MIT Technology Review

Humans and Technology

Machine learning has been used to automatically translate long-lost languages

Some languages that have never been deciphered could be the next ones to get the machine translation treatment.

by **Emerging Technology from the arXiv**   Jul 1, 2019

In 1886, the British archaeologist Arthur Evans came across an ancient stone bearing a curious set of inscriptions in an unknown language. The stone came from the Mediterranean island of Crete, and Evans immediately traveled there to hunt for more evidence. He quickly found numerous stones and tablets bearing similar scripts and dated them from around 1400 BCE.

that its linear form was clearly derived from rudely scratched line pictures belonging to the infancy of art, thereby establishing its importance in the history of linguistics.

He and others later determined that the stones and tablets were written in two different scripts. The oldest, called Linear A, dates from between 1800 and 1400 BCE, when the island was dominated by the Bronze Age Minoan civilization.

Sign up for The Download — your daily dose of what's up in emerging technology

| Enter your email | Sign up |

Also stay updated on MIT Technology Review initiatives and events? ○ Yes ○ No

The other script, Linear B, is more recent, appearing only after 1400 BCE, when the island was conquered by Mycenaeans from the Greek mainland.

Evans and others tried for many years to decipher the ancient scripts, but the lost languages resisted all attempts. The problem remained unsolved until 1953, when an amateur linguist named Michael Ventris cracked the code for Linear B.

His solution was built on two decisive breakthroughs. First, Ventris conjectured that many of the repeated words in the Linear B vocabulary were names of places on the island of Crete. That turned out to be correct.

His second breakthrough was to assume that the writing recorded an early form of ancient Greek. That insight immediately allowed him to decipher the rest of the language. In the process, Ventris showed that ancient Greek first appeared in written form many centuries earlier than previously thought.

Ventris's work was a huge achievement. But the more ancient script, Linear A, has remained one of the great outstanding problems in linguistics to this day.

You've read **1/3** of your free monthly feature stories. Subscribe for unlimited access.

Subscribe now   Sign in

First some background. The big idea behind machine translation is the understanding that words are related to each other in similar ways, regardless of the language involved.

So the process begins by mapping out these relations for a specific language. This requires huge databases of text. A machine then searches this text to see how often each word appears next to every other word. This pattern of appearances is a unique signature that defines the word in a multidimensional parameter space. Indeed, the word can be thought of as a vector within this space. And this vector acts as a powerful constraint on how the word can appear in any translation the machine comes up with.

These vectors obey some simple mathematical rules. For example: king – man + woman = queen. And a sentence can be thought of as a set of vectors that follow one after the other to form a kind of trajectory through this space.

The key insight enabling machine translation is that words in different languages occupy the same points in their respective parameter spaces. That makes it possible to map an entire language onto another language with a one-to-one correspondence.

In this way, the process of translating sentences becomes the process of finding similar trajectories through these spaces. The machine never even needs to "know" what the sentences mean.

This process relies crucially on the large data sets. But a couple of years ago, a German team of researchers showed how a similar approach with much smaller databases could help translate much rarer languages that lack the big databases of text. The trick is to find a different way to constrain the machine approach that doesn't rely on the database.

Now Luo and co have gone further to show how machine translation can decipher languages that have been lost entirely. The constraint they use has to do with the way languages are known to evolve over time.

The idea is that any language can change in only certain ways—for example, the symbols in related languages appear with similar distributions, related words have the same order of

✕

That's impressive work that takes machine translation to a new level. But it also raises the interesting question of other lost languages—particularly those that have never been deciphered, such as Linear A.

In this paper, Linear A is conspicuous by its absence. Luo and co do not even mention it, but it must loom large in their thinking, as it does for all linguists. Yet significant breakthroughs are still needed before this script becomes amenable to machine translation.

For example, nobody knows what language Linear A encodes. Attempts to decipher it into ancient Greek have all failed. And without the progenitor language, the new technique does not work.

But the big advantage of machine-based approaches is that they can test one language after another quickly without becoming fatigued. So it's quite possible that Luo and co might tackle Linear A with a brute-force approach—simply attempt to decipher it into every language for which machine translation already operates.

If that works, it'll be an impressive achievement, one that even Michael Ventris would be amazed by.

Ref: arxiv.org/abs/1906.06718 : Neural Decipherment via Minimum-Cost Flow: from Ugaritic to Linear B

Ͳ

Share   f   🐦   🔴   in   📱   ✉

Author   Emerging Technology from the arXiv

✕

Popular

01.

London police's face recognition system gets it wrong 81% of the time

02.

A third CRISPR baby may have already been born in China

03.

South Pole sea ice is now vanishing at an alarming rate, too

Special issue: space　July/August 2019

Does the world need a 3D-printed rocket?

Relativity Space, a well-funded startup, is going all-in on additive manufacturing. But is that too much of a good thing?

✕

Artificial Intelligence 6h

London police's face recognition system gets it wrong 81% of the time

The first independent evaluation of the Metropolitan police's use of face recognition warned it is "highly possible" the system would be ruled unlawful if challenged in court....

Expand

Humans and Technology Jul 3

TikTok is being investigated over its use of children's data (again)

You've read **1/3** of your free monthly feature stories. Subscribe for unlimited access.

Subscribe now   Sign in

February....

Expand

Computing Jul 3

Huawei is giving $300 million a year to universities with no strings attached

TR 35 2019

Innovators Under 35 | 2019

It's part of our ethos that technology can and should be a force for good. In these profiles ✕

You've read **1/3** of your free monthly feature stories. Subscribe for unlimited access.

Subscribe now | Sign in

Rediet Abebe   Tim Ellis   Ritu Raman

Silicon Valley Jul 3

Google's internet balloon project is about to start its first commercial trial

It'll be a crucial test of whether the technology can sustainably make money....

Expand

✕

Produced in association with Adobe

Biotechnology Jul 3

A third CRISPR baby may have already been born in China

Read more

Tech Policy Jul 2

Chinese border guards are putting a surveillance app on tourists' phones

The spyware gathers everything from text messages to people's contacts, and also looks for content China considers threatening....

✕

Sign up for **The Download** — your daily dose of what's up in emerging technology

Enter your email	Sign up

Also stay updated on MIT Technology Review initiatives and events? ◯ Yes ◯ No

---

Computing Jul 2

Military satellites are still worryingly vulnerable to cyberattack

A new report says hackers could wreak havoc by interfering with space-based communications and navigation services that NATO armies rely on....

Expand

Special issue: space   July/August 2019

Soon, satellites will be able to watch you everywhere all the time

Can privacy survive?

✕

course.	decade	future of space
April 2019	July 2019	March 2019

◀　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　▶

Space Jul 2

NASA just tested how its Orion crew capsule will keep astronauts safe in an emergency

Today's test was to see if NASA's new deep-space crew capsule can still keep its crew safe, even when everything goes wrong....

Expand

✕

BBVA   Produced in association with BBVA

Climate Change   Jul 1

South Pole sea ice is now vanishing at an alarming rate, too

Just a few months ago, Augustín Carstens, the general manager for the Bank for International Settlements (BIS), the so-called central bank for central banks, said his organization saw no value in the...

Expand

Biotechnology Jul 1

The CRISPR books racing to be the technology's definitive guide

Special issue: space July/August 2019

Spaceplanes, giant rockets, tethers and catapults

Five schemes for cheaper space launches—and five cautionary tales

Climate Change Jul 1

We've already built too many power plants and cars to prevent 1.5 ˚C of warming

✕

Space Jul 1

SpaceX has lost communication with three of its 60 Starlink satellites

They will continue to orbit Earth but will eventually be brought down by gravity, burning up in the

You've read **1/3** of your free monthly feature stories. Subscribe for unlimited access.

Subscribe now   Sign in