않는다. 안드로이드와 아이폰에서 모두 무료로 쓸 수 있다[6].

2022년 10월 26일에 나온 1.1판부터 인식 모델이 루리로 바뀌었다[5]. 같은 달 이 앱은 2022년도 굿디자인상을 받았다[7]. 이용 기록은 꾸준히 쌓였다. 공개 이후 인식한 이미지가 2026년 9월 3일 기준으로 3,523,373건이다[6]. 도서관과 사찰과 지역 문서고에서 사람들이 옛 문서를 찍고 있다는 뜻이다.

시민이 함께 읽는다

기계가 먼저 읽고 사람이 고치는 방식은 다른 곳에서도 쓰인다. 민나데혼코쿠라는 온라인 번각 사업이 있다. 번각은 옛 글씨를 요즘 글자로 옮겨 적는 일이다. 표기는 みんなで翻刻이고 뜻은 다 함께 옮겨 적기다. 2017년에 교토대학 고지진연구회에서 시작했다. 지금은 국립역사민속박물관과 도쿄대학 지진연구소가 함께 운영한다[8].

참여 방법은 열려 있다. 인터넷 브라우저만 있으면 누구나 옛 문서를 옮겨 적을 수 있다. 화면에는 인공지능이 먼저 읽은 초벌 결과가 뜬다. 사람은 틀린 글자만 고치면 된다. 이 사이트는 인문학 오픈데이터 공동이용센터의 쿠즈시지 인식 서비스를 끌어다 쓴다. 지금까지 참가자 1만 800여 명이 4,800만 자가 넘는 글을 옮겨 적었다[8].

옮겨 적은 글은 그냥 쌓이지 않는다. 사람이 고친 정답이 다시 학습 자료가 된다. 학습 자료가 늘면 모델이 좋아진다. 모델이 좋아지면 초벌 결과가 정확해지고 사람의 손이 덜 간다. 이 순환이 돌기 시작하면 속도가 붙는다.

읽어 낸 문서의 쓰임도 넓다. 이 사업은 지진 연구자들이 시작했다. 옛 일기와 관청 문서에는 지진과 해일과 흉년의 날짜와 피해가 적혀 있다. 흘림 글씨 때문에 손대지 못하던 기록이 데이터로 바뀌면 과거의 재난 주기를 셀 수 있다. 오래된 붓글씨가 오늘의 방재 자료가 된다.

여기까지가 일본 종이 고문서의 이야기다. 쿠로넷과 루리와 미오는 모두 종이에 쓴 쿠즈시지를 읽는 모델이다. 진흙 속 목간을 이 모델이 자동으로 읽어 낸 사례는 아직 확인되지 않는다. 목간은 글자 수가 적고 바탕이 나무이며 먹이 많이 지워졌다. 학습에 쓸 정답 자료도 종이 고문서만큼 쌓이지 않았다. 한국과 일본의 목간 데이터가 더 모이면 사정이 달라질 수 있다. 지금은 기대이지 성과가 아니다.

참고 자료

[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/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[7] 国立情報学研究所. (2022, 10월 26일). 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 (2026년 9월 3일 접속)

[8] みんなで翻刻. みんなで翻刻について. https://honkoku.org/ (2026년 9월 3일 접속)

4. 밝혀진 진실

월지 목간이 알려 준 왕실 살림

경주 동궁과 월지는 통일신라 왕실의 별궁과 연못이다. 1975년 월지 발굴에서 목간이 처음 나왔다. 그 뒤로도 조사가 이어지면서 수량이 늘었다. 국립경주박물관은 이곳에서

나온 유물을 월지관에서 전시한다[1].

월지 목간은 왕실의 하루를 보여 준다. 어떤 목간은 항아리에 매달던 꼬리표다. 꼬리표 목간은 끝에 좌우로 홈이 파여 있다. 그 홈에 끈을 감아 물건에 묶었다. 오늘날의 택배 송장과 쓰임이 비슷하다. 월지에서 나온 목간에는 녹갑과 옹이라는 글자가 있다. 녹갑은 사슴으로 만든 젓갈이고 옹은 항아리다. 신라 왕실이 젓갈의 종류와 담근 때를 적어 관리했다는 뜻이다[2].

이런 기록은 역사책에 나오지 않는다. 삼국사기는 왕과 전쟁과 제도를 적었다. 부엌에 무엇이 들어왔는지는 적지 않았다. 목간은 그 틈을 메운다. 진흙 속 나무판 하나가 궁궐의 살림 장부가 된다.

한국목간학회와 국립경주박물관은 2026년 4월 3일과 4일에 국제학술대회를 열었다. 주제는 신라 왕경 경주의 공간과 문자 생활이었다. 경주화백컨벤션센터에서 이틀 동안 진행했다. 월성 해자와 동궁과 월지에서 나온 자료를 함께 놓았다. 목간과 비석과 묵서로 신라 왕경의 모습을 다시 그리려는 시도였다. 한국과 일본의 연구자가 함께 발표했다[3].

성산산성 목간의 처벌 보고

함안 성산산성은 신라가 6세기에 쌓은 산성이다. 성벽 안쪽 저습지에서 목간이 계속 나왔다. 이곳에서 나온 묵서 목간은 2025년 8월 기준으로 247점이다. 우리나라에서 나온 고대 목간의 절반이 넘는 수량이다[4].

제18차 발굴조사에서 새로 판독한 두 점은 소나무류로 만들었다. 한 점은 네 면에 글을 쓴 다면 목간이다. 다른 한 점은 양면 목간이다[4]. 다면 목간은 네 면 가운데 세 면에 행정 내용을 담고 있었다. 사람에게 처벌을 내린 기록이다[5]. 만든 시기는 6세기 중반으로 본다[4].

이 판독은 앞에서 본 초분광 영상으로 얻은 결과다. 눈으로도 일반 사진으로도 알아볼 수 없던 문자가 살아났다[4]. 천 년 넘게 진흙 속에 있던 신라 하급 관리의 보고서가 문장으로 돌아왔다. 성산산성 목간에는 곡식을 보낸 고을 이름도 많이 적혀 있다. 6세기 신라가 지방을 어떻게 나누어 다스렸는지 보여 주는 자료다. 세금이 어느 길로 성까지 왔는지도 짚어 볼 수 있다.

백제 왕궁의 문서 더미

부여 관북리 유적은 백제 사비기 왕궁터로 알려진 곳이다. 국립부여문화유산연구소는 2024년과 2025년에 제16차 발굴조사를 했다. 여기서 목간류 329점이 나왔다. 글이 적힌 목간이 82점이고 삭설이 247점이다[6]. 국내 단일 유적 최대 수량이고 사비기 목간 가운데 이른 시기의 것이다[7].

목간에는 연대를 적은 것이 섞여 있다. 경신년과 계해년이라는 간지가 나온다. 각각 540년과 543년으로 본다. 백제가 웅진에서 사비로 도읍을 옮긴 직후의 기록이라는 뜻이다. 관리의 인사 기록과 나라 살림을 적은 장부와 관직 이름이 함께 확인되었다[7]. 대나무로 만든 백제 관악기 실물도 처음 나왔다[8].

2026년 7월 22일과 23일에는 국제학술대회가 열렸다. 국립부여문화유산연구소와 한국목간학회와 백제학회가 함께 마련했다. 주제는 부여 관북리유적 악기와 목간이었다. 부산 해운대구에서 열렸다. 박순발 충남대 명예교수의 기조 강연으로 시작했다. 세 분과에서 열두 편이 발표되었다. 한국과 중국과 일본의 연구자가 관북리 목간의 뜻을 함께 검토했다[8].

비석 앞에서는 조심한다

목간에서 통한 방법이 옛 글자 모두에 통하지는 않는다. 바탕이 돌로 바뀌면 사정이 달라진다. 광개토대왕비가 그 사례다. 414년에 세운 이 비석에는 고구려의 역사가 새겨져 있다. 비면이 오래 비바람을 맞아 글자가 닳았다. 여기서는 먹이 아니라 파인 홈을 읽어야 한다.

그래서 연구자들은 탁본을 견준다. 원석탁본은 비면에 석회를 바르기 전에 뜬 탁본이다. 원래 돌에 종이를 대고 직접 두드려 떠냈다. 지금까지 확인된 것이 열 본 남짓이다. 19세기 후반에는 다른 것이 나왔다. 글자를 또렷하게 보이려고 비면에 석회를 발라 뜬 석회탁본이다. 쌍구가묵본이라는 것도 여럿 만들어졌다. 비문에 종이를 대고 글자 둘레에 선을 그린 것이다. 그 여백을 진한 먹으로 메워 탁본처럼 보이게 만들었다. 이것은 두드려 뜬 탁본이 아니라 베껴 그린 복제본이다. 만드는 사이에 글자가 잘못 옮겨질 수 있다[9]. 그래서 같은 글자가 탁본마다 달라 보이는 일이 생겼다.

논쟁의 중심은 신묘년조 문장이다. 이 문장의 해석이 고대 한일 관계사와 얽히면서 오랫동안 다투어졌다. 글자가 사람 손으로 바뀌었는지가 쟁점이었다[10]. 본래 글자를 알

아내려면 이른 원석탁본과 여러 석회탁본을 견주어야 한다. 각 탁본을 가려낼 기준을세우는 작업이 먼저다[9].

3차원 계측과 인공지능이 이 논쟁에 답을 주었다는 이야기가 돌아다닌다. 확인해 보면 근거로 제시된 연구를 찾을 수 없는 경우가 많다. 목간에서 통했던 방법이 비석에서도 통할 것이라는 기대와 실제 연구 결과는 다르다. 앞에서 본 기술은 안 보이던 것을 보이게 만든다. 보이게 만든 것의 뜻을 정하는 일은 그다음이다.

그래서 결론은 두 겹이다. 적외선과 초분광 영상은 신라와 백제 목간에서 사라진 먹을 되살렸다. 무분할 인식 모델은 일본 종이 고문서의 나눌 수 없던 흘림 글씨를 읽었다. 두 기술이 다룬 대상은 다르지만 방향은 같다. 사람 눈이 놓친 획을 기계가 먼저 집어 준다. 그와 동시에, 기계가 그린 획과 원본에 남은 획을 구분하는 규율이 없으면 같은 기술이 없던 역사를 만든다. 붓글씨를 깨운 것은 기계이고 그 글을 책임지는 쪽은 사람이다.

참고 자료

[1] 국립경주박물관. 월지관. 국립경주박물관. https://gyeongju.museum.go.kr/kor/html/sub02/0201.html (2026년 9월 3일 접속)

[2] 국사편찬위원회. 목간. 우리역사넷. https://contents.history.go.kr/mobile/tt/view.do?levelId=tt_b36 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[4] 경남일보. (2025, 8월 7일). 함안 성산산성서 새로운 목간 2점 명문 확인. 경남일보. https://www.gnnews.co.kr/news/articleView.html?idxno=616028 (2026년 9월 3일 접속)

[5] 경남도민일보. (2025, 8월 7일). 누군가를 처벌한 신라 행정 보고, 함안 성산산성 목간서 확인. 경남도민일보. https://www.idomin.com/news/articleView.html?idxno=943461 (2026년 9월 3일 접속)

[6] 파이낸셜뉴스. (2026, 2월 5일). 부여 관북리 '백제 피리' 첫 확인..목간 329점 무더기 발견. 파이낸셜뉴스. https://www.fnnews.com/news/202602051448564101 (2026년 9월 3일 접속)

[7] 경향신문. (2026, 2월 17일). 부여 관북리서 발견된 목간 329점, '백제 피리'만큼이나 중요한 이유. 경향신문. https://www.khan.co.kr/article/202602171427001 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[9] 국립중앙박물관. 광개토대왕비 탁본. 국립중앙박물관 큐레이터 추천 소장품. https://www.museum.go.kr/MUSEUM/contents/M0501000000.do?schM=view&relicRecommendId=254462 (2026년 9월 3일 접속)

[10] 국사편찬위원회. 광개토대왕릉비. 우리역사넷. https://contents.history.go.kr/mobile/kc/view.do?levelId=kc_r100300 (2026년 9월 3일 접속)

제 5 부

침묵하는 문자들의 최후 반격: 미해독 문자의 전산 언어학

09 CHAPTER

제9장. 에게해의 잃어버린 수수께끼: 선형문자 A와 기하학적 격자

선형문자 A를 소리의 격자에 얹어 풀어 보려는 시도

1. 사건 파일

크레타섬 궁전에서 나온 세 가지 글자

에게해 남쪽에 크레타섬이 있다. 청동기 시대에 이 섬에서 미노아 문명이 자랐다. 미노아 사람들은 배를 만들어 지중해를 오갔다. 그들은 올리브기름과 포도주와 곡물을 실어 날랐다. 궁전은 이 물건을 거두었다가 사람들에게 나누어 주었다.

1900년 무렵 영국 고고학자 아서 에번스가 크노소스 궁전 터를 팠다. 그는 땅속에서 글자가 새겨진 점토판을 찾아냈다. 글자는 한 종류가 아니었다. 에번스는 찾아낸 글자를 셋으로 나누었다. 크레타 상형문자, 선형문자 A, 선형문자 B다[1].

미노아 문명은 천 년 넘게 이어졌다. 그러다 청동기 시대 후반에 큰 변화를 겪었다. 북쪽 산토리니섬에서 화산이 터진 것이 그 하나다. 터진 때는 기원전 17세기에서 16세기 사이로 본다. 정확한 연대는 지금도 논쟁 중이다[7]. 화산재가 크레타섬까지 날아왔고 해일이 해안을 때렸다고 본다. 기원전 1450년 무렵에는 미케네 그리스인이 섬을 차지했다[1]. 화산과 미케네의 지배 사이에는 백 년이 넘는 시간이 놓여 있다. 두 사건을 곧바로 원인과 결과로 묶기는 어렵다. 그 뒤로 미노아 사람들의 말은 기록에서 사라졌다.

세 글자는 쓰인 시기가 서로 다르다. 선형문자 A는 기원전 1750년 무렵부터 기원전 1450년 무렵까지 쓰였다. 선형문자 B는 그보다 늦게 나타났다. 미케네 그리스인이 크레타섬을 차지한 뒤에 쓰인 글자다. 두 글자는 모양이 닮은 기호를 여럿 나눠 가지고 있다[1].

점토판이 남은 것은 우연이었다. 미노아 사람들은 점토판을 굽지 않았다. 젖은 흙에 글자를 새겨 말린 뒤 한 해쯤 쓰다 버렸다. 그런데 궁전에 불이 나면서 흙판이 도자기처럼

구워졌다. 불에 구워진 판만 삼천오백 년을 버텼다. 오늘 남은 자료는 불이 남긴 자료다.

에번스가 나눈 세 글자 가운데 크레타 상형문자가 제일 먼저 나타났다. 사람 머리나 그릇 모양이 그대로 살아 있는 글자다. 선형문자 A는 그 그림을 선으로 줄여 놓은 모습에 가깝다. 두 글자는 이백 년 가까이 같은 시기에 함께 쓰였다. 한쪽이 사라진 뒤에 다른 쪽이 생긴 것이 아니다. 선형문자 A가 크레타 상형문자를 본떠 만들어졌다고 보는 견해가 있다. 그러나 세 글자가 곧은 한 줄로 이어진다고 단정하지는 못한다. 기호를 나눠 가진 것은 분명하지만 이어진 길은 아직 밝혀지지 않았다[1].

선형문자 B는 읽혔다

선형문자 B는 1950년대에 읽혔다. 영국의 건축가 마이클 벤트리스가 해냈다. 그는 선형문자 B가 적어 놓은 말이 그리스어의 옛 형태임을 밝혔다. 고전학자 존 채드윅이 그의 연구를 도왔다[1].

벤트리스는 표를 하나 그려 놓고 시작했다. 가로줄에는 자음을, 세로줄에는 모음을 놓은 표다[8]. 이런 표를 음절 격자라고 부른다. 소리마디 하나하나가 들어갈 칸을 미리 마련해 둔 그림이다. 벤트리스는 어떤 기호가 어느 칸에 들어갈지를 통계로 좁혀 나갔다[1].

그는 좁혀 놓은 칸에 크노소스와 파이스토스 같은 지명을 끼워 맞추었다. 지명이 들어맞자 나머지 기호도 줄줄이 풀렸다. 학자들은 같은 방법을 선형문자 A에 넣어 보았다. 선형문자 B의 소리 값을 닮은 기호에 그대로 옮겨 본 것이다[3]. 소리는 나왔다. 그러나 그 소리가 어떤 말인지 아무도 알아듣지 못했다[1].

소리 값을 옮겨 놓고도 뜻이 나오지 않는 상황은 낯설다. 글자를 읽는 일과 말을 알아듣는 일이 다르기 때문이다. 한글을 아는 사람은 낯선 외국어를 한글로 적어도 읽을 수 있다. 소리는 나오지만 뜻은 모른다. 선형문자 A 앞에 선 학자들이 놓인 처지가 이와 같다.

두 언어로 적힌 돌이 없다

옛 글자를 읽어낼 때 학자들은 대개 짝이 되는 문서를 찾는다. 같은 내용을 두 언어로 나란히 적어 둔 문서다. 이런 문서를 이중 언어 비문이라고 부른다. 이집트 상형문자는 로제타석 덕분에 풀렸다. 로제타석에는 같은 법령이 세 가지 글자로 새겨져 있었다. 그중 하나가 학자들이 이미 읽을 줄 아는 고대 그리스어였다.

메소포타미아 쐐기문자도 비슷한 길을 걸었다. 이란의 베히스툰 절벽에 세 언어로 새긴 비문이 남아 있었다. 이 비문은 세 언어 모두 처음에는 읽을 수 없었다. 그중 고대 페르시아어가 먼저 풀렸다. 기호가 적고 낱말 사이가 나뉘어 있었기 때문이다. 학자들은 먼저 푼 고대 페르시아어를 발판으로 나머지를 읽었다.

크레타섬에서는 이런 문서가 나오지 않았다. 선형문자 A 옆에 그리스어 대역을 적어 둔 유물이 없다. 미노아 사람들이 남긴 글은 미노아 사람들의 말로만 적혀 있다. 비교할 짝이 없으니 뜻을 확인할 길이 막힌다[1].

단서가 될 만한 표시도 보이지 않는다. 이집트 상형문자를 읽은 샹폴리옹에게는 카르투슈가 있었다. 왕의 이름을 둘러싼 길쭉한 테두리다. 이 테두리 안에 있는 글자는 사람 이름이라는 사실이 미리 정해져 있었다. 샹폴리옹은 여기서 소리 값 몇 개를 건져 나머지를 풀었다. 선형문자 A 점토판에는 이런 테두리가 없다[1].

문제는 하나 더 있다. 선형문자 A가 적은 말의 뿌리를 모른다는 점이다. 학자들은 이 말을 미노아어라고 부른다. 미노아어가 어느 어족에 속하는지는 밝혀지지 않았다. 인도유럽어족이라는 주장도 있고, 서아시아 계통이라는 주장도 있다. 그중 근거가 낫다고 꼽히는 것은 셈어 계통과 리키아어 계통이다. 리키아어는 지금의 튀르키예 남서쪽에서 쓰이던 옛말이다[1]. 어느 쪽도 학계의 동의를 얻지 못했다.

선형문자 A는 크레타섬에서만 나오지 않는다. 에게해의 여러 섬과 그리스 본토에서도 발견되었다[1]. 화산에 묻힌 산토리니섬 유적에서도 나왔다. 미노아 사람들이 오간 뱃길을 따라 글자도 퍼진 것이다. 청동기 시대 지중해가 얼마나 넓게 이어져 있었는지를 보여 주는 자료다.

학자들이 선형문자 A에 매달리는 까닭도 여기에 있다. 선형문자 A는 유럽 땅에서 나온 오래된 글자에 든다. 미노아 사람들이 자기 말로 남긴 기록은 이것뿐이다. 선형문자 A가 풀리면 그들이 무엇을 믿고 무엇을 사고팔았는지 드러난다. 그리스 신화에 남은 미노타우로스 이야기의 뿌리도 여기에 있다.

1,400점의 짧은 장부

남아 있는 자료의 양도 넉넉하지 않다. 명문은 유물에 새기거나 그려 넣은 글을 말한다. 지금까지 확인된 선형문자 A 명문은 1,370점가량이다. 거기에 적힌 기호를 다 합쳐도 7,400개 안팎이다[1]. 이집트나 메소포타미아의 기록물과는 견줄 수 없는 양이다.

글의 길이도 짧다. 점토판 대다수는 창고 장부다. 무엇이 몇 개 들어왔고 누구에게 얼마가 나갔는지를 적었다. 사람 이름과 물건 이름과 숫자가 줄지어 있을 뿐이다. 문장다운 문장이 드물다. 문장이 없으면 문법을 알아낼 방법도 줄어든다.

기호의 수를 셀 때는 층을 나누어야 한다. 소리마디를 적는 기호는 97개다. 물건 이름을 통째로 나타내는 기호는 50개쯤이다. 수를 적는 기호가 4개, 분수를 적는 기호가 17개 있다. 두 기호를 겹쳐 만든 조합도 150개가량 알려져 있다[6]. 이것을 다 모은 표준 기호 목록은 300개에 이른다[2]. 이 가운데 자주 나오는 기호는 얼마 되지 않는다. 나머지는 몇 번 나오다 마는 기호다. 컴퓨터에게 통계를 맡기려면 같은 기호가 여러 번 나와 주어야 한다. 몇 번 나오지 않는 기호는 통계가 잡아내지 못한다.

그래서 연구자들은 자료 자체를 다시 만들기로 했다. 손으로 모사한 그림을 컴퓨터가 다룰 수 있는 형태로 바꾸는 작업이다[2]. 2026년에는 새김 자국을 빛으로 다시 읽는 방법도 나왔다[4]. 눈으로 놓치던 얕은 홈까지 드러내는 방법이다[5]. 에게해의 침묵을 깨우는 일은 이렇게 자료를 다듬는 데서 다시 시작되었다.

참고 자료

[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/ (2026년 9월 3일 접속)

[3] Younger, J. G. (n.d.). Linear A texts and inscriptions in phonetic transcription. University of Kansas. http://people.ku.edu/~jyounger/LinearA/ (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[5] GreekReporter. (2026, 3월 14일). 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/ (2026년 9월 3일 접속)

[6] Del Freo, M. (2022). Linear A. Mnamon: Ancient writing systems in the Mediterranean. Scuola Normale Superiore. https://doi.org/10.25429/sns.it/lettere/mnamon022 (2026년 9월 3일 접속)

[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

2. 과학 수사대의 추적: 비지도 음운 격자 맞추기

사전 없이 짝을 찾는 기계

번역기를 만들려면 보통 대역 문장이 필요하다. 한국어 문장과 영어 문장을 나란히 쌓아 놓고 기계에게 보여 준다. 이런 자료를 병렬 말뭉치라고 부른다. 선형문자 A에는 병렬 말뭉치가 없다[3].

그래서 연구자들은 정답표 없이 배우는 방법을 골랐다. 정답을 미리 주지 않고 통계만으로 규칙을 찾게 하는 방식이다. 이런 방식을 비지도 학습이라고 부른다. 기계는 기호가 어떤 순서로 붙어 나오는지를 센다. 그리고 다른 글자 체계의 통계와 견주어 짝을 찾아본다.

2019년에 미국 매사추세츠공과대학 연구진이 이 방법을 내놓았다. 지아밍 루오와 위안 차오와 레지나 바르질라이가 함께 만든 모형이다. 이들은 두 글자 체계를 잇는 일을 최소 비용 흐름 문제로 바꾸었다. 기호와 기호 사이에 물길을 놓고, 물이 흐르는 값을 매기는 방식이다. 값이 제일 적게 드는 물길이 올바른 짝일 확률이 높다[1].

연구진은 이 모형을 우가리트 문자에 걸어 보았다. 앞선 최고 성적을 5퍼센트포인트 앞섰다. 선형문자 B에도 걸어 보았다. 그리스어와 뿌리가 같은 낱말 가운데 67.3퍼센트를 옳게 옮겼다[1]. 다만 선형문자 B에는 답이 있었다. 그 글자가 그리스어를 적었다는 사실을 연구진이 미리 알고 있었다.

연구진이 선형문자 B를 시험대로 고른 데는 까닭이 있다. 답을 아는 문제라야 모형의 성적을 잴 수 있다[1]. 답이 없는 문제에 바로 걸면 결과가 맞는지 알 방법이 없다. 그래서 이미 읽힌 글자로 먼저 성능을 재고 미해독 글자로 옮겨 간다[3].

모형이 내놓는 것은 완성된 번역문이 아니다. 기호마다 붙을 만한 소리 값의 후보 목록이다[1]. 사람이 이 목록을 받아 유물과 견주며 걸러 낸다. 컴퓨터는 사람이 볼 분량을 줄여 주는 일을 맡는다[3].

소리의 생김새를 좌표로 바꾸다

선형문자 A는 사정이 다르다. 어떤 말과 견주어야 하는지부터 밝혀지지 않았다[3]. 2021년에 같은 연구실이 이 조건을 다루는 모형을 내놓았다. 루오와 프레데리크 하르트만과 엔리코 산투스가 바르질라이, 차오와 함께 만들었다[2].

이 모형은 두 가지 어려움을 한꺼번에 다룬다. 첫 어려움은 낱말 경계가 없다는 점이다. 옛 글에는 띄어쓰기가 없는 경우가 많다. 둘째 어려움은 친척 언어가 정해지지 않았다는 점이다[2].

연구진은 소리의 생김새를 숫자 좌표로 바꾸어 넣었다. 국제음성기호를 바탕으로 자음과 모음의 성질을 좌표에 놓은 것이다. 입술로 내는 소리끼리는 좌표가 가깝다. 혀뿌리로 내는 소리와는 좌표가 멀다. 이 좌표가 있으면 기계가 엉뚱한 짝을 덜 만든다[2].

연구진은 이 모형을 고트어와 우가리트 문자로 시험했다. 그다음 아직 뜻이 풀리지 않은 이베리아어에 걸었다. 이베리아 문자는 기호의 소리 값이 꽤 알려져 있다. 풀리지 않은 것은 그 소리가 적어 놓은 말이다. 연구진은 이베리아어가 바스크어와 친척이라는 증거를 찾지 못했다. 이 결과는 학계의 주류 견해와 어긋나지 않았다. 어떤 언어가 친척일 만한지 후보를 견주는 절차도 함께 보였다[2]. 같은 방법은 선형문자 A 연구에도 그대로 쓸 수 있다[3].

빈칸이 알려 주는 규칙

기계가 견주는 것은 기호의 모양만이 아니다. 기호가 이웃과 어울리는 버릇도 계산에 들어간다. 어떤 기호는 낱말 앞머리에만 나온다. 어떤 기호는 끝에만 붙는다. 두 기호가 나란히 붙는 일이 아예 없기도 하다[3].

이런 통계를 함께 나오는 정도, 곧 공기 통계라고 부른다. 공기 통계를 모으면 표의 빈칸이 드러난다. 같은 가로줄에 놓인 기호들은 자음이 같을 확률이 높다. 같은 세로줄에 놓인 기호들은 모음이 같을 확률이 높다[8]. 벤트리스가 손으로 하던 일을 컴퓨터가 훨씬 빠르게 되풀이하는 것이다[3].

컴퓨터는 이 계산을 사람보다 훨씬 빨리 끝낸다. 기호가 300개면 짝의 수는 수만 가지가 된다. 사람이 종이에 적어 가며 세기에는 벅찬 양이다. 벤트리스는 이 일을 여러 해에 걸쳐 해냈다[3].

여기까지는 소리의 짜임을 그리는 일이다. 짜임을 그려도 그 소리가 무엇을 뜻하는지는 나오지 않는다. 격자를 채우는 일과 말을 알아듣는 일은 별개다[3].

글자 모양을 데이터로 만들다

기계가 통계를 내려면 자료가 기계가 읽을 수 있는 꼴이어야 한다. 종이책에 실린 그림만으로는 계산을 시킬 수 없다. 이 일을 맡은 것이 SigLA라는 자료 창구다. 케임브리지의 에스테르 살가렐라와 렌의 시몽 카스텔랑이 함께 만들었다[4].

SigLA에는 표준 기호 300개와 손으로 모사한 명문 400여 점이 들어 있다. 명문 안에서 낱개 기호 3,000개 이상을 찾아볼 수 있다. 자료와 그림은 누구나 쓸 수 있게 공개되어 있다. 기호 하나의 획을 따로 떼어 견줄 수도 있다[4].

2026년 3월에 나온 연구는 자료의 결을 한 겹 더 넓혔다. 라비니아 조르지와 사라 로페스가 반사 변환 이미징을 에게해 글자에 걸었다. 같은 유물에 빛을 여러 방향에서 비추어 사진을 여러 장 찍는 방법이다. 이 사진을 합치면 표면의 자잘한 굴곡이 드러난다. 새김칼이 어느 획을 먼저 그었는지까지 따져 볼 수 있다[7].

이런 자료가 쌓이면 다른 문자에도 쓸모가 생긴다. 2022년에 볼로냐대학 연구진이 키프로미노아 문자를 다시 들여다보았다. 미켈레 코라차와 파비오 탐부리니와 미겔 발레리오와 실비아 페라라가 함께한 연구다. 이들은 사전 지식을 넣지 않은 신경망으로 기호 모양을 견주었다. 그 결과 세 무리로 나누어 오던 기존 분류를 다시 검토해야 한다는 결론을 얻었다. 하나의 글자 체계로 보는 편이 통계에 맞는다는 것이다[5].

컴퓨터가 내놓은 결과는 곧바로 논문이 되지 않는다. 고고학자와 문헌학자가 다시 들여다보는 과정을 거친다[3]. 어느 점토판이 어느 방에서 나왔는지를 맞춰 본다. 같은 서기가 쓴 글씨인지 획의 버릇을 견준다[7]. 이 검토를 통과한 뒤라야 학계의 자료로 들어간다[3].

에게해 세 문자를 함께 다루는 연구도 이어지고 있다. 크레타 상형문자와 선형문자 A와 키프로미노아 문자를 한 자리에 놓는 방식이다. 세 문자를 견주면 한 문자만 볼 때 안 보이던 규칙이 드러난다. 컴퓨터는 이런 대량 비교에서 힘을 낸다[3].

기계가 대신해 주지 못하는 일

이런 방법이 선형문자 A의 뜻까지 밝혀 주지는 않았다. 넘어야 할 벽이 셋 남아 있다.

첫째는 견줄 상대다. 비지도 정렬은 미해독 글자와 아는 글자 사이에 친연성이 있을 때 힘을 낸다[1]. 미노아어에는 살아남은 친척 언어가 없다[3]. 좌표는 그려지지만 그 좌표를 고정해 줄 기준점이 없다.

둘째는 자료의 양이다. 명문 1,370점은 기계학습에 쓰기에는 적다[3]. 자료가 적으면 큰 신경망이 그럴듯한 헛것을 지어낸다. 이런 오류를 환각이라고 부른다.

셋째는 검증이다. 2026년 7월 더컨버세이션에 실린 글이 이 문제를 짚었다. 더블린시티대학의 제인 애드킨스가 쓴 글이다. 그는 통계로 패턴을 맞추는 일이 없는 뜻을 만들어 내지는 못한다고 적었다. 아는 언어나 대역 문서 같은 발판이 있어야 우연한 일치와 진짜 규칙을 갈라낼 수 있다는 지적이다. 기계는 후보를 줄여 줄 뿐이고, 옳은지 가리는 일은 사람의 몫으로 남는다[6].

참고 자료

[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/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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, 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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[8] Stavrakakis, E. (2024). Ventris' Grids. Bulletin of the Institute of Classical Studies, 67(1), 86-103. https://doi.org/10.1093/bics/qbae033

3. 밝혀진 진실

숫자와 분수는 풀렸다

선형문자 A에서 먼저 풀린 쪽은 수를 세는 방식이다. 미노아 사람들은 열을 묶어 세는 십진법을 썼다. 세로줄 하나가 1이고 가로줄 하나가 10이다. 동그라미는 100이고, 빛살이 뻗은 동그라미는 1000이다[3]. 이 규칙은 그림만 보아도 짐작할 수 있었다.

어려웠던 것은 분수 기호였다. 점토판에는 열일곱 개의 분수 기호가 나온다[8]. 두 기호를 겹쳐 쓴 것까지 세면 서른 개 가까이 된다[9]. 어느 기호가 얼마를 뜻하는지가 오래 풀리지 않았다.

2021년에 볼로냐대학 연구진이 이 문제를 다루었다. 미켈레 코라차와 실비아 페라라와 바르바라 몬테키와 파비오 탐부리니와 미겔 발레리오가 함께한 연구다. 이들은 기호의 생김새를 살피는 옛 방법과 컴퓨터 계산을 함께 걸었다[1].

연구진은 먼저 기호들이 점토판에서 지키는 규칙을 정리했다. 어떤 기호끼리 붙어 나오고 어떤 기호끼리는 붙지 않는지를 셌다. 그다음 값의 조합을 컴퓨터로 따져 보았다. 후보는 400만 가지에 가까웠다. 통계 검정과 다른 문명의 분수 체계 비교로 이 후보를 좁혀 나갔다[1].

남은 답은 이랬다. 미노아 분수 체계에서 제일 작은 값은 60분의 1이었다. 이 체계로 60을 분모로 삼는 값 대다수를 적을 수 있었다. 10분의 1을 뜻하는 기호는 선형문자 B에도 그대로 이어졌다. 선형문자 B에서 이 기호는 마른 곡물을 재는 단위로 쓰였다. 두 문자의 기록 사이에는 250년쯤 되는 시차가 있다. 그 시차를 건너뛰고도 셈법은 살아남았다[1].

셈법이 풀리자 장부의 짜임도 보이기 시작했다. 점토판은 품목 기호와 수량을 짝지어 적는 방식을 지킨다. 같은 서식이 여러 유적에서 되풀이된다. 크레타섬의 궁전들이 같은 회계 규칙을 나누어 쓰고 있었다는 뜻이다[3]. 뜻을 몰라도 문서의 성격은 이렇게 드러난다.

지명은 읽히지만 뜻은 풀리지 않았다

소리 쪽에서도 얻은 것이 있다. 미국 캔자스대학의 존 영거는 선형문자 A 명문을 소리로 옮긴 목록을 공개해 왔다. 선형문자 B에서 같은 모양으로 쓰이는 기호에 같은 소리 값을 넣어 본 전사다[2].

이 전사에서 크레타섬의 옛 지명으로 보이는 낱말이 여럿 나온다. 크노소스와 파이스토스처럼 나중 시대에도 남은 이름이다[2]. 지명이 들어맞는다는 것은 소리 값 가설이 아주 틀리지는 않았다는 뜻이다[3].

소리 값 가설에는 한계도 뚜렷하다. 두 글자에서 모양이 같아 보여도 소리까지 같았다는 보장이 없다. 선형문자 B를 만든 사람들은 남의 글자를 빌려 자기 말을 적었다. 빌려 쓰는 과정에서 소리가 옮겨 갔을 가능성이 있다[3]. 그래서 전사문은 잠정적인 읽기로만 쓴다[2].

그러나 여기까지다. 지명을 빼고 나면 나머지 낱말의 뜻이 이어지지 않는다. 소리 나는 대로 읽을 수는 있어도 무슨 말인지 모른다. 소리와 뜻 사이의 다리가 놓이지 않은 상태다[3].

2026년의 소동

2026년 여름에 선형문자 A를 읽었다는 발표가 나왔다. 인공지능 기술자 톰 디 미노가 내놓은 주장이다. 그는 언어학을 정식으로 배운 사람이 아니다[4].

디 미노는 파이썬 스크립트를 짜서 두 자료 창구를 뒤졌다고 밝혔다. 손으로 모사한 자료를 모은 GORILA와 앞에서 본 SigLA다. 처음 발표에서 그는 기호 40개에 소리 값을 매기고 낱말 408개를 목록으로 만들었다고 했다[4]. 8월에는 41쪽짜리 글을 내놓으며 목록을 508항목으로 늘렸다. 기호 42개의 읽기와 명문 443건의 번역을 함께 실었다. 그는 미노아어가 히브리어와 아람어가 속한 셈어족에 속한다고 주장했다[5].

학계는 이 주장을 받아들이지 않았다. 심사를 거친 논문이 나온 적이 없기 때문이다[4]. 러트거스대학과 케임브리지대학에서 검토가 진행 중이라고 전해졌다. 그러나 결론은 아직 나오지 않았다[5]. 선형문자 A를 읽었다는 주장은 지난 백 년 동안 여러 번 나왔다. 그래서 연구자들은 새 주장에 조심스럽게 접근한다[6].

같은 8월에 다른 해석도 나왔다. 유크타스산 꼭대기 성소에서 나온 돌 탁자의 명문에 관한 것이다. 신에게 술을 부어 바치던 헌주 탁자다. 이 돌에 새겨진 두 이름이 나중 그

리스 신들과 이어진다는 제안이었다. 이 제안 역시 가설에 머물러 있다. 선형문자 A가 확실하게 풀리지 않았기 때문이다[7].

이 이야기는 인공지능 연구 전체에도 남는 것이 있다. 기계는 규칙이 있는 곳에서 규칙을 잘 찾아낸다. 분수 기호처럼 앞뒤가 맞아야 하는 자료에서 성과가 났다[1]. 반면 규칙만으로는 닿지 않는 곳이 있다. 어떤 소리가 어떤 뜻과 묶이는지는 약속의 문제다. 그 약속을 아는 사람이 모두 사라지면 계산으로 되살릴 길이 좁아진다[6].

남은 자료를 열어 두는 일

연구 방식도 달라졌다. 예전에는 명문 그림이 값비싼 도록에만 실려 있었다. 도서 몇 권을 구한 사람만 통계를 낼 수 있었다. 지금은 자료가 인터넷에 공개되어 있다. 학생도 내려받아 기호를 세어 볼 수 있다[3].

자료가 열리면 검증도 빨라진다. 누군가 새 읽기를 주장하면 다른 사람이 같은 자료로 확인해 볼 수 있다[6]. 2026년 여름의 소동에서도 검증 과정이 그대로 굴러갔다[4]. 주장하는 쪽과 검토하는 쪽이 같은 자료를 보고 있었다[5].

자료가 공개된 덕에 연구의 문턱도 낮아졌다. 대학에 소속되지 않은 사람도 기호 통계를 내볼 수 있다. 새 눈이 늘어나면 놓치던 규칙이 잡히기도 한다. 그 대신 검증되지 않은 주장도 함께 늘어난다. 공개와 검증은 짝으로 굴러가야 한다[6].

아직 읽지 못한 글자

인공지능이 선형문자 A에서 해낸 일은 넷이다. 기계는 흩어진 자료를 한자리에 모아 주었다. 기호의 모양과 획을 숫자로 바꾸어 견줄 수 있게 했다[3]. 분수 기호의 값은 컴퓨터 계산이 사람의 판독을 도와 좁혀 냈다. 후보가 수백만 개인 문제를 사람이 검토할 만한 크기로 줄여 주었다[1].

기계가 하지 못한 일도 분명하다. 미노아어가 어떤 말인지는 여전히 밝혀지지 않았다. 로제타석 노릇을 할 문서는 아직 땅속에 있거나 없다[3]. 소리를 옮겨 적어도 그 말이 무엇을 가리키는지는 나오지 않는다[6].

키프로미노아 문자의 사례가 이 차이를 잘 보여 준다. 신경망은 기존 분류가 흔들린다는 결과를 내놓았다[10]. 그렇다고 그 문자를 읽어낸 것은 아니다. 기계가 정리해 준 것은 글자의 짜임이었지 말의 뜻이 아니었다.

남은 길은 둘로 보인다. 하나는 새 유물이 나오는 것이다. 두 언어로 적힌 문서 한 점이면 판이 바뀐다. 다른 하나는 자료를 더 촘촘히 다듬는 것이다. 획 하나까지 기록해 두면 나중 연구가 그 위에서 시작한다[3].

선형문자 A는 지금도 읽지 못한 글자다. 기계는 그 앞에서 문을 열어 준 것이 아니라 자물쇠의 생김새를 그려 주었다. 문을 여는 일은 새 유물과 사람의 판단에 달려 있다[6]. 에게해의 점토판은 아직 자기 말을 하지 않고 있다.

참고 자료

[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/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[8] Del Freo, M. (2022). Linear A. Mnamon: Ancient writing systems in the Mediterranean. Scuola Normale Superiore. https://doi.org/10.25429/sns.it/lettere/mnamon022 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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

10 CHAPTER

제10장. 태양의 주기와 동물 인장의 암호를 해킹하다: 인더스 문자와 롱고롱고의 소리

인더스 문자와 롱고롱고에 정보 엔트로피를 들이대다

1. 사건 파일

도장에 새겨진 네 글자

인더스 문명은 기원전 2600년 무렵부터 기원전 1900년 무렵까지 이어졌다[1]. 터전은 오늘날 파키스탄과 인도 북서부다. 도시의 크기와 수를 보면 청동기 시대에서 손꼽히는 문명이다. 청동기 시대는 쇠보다 청동을 많이 쓰던 옛 시대다. 도시와 마을이 퍼진 땅은 100만 제곱킬로미터가 넘는 것으로 알려져 있다[2].

1920년대에 하라파와 모헨조다로에서 큰 발굴이 시작되었다[1]. 이 일을 총괄한 사람은 영국 고고학자 존 마셜이다. 그는 인도고고조사국의 책임자였다. 하라파 현장은 1921년에 다야 람 사니가 맡아 팠다. 모헨조다로 현장은 1922년에 라칼다스 바네르지가 맡았다. 마셜은 이 발굴들을 묶어 새 문명의 존재를 세상에 알렸다[3].

땅속에서 곧은 벽돌 길과 배수 시설이 나왔다. 그리고 손바닥보다 작은 네모난 도장이 잔뜩 나왔다[1].

도시의 모습만으로도 알 수 있는 것이 많다. 길은 바둑판처럼 곧게 뻗었고 집집마다 하수구가 이어졌다. 벽돌은 크기를 맞춰 구웠다. 무게를 재는 추도 규격이 있었다. 계획을 세우고 규칙을 지키는 사회였다는 뜻이다. 그런데 그들이 남긴 글은 읽히지 않는다.

도장은 활석이라는 무른 돌로 만들었다. 활석은 손톱으로도 자국이 나는 돌이다. 도장 아래쪽에는 동물 그림이 파여 있다. 뿔이 하나 달린 짐승, 혹이 솟은 소, 코끼리 같은 그림이다. 그림 위 빈자리에는 기호가 한 줄로 늘어서 있다[1].

도장은 눌러 찍는 물건이다. 무른 진흙에 대고 누르면 그림과 기호가 그대로 옮겨 간다. 짐을 묶은 끈 위에 진흙을 붙이고 도장을 찍으면 봉인이 된다. 인더스 도시들에서는 도

장을 찍은 진흙 봉인도 함께 나왔다[1].

학자들은 이 기호 무리를 인더스 문자 또는 하라파 문자라고 부른다[1]. 이름에 문자가 들어가지만 아직 읽지 못한다. 소리도 뜻도 모른다. 발견된 지 백 년이 넘도록 그렇다[2].

읽지 못한다는 말에는 두 가지 뜻이 겹쳐 있다. 기호의 소리를 모른다는 뜻이 하나다. 기호의 의미를 모른다는 뜻이 다른 하나다. 지금은 둘 다 모르는 상태다.

문제는 길이다. 도장 하나에 새겨진 기호는 보통 네다섯 개뿐이다[1]. 한 면에 새겨진 긴 기록도 열일곱 자를 넘지 않는다. 세 면에 나누어 새긴 것을 다 세어야 스물여섯 자가 된다[4]. 교실에서 쓰는 공책 한 줄도 채우지 못하는 분량이다.

도장이나 그릇에 새긴 이런 글을 명문이라고 부른다. 지금까지 모은 인더스 명문은 3천 점 안팎이다[5]. 점수는 적지 않지만 하나하나가 너무 짧다. 짧은 글만 모아서는 문장의 짜임을 알아내기 어렵다. 주어와 서술어가 어떻게 놓이는지 볼 기회가 없기 때문이다.

바다 건너 나무판의 기호

남태평양 한가운데에 라파누이라는 섬이 있다. 우리에게는 이스터섬이라는 이름이 익숙하다. 사람이 사는 이웃 땅까지 2천 킬로미터가 넘는다. 커다란 돌사람상 모아이로 알려진 곳이기도 하다. 이 섬에서도 뜻 모를 기호가 적힌 나무판이 나왔다. 이 기록은 오늘날 롱고롱고라고 불린다[6].

기호는 붓으로 쓴 글씨가 아니다. 날카로운 흑요석 조각이나 뼈로 만든 도구로 나무를 파서 새겼다. 새겨진 모양은 사람, 새, 물고기를 닮았다[6]. 손발을 든 사람 모양이 자주 나온다.

읽는 방향도 낯설다. 읽는 사람은 판 아래쪽 왼쪽 끝에서 시작한다. 한 줄을 읽고 나면 판을 180도 돌린다. 위아래가 뒤집힌 상태에서 다음 줄을 읽는다. 줄마다 판을 뒤집는 이 방식을 역우행서체라고 부른다[6]. 쉬운 말로는 뒤집어 읽는 방식이다. 옛 그리스에도 줄마다 방향을 바꾸는 글이 있었다. 그쪽은 판을 돌리지 않고 글자의 방향만 바꾼다.

이 방식은 연구자를 한 번 더 시험한다. 어느 줄이 첫 줄인지 정해야 순서가 잡힌다[6]. 순서를 잘못 잡으면 기호가 이어지는 차례가 통째로 뒤바뀐다. 컴퓨터에 자료를 넣기

전에 사람이 먼저 판단해야 하는 대목이다.

남은 유물은 얼마 되지 않는다. 세계의 박물관과 개인 소장가에게 흩어져 있는 진품은 26점이다. 모두 라파누이섬 밖에 있다[6]. 거기 새겨진 기호를 다 합쳐도 약 1만 5천 자다[7]. 책 한 권에도 못 미치는 양이다.

유물이 이렇게 적은 데는 이유가 있다. 19세기에 섬의 인구가 크게 줄었고 나무판도 많이 사라졌다. 남은 것들은 유럽의 박물관과 수도원으로 흩어졌다. 그중 네 점이 로마의 한 수도회에 보관되어 있다[8].

이 나무판이 언제 만들어졌는지도 오래 다투던 문제였다. 유럽 사람이 섬에 들어온 뒤 흉내 내어 만들었다는 의심이 있었다. 2024년에 실비아 페라라 연구진이 로마의 나무판 네 점의 나이를 쟀다[8]. 나무에 남은 탄소의 양을 재어 나이를 알아내는 방법을 썼다. 이 방법을 방사성 탄소 연대 측정이라고 부른다.

네 점 가운데 세 점은 19세기 나무였다. 나머지 한 점은 훨씬 오래된 나무로 나왔다. 유럽 배가 섬에 닿은 1720년대보다 200년 넘게 앞선다[8]. 라파누이 사람들이 스스로 글자를 만들었을 가능성이 커졌다.

연구진은 조건을 하나 달아 두었다. 나무를 벤 때가 기호를 새긴 때와 같지는 않다. 나무의 나이는 새김이 그보다 뒤라는 사실만 알려 준다. 오래된 나무를 나중에 깎아 썼을 수도 있기 때문이다. 이른 연대가 나온 나무판이 한 점뿐이라는 점도 남는다[8].

이것이 글자이기는 한가

두 기호 체계를 두고 학계는 오래 다투었다. 다툼은 해독에서 시작하지 않았다. 이것이 말을 받아 적은 글자가 맞느냐를 먼저 따졌다[1][9].

2004년에 스티브 파머, 리처드 스프롯, 마이클 위첼이 함께 논문을 냈다. 이들은 인더스 기호가 글자가 아니라고 주장했다. 기록이 너무 짧고, 긴 문서가 하나도 없다는 점을 근거로 들었다. 그러니까 이 기호는 집안의 표식이나 물건의 소유 표시, 또는 부적에 가깝다는 뜻이다[9].

반대편에서는 통계로 답을 찾자고 나섰다. 기호가 나오는 빈도와 앞뒤로 붙는 규칙을 세어 보자는 것이다[5]. 사람의 말에는 말버릇 같은 통계 습관이 있다. 그 습관이 기호에도 있는지 재 보자는 접근이다.

이 방법의 장점은 뜻을 몰라도 쓸 수 있다는 데 있다. 기호를 번호로 바꿔 놓고 순서만 세면 된다. 컴퓨터가 잘하는 일이기도 하다.

해독을 가로막는 세 개의 벽

두 기호 체계 모두 아직 해독되지 않았다. 앞을 막는 벽은 세 개다.

대조할 글이 없다. 이집트 상형문자는 로제타석 덕분에 풀렸다. 로제타석에는 같은 내용이 그리스어로도 적혀 있었다. 프랑스 학자 샹폴리옹은 아는 글과 모르는 글을 나란히 놓고 짝을 맞췄다. 같은 내용을 두 가지 말로 나란히 적은 글을 대역본이라고 부른다. 인더스 유적과 라파누이섬에서는 그런 기록이 나오지 않았다[2][6].

어떤 말을 적은 것인지도 모른다. 기호가 받아 적은 실제 말을 바탕 언어라고 부른다. 인더스 문자의 바탕 언어를 두고 드라비다어 계통설과 인도아리아어 계통설이 맞선다. 알려진 어느 어족에도 속하지 않는다는 주장도 있다[2].

남은 벽은 자료의 양이다. 오늘날 인공지능은 방대한 글을 먹고 자란다. 그런데 여기 있는 것은 인더스 도장 몇천 점과 나무판 26점뿐이다[1][6]. 자료가 모자라면 인공지능은 그럴듯한 거짓말을 만들어 낸다. 이 현상을 환각이라고 부른다.

이 세 개의 벽은 서로를 붙들고 있다. 바탕 언어를 알면 소릿값을 맞춰 볼 수 있다. 대역본이 있으면 바탕 언어를 알아낼 수 있다. 자료가 많으면 통계로 실마리를 잡을 수 있다. 어느 하나가 풀리지 않으니 나머지도 제자리다.

2025년 1월 5일에 인도 타밀나두주가 상금을 걸었다. 첸나이에서 열린 인더스 문명 발견 100주년 국제 학술 대회 개막식에서 나온 발표다[10]. 인더스 문자를 고고학 전문가가 인정할 만큼 풀어내는 사람에게 100만 달러를 준다는 내용이다[10][11]. 2026년 9월까지 알려진 수상자는 없다.

참고 자료

[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 (2026년 9월 3일 접속)

[4] Farmer, S. (n.d.). Claims concerning the longest Indus "inscription". safarmer.com. https://safarmer.com/indus-longestinscription/ (2026년 9월 3일 접속)

[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, 5월 28일). Kohau Rongorongo: Talking tablet. australian.museum. https://australian.museum/learn/cultures/pasifika-collections/rapa-nui-collections/kohau-rongorongo-talking-tablet/ (2026년 9월 3일 접속)

[7] Ross, L. (2026, 3월 31일). Reading Rongorongo (Version 2). Zenodo. https://doi.org/10.5281/zenodo.19362491 (2026년 9월 3일 접속)

[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, 1월 6일). 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/ (2026년 9월 3일 접속)

[11] Smithsonian Magazine. (2025, 1월 30일). 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/ (2026년 9월 3일 접속)

2. 과학 수사대의 추적: 정보 엔트로피 측정기

글자의 지문을 재는 자

인더스 도장과 라파누이의 나무판은 서로 만난 적이 없다. 그런데 두 곳의 연구자가 같은 도구를 집어 들었다. 뜻을 모르는 기호 줄을 앞에 두고 할 수 있는 일이 비슷했기 때문이다.

읽지 못하는 글자를 앞에 두고 연구자들은 다른 질문을 골랐다. 무슨 뜻이냐고 묻는 대신 어떻게 배열되어 있느냐고 물었다. 뜻을 몰라도 순서는 셀 수 있기 때문이다.

이때 쓰는 도구가 정보 엔트로피다. 엔트로피는 다음에 무엇이 올지 얼마나 알기 어려운가를 재는 값이다. 값이 크면 다음이 예측하기 어렵다는 뜻이다. 값이 작으면 다음이

거의 정해져 있다는 뜻이다.

이 개념은 통신 공학에서 나왔다. 전화선으로 신호를 보낼 때 얼마만큼의 정보가 실리는지 재려고 만든 자다. 뜻과 상관없이 순서만 보고 계산한다. 읽지 못하는 글자에 댈 수 있는 이유가 여기에 있다.

교실 좌석을 떠올리면 이해가 쉽다. 제비뽑기로 앉으면 옆에 누가 올지 알 수 없다. 이때 엔트로피가 높다. 번호순으로 앉으면 옆에 올 사람이 정해진다. 이때 엔트로피가 낮다.

사람의 말은 두 극단 사이에 있다. 문법이 순서를 묶어 두지만 할 말은 자유롭게 고른다. 그래서 사람이 쓰는 말의 엔트로피는 중간 지대에 놓인다. 이 중간 지대를 자로 삼자는 것이 연구의 출발점이다.

우리말로도 확인할 수 있다. "학교에" 다음에 올 말은 여러 가지다. 간다, 갔다, 다녀왔다가 모두 가능하다. 그러나 아무 말이나 오지는 못한다. 선택지가 있되 좁혀져 있는 이 상태가 언어의 특징이다.

인더스 기호를 재 보다

2009년에 라제시 라오와 니샤 야다브를 포함한 연구진이 이 자를 인더스 기호에 댔다. 이들은 앞 기호를 알고 있을 때 다음 기호의 불확실성을 계산했다[1]. 앞 기호를 알 때 다음 기호가 얼마나 예상되는지 재는 이 값을 조건부 엔트로피라고 부른다.

비교 상대도 함께 준비했다. 한쪽에는 영어, 산스크리트어, 수메르어 같은 사람의 말을 놓았다. 다른 쪽에는 사람의 말이 아닌 것을 놓았다. 생물의 유전자 서열, 컴퓨터 프로그램 코드, 무작위로 만든 기호열이 그것이다[1].

결과는 인더스 기호가 사람 말들 쪽에 붙어 나온 것이었다. 무작위 기호열보다 규칙이 뚜렷하고 컴퓨터 코드보다 자유로웠다. 연구진은 이 결과를 사이언스에 실었다[1].

같은 연구진은 확률 모형도 함께 내놓았다. 앞 기호가 정해지면 다음 기호가 나올 확률을 표로 만드는 방식이다. 이렇게 앞 기호만 보고 다음을 셈하는 모형을 마르코프 모형이라고 부른다. 표를 들여다보면 어떤 기호가 글머리에 잘 오는지 드러난다. 어떤 기호가 글 끝에 잘 오는지도 드러난다[2].

같은 연구진은 다음 해에 기호의 빈도도 조사했다. 약 3천 점의 명문을 모으고 그중 상태가 좋은 1,548점을 골랐다. 소수의 기호가 쓰임의 대부분을 차지하고 나머지 기호는 드물게 나왔다. 이 모양은 사람의 언어에서 늘 보이는 분포다[3].

이 분포에는 이름이 있다. 지프의 법칙이라고 부른다. 자주 쓰는 낱말 몇 개가 글의 대부분을 채운다는 규칙이다. 우리말도 은, 는, 이, 가 같은 낱말이 앞줄을 차지한다. 반대로 한 번만 나오고 마는 낱말은 아주 많다. 인더스 기호도 같은 모양을 보였다[3].

놓이는 순서에서도 규칙이 드러났다. 기호 23개가 글의 끝 80퍼센트를 차지했다. 반대로 글머리 80퍼센트를 채우려면 기호 82개가 필요했다[3]. 끝에 오는 기호가 훨씬 한정되어 있다는 뜻이다. 우리말에서 문장 끝에 조사나 어미가 몰리는 모습과 비슷하다.

연구진은 앞 기호로 다음 기호를 맞히는 모형도 만들었다[2]. 이 모형은 마모되어 읽기 어려운 기호를 약 75퍼센트 확률로 되살렸다[3]. 규칙이 있어야 나올 수 있는 성적이다.

되살린다는 말을 오해하면 안 된다. 모형이 채워 넣는 것은 기호의 모양뿐이다. 그 기호가 무슨 소리였는지는 여전히 모른다. 빠진 퍼즐 한 칸을 채워 준 것이지 그림을 읽어 준 것이 아니다.

반박과 되받는 반박

이 모형에는 한계도 있다. 앞의 한두 기호만 보고 다음을 맞히는 방식이기 때문이다[2]. 사람의 문장은 훨씬 멀리 떨어진 말끼리 호응한다. 자료가 짧으면 그 먼 호응을 배울 수 없다.

이 결과에 대해 곧바로 반박이 나왔다. 리처드 스프롯이 2010년에 비판 글을 실었다. 엔트로피 값이 비슷하다는 사실만으로는 글자라고 못 박을 수 없다는 지적이다. 사람 말이 아닌 기호 체계도 비슷한 값을 낼 수 있다는 것이다[4].

라오 연구진은 같은 해에 같은 학술지에서 답을 냈다. 자신들은 인더스 기호가 글자임을 증명한 것이 아니라고 밝혔다. 글자일 가능성을 통계로 뒷받침했을 뿐이라는 설명이다[5].

스프롯은 2014년에 비판을 한 걸음 더 밀고 나갔다. 그는 사람 말이 아닌 기호 자료를 여러 벌 모았다. 메소포타미아의 신 표시, 북아메리카의 토템 기둥, 헛간 벽에 그리는 별무늬가 그것이다. 날씨 예보에 쓰는 그림도 넣었다. 이 자료들과 인더스 기호를 같은

잣대로 견주었다[6].

그가 만든 판정 절차는 인더스 기호를 사람 말이 아닌 쪽으로 분류했다. 스프롯은 통계만으로는 부족하다고 적었다. 해독에 성공하거나 그 문화가 글을 썼다는 고고학 증거가 따로 있어야 한다는 주장이다[6]. 이 논쟁은 아직 끝나지 않았다.

2026년 4월에 나온 새 분석도 같은 물음 앞에 선다. 아시시 나이르는 인더스 명문 1,916점을 놓고 검사를 다시 설계했다. 그는 사람 말이 아닌 기호 체계를 흉내 낸 가짜 자료를 컴퓨터로 만들었다. 진짜 인더스 자료와 가짜 자료의 통계를 나란히 놓고 견주었다[7].

인더스 자료는 어느 가짜 자료와도 딱 맞아떨어지지 않았다. 두 가짜 자료 사이의 중간에 놓였다. 실제로 존재하는 어떤 비언어 기호 체계도 인더스 통계를 그대로 재현하지 못했다. 나이르는 이것이 글자인지 아닌지 판정을 미뤄야 한다고 적었다. 이 글은 학술지 심사를 거치기 전에 인터넷에 먼저 올린 사전 공개본이다[7].

기호를 알아보는 눈

통계 이전에 해야 할 일이 있다. 돌에 새겨진 흐릿한 자국을 기호로 옮겨 적는 일이다. 이 작업은 오랫동안 사람의 눈과 손에 기댔다.

옮겨 적기는 생각보다 어렵다. 도장은 4천 년 동안 닳았다. 같은 기호라도 새긴 사람에 따라 모양이 조금씩 다르다. 두 개의 기호로 볼지 하나의 기호로 볼지 판단이 갈리기도 한다. 이 판단이 달라지면 뒤따르는 통계도 달라진다.

2017년에 사티시 팔라니아판과 로노조이 아디카리가 이 일을 컴퓨터에 맡겼다. 도장 사진을 넣으면 기호 목록이 나오는 처리 흐름을 만들었다[8].

2025년에는 바이슈나비 딕시트 연구진이 이 흐름을 이어받았다. 인장 사진에서 기호를 알아보는 신경망과 동물 그림을 알아보는 신경망을 따로 만들었다. 자료를 다섯 몫으로 나눠 돌아가며 시험했다. 두 신경망 모두 평균 85퍼센트 안팎의 정확도를 보였다[9].

이 문제를 줄이려면 판단 기준을 공개하고 서로 견주어야 한다. 2025년에 무자파르 알리 칸 연구진이 그 작업을 정리했다. 2009년부터 2025년까지 나온 인더스 문자 관련 인공지능 연구를 모아 유형별로 나누었다. 자주 되풀이되는 실수도 함께 적었다[10].

2025년에 아툴 샤르마와 슈바지트 로이 초두리가 도구를 하나 공개했다. 이름은 AI-EPIGRAPHY다. 사용자가 기호 순서를 넣으면 확률이 높은 해석 후보를 돌려준다. 빈도 분석과 연쇄 확률 계산을 묶어 놓은 도구다[11].

2026년 5월에는 자료 기반 자체를 다시 짜는 시도가 발표되었다. 베아타 메제시 연구진은 희귀하고 읽지 못하는 문자를 한데 모으는 데이터베이스를 만들었다. 고화질 사진, 옮겨 적은 기호, 주석, 출처 정보를 하나의 틀에 담았다. 흩어진 자료를 모으는 일이 해독보다 앞선 숙제라는 판단이다[12].

참고 자료

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

3. 밝혀진 진실

달을 세는 나무판

라파누이의 나무판 가운데 마마리라는 이름이 붙은 것이 있다. 이 나무판에는 다른 곳과 생김새가 다른 구간이 있다. 초승달 모양 기호가 줄줄이 이어진다[1].

독일 학자 토마스 바르텔이 20세기 중반에 이 구간을 주목했다. 그는 이것이 달의 날짜를 적은 기록이라고 보았다[1]. 1990년에 자크 기가 자료를 다시 정리해 같은 결론을 폈다[2].

2011년에 폴 홀리가 이 해석을 더 다졌다. 마마리 나무판의 달력은 서른 밤으로 이루어져 있다. 이 숫자는 실제 달의 주기와 맞는다. 폴리네시아 여러 섬의 달력과도 맞는다[1].

홀리는 타히티와 뉴질랜드의 달력도 함께 놓고 견주었다. 밤마다 붙는 이름이 섬마다 비슷했다. 라파누이의 달력은 다른 섬들보다 사흘쯤 밀려 있었다. 섬이 멀리 떨어져 있는 동안 생긴 차이로 보인다[1].

기호 배치도 규칙을 따랐다. 초승달 기호 사이사이에 다른 기호가 끼어들어 여덟 무리로 나눈다. 마지막 두 개의 초승달은 달이 보이지 않는 밤을 가리킨다[1].

같은 배열이 나무판 밖에서도 나온다. 섬의 바위그림 가운데 거북 무늬와 초승달이 함께 새겨진 것이 있다. 거기서도 마지막 두 초승달이 거북의 등딱지 안에 놓여 있다. 나무판이 달 없는 밤을 따로 떼어 그린 방식과 같다[1]. 나무판과 바위가 같은 셈법을 나눠 가졌다.

이 해석은 롱고롱고 연구에서 폭넓게 인정받는다. 그렇더라도 읽어 낸 것은 날짜의 순서다. 라파누이 사람들이 그 밤을 무엇이라 불렀는지는 모른다.

인공지능이 낸 답과 그 무게

달을 세는 일과 해를 세는 일은 붙어 있다. 일식은 해와 달이 겹칠 때 일어난다. 그래서 달력을 읽어 냈다는 주장은 곧 하늘 전체를 읽어 냈다는 주장으로 번진다.

2026년 3월에 로건 로스가 사전 공개 논문을 냈다. 그는 개인 컴퓨터에서 언어 모형을 돌려 롱고롱고를 분석했다. 기호 89개로 이루어진 사전을 만들었고, 이것으로 남은 기록의 97퍼센트 가까이를 덮었다고 주장했다. 마마리 나무판이 일식과 월식을 예측하는 표라는 주장도 폈다[3].

이 주장은 아직 학계의 검증을 거치지 않았다. 로스의 글은 학술지 심사를 통과한 논문이 아니다. 인터넷 저장소에 먼저 올린 사전 공개본이다[3]. 사전 공개본은 누구나 올릴 수 있다. 내용이 맞는지는 다른 연구자들이 뒤이어 따져 보아야 한다.

이런 주장을 다룰 때는 검사 방법을 함께 보아야 한다. 남은 롱고롱고 기록은 26점, 약 1만 5천 자다[3]. 자료가 이만큼뿐인 곳에서는 우연히 들어맞는 결과가 나오기 쉽다. 기호 89개에 소릿값을 자유롭게 배정하면 어떤 문장이든 만들어 낼 수 있다. 그래서 학계는 다른 사람이 같은 방법으로 같은 결과를 얻는지를 먼저 묻는다.

진짜 해독이라면 통과해야 하는 시험이 있다. 해독안을 만들 때 쓰지 않은 다른 나무판을 가져와 읽어 보는 것이다. 거기서도 뜻이 통하는 문장이 나와야 한다. 미케네 그리스어를 적은 선형문자 B가 그 시험을 통과한 예다[4]. 롱고롱고와 인더스 문자는 아직 그 문턱을 넘지 못했다.

기계가 하는 일과 못 하는 일

지금까지 인공지능이 이 두 문자에서 해낸 일은 분명하다.

기계는 흐릿한 자국에서 기호를 알아본다[5]. 사람이 눈으로 대조하려면 몇 달 걸릴 일을 몇 시간에 끝낸다. 마모되어 읽기 어려운 기호를 앞뒤 맥락으로 되살리기도 한다[6]. 흩어진 유물의 사진과 기록을 한 곳에 모아 검색할 수 있게 만든다[7].

자료를 모으는 일은 눈에 잘 띄지 않는다. 그러나 이 일이 없으면 뒤따르는 계산도 없다. 사진의 밝기와 각도가 제각각이면 기호를 견줄 수 없다. 옮겨 적는 규칙이 연구자마다 다르면 통계도 어긋난다. 2026년에 나온 데이터베이스는 그 기준을 하나로 맞추려는 시도다[7].

통계 쪽에서도 성과가 있다. 인더스 기호가 무작위 배열이 아니라는 사실은 여러 연구가 확인했다[6][8]. 배열에 규칙이 있다는 결론은 되풀이해 재현되었다. 그렇다고 이것이 글자라는 결론은 아니다. 규칙이 있다는 말과 말을 받아 적었다는 말은 다르다.

그러나 기계가 하지 못한 일이 더 크게 남아 있다. 인더스 기호 하나하나의 소릿값은 여전히 모른다[9]. 롱고롱고 기호의 소릿값도 모른다[3]. 두 문자 모두 뜻을 확정한 낱말이 아직 없다. 인더스 도장에 적힌 것이 사람 이름인지 물건 이름인지도 정해지지 않았다[9].

이유는 자료의 성질에 있다. 대역본이 있으면 해독이 훨씬 쉬워진다. 인더스 유적에도 라파누이섬에도 그런 글은 없다[9]. 대역본 없이 풀린 예가 없지는 않다. 선형문자 B가 그렇게 풀렸다. 대신 그때는 바탕 언어를 짐작할 실마리가 있었고 자료도 넉넉했다[4]. 인더스 문자는 바탕 언어가 무엇인지도 정해지지 않았다[9]. 롱고롱고는 남은 글이 너무 적다[3].

2026년 4월의 분석은 이 사정을 숫자로 다시 보여 주었다. 인더스 명문 1,916점은 사람 말의 통계와도, 사람 말이 아닌 기호의 통계와도 딱 맞지 않았다[10]. 백 년 넘게 이어진 물음이 아직 열려 있다는 뜻이다.

기다리는 쪽에 서서

읽지 못한 글자를 다루는 연구는 두 가지 태도를 요구한다. 하나는 새 도구를 끝까지 밀어붙이는 태도다. 다른 하나는 나온 결과를 의심하는 태도다.

인공지능은 앞의 일을 잘한다. 세고, 견주고, 후보를 늘어놓는다. 뒤의 일은 사람의 몫이다. 후보 가운데 무엇이 그 시대에 있을 법한지 판단해야 한다. 그 판단에는 유적의 위치, 무덤에서 나온 물건, 남아 있는 구전이 함께 들어간다.

라파누이의 달력 해석이 그 예다. 초승달 기호를 세는 일은 계산이다. 그 서른 밤을 폴리네시아 여러 섬의 달력과 견주는 일은 고증이다[1]. 두 작업이 맞물렸을 때 해석이 설득력을 얻었다.

인더스 도장의 동물 그림도 그런 자료다. 뿔 하나 달린 짐승이 무엇을 뜻했는지 우리는 모른다. 마마리 나무판의 서른 밤도 마찬가지다. 순서는 알아냈지만 그 밤에 사람들이 무엇을 했는지는 모른다.

모른다고 말하는 일에도 값어치가 있다. 어디까지 밝혀졌고 어디부터 추측인지 선을 그어야 다음 연구가 시작된다. 선을 흐리면 그럴듯한 이야기가 사실 행세를 한다. 읽지 못한 글자를 다루는 연구에서 이 선 긋기는 결과를 내는 일만큼 중요하다.

이 장에서 다룬 두 문자는 서로 만난 적이 없다. 시대도 3천 년 넘게 떨어져 있고 사는 곳도 지구 반대편이다. 그런데 연구자들이 겪는 어려움은 거의 같다. 짧은 자료, 없는 대역본, 모르는 바탕 언어다. 같은 도구가 두 곳에서 함께 쓰이는 까닭이 여기에 있다.

타밀나두주가 내건 100만 달러도 그 사정을 비춘다[11]. 상금이 걸린 지 한 해가 넘도록 주인이 나오지 않았다. 계산기는 빨라졌지만 열쇠는 아직 나오지 않았다.

참고 자료

[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, 3월 31일). Reading Rongorongo (Version 2). Zenodo. https://doi.org/10.5281/zenodo.19362491 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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, 4월 20일). How non-linguistic is the Indus sign system? A synthetic-baseline scorecard. arXiv:2604.17828. https://arxiv.org/abs/2604.17828 (2026년 9월 3일 접속)

[11] The Tribune. (2025, 1월 6일). Stalin announces $1 mn prize for deciphering Indus Valley script. tribuneindia.com. https://www.tribuneindia.com/news/india/stalin-announces-1-mn-prize-for-decip

hering-indus-valley-script/ (2026년 9월 3일 접속)

제 6 부

디지털 알렉산드리아 도서관과 미래의 고고학

11 CHAPTER

제11장. 인공지능, 깨진 그리스 비문을 다시 읽다: 파운데이션 모델과 이타카

깨진 그리스 비문의 빈칸을 채우는 이타카와 아폴로

1. 사건 파일

돌에 새긴 나라의 기록

고대 그리스 사람들은 중요한 일을 돌에 새겼다. 새 법을 만들면 그 내용을 돌기둥에 새겨 광장에 세웠다. 세금을 걷은 장부도 돌에 남겼다. 신전에 바친 물건 목록도 돌에 남겼다. 돌이나 청동에 새긴 이런 글을 비문이라고 부른다. 비문을 모으고 읽는 학문은 금석학이다. 돌이나 쇠붙이에 새긴 글을 연구하는 학문이라는 뜻이다[4].

비문은 대체로 그 시대 사람이 직접 남긴 기록이다. 후대 사람이 옮겨 적으며 고친 흔적이 드물다. 그래서 역사학자에게는 값진 자료다. 다만 늘 처음 모습 그대로는 아니다. 뒤에 글자를 지우거나 고쳐 새긴 비석도 있다. 돌을 갈아 내고 다시 쓴 사례도 남아 있다. 아테네가 어느 도시와 언제 손을 잡았는지도 비문이 알려 준다[4].

비문은 여러 사람이 보라고 만든 글이다. 광장이나 신전 앞에 세워 두고 지나는 사람이 읽게 했다. 학교 복도에 붙은 안내문과 쓰임이 비슷하다. 그래서 비문에는 그날의 관리 이름과 날짜가 또박또박 들어간다. 역사학자는 이 이름과 날짜를 실마리로 삼는다[4].

지워진 글자, 결손부

문제는 시간이다. 비석은 수천 년 동안 비와 바람 앞에 서 있었다. 돌 표면은 조금씩 닳았고 글자 홈은 얕아졌다. 땅에 묻힌 비석은 습기와 미생물에 갉혔다. 비석이 깨져 한쪽만 남기도 했다[1].

그 결과 비문에는 글자가 사라진 빈칸이 생긴다. 학자들은 이 빈칸을 결손부라고 부른다. 라틴어로는 라쿠나다[1]. 결손부는 문장 한가운데를 뚫어 놓는다. 사람 이름이 사라지면 누가 한 일인지 알 수 없다. 숫자가 사라지면 얼마를 냈는지 알 수 없다. 도시 이름

이 사라지면 어느 나라와 맺은 약속인지 알 수 없다.

학자들은 오래전부터 이 빈칸을 채우려 했다. 비슷한 시기, 비슷한 내용의 다른 비문을 찾아 견주었다. 그리스 비문에는 정해진 문장 틀이 있었다. 법을 알리는 문서는 거의 같은 순서로 문장을 늘어놓았다. 그래서 앞뒤가 남아 있으면 사라진 낱말을 짐작할 수 있다[1].

비석은 무거워서 연구실로 옮길 수 없다. 그래서 학자들은 젖은 종이를 비석에 대고 눌렀다. 종이가 마르면 글자 홈이 그대로 떠올랐다. 이 종이 사본을 탁본이라고 한다. 학자들은 탁본을 들고 돌아와 책상에서 글자를 읽었다. 오래된 인쇄 도록에는 사진 대신 옮겨 적은 글자만 실렸다. 그래서 뒷사람은 돌의 생김새를 보지 못한 채 글자만 읽어야 했다[4].

짐작에는 규칙이 있다. 빈칸이 몇 글자였는지 먼저 센다. 돌에 남은 글자 사이 간격을 재면 대강 알 수 있다. 그다음 글자 수가 맞는 낱말만 후보로 올린다. 이렇게 채워 넣은 글자는 반드시 대괄호 안에 적는다. 형태가 무너져 읽기 어려운 글자 밑에는 작은 점을 찍는다. 이 약속을 레이덴 표기법이라고 한다. 돌에 남은 글자와 사람의 추측을 구분하려고 만든 규칙이다[1, 4].

시그마 세 획을 둘러싼 다툼

비문을 읽으려면 언제 새긴 글인지도 알아야 한다. 오래된 방법은 글자 모양을 보고 시기를 재는 것이다. 글자 모양이 시대에 따라 조금씩 바뀌기 때문이다[2].

아테네 비문에서 자주 나오는 예가 시그마다. 시그마는 그리스 알파벳의 한 글자다. 이른 시기에는 획이 세 개인 시그마를 새겼다. 뒤에는 획이 네 개인 시그마가 널리 쓰였다. 오랫동안 학자들은 획이 세 개인 시그마를 기준으로 삼았다. 그런 글자가 있으면 기원전 446년 무렵보다 앞선 비문으로 보았다[2].

그런데 이 기준을 의심하는 학자들이 나왔다. 글자체가 바뀌는 속도는 지역마다 달랐기 때문이다. 옛 글자체를 오래 고집한 지역도 있었다. 한 도시 안에서도 장인마다 새기는 버릇이 달랐다[2].

그래서 아테네가 다른 도시와 맺은 조약 비문의 연대를 두고 다툼이 이어졌다. 한쪽은 기원전 446년 무렵보다 앞선다고 했다. 다른 쪽은 기원전 420년대에 새긴 글이라고 했

다[2]. 기원전 420년대는 펠로폰네소스 전쟁이 한창이던 때다. 이 전쟁은 기원전 431년에 시작했다. 연대가 수십 년 달라지면 역사의 줄거리도 달라진다. 아테네가 힘으로 동맹 도시를 눌렀던 시점이 통째로 옮겨 가기 때문이다.

빈칸을 채우는 일에는 위험이 따른다. 한 학자가 채워 넣은 낱말이 다음 세대에는 원문처럼 읽히기 때문이다. 대괄호를 빠뜨린 인쇄본 하나가 추측을 사실로 바꿔 놓기도 한다[1]. 그래서 금석학자들은 자기 추측에 늘 표시를 남겼다.

비문이 원래 어디에 서 있었는지도 알아내야 한다. 그리스 도시들은 저마다 다른 사투리를 썼다. 같은 낱말도 도시마다 철자가 달랐다[2]. 학자들은 이 철자 차이와 관직 이름을 보고 출토지를 짐작했다[4]. 출토지는 그 비석이 땅에서 나온 곳이다.

사람 손으로는 다 읽을 수 없다

읽어야 할 비문은 계속 쌓인다. 팩커드 인문학 연구소가 모은 그리스어 비문은 178,551점이다[2]. 로마 시대 라틴어 비문 자료도 176,000점이 넘는다[3]. 반면 비문을 읽을 수 있는 학자는 세계에 얼마 되지 않는다[4].

숫자를 다듬는 일부터 쉽지 않다. 이 자료를 기계가 읽을 수 있게 손보자 78,608점이 남았다. 절반 넘는 자료가 너무 짧거나 상태가 나빠 빠졌다. 기계에게 먹일 자료를 고르는 일만으로도 큰 작업이었다[2].

빈칸을 채우는 일은 사람에게도 어렵다. 훈련받은 금석학자들에게 글자가 지워진 비문을 주고 채우게 한 시험이 있다. 사람이 첫 번째로 내놓은 답이 맞은 비율은 25퍼센트였다[2]. 네 번에 세 번은 틀렸다는 뜻이다. 그보다 앞선 실험에서는 사람이 채운 글자의 오류율이 57.3퍼센트로 나왔다[1].

파피루스 사정은 더 나쁘다. 파피루스는 이집트 물가에서 자라는 풀의 이름이다. 옛사람은 그 풀의 속을 얇게 저몄다. 저민 것을 가로세로로 겹쳐 두드리고 눌러 말렸다. 그렇게 만든 기록 재료도 파피루스라고 부른다. 이집트의 마른 모래 속에서 수십만 장이 나왔다. 아직 아무도 읽지 않은 그리스어 파피루스는 백만 점으로 헤아려진다[5]. 이 종이에는 세금 영수증과 편지와 계약서가 적혀 있다. 옛사람의 하루가 그대로 담긴 기록이다.

읽는 일은 상자를 여는 것으로 끝나지 않는다. 종이가 찢기고 잉크가 지워진 상태이기 때문이다. 보존 처리를 하고 흩어진 파편을 맞추는 일이 먼저다. 그 뒤에야 글자를 판독할 수 있다. 이 모든 일을 할 사람이 모자란다[4].

자료를 읽는 방식도 문제였다. 오랫동안 비문 자료는 책 속의 글자 목록으로만 쌓였다. 돌의 사진이나 재질 정보는 함께 담기지 않았다[4]. 기계에 넣을 수 있는 정보가 글자뿐이라는 뜻이다.

발굴은 지금도 이어진다. 해마다 새 비석과 새 파피루스가 땅에서 나온다. 읽어야 할 자료는 늘어나는데 읽을 사람은 그대로다[4]. 이 격차는 해가 갈수록 벌어졌다.

그래서 많은 기록이 읽히지 못한 채 박물관 상자 안에 남았다. 돌과 종이는 있는데 그 위의 문장을 되살릴 사람이 모자랐다. 연구자들은 이 막힌 곳을 뚫으려고 기계를 불러들였다[4].

참고 자료

[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, 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 (2026년 9월 3일 접속)

2. 과학 수사대의 추적: 고대 텍스트 자동완성 장치 Ithaca와 Apollo

빈칸 문제로 배우는 기계

인공지능이 옛 글을 읽는 방법은 국어 시험의 빈칸 문제와 닮았다. 기계에게 온전한 비문을 잔뜩 보여 준다. 그다음 글자 몇 개를 가리고 무엇이 들어갈지 맞히게 한다. 답이 틀리면 기계 안의 숫자를 조금씩 고친다[4]. 이 과정을 수백만 번 되풀이하면 기계는 그리스어 문장의 버릇을 익힌다.

이렇게 배운 기계는 낱말을 외우지 않는다. 어떤 낱말 뒤에 어떤 낱말이 잘 오는지를 확률로 기억한다[4]. 그래서 처음 보는 비문에서도 빈칸에 어울리는 후보를 꺼낼 수 있다.

피티아에서 이타카로

2019년 연구자들이 피티아라는 모델을 내놓았다. 야니스 아사엘, 테아 좀머실드, 조너선 프라그가 만들었다. 피티아는 지워진 그리스어 비문의 글자를 예측하는 신경망이다[1]. 신경망은 자료를 보고 규칙을 스스로 익히는 계산 구조다. 사람 뇌의 신경 세포 연결을 본떠 만들었다.

피티아는 글자 단위와 낱말 단위를 함께 읽는다. 낱말 절반이 부서진 자료를 다루려면 두 눈이 필요하기 때문이다. 사람이 채운 글자의 오류율이 57.3퍼센트일 때 피티아는 30.1퍼센트를 기록했다. 정답이 피티아의 후보 20개 안에 든 경우는 73.5퍼센트였다[1].

2022년에는 이타카가 나왔다. 구글 딥마인드와 여러 대학이 함께 만들었다. 옥스퍼드 대학교, 베네치아 카포스카리 대학교, 아테네 경제경영대학교가 참여했다. 연구 결과는 2022년 3월 학술지 네이처에 실렸다[2]. 이름은 서사시 오디세이아에 나오는 섬 이타카에서 따왔다. 오랜 떠돌이 끝에 집으로 돌아간다는 뜻을 담았다.

이타카는 어떻게 판단하는가

이타카는 트랜스포머라는 구조를 쓴다. 트랜스포머는 문장 안의 낱말들이 서로 얼마나 관련 있는지 따지는 신경망이다. 글자가 지워진 곳은 빈칸을 뜻하는 기호로 바뀌어 들어간다. 모델은 그 기호를 대신할 글자의 확률을 계산한다[2].

이타카가 하는 일은 세 가지다. 지워진 글자를 채우고, 비문이 원래 어디 있었는지 맞히고, 언제 새긴 글인지 추정한다. 세 가지를 따로 배우지 않고 한꺼번에 배운다. 방언과 철자 습관은 장소를 알려 주는 단서다. 낱말 쓰임과 문장 틀은 시기를 알려 주는 단서다. 한 신경망이 이 단서들을 함께 다루면 서로를 돕는다[2].

이타카는 돌 사진을 보지 않는다. 옮겨 적은 글자만 읽고 판단한다[2]. 획의 생김새로 시기를 재던 옛 방법과는 길이 다르다.

연대를 내놓는 방식도 다르다. 이타카는 연도 하나를 딱 집어 말하지 않는다. 기원전 800년부터 서기 800년까지를 열 해 단위로 나눈다. 그리고 각 구간에 확률을 매겨 그래프로 보여 준다[2]. 학자는 그래프의 봉우리를 보고 어느 시기가 유력한지 판단한다.

학습에 쓴 자료는 팩커드 인문학 연구소의 비문 모음이다. 앞에서 본 78,608점이 그 자료다. 열 번 넘게 나오는 낱말만 세어도 35,884개였다[2].

연구진은 이타카를 만들 때 세 가지를 목표로 삼았다. 학자와 함께 일할 것, 판단을 도울 것, 판단 과정을 보여 줄 것이다. 사람을 대신하는 기계가 목표가 아니었다. 학자의 일하는 방식 안으로 들어가는 도구를 겨냥했다[2].

이타카는 답을 하나만 내놓지 않는다. 빈칸에 들어갈 후보를 확률과 함께 여러 개 보여 준다. 어느 글자를 보고 그렇게 판단했는지도 화면에 표시해 준다. 학자는 기계가 무엇에 기대어 답했는지 눈으로 확인할 수 있다. 최종 판단은 사람 몫으로 남긴다. 연구진은 이타카의 코드와 자료를 공개했다[2].

후보 스무 개의 뜻

기계가 내놓는 답은 하나가 아니다. 확률이 높은 순서로 후보를 줄 세운다. 첫 번째 후보가 맞은 비율을 상위 1 정확도라고 부른다. 정답이 앞의 스무 개 안에 든 비율은 상위 20 정확도다[1].

이 구분이 중요하다. 학자에게 필요한 것은 정답 하나가 아니다. 검토할 만한 후보 목록이다. 스무 개 안에 정답이 있으면 학자는 그 안에서 고를 수 있다[1]. 자기가 떠올리지 못한 낱말을 그 목록에서 발견하기도 한다[2].

라틴어로 건너간 아이네아스

같은 연구진은 라틴어 비문으로 방향을 넓혔다. 2025년 7월 네이처에 아이네아스라는 모델을 발표했다[3]. 이름은 로마 건국 이야기의 주인공 아이네아스에서 따왔다[5].

아이네아스는 라틴어 비문 176,000점이 넘는 자료로 배웠다. 로마, 하이델베르크, 클라우스 슬라비의 세 비문 데이터베이스를 합친 자료다. 열 글자 이내가 지워진 곳에서 정답이 후보 20개 안에 든 비율은 73퍼센트였다. 몇 글자가 지워졌는지 모르는 경우에도 58퍼센트를 기록했다[3]. 지워진 길이를 모른 채 답하는 일은 사람에게도 어려운 문제다.

비문이 로마의 어느 속주에서 나왔는지도 맞혔다. 속주는 로마가 다스리던 지방 땅이다. 62개 선택지 가운데 72퍼센트를 맞혔다[5]. 새긴 시기는 역사학자들이 정한 범위에서 평균 13년 안으로 좁혔다[3, 5].

아이네아스에는 새 기능이 있다. 비문의 사진을 함께 넣어 판단할 수 있다[3]. 옮겨 적은 글자만 보지 않고 돌의 생김새까지 함께 본다는 뜻이다. 비슷한 문장을 가진 다른 비문을 찾아 나란히 보여 주기도 한다. 학자가 며칠 걸려 찾던 비슷한 사례를 몇 초 만에 꺼내 준다[5].

연구진은 이 도구를 로마 황제 아우구스투스의 기록에 대 보았다. 아이네아스는 그 글이 새겨진 시기로 두 구간을 함께 내놓았다. 하나는 기원전 10년에서 기원전 1년 사이였다. 다른 하나는 서기 10년에서 서기 20년 사이였고 확률이 더 높았다[5]. 학계에서 갈리던 두 주장이 확률 그래프 위에 그대로 드러났다.

연구진은 역사학자 23명에게 이 기능을 써 보게 했다. 학자들은 비문 열 개 가운데 아홉 개에서 새 연구 아이디어를 얻었다고 답했다[3, 5]. 이 모델은 누구나 쓸 수 있게 공개되었다[5].

그리스어를 통째로 배우는 아폴로

2026년 3월 오스트리아 과학아카데미가 새 모델을 알렸다. 이름은 아폴로다. 인공지능 기업 미스트랄 AI와 정보기술 기업 세일 리플라이가 함께 만들고 있다[7]. 이타카가 비문 한 점을 다루는 도구라면, 아폴로는 그리스어 전체를 배우는 모델을 목표로 삼는다[6].

이런 모델을 파운데이션 모델이라고 부른다. 여러 일에 두루 쓰라고 넓게 배워 둔 기본 모델이라는 뜻이다. 한 가지 일만 하도록 만든 도구가 아니다[4]. 비문 복원에도, 문헌 검색에도, 글씨 판독에도 같은 바탕을 쓴다[6].

아폴로가 배우는 자료는 약 6억 낱말 분량의 그리스어 문헌이다. 여기에 이미 출판된 비문과 파피루스 수만 점이 더해진다[6, 7]. 연구진은 이것이 지금까지 모은 그리스어 자료 가운데 규모가 크다고 밝혔다[7].

오스트리아 국립도서관은 파피루스를 많이 가지고 있다. 이 도서관에만 수만 점이 보관되어 있다[7]. 아직 읽히지 않은 문서가 그 안에 섞여 있다.

연구를 이끄는 사람은 오스트리아 고고학연구소의 안나 돌가노프다. 그는 아직 아무도 읽지 않은 그리스어 파피루스가 세계에 백만 점 있다고 말했다[7]. 아폴로는 이런 자료에서 글자를 복원하고 내용을 찾아 주는 일을 맡는다. 손으로 쓴 글씨를 읽어 내는 일도 목표에 들어 있다[6, 7]. 연구진은 이 작업이 몇 해가 아니라 몇 시간 만에 끝날 수 있다고 밝혔다[7].

참고 자료

[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, 7월 23일). Aeneas transforms how historians connect the past. deepmind.google. https://deepmind.google/blog/aeneas-transforms-how-historians-connect-the-past/ (2026년 9월 3일 접속)

[6] Mistral AI. (2026). Austrian Academy of Sciences. mistral.ai. https://mistral.ai/customers/austrian-academy-of-sciences/ (2026년 9월 3일 접속)

[7] 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 (2026년 9월 3일 접속)

3. 밝혀진 진실

62퍼센트와 72퍼센트

연구진은 이타카의 실력을 숫자로 확인했다. 글자를 일부러 가린 비문을 주고 채우게 하는 시험이다. 같은 시험을 훈련받은 금석학자들에게도 냈다[1].

사람이 혼자 풀었을 때 첫 답이 맞은 비율은 25퍼센트였다. 이타카가 혼자 풀었을 때는 62퍼센트였다. 두 배가 넘는 차이다[1].

이 시험에는 세 번째 조건이 있었다. 사람이 이타카의 후보 목록을 보며 함께 푸는 조건이다. 그러자 정확도가 72퍼센트로 올라갔다[1]. 사람 혼자보다 높고 기계 혼자보다도 높은 값이다.

이 숫자가 이 연구의 핵심이다. 기계는 78,608점을 다 읽은 기억으로 확률을 계산한다. 사람은 그 비문이 어떤 역사 속에 있는지 안다. 기계가 후보 스무 개를 늘어놓으면 사람이 역사 지식으로 걸러 낸다. 두 가지가 합쳐질 때 정확도가 제일 높았다[1, 2].

장소와 시기를 맞히는 실력도 확인했다. 이타카는 그리스 세계를 84개 지역으로 나누어 배웠다. 출토지를 모르는 비문의 원래 고향을 71퍼센트 정확도로 맞혔다. 새긴 시기는 역사학자들이 정한 범위에서 평균 29.3년 떨어진 값을 내놓았다. 사람의 눈으로 글자체를 재던 방법은 오차가 백 년에 이르기도 했다[1].

기원전 421년이라는 답

이타카는 오래된 다툼에도 답을 내놓았다. 앞에서 본 아테네 조약 비문의 연대 문제다. 획이 세 개인 시그마 때문에 기원전 446년 무렵보다 앞선 것으로 보던 비문들이다[1].

연구진은 이 비문들을 이타카에 넣었다. 이타카는 글자 모양이 아니라 문장 자체를 보고 판단한다. 어떤 낱말이 어떤 낱말과 붙어 쓰이는지를 따진다. 어떤 관직 이름과 어떤 표현이 함께 나오는지도 따진다[1].

이타카가 내놓은 평균 연대는 기원전 421년이었다. 옛 기준보다 한 세대 뒤다. 이 값은 새로 나온 다른 증거와 방향이 같았다. 최근 학자들이 다시 매긴 연대와는 평균 5년 차이였다. 옛 기준으로 매긴 연대는 27년이나 벌어져 있었다[1].

기계가 혼자 역사를 다시 쓴 것은 아니다. 오래 갈리던 두 주장 가운데 한쪽에 무게를 실어 주었다[1, 2]. 학자들은 글자체 하나로 연대를 못 박던 습관을 다시 살펴보게 되었다. 도구 하나가 백 년 묵은 논쟁의 방향을 바꿔 놓았다.

기계가 지어내는 그럴듯한 문장

이런 성과에도 남은 문제가 있다. 배운 자료가 한쪽으로 치우쳐 있다는 점이다. 그리스어 비문 자료에는 아테네 것이 많다. 시대도 고전기에 몰려 있다. 자료가 적은 변두리 지역 비문은 그만큼 덜 정확하다[2].

기계가 왜 그렇게 답했는지 알기 어렵다는 문제도 있다. 신경망 안에는 수백만 개의 숫자가 들어 있다. 어느 글자에 더 집중했는지는 보여 주지만, 그 답을 고른 이유를 말로 설명하지는 못한다[2].

빈칸이 길어질 때도 문제가 생긴다. 몇 글자를 채우는 일은 잘하지만 여러 낱말이 통째로 사라진 곳은 어렵다. 아이네아스 시험에서도 지워진 길이를 모를 때 성적이 73퍼센트에서 58퍼센트로 떨어졌다[3]. 이때 모델은 문법에 맞고 그럴듯하지만 실제로는 없던 문장을 만들어 낼 수 있다. 이렇게 지어낸 답을 환각이라고 부른다. 기계가 사실처럼 꾸며 낸 틀린 답이라는 뜻이다. 환각은 옛 기록을 다룰 때 위험하다. 없던 역사가 사실처럼 굳어질 수 있기 때문이다[2].

2026년 8월 공개된 한 연구는 이 한계를 정면으로 다루었다. 여러 모델과 여러 조건에서 손상된 옛 문서를 복원해 본 실험이다. 연구진은 복원을 완전히 자동으로 맡길 수 없다고 결론지었다. 문서의 어느 곳을 채우느냐에 따라 성적이 크게 달라졌다. 정해진 문구가 되풀이되는 문서는 잘 채웠다. 사람마다 다르게 쓴 문장은 어려웠다. 지워진 길이를 아느냐 모르느냐에 따라서도 결과가 갈렸다. 연구진은 기계를 고문서 학자를 돕는 도구로 만들어야 한다고 밝혔다[4].

연구자들은 이 위험을 줄이려고 규칙을 두었다. 기계가 채운 글자에도 대괄호를 씌운다. 확률이 낮은 후보는 학자가 걸러 낸다. 옛 금석학자들이 지키던 표기 약속을 기계에도 그대로 적용한다[2].

그래서 이타카가 내놓는 답은 역사 사실이 아니다. 검토해 볼 가설 목록이다. 학자가 다른 자료와 견주어 확인해야 비로소 사료가 된다[2]. 사료는 역사를 밝히는 근거 자료다.

칠 년 사이에 일어난 일

이 분야의 변화 속도는 숫자로 견줄 수 있다. 2019년 피티아는 사람의 오류율 57.3퍼센트를 30.1퍼센트로 낮췄다[5]. 2022년 이타카는 첫 답 정확도를 62퍼센트로 끌어올렸다[1]. 2025년 아이네아스는 라틴어로 넘어가 연대 오차를 평균 13년까지 좁혔다[3]. 2026년에는 그리스어 전체를 배우는 아폴로가 시작되었다[6].

칠 년 사이에 대상 언어가 늘고 정확도가 올랐다. 다루는 자료도 글자에서 사진으로 넓어졌다[3]. 그동안 바뀌지 않은 것이 하나 있다. 마지막 판단은 사람이 한다는 원칙이다[2].

사진과 문맥, 다음 걸음

연구는 계속 넓어지고 있다. 2026년 나온 논문 하나는 널리 쓰이는 큰 언어 모델을 그리스어 자료로 다시 학습시켰다. 메타가 공개한 라마 3.1 모델을 쓴 실험이다. 파피루스 복원에서 첫 답이 맞은 비율은 73.5퍼센트였다. 비문 복원에서는 63.7퍼센트가 나왔다. 정답이 후보 20개 안에 든 비율은 각각 86.0퍼센트와 83.0퍼센트였다. 비문 전용으로 만든 모델이 아니어도 이 일을 해낼 수 있다는 뜻이다[7].

글자 이미지를 되살리는 연구도 나왔다. 2026년 4월 공개된 메사는 비문 사진에서 지워진 획을 채우는 방법이다. 같은 비석에 잘 남은 글자를 본보기로 삼아 닳은 곳의 무늬를 되살린다[8]. 문장을 채우는 모델과 사진을 채우는 모델이 나란히 자라고 있다.

잃어버린 비극도 실제로 돌아오고 있다. 2022년 이집트 필라델피아 유적에서 파피루스 한 장이 나왔다. 서기 3세기에 쓴 것으로 보이는 종이다. 2024년 학자들이 그 내용을 읽어 발표했다. 에우리피데스가 쓴 비극 폴리이도스와 이노의 대사 97행이었다. 두 작품은 그때까지 짧은 인용문과 줄거리로만 알려져 있었다[9].

이 복원은 사람의 손과 눈이 해냈다. 기계가 한 일이 아니다. 그런 종이가 세계에 백만 점 더 남아 있다. 사람의 눈만으로는 이 종이 더미를 다 읽을 수 없다. 아폴로 같은 모델이 겨냥하는 곳이 바로 여기다[6].

도구가 공개되면서 연구 방식도 바뀌었다. 이타카와 아이네아스는 코드와 자료가 공개되어 있다[1, 10]. 대학 연구실이 아니어도 누구나 웹에서 써 볼 수 있다. 비문 사진 한 장을 올리면 후보 목록과 지도와 연대 그래프가 나온다[10]. 예전에는 몇 해가 걸리던 비교 작업이다[3].

남은 일은 여전히 많다. 그리스어 파피루스만 백만 점이 읽히지 않은 채로 있다[6]. 비문 자료는 지역과 시대가 고르지 않다[2]. 기계가 잘 다루지 못하는 긴 결손부도 그대로 남아 있다[3]. 이 문제들은 더 좋은 자료와 더 많은 검증으로 하나씩 줄여 가야 한다[4].

기계가 후보를 내놓고 사람이 고른다. 이타카를 62퍼센트에서 72퍼센트로 끌어올린 방법이 그 답이다[1]. 침묵하던 돌과 종이가 다시 말을 하려면 두 쪽이 함께 있어야 한다.

참고 자료

[1] 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

[2] 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

[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] Zhang, S., Caraffa, E., Zuffrano, A., Modesti, M., & Colavizza, G. (2026, 8월 28일). Text restoration of ancient documents with language models (arXiv:2608.28170). arXiv. https://arxiv.org/abs/2608.28170 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

12 CHAPTER

제12장. 영원히 타지 않는 가상 도서관: 초고해상도 다분광 아카이브와 미래

다분광 촬영과 지식 그래프가 짓는 불타지 않는 도서관

1. 사건 파일

알렉산드리아 도서관이 사라진 방식

이집트 알렉산드리아에는 고대 세계 최대의 도서관이 있었다 . 프톨레마이오스 왕조가 기원전 3세기 무렵에 세웠다 . 왕들은 세상의 책을 한자리에 모으려 했다 . 항구에 들어온 배의 책은 거두어 베껴 쓰게 했다 . 사본을 주인에게 돌려주고 원본은 도서관이 가졌다 . 옛 기록은 두루마리가 수십만 개였다고 전한다 [1].

많은 사람은 이 도서관이 큰불 하나로 사라졌다고 생각한다. 그러나 연구자들은 서서히 무너졌다고 본다 . 기원전 145년에는 왕이 외국 학자들을 내쫓았다 . 학자가 떠나자 책을 돌보는 손도 함께 떠났다 . 로마 시대에는 건물이 여러 차례 부서졌다 . 새 책을 사들일 돈도 끊겼다 [1].

고대의 글은 원래도 잘 남지 않는다. 그때의 서기들은 새 글을 짓기보다 있는 책을 베끼는 일을 많이 했다. 손을 많이 탄 파피루스 두루마리는 백 년쯤 지나면 낡아서 읽기 어려웠다. 건조한 이집트 땅속에서는 훨씬 오래 남기도 했다. 그래서 누군가 때맞춰 다시 베껴야 글이 이어졌다. 베끼는 손이 끊기면 그 책은 세상에서 지워졌다.

우리가 오늘 읽는 고대 문헌은 그 시절 글의 일부다. 이름만 전해지고 본문은 사라진 책이 훨씬 많다. 고대 그리스 비극 작가들이 쓴 작품 대부분도 그렇게 없어졌다. 남은 책들은 운이 좋아 여러 곳에 사본이 흩어져 있던 것들이다.

파괴는 지금도 이어진다

기록을 부수는 힘은 옛날에만 있던 것이 아니다. 전쟁은 지금도 도서관과 기록보관소를 무너뜨린다. 유네스코는 우크라이나에서 부서진 문화 시설을 하나씩 확인해 세고 있다

. 2026년 7월 29일 기준으로 확인된 곳은 545곳이다 . 그 안에는 박물관 43곳, 도서관 22곳, 기록보관소 5곳이 들어 있다 . 종교 시설과 오래된 건물도 수백 곳이 함께 상했다 [2].

전쟁은 유물을 부수기만 하지 않는다. 관리가 멈춘 유물 보관 창고에서 유물이 몰래 빠져나가기도 한다. 어디서 나왔는지 기록이 지워진 유물은 연구 가치를 크게 잃는다. 점토판 한 장이 어느 방에서 어떤 판들과 같이 있었는지가 곧 정보이기 때문이다.

기후도 기록을 위협한다. 바닷가에 있는 유적은 파도에 깎여 조금씩 무너진다. 땅속 유물은 비가 늘거나 지하수가 차면 더 빨리 상한다. 발굴하는 속도보다 사라지는 속도가 빠른 곳도 있다.

부서진 유물이 다 없어지는 것은 아니다. 형태가 남아 있으면 다시 세울 길이 있다. 그래서 사고가 나기 전에 미리 재어 두는 일이 중요하다. 전쟁이 이어지는 지역에서는 건물과 유물을 3차원으로 재 두는 작업이 진행된다.

재료마다 수명이 다르다

기록을 담는 재료는 저마다 다른 방식으로 망가진다. 메소포타미아의 점토판은 진흙을 눌러 글자를 새긴 판이다. 궁전이 불탈 때 생긴 열은 점토판을 도자기처럼 단단하게 구웠다. 불을 만나지 못한 점토판은 물이 들어오면 다시 진흙으로 풀어진다.

점토판에는 다른 한계도 있다. 서기들은 다 쓴 판을 물에 적셔 겉면을 밀어내고 다시 썼다. 그래서 오래된 판은 대개 다시 진흙으로 돌아갔다. 왕국이 무너지기 직전의 몇 해만 남은 경우가 많은 까닭이다. 궁전이 불탈 때 서고에 있던 마지막 판들만 단단하게 구워졌기 때문이다.

파피루스와 가죽과 종이는 더 약하다. 이 재료들은 식물과 동물에서 나온 유기물이다. 덥고 습한 곳에서는 몇백 년을 버티지 못하고 썩는다. 로마 군인들이 쓰던 밀랍 판에서는 밀랍이 통째로 사라졌다. 남은 것은 철필이 나무판까지 긁고 지나간 얕은 흠뿐이다.

단단한 재료라고 안전하지도 않다. 거북 배딱지와 소 어깨뼈에 새긴 갑골문도 땅속에서 눌려 깨졌다 . 오늘 남은 갑골은 대부분 손바닥보다 작은 조각이다 . 한 문장이 세 파편으로 흩어져 다른 창고에 들어간 경우도 있다 [3].

불이 살린 도서관

기록이 살아남는 방식은 때로 뜻밖이다. 아시리아 왕 아슈르바니팔은 니네베에 큰 서고를 두었다 . 그는 기원전 669년 무렵부터 나라를 다스렸다 . 서고에는 신화와 점술과 의학과 행정 문서가 모여 있었다 [4].

기원전 612년 무렵 니네베는 불탔다 . 종이로 된 책이었다면 그때 모두 사라졌을 것이다 . 그런데 점토판은 불에 구워져 오히려 더 단단해졌다 . 이 판들은 1840년대부터 1930년대까지 여러 차례 발굴되었다 . 지금 대영박물관에는 그 파편이 3만 점 넘게 있다 [4].

그렇다고 온전한 상태로 나온 것은 아니다 . 도시가 무너질 때 판들은 일부러 깨져 궁전 곳곳에 버려졌다 . 발굴자들이 만난 것은 정리된 서가가 아니라 흩어진 파편 더미였다 . 한 작품의 앞뒤가 서로 다른 방에서 나오는 일이 흔했다 [4]. 이 파편들을 다시 맞추는 일이 오늘날 전산 고고학의 큰 숙제다. 전산 고고학은 컴퓨터로 옛 유물을 연구하는 분야다.

이미 되살아난 글자들

디지털로 옮기는 일에도 벽이 있다. 세상의 유물은 사람 손보다 훨씬 많다. 앞서 본 니네베 파편 더미가 그런 예다 [4]. 한 점씩 찍고 읽어 낸 글까지 붙이는 데는 오랜 시간이 든다. 그래서 기계가 사람의 속도를 늘려 주어야 한다.

사라져 가는 유물을 살리는 방법은 두 방향으로 나뉜다. 하나는 물건 자체를 오래 지키는 보존이다. 다른 하나는 물건을 건드리지 않고 속을 읽어 컴퓨터에 옮기는 일이다. 뒤쪽 방법은 이미 여러 차례 성과를 냈다.

이스라엘 엔게디에서 나온 두루마리는 불에 타 숯덩이가 되어 있었다 . 손으로 펴려 하면 그대로 부서질 상태였다 . 연구팀은 물건을 자르지 않고 속을 보는 CT 촬영으로 안을 들여다보았다 . 컴퓨터는 말려 있는 가죽 층을 수학으로 따라가 평면으로 폈다 . 그 위에서 금속 성분이 섞인 잉크가 드러났고 레위기 첫머리가 읽혔다 [5].

베수비오 화산재에 묻힌 헤르쿨라네움 두루마리도 길이 열렸다. 2026년 6월 25일 연구진은 두루마리 하나를 처음부터 끝까지 읽어 냈다고 밝혔다 . 이 두루마리의 이름은 PHerc. 1667이다 . 안에는 그리스어 스물두 단 분량의 윤리 이야기가 담겨 있었다 [6]. 도전 과제로 지정된 봉인 두루마리 열세 개에는 새로운 상금이 걸렸다 . 마감은 2027년 6월 25일이다 [7].

가상 도서관은 사진을 잔뜩 모아 둔 창고가 아니다. 유물의 모양과 잉크 성분과 읽어 낸 글이 한곳에 묶여 있어야 한다. 그래야 뒷날 새 방법이 나왔을 때 옛 자료를 다시 계산할 수 있다. 원본이 사라진 뒤에도 검증이 이어지려면 이 묶음이 필요하다.

촬영과 판독에는 돈과 사람이 든다. 어떤 유물을 먼저 찍을지 정하는 일도 쉽지 않다. 위험한 지역의 유물은 촬영 자체가 어렵다. 그래서 어느 나라나 순서를 정해 놓고 일한다.

이 사건들은 하나의 물음을 남긴다. 물건이 끝내 부서져도 그 안의 글은 남길 수 있는가. 그 답을 찾는 곳이 가상 도서관이다.

참고 자료

[1] Mark, J. J. (2023, 7월 25일). Library of Alexandria. World History Encyclopedia. https://www.worldhistory.org/Library_of_Alexandria/ (2026년 9월 3일 접속)

[2] UNESCO. (2026). Damaged cultural sites in Ukraine verified by UNESCO. UNESCO. https://www.unesco.org/en/ukraine-war/damaged-cultural-sites (2026년 9월 3일 접속)

[3] Ma, Y. R. (2026). Towards Computational Chinese Paleography. arXiv:2601.06753. https://arxiv.org/abs/2601.06753 (2026년 9월 3일 접속)

[4] The British Museum. What was Ashurbanipal's Library? The British Museum. https://www.britishmuseum.org/research/projects/what-was-ashurbanipals-library (2026년 9월 3일 접속)

[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, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

[7] Vesuvius Challenge. (2026). Open Prizes. scrollprize.org. https://scrollprize.org/prizes (2026년 9월 3일 접속)

2. 과학 수사대의 추적: 지식 그래프와 빛의 스펙트럼 분석

유물을 잇는 커다란 지도

옛 연구실에서는 유물의 모양과 글을 따로 다루었다. 모양은 보존 과학자가 재고 글은 문헌학자가 읽었다. 두 결과는 서로 다른 서랍에 들어갔다. 그래서 같은 유물을 두 번 연구하면서도 답을 합치지 못했다.

전산 고고학은 이 둘을 한 지도 위에 올린다. 그 지도의 이름이 지식 그래프다. 지식 그래프는 정보를 점과 선으로 그린 그림이다. 점은 사물이고 선은 사물 사이의 관계다.

이 지도에서는 점토판 하나가 점이 된다. 그 판을 쓴 서기도 점이 되고 판이 나온 도시도 점이 된다. 잉크의 성분과 3차원 모양 정보도 각각 점이 된다. 판에 적힌 사람 이름과 물건 이름도 점으로 뽑아 놓는다. 컴퓨터는 이 점들을 잇는 선을 찾아 학습한다.

지도가 만들어지면 질문의 모양이 달라진다. 예전에는 판 하나를 놓고 무슨 글자인지 물었다. 이제는 이 잉크를 쓴 다른 문서가 어디 있는지 물을 수 있다. 같은 서기가 쓴 판이 어느 도시에서 나왔는지도 한 번에 찾는다.

2026년 2월 폴란드 야기엘로니안 대학교 연구팀이 이런 지도를 공개했다 . 이 지도에는 사람과 사건과 물건의 관계가 약 320만 줄로 들어 있다 . 연구팀은 기계가 관계를 추천하고 그 이유까지 설명하게 만들었다 . 전문가들이 그 설명을 읽고 쓸 만한지 확인했다 [1].

모양을 재는 층

지도의 첫 층은 유물의 겉모양이다. 오래된 비석과 점토판은 겉이 닳아 글자 홈이 얕아진다. 눈으로는 홈이 남았는지 알기 어렵다. 그래서 연구원들은 3차원 스캐너로 표면을 훑는다.

스캐너는 표면을 수백만 개의 점으로 바꾸어 저장한다. 컴퓨터는 그 점들을 이어 삼각형 그물로 만든다. 이 그물 위에서 가상의 빛을 여러 방향으로 비출 수 있다. 빛의 방향을 바꾸면 얕은 홈에도 그림자가 생겨 글자가 드러난다.

기계는 홈의 단면 모양도 배운다. 사람이 도구로 판 자국은 단면이 고르고 방향이 일정하다. 자연히 생긴 금은 그렇지 않다. 이 차이를 배운 모델은 글자와 흠집을 갈라낸다.

같은 방법으로 깨진 단면도 잰다. 두 파편의 단면 굴곡이 서로 맞물리는지 기계가 대조한다. 이 계산은 사람이 파편을 직접 들고 맞춰 보는 일을 대신한다. 손이 닿지 않는 다른 나라 박물관의 파편과도 맞춰 볼 수 있다.

말려 있는 두루마리에는 또 다른 계산이 쓰인다 . 엔게디 두루마리에서 연구팀은 가죽의 층을 컴퓨터 안에서 한 겹씩 갈라냈다 . 그리고 각 층을 평면으로 펴서 글자를 세웠다 [2].

빛으로 잉크를 읽는 층

지도의 두 번째 층은 잉크다. 잉크는 재료마다 빛을 다르게 받아들인다. 다분광 촬영은 눈에 보이는 빛뿐 아니라 적외선과 자외선까지 나누어 찍는 방법이다. 어떤 파장에서는 바탕이 어둡고 잉크가 밝게 뜬다.

이 차이 덕분에 지워진 글씨도 되살아난다. 옛 사람들은 값비싼 양피지를 아끼려고 글을 지우고 다시 썼다. 지운 표면에는 잉크 성분이 조금 남는다. 여러 파장으로 찍어 겹치면 그 흔적이 밑에서 떠오른다. 이렇게 두 번 쓴 문서를 팔림프세스트라고 부른다.

이집트 시나이산의 성 카타리나 수도원에는 이런 문서가 많다 . 미국과 그리스의 연구진이 여기서 74권을 다분광으로 찍었다 . 촬영한 면은 6,800쪽에 이른다 . 그 밑에서 지워진 글 305편이 다시 드러났다 . 그 글들은 그리스어와 시리아어를 비롯한 열 개 언어로 쓰여 있었다 [3].

성분 자체를 재는 방법도 있다. 라만 분광법은 잉크에 빛을 쏘고 되튀어 나오는 빛의 무늬를 읽는다. 무늬는 물질마다 달라서 지문 구실을 한다. 인공지능은 이 무늬를 보고 그을음으로 만든 먹인지 쇠 성분이 든 잉크인지 가른다. 잉크의 배합은 공방마다 조금씩 달라 출신지를 짐작하는 단서가 된다.

헤르쿨라네움 두루마리에는 더 센 빛이 쓰였다 . 연구진은 아주 강한 X선을 만드는 방사광 가속기로 두루마리 속을 찍었다 . 탄 파피루스와 탄소 먹은 밀도가 비슷해 그냥은 구별되지 않는다 . 딥러닝 모델이 미세한 결의 차이를 배워 글자가 있던 곳을 찾아냈다 [4].

글과 글을 잇는 층

지도의 세 번째 층은 문서들 사이의 관계다. 고대의 책은 베껴 쓰며 퍼졌다. 베낄 때마다 작은 오류가 하나씩 끼어들었다. 학자들은 이 오류를 따라가 사본의 족보를 그린다.

컴퓨터는 이 작업에 생물학에서 쓰던 계산법을 빌려 온다. 유전자 서열을 비교해 생물의 계통도를 그리는 방식과 닮았다. 문서의 차이를 세어 어느 사본이 어느 사본에서 나왔는지 배열한다. 이렇게 그린 족보는 잃어버린 원본의 모습을 짐작하게 해 준다.

글씨체도 관계를 잇는 실마리다. 서기마다 붓을 누르는 힘과 획의 각도가 다르다. 인공지능은 이 버릇을 숫자로 바꾸어 저장한다. 그 숫자가 가까운 문서끼리 묶으면 서기의 관계망이 나온다.

이미 돌아가고 있는 아카이브

이런 층들을 모아 실제로 돌아가는 아카이브가 여럿 있다. 트리스메기스토스는 고대 세계의 문서에 번호를 붙이는 사업이다 . 이곳은 100개가 넘는 언어의 자료 96만 건 남짓을 한 체계로 묶었다 . 같은 유물을 나라마다 다르게 부르던 문제가 이 번호로 풀렸다 [5].

펠라기오스 네트워크는 흩어진 자료를 서로 연결하는 공동체다 . 이들은 장소와 사람과 시간을 잇는 공통 방식을 정한다 . 옛 지명이 시대마다 바뀌어도 한 좌표로 묶이게 만든다 [6]. 뮌헨에서 운영하는 전자 바빌로니아 도서관이 있다 . 이곳은 세계 곳곳의 쐐기문자 파편을 한 데이터베이스에 모은다 . 학자들은 여기서 파편을 검색해 잃어버린 문장을 되찾는다 [7].

공통 번호는 지루해 보이지만 힘이 세다 . 번호가 같아야 서로 다른 기관의 자료가 한 줄로 이어진다 [5]. 사진은 박물관에, 읽어 낸 글은 대학에, 발굴 기록은 연구소에 있는 경우가 흔하다. 번호가 이 셋을 묶으면 흩어진 정보가 한 유물의 이야기로 모인다.

자료를 공개하는 방식도 달라졌다. 예전에는 사진과 읽어 낸 글을 논문이 나온 뒤에야 볼 수 있었다. 요즘 연구팀은 스캔 자료와 프로그램을 함께 공개한다 . 그러면 장비가 없는 나라의 연구자도 같은 자료로 다시 계산해 볼 수 있다 [8].

아직 남은 숙제

베수비오 챌린지는 2026년 7월 10일에 남은 문제를 정리해 공개했다 . 눌려 뭉개진 곳에서는 두루마리 층이 서로 나뉘지 않는다 . 컴퓨터가 층을 따라가다 옆 층으로 잘못 건너뛰기도 한다 . 잉크가 표면에서 얼마나 깊이 스몄는지도 아직 확실하지 않다 . 한 두루마리로 배운 모델이 다른 두루마리에서는 힘을 잃는다 . 자료의 크기가 커서 보통 컴퓨터로는 다루기 어렵다 . 연구진이 잡은 다음 목표 시점은 2027년 6월이다 [8].

옛 글자를 다루는 다른 분야도 같은 벽을 만난다. 중국 고문자 연구를 정리한 2026년 논문은 자료 부족을 큰 난제로 꼽았다 . 그리고 사람의 판단을 대신하기보다 돕는 방향을 권했다 [9].

참고 자료

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[4] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

[5] KU Leuven. Trismegistos portal. Trismegistos+. https://tmplus.kuleuven.be/trismegistos-portal (2026년 9월 3일 접속)

[6] Pelagios Network. Pelagios Network. https://pelagios.org/ (2026년 9월 3일 접속)

[7] Ludwig-Maximilians-Universität München & Bayerische Akademie der Wissenschaften. electronic Babylonian Library. https://www.ebl.badw.de/ (2026년 9월 3일 접속)

[8] Vesuvius Challenge. (2026, 7월 10일). Open problems: Why reading every Herculaneum scroll is still a challenge. scrollprize.org. https://scrollprize.org/2026_open_problems (2026년 9월 3일 접속)

[9] Ma, Y. R. (2026). Towards Computational Chinese Paleography. arXiv:2601.06753. https://arxiv.org/abs/2601.06753 (2026년 9월 3일 접속)

3. 에필로그

열한 번의 사건을 되짚는다

이 책은 열한 개의 사건을 따라왔다. 시작은 화산재에 묻힌 헤르쿨라네움의 두루마리였다. 손대면 부서지는 숯덩이를 컴퓨터가 가상으로 펼쳤다 . 탄 파피루스와 탄소 먹은 X선으로 보면 거의 같아 보인다 . 기계는 표면에 남은 미세한 결의 차이를 배워 글자를 찾아냈다 [1].

엔게디의 레위기 두루마리도 가상 펼침으로 첫 구절을 되찾았다 . 이쪽에서는 금속 성분이 든 잉크가 X선에 뚜렷이 잡혔다 [2]. 이사야 두루마리에서는 글씨체가 단서였다. 인공지능은 획의 미세한 버릇을 재어 서기가 한 명이 아님을 보였다 . 사람의 눈으로는 같은 글씨로만 보이던 부분이었다 [3].

네 번째 사건부터는 조각 맞추기가 이어졌다. 흩어진 점토판과 양피지 파편이 단면 모양으로 다시 붙었다. 위조된 유물을 가려내는 일에도 같은 계산이 쓰였다. 길가메시 서사시의 빠진 줄은 파편을 이어 붙여 채웠다.

메소포타미아의 영수증은 기계 번역으로 말문이 트였다. 수천 년 전의 곡식 장부와 편지가 오늘 말로 옮겨졌다. 거북 배딱지의 갑골문은 3차원 스캔과 전용 모델로 이어 붙었다 [4]. 신라의 목간도 글자 인식 모델이 읽어 냈다. 일본의 초서체 문서도 같은 방식으로 풀렸다 [5].

아직 열리지 않은 문도 보았다. 선형문자 A와 인더스 문자와 롱고롱고는 여전히 답이 없다. 기계는 글자 배열의 규칙성을 재어 그것이 언어인지 아닌지 따졌다. 뜻을 읽는 일은 다음 세대의 몫으로 남았다.

마지막 사건은 고대 비문 자동완성이었다. 이타카는 그리스어 비문의 빠진 글자를 채우는 모델이다 . 이 모델은 글자를 채우면서 비문이 만들어진 곳과 시기도 함께 짚는다 [6]. 뒤이어 나온 아이네아스는 라틴어 비문을 다룬다 . 아이네아스는 옮겨 적은 글자와 비문 사진을 함께 본다 [7].

열한 번의 사건에는 공통점이 있다. 어느 것도 기계 혼자 푼 문제가 아니었다. 물리학자가 촬영을 맡고 전산학자가 계산을 맡았다. 고대어를 읽는 학자가 결과를 확인하고 뜻을 붙였다. 세 분야가 같은 유물을 두고 만나야 한 줄이 살아났다.

기계가 아직 못 하는 일

여기까지 오는 동안 기계의 약점도 함께 드러났다. 첫 번째 벽은 자료 부족이다. 선형문자 A는 남은 기호를 다 합쳐도 7,400개 남짓이다 . 이만한 양으로는 기계가 규칙을 배우기 어렵다 [8]. 오늘의 언어 모델은 수십억 낱말을 읽고 자란다. 사라진 옛말의 자료는 그 크기에 한참 못 미친다.

자료가 적은 곳에서 무리하게 답을 내면 없는 낱말을 지어낸다. 빈칸이 길어질수록 답의 신뢰도는 떨어진다. 문맥에 그럴듯하지만 역사에 없던 문장이 나올 수 있다. 그런 문장이 한 번 논문에 실리면 다음 연구가 그것을 사실로 받는다.

학습 자료가 한쪽에 쏠려도 답이 기운다. 어느 지역 비문이 많으면 다른 지역 비문도 그 지역 양식으로 읽어 버린다. 기계는 왜 그렇게 답했는지 스스로 설명하지 못할 때가 많다. 그래서 요즘 연구는 근거를 함께 내놓는 쪽으로 옮겨 가고 있다 [9].

두루마리 읽기 연구팀도 남은 문제를 목록으로 공개했다 . 눌려 뭉개진 곳에서는 층이 서로 나뉘지 않는다 . 한 두루마리로 배운 모델이 다른 두루마리에서는 힘을 잃기도 한

다 . 무엇을 못 하는지 밝히는 일이 다음 걸음을 만든다 [10].

기계가 내놓은 글자를 다루는 규칙도 필요하다. 확실히 읽은 글자와 추측한 글자는 표시가 달라야 한다. 어떤 모델이 어떤 자료로 그렇게 읽었는지도 함께 적어야 한다. 그래야 나중에 다른 연구자가 같은 계산을 다시 해 볼 수 있다.

원본은 원본대로 남는다

디지털 복원이 아무리 정교해도 원본을 대신하지는 못한다. 학계는 스캔한 자료를 원본이 아니라 디지털 분신으로 본다. 촬영에는 사람의 선택이 들어간다. 어떤 빛을 쓰고 어떤 각도로 찍을지 사람이 정하기 때문이다. 그 선택이 달라지면 결과 이미지도 달라진다.

실물에는 스캐너가 담지 못한 정보가 남아 있다. 점토판 표면의 흙 알갱이 구조가 그렇다. 거북 배딱지의 미세한 구멍 배열도 실물에서만 볼 수 있다. 잉크가 파피루스 속으로 얼마나 스몄는지도 실물에서만 잰다 [10].

새 측정 기술은 앞으로도 나온다. 지금 버린 유물에서는 그 기술을 쓸 수 없다. 그래서 유물을 자르거나 억지로 펴는 일은 삼가야 한다. 가상 도서관의 목적은 원본을 버리는 것이 아니다. 원본을 덜 만지고도 읽게 만드는 데 목적이 있다 [2]. 유물을 만지는 횟수가 줄면 원본의 수명은 늘어난다.

함께 읽는 미래

인공지능은 역사가를 밀어내는 도구가 아니다. 이타카 연구는 이 점을 숫자로 보여 주었다 . 이타카가 혼자 빠진 글자를 복원했을 때 정확도는 62퍼센트였다 . 역사가들이 혼자 했을 때는 25퍼센트였다 . 두 쪽이 함께 하자 72퍼센트로 올랐다 [6].

이 숫자는 두 가지를 함께 말해 준다. 이 실험에서 기계 혼자는 사람보다 나았다. 그리고 사람과 함께했을 때 성적이 더 올랐다. 어느 한쪽만으로는 여기까지 오지 못했다.

기계는 후보를 빠르게 많이 내놓는다. 사람은 그 후보가 그 시대에 있을 법한지 따진다. 이렇게 사람과 기계가 번갈아 확인하는 방식을 학계는 사람이 고리 안에 있는 협업이라고 부른다. 이 고리가 돌아가려면 옛 글을 읽을 줄 아는 사람이 계속 있어야 한다. 읽을 사람이 없으면 기계가 뽑아낸 글자도 또 하나의 수수께끼가 된다.

가르치는 일도 연구만큼 중요하다. 고대어를 읽는 법은 사람에게서 사람으로 이어져 왔다. 그 사슬이 끊기면 기계가 뽑은 글자를 검증할 사람이 없다. 새 세대는 옛 글자와 컴퓨터를 함께 배워야 한다.

이 고리는 전문가만의 것이 아니다. 일본에서는 초서체를 읽어 주는 무료 앱 미오가 나와 있다 . 사진을 찍으면 화면에 오늘날 글자가 뜬다 . 옛 문서를 처음 보는 사람도 첫 줄을 읽을 수 있다 [5]. 박물관 유리장 앞에서 스마트폰을 드는 일이 연구의 시작이 되기도 한다.

앞으로 할 일은 아직 많다. 헤르쿨라네움에는 봉인된 두루마리가 아직 남아 있다 [11]. 우크라이나에서는 지금도 도서관과 기록보관소가 부서지고 있다 [12]. 세상의 옛 글은 여전히 사라지는 쪽이 더 빠르다.

이 책을 읽은 사람이 할 수 있는 일도 있다. 공개된 스캔 자료는 누구나 내려받아 볼 수 있다 [10]. 파편을 맞추는 대회에는 학생도 참가한다 [11]. 옛 문서를 옮겨 적는 일에 시민이 함께하는 곳도 많다. 한 사람이 한 줄씩 맡아도 자료는 쌓인다.

기록을 지키는 일은 옛날 이야기만이 아니다. 오늘 우리가 만드는 자료도 언젠가 옛 자료가 된다. 파일 형식이 바뀌면 지금의 문서도 읽히지 않는다. 그래서 도서관들은 자료를 새 형식으로 옮기는 일을 계속한다. 가상 도서관을 세우는 일에는 관리하는 손이 늘 따라붙는다.

그래도 방향은 정해졌다. 2000년 전 화산재에 묻힌 두루마리가 다시 읽혔다 [1]. 한 두루마리에서 얻은 방법은 다음 두루마리 연구의 출발점이 된다 [11]. 자료와 코드를 공개해 두면 다른 나라 연구자도 같은 길을 간다 [10]. 유물은 언젠가 흙으로 돌아가더라도 그 안의 문장은 남는다. 불에 타지 않는 도서관은 그렇게 한 줄씩 세워지고 있다.

참고 자료

[1] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[6] 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

[7] Assael, Y., Sommerschield, 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 (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[11] Vesuvius Challenge. (2026). Open Prizes. scrollprize.org. https://scrollprize.org/prizes (2026년 9월 3일 접속)

[12] UNESCO. (2026). Damaged cultural sites in Ukraine verified by UNESCO. UNESCO. https://www.unesco.org/en/ukraine-war/damaged-cultural-sites (2026년 9월 3일 접속)

APPENDIX 부록

부록. 나만의 AI 고고학 여행기: 대중을 위한 글로벌 탐사 포털

누구나 집에서 따라 해 보는 고대 문자 탐사 안내

1. Vesuvius Challenge 공식 데이터 아카이브

무엇을 볼 수 있나

베수비오 챌린지는 국제 대회다. 불에 탄 두루마리를 컴퓨터로 읽는 일을 겨룬다. 두루마리가 나온 곳은 헤르쿨라네움이다. 베수비오 화산 옆에 있던 옛 로마 도시다. 두루마리는 파피루스로 만들었다. 파피루스는 풀을 눌러 만든 옛날 종이다. 공식 주소는 scrollprize.org다[1]. 이 사이트는 대회에서 쓰는 자료를 누구에게나 공개한다[2]. 회원 가입도 필요 없고 돈도 들지 않는다[2][5].

첫 화면 위쪽에는 메뉴가 줄지어 있다. Prizes, Problems, Tutorials & In-depth, Data가 앞에 온다. 그 뒤로 Milestones, Community, Press, Donate가 이어진다. 2026년 9월 현재 두루마리와 파피루스 파편 45점이 스캔되어 있다. 그중 한 점은 처음부터 끝까지 읽혔다[1].

2026년 6월 25일에 대회 팀은 그 한 점의 소식을 알렸다. PHerc. 1667이라고 부르는 두루마리다[3]. 두루마리를 펴지 않고 끝까지 읽어 낸 첫 사례다[3][4]. 펼쳐진 파피루스 길이는 약 1.4미터다. 옛 책은 글을 세로 칸으로 나누어 적었다. 이 칸 하나를 단이라고 부른다. 이 두루마리에서는 그리스 문자로 적은 단이 스물두 개쯤 나왔다. 내용은 스토아 학파의 윤리를 다룬 글이다. 사람의 본성과 충동, 도덕적 성장을 이야기한다. 기원전 2세기 무렵의 글로 본다[3].

두루마리 전체가 온전히 남은 것은 아니다. 지금 남은 것은 두루마리 안쪽 심지다. 바깥 겹은 19세기와 1980년대에 손으로 펴다가 부서졌다. 표면이 떨어져 나간 곳은 글자도 함께 사라졌다. 그래서 읽어 낸 글에는 군데군데 빈 곳이 있다[3].

스캔은 프랑스 그르노블에서 했다. 유럽 싱크로트론 방사광 시설 ESRF다. 아주 밝은 엑스선을 만드는 커다란 연구 장치다. 이 가운데 BM18 빔라인을 썼다[4]. 빔라인은 엑스선을 끌어와 물체에 쬐는 실험 통로다.

집에서 따라 하는 순서

먼저 scrollprize.org에 들어가 Get Started 단추를 누른다. 이 화면은 처음 오는 사람에게 세 단계를 안내한다[5].

첫 단계는 잉크 찾기다. 눈에 보이는 잉크를 화면에서 칠해 준다. 그러면 모델이 그 표시를 배워 나머지를 예측한다[5].

둘째 단계는 가상 펼치기다. 컴퓨터로 두루마리를 펴 보는 일이다. CT 단면 사진에서 파피루스 한 겹을 따라간다. CT 단면 사진은 속을 얇게 잘라 찍듯이 찍은 사진이다. 그 겹을 평평하게 펴서 결과를 확인한다[5].

셋째 단계는 전체 과정 익히기다. 스캔과 펼치기와 읽기를 작은 크기로 이어서 해 본다[5].

체험 순서와 실제 복원 순서는 다르다는 점을 알아 두어야 한다. 체험은 쉬운 잉크 찾기부터 시작한다. 실제 복원은 스캔이 먼저다. 그다음 두루마리를 가상으로 편다. 잉크를 읽는 일은 맨 마지막이다[5].

자료를 직접 내려받으려면 scrollprize.org/data로 간다. 자료는 공개 저장소 s3://vesuvius-challenge-open-data/에 있다. 웹 브라우저로도 열어 볼 수 있다. 쓰는 법을 보여 주는 예제 파일은 깃허브에 있다[2].

도전자에게는 상금도 걸려 있다. 2026년 9월 기준 남은 상금은 214만 달러다. 2027 그랜드 프라이즈는 100만 달러 규모다. 1등이 80만 달러를 받는다. 마감은 2027년 6월 25일이다. 아직 글자를 찾지 못한 두루마리에도 상금이 걸려 있다. 넓이 4제곱센티미터 안에서 글자 열 개를 읽어 내면 된다. 이때 두루마리 한 점마다 5만 달러를 준다. 달마다 주는 프로그레스 프라이즈도 있다. 그달에 뽑힌 으뜸 제출물에는 2만 달러가 간다[6].

주의할 점

자료는 CC BY-NC 4.0 조건으로 공개된다. 돈을 버는 목적으로는 쓸 수 없다. 쓸 때는 출처를 밝혀야 한다. 옛 자료에는 인용할 논문이 따로 있다. 두루마리 1번에서 4번, 파편 1번에서 6번이 그렇다[2]. 이때는 Parsons 등의 2023년 논문을 적는다[2][7].

파일 크기가 크다는 점도 미리 알아 두어야 한다. 3차원 스캔 자료는 전체가 수 테라바이트에 이른다[2]. 딥러닝을 돌리려면 성능이 좋은 그래픽카드가 필요하다. 집 컴퓨터가 약하면 자료를 눈으로 둘러보는 일부터 하면 된다.

상금에 도전할 때는 조건이 붙는다. 대회 디스코드 방에 가입해야 한다. 코드는 MIT 같은 자유로운 조건으로 공개해야 한다. 수상 발표 전까지는 결과를 밖에 알리지 않는다[6].

참고 자료

[1] Vesuvius Challenge. (2026). Vesuvius Challenge: Reading the Herculaneum Scrolls with machine learning. scrollprize.org. https://scrollprize.org/ (2026년 9월 3일 접속)

[2] Vesuvius Challenge. (2026). Data. scrollprize.org. https://scrollprize.org/data (2026년 9월 3일 접속)

[3] Vesuvius Challenge. (2026, 6월 25일). An entire Herculaneum scroll has been read for the first time. scrollprize.org. https://scrollprize.org/firstscroll (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[5] Vesuvius Challenge. (2026). Get Started. scrollprize.org. https://scrollprize.org/get_started (2026년 9월 3일 접속)

[6] Vesuvius Challenge. (2026). Open Prizes. scrollprize.org. https://scrollprize.org/prizes (2026년 9월 3일 접속)

[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

2. eBL 전자 바빌로니아 문학 포털

포털의 얼개

eBL은 electronic Babylonian Library의 줄임말이다. 우리말로는 전자 바빌로니아 도서관이다. 옛 메소포타미아 사람들은 쐐기문자로 글을 적었다. 쐐기문자는 젖은 점토판에 눌러 새긴 못 모양 글자다. 이 포털은 그 점토판 파편을 한자리에 모은다. 모으는 대상은 바빌로니아 문학을 적은 파편이다. 세상의 모든 쐐기문자 자료를 담는 곳은 아

니다[1].

포털은 독일 뮌헨 루트비히 막시밀리안 대학교에 있다. 엔리케 히메네스 교수가 이끈다. 사업은 2018년 4월에 시작했다. 알렉산더 폰 훔볼트 재단의 소피아 코발레프스카야 상 지원금으로 운영한다[2].

주소는 www.ebl.lmu.de다. 로그인 없이 들어갈 수 있고 이용료도 없다. 화면 왼쪽에는 여덟 개의 항목이 있다. eBL Project, Library, Corpus, Signs가 앞에 온다. 그 뒤로 Akkadian Dictionary와 Bibliography가 온다. 끝에는 News와 Archaeology가 있다[1].

Library는 파편 하나하나를 모아 둔 곳이다. 예전에는 Fragmentarium이라고 불렀다. eBL 연구원들은 이곳에서 파편을 이어 붙인 사례를 약 1,200건 찾아냈다[1]. Corpus는 작품 단위로 본문을 보여 준다. 길가메시 서사시는 약 3,150행으로 정리되어 있다. 창조 신화 에누마 엘리시는 1,096행이다[3]. 2024년에 나온 자료 논문이 규모를 알려 준다. 이 포털은 점토판 약 25,000점을 담는다. 전사문은 35만 행이 넘는다[4]. 전사문은 쐐기문자를 라틴 글자로 옮겨 적은 글이다.

화면에서 무엇을 해 볼 수 있나

Library를 누르면 검색 상자가 나온다. 이름, 연도, 제목으로 찾을 수 있다. 낱말 단위 검색과 고급 검색도 있다. 무엇을 찾을지 모르겠으면 I'm feeling lucky 단추를 누른다. 그러면 아무 파편이나 하나 열린다[5].

파편 하나를 열면 화면이 여러 칸으로 나뉜다. 위쪽에는 소장 기관, 크기, 나온 곳이 적힌다. 글자체와 문서 유형도 함께 나온다. 예를 들어 K.5743은 대영박물관의 쿠윤지크 수집품이다. 쿠윤지크는 옛 도시 니네베가 있던 언덕이다. 그 언덕에서 나온 유물을 한데 묶어 부르는 이름이다. 크기는 3.81센티미터에 2.54센티미터다. 글자체는 신아시리아식이다[6]. 새 아시리아 제국 시대에 쓰던 글자 모양이다.

가운데에는 전사문이 행 번호와 함께 나온다. 아래쪽에는 점토판 사진이 붙는다. 어느 파편과 이어 붙었는지도 함께 적힌다[6].

글을 인용할 때 쓰는 방식도 사이트가 알려 준다. 파편 번호와 eBL edition을 적는다. 그 뒤에 주소와 접속 날짜를 붙인다[1].

새 소식과 주의할 점

News 항목에서 소식지를 볼 수 있다. 2026년 5월 28일자 22호가 맨 나중 것이다. 이 호에서 포털은 새 화면 구성을 내놓았다. 소장 기관도 여러 곳 늘었다. 인천에 있는 국립세계문자박물관도 그중 하나다. 대영박물관 사진 3,100장 이상이 새로 올라왔다. 예일 피바디 박물관 사진도 1,000장 이상 올라왔다. 사전에 새 낱말을 직접 등록하는 기능도 생겼다[7].

쓰기 전에 알아 둘 점이 있다. Library의 전사문은 완성본이 아니라 임시본이다. 연구진이 지금도 계속 고치고 있다. Corpus에 정리된 작품 본문도 마찬가지다. 지금까지 모은 파편으로 짜 맞춘 중간 결과다. 새 파편이 붙으면 행 수와 내용이 바뀐다. 그러니 학교 과제나 글에 쓸 때는 접속 날짜를 함께 적는다[1].

전사문은 CC BY-NC-SA 4.0 조건으로 공개된다. 사진은 조건이 다르다. 사진을 다시 쓰려면 허락을 따로 받아야 한다[1]. 점토판을 가진 박물관이 허락하는 쪽이다[1][4]. 화면 아래 사진 설명에 붙은 안내를 먼저 읽어야 한다[1].

참고 자료

[1] electronic Babylonian Library. (2026). About: Library. ebl.lmu.de. https://www.ebl.lmu.de/about/library (2026년 9월 3일 접속)

[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 (2026년 9월 3일 접속)

[3] electronic Babylonian Library. (2026). Corpus. ebl.lmu.de. https://www.ebl.lmu.de/corpus (2026년 9월 3일 접속)

[4] Cobanoglu, Y., Laasonen, J., Simonjetz, F., Khait, I., Cohen, S., Földi, Z., Hätinen, 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 (2026년 9월 3일 접속)

[6] electronic Babylonian Library. (2026). K.5743, eBL edition. ebl.lmu.de. https://www.ebl.lmu.de/library/K.5743 (2026년 9월 3일 접속)

[7] electronic Babylonian Library. (2026, 5월 28일). eBL Newsletter 22. ebl.lmu.de. https://www.ebl.lmu.de/news (2026년 9월 3일 접속)

3. Miwo 사용 안내

어떤 앱인가

일본의 옛 책과 문서에는 흘림체가 많이 쓰였다. 이 글자를 쿠즈시지라고 부른다. 에도 시대 인쇄본과 옛 문서에서 자주 보인다[3]. 오늘날 일본 사람 대부분은 이 글자를 읽지 못한다.

미오는 이 흘림체를 사진으로 찍어 오늘날 글자로 바꿔 주는 스마트폰 앱이다. 일본의 인문학 오픈데이터 공동이용센터 CODH가 만들었다[1]. 값은 무료다[2]. 앱 이름은 겐지 이야기 열네 번째 권 미오쓰쿠시에서 따 왔다. 미오쓰쿠시는 뱃길을 알려 주는 표지다. 옛 문서의 바다에서 길잡이가 되겠다는 뜻이다[3].

앱은 2021년 8월 30일에 처음 나왔다. 2022년 10월 26일에 1.1판으로 올라갔다. 이 때 루리라는 새 인식 모델이 들어갔다. 루리는 글자 배경에 무늬나 색이 있어도 잘 찾아낸다[1]. 학습에는 일본 고전적 쿠즈시지 데이터셋을 썼다. 기계 학습을 위해 만든 옛 글자 사진 모음이다. 글자 백만 자가 넘는 자료다. 이 앱은 2022년 굿디자인상을 받았다[3]. 2026년 9월 기준으로 이 앱이 인식한 사진은 350만 장이 넘는다[1].

설치와 사용 순서

먼저 앱 장터를 연다. 아이폰과 아이패드는 App Store를 쓴다. 안드로이드 기기는 Google Play를 쓴다. 검색창에 みを 또는 miwo를 넣는다[1]. App Store 기준으로 앱 크기는 약 52메가바이트다. 아이폰과 아이패드는 iOS 11.0 이상이면 된다[2].

앱을 켜면 사진 찍는 화면이 나온다. 옛 문서를 화면에 담고 촬영 단추를 누른다. 이미 찍어 둔 사진을 불러와도 된다. 그다음 인식 단추를 누르면 인공지능이 글자를 읽는다[1].

결과 화면에는 여러 기능이 있다. 화면에는 조절 막대가 하나 놓여 있다. 이 막대를 좌우로 밀면 두 화면이 겹쳐 보인다. 원본 사진과 읽어 낸 글자를 나란히 견줄 수 있다. 일본어 가나 글자를 누르는 기능도 있다. 그 글자가 원래 어떤 한자에서 나왔는지 알려 준다. 읽어 낸 글은 복사 단추로 다른 앱에 옮길 수 있다[1]. 결과를 저장했다가 나중에 다시 꺼내 볼 수도 있다[1][3].

인터넷이 켜져 있어야 한다는 점을 기억해야 한다. 사진은 인식을 위해 CODH 서버로 보내진다. 보낸 사진은 서버에 남기지 않는다[1].

잘 읽는 자료와 못 읽는 자료

앱을 쓰기 전에 한계를 알아 두면 실망이 적다. 인공지능은 에도 시대 인쇄본을 중심으로 학습했다. 그 시기 책은 목판 인쇄본이 많다. 나무판에 글자를 새겨 종이에 찍어 낸 책이다. 그래서 그 시기 인쇄본은 잘 읽는다. 다른 시대 자료나 손으로 쓴 문서는 정확도가 떨어진다[3].

종이 상태도 결과를 바꾼다. 얼룩이 지거나 벌레가 갉아먹은 곳은 읽기 어렵다. 종이의 무늬, 조명, 그림자도 방해가 된다[3]. 밝고 고른 빛 아래에서 찍는 편이 낫다.

읽지 못하는 대상도 있다. 돌비석이나 간판에 새긴 흘림체는 지금 지원하지 않는다[3].

개발진은 인식 결과에 틀린 곳이 있을 수 있다고 밝힌다. 인공지능 흘림체 인식은 완벽해지지 않는다고 못 박는다. 그래서 중요한 자료는 사람이 다시 확인해야 한다[3].

옛 글자를 오늘날 글자로 옮겨 적는 일을 번각이라고 한다. 이 일을 여럿이 나눠 하는 곳이 따로 있다. 민나 데 혼코쿠라는 시민 참여 번각 사이트다. 우리말로 옮기면 모두의 번각이다. 미오를 만든 곳과는 다른 기관이 꾸린다. 앱이 읽어 준 글을 사람이 다듬는 일과 성격이 같다. 주소는 honkoku.org다. 지금까지 1만 800명이 넘는 사람이 참여했다. 옮겨 적은 글자는 4,800만 자를 넘겼다. 2025년 8월에 두 번째 새 단장을 했다. 새 판은 시험 운영 중이다. 일본 국립역사민속박물관이 참여한다. 도쿄대학 지진연구소와 교토대학 연구진도 함께 꾸린다[4].

참고 자료

[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/ (2026년 9월 3일 접속)

[2] Clanuwat, T. (2022, 10월 27일). miwo (Version 1.1) [Mobile app]. App Store. https://apps.apple.com/jp/app/miwo/id1581794085 (2026년 9월 3일 접속)

[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/ (2026년 9월 3일 접속)

[4] みんなで翻刻. (2026). みんなで翻刻 Minna de Honkoku. honkoku.org. https://honkoku.org/ (2026년 9월 3일 접속)

AI 침묵의 글자를 깨우다

인공지능이 되살린 고대의 기록

ISBN |

저 자 | 김경진
펴낸이 | 김경진
펴낸곳 | 김경진 변호사 출판사
출판사등록 | 2025. 3. 10. (제2025-000015호)
주 소 | 서울특별시 동대문구 전농로 91, 백일빌딩 304호
전 화 | 02-6338-1905
이메일 | kimkj008@gmail.com

© 김경진 2026
본 책은 저작자의 지적 재산으로서 무단 전재와 복제를 금합니다.

참고) 이 책의 일부 내용과 편집 과정에는 인공지능 도구의 도움을 받았습니다.

이 책을 잘 읽으셨으면 그리고 새로운 가치있는 지식을 얻으셨다고 판단되시면
농협 302-1096-0948-81 (예금주 김경진) 에 자발적 후원 부탁드립니다.