

Machine translation could make English-only science accessible to all

August 17 2022



**Translate
your work**

TO THE LOCAL
LANGUAGE

this helps to facilitate outreach to local communities
for political and environmental decision making

Machine learning using artificial intelligence has improved computer translation over the past decade, but scientific articles employing specialized jargon are still a challenge for machine translation. Nevertheless, scientists should prioritize translating articles into multiple languages to provide an equitable landscape for budding scientists worldwide, UC Berkeley researchers argue. Credit: Valeria

Ramírez-Castañeda, UC Berkeley

While still in high school, Xinyi Liu worked briefly in a lab at Beihang University in Beijing and was surprised to see Chinese researchers routinely using Google Translate to generate the first English draft of scientific papers. Translation is a must if scientists want to submit to high-profile journals, almost all of which are in English.

"It was normal for postdocs to just use Google Translate to first translate everything and then to modify and polish it. But after the first translation, the whole paper didn't make sense," said Liu, a rising junior at the University of California, Berkeley, who is majoring in molecular and cell biology. "Literally, all the words, all the terms were stuck together just randomly."

There had to be a better way, she thought.

So last year, when she saw a new seminar being taught by Rebecca Tarvin about breaking [language](#) barriers in science, she signed up.

That class, which will be taught at UC Berkeley for a third time in spring 2023, was a trial balloon for Tarvin, an assistant professor of integrative biology. With renewed campus-wide interest in diversity, equity and inclusion, she and working groups within her department thought that the class could help UC Berkeley address a long-standing issue in science: English, the dominant language of science, is a major obstacle to scientists who are not native English speakers.

It's not just [foreign students](#) and scientists who are at a disadvantage when science is communicated primarily in English. So are many American-born students. In fall 2020, about 40% of entering UC

Berkeley freshmen were first-generation college students, and within the 10-campus University of California system, 39% of first-generation students grew up with a language other than English as their first language.

"Many of our students from California grew up translating for their parents," Tarvin said. "Translation has been a part of their life since they were very young."

For Tarvin, the class—Breaking Language Barriers in Evolution and Ecology—was an "opportunity to both teach students skills in translation literacy, as well as encourage students to be activists in this realm of structural change. And in fact, I have seen a really positive reception of this sort of activism from the students, as they all seem to agree that addressing [language barriers](#) is really important after taking the course."

The class led Tarvin and some graduate students at UC Berkeley, along with collaborators in Canada, Israel and Hungary, to write a [scientific paper](#) evaluating new machine translation tools that can be used by people worldwide to make their scientific articles accessible to non-English speakers. The paper appeared online this month in the journal *BioScience*. Translations into Spanish, French, Portuguese and Hungarian, the languages of the co-authors, are also online.

"The idea here is that we're trying to give people the tools and motivation to translate their own scientific research," Tarvin said.

"Science doesn't need to be based on a single language. And there's lots of additional benefits that come from incorporating multilingual approaches in every phase of science. For example, publishing in multiple languages will benefit society because of better science communication."

"Language can be a barrier, as well as a fantastic tool, to bring people

together," emphasized Emma Steigerwald, who is first author of the paper and a UC Berkeley graduate student in environmental science, policy and management. "It's a barrier that we can surmount using this new technology. We explain about the technology and how it can be implemented and the things that we need to be aware of when we use the technology, and all the wonderful and positive ways that science communication can be transformed by bringing this new technology to bear."

Toward a multilingual scientific network

Until recently, computer translation was the butt of jokes. People shared amusing examples of mistranslations, often seeming to disparage languages other than English, and by implication, other cultures.

But [machine learning](#), or artificial intelligence, has dramatically increased translation accuracy to the extent that tourists use Internet services like Google Translate to communicate with people in the countries they visit.

But for text that contains lots of jargon—much of it scientific, but from many other academic fields, as well—Google Translate is woefully inadequate.

"The translation quality is not for a journal," said Ixchel Gonzalez Ramirez, one of the graduate student mentors for the course. "Many times, people have to pay for getting a professional translator to translate their work, and that's very expensive."

The new paper highlights some of the numerous services—most of them free—that can convert English scientific writing into other languages. Besides the well-known Google Translate platform, these include DeepL, which uses neural networks and claims to be many times more

accurate than competitors when translating English into Chinese, Japanese, Romance languages or German, and vice versa; Baidu Translate, a service by the Chinese Internet company Baidu that initially focused on translating between English and Chinese; Naver Papago, a multilingual translator created by a company in South Korea; and Yandex.Translate, which uses statistical machine translation and focuses mostly on Russian and English.

"Translation is becoming more and more in reach of any person. Whether or not you are an expert, and whether or not you even are bilingual, the ability to translate is just so expedited by so many of the technologies we have available today," Steigerwald said. "And so how can we integrate this into our workflow as scientists, and how does this change the expectations that surround scientific communication?"



Traduce tu trabajo

AL IDIOMA LOCAL

El aprendizaje automatizado que usa tecnologías de inteligencia artificial ha mejorado la traducción en computador en la última década. Sin embargo, los artículos científicos que emplean terminología especializada siguen siendo un reto para la traducción automática. No obstante, la comunidad científica debería dar prioridad a la traducción de artículos en varios idiomas para ofrecer un panorama equitativo a los científicos y las científicas en formación de todo el mundo, afirman los investigadores de la UC Berkeley. Credit: Valeria Ramírez-Castañeda, UC Berkeley

English is the lingua franca of science

Tarvin's interest in translation arose from one of her graduate students,

Valeria Ramírez Castañeda, who in 2020 [published a paper](#) describing the costs incurred by her fellow Colombian doctoral students who wanted to publish or interact with colleagues in a world dominated by English.

As an evolutionary biologist interested in how some animals came to use poison, Tarvin decided to focus her new seminar on translating papers in the fields of evolution and ecology, though students who signed up eventually charted their own courses. She particularly sought out students, like Liu, and mentors, like Gonzalez Ramirez, who are bilingual or multilingual.

"Everyone in the class has had some kind of family-related relationship with language," Tarvin said.

Tarvin also asked Mairi-Louise McLaughlin, UC Berkeley professor of French and linguistics and an expert on journalistic and literary translation, to talk to the class about how professionals approach translation and how translation affects meaning. That subject resonated with the students when they tried their hand at translating scientific abstracts and sometimes whole papers.

Ruoming Cui, a rising sophomore who took the course in spring 2022, chose Baidu to translate scientific abstracts. She immediately discovered that English's long, complex sentences and use of multiple words to describe a concept seemed redundant when rendered into Chinese.

"We don't usually do that in Chinese because it will make every sentence extra-long, and it's very tedious," she said.

Liu added that without considerable polishing, many English translations get garbled, she said.

"I heard the saying that even though your result is amazing, if you write a confusing paper due the translation, people will get annoyed because they cannot understand what you are doing," Liu said. "And that will greatly affect how people validate the research or whether they will even read it. I think that's a big barrier in the scientific world."

Steigerwald, Tarvin and their co-authors also realized that writing scientific papers in plainer English—something nonscientists have been encouraging for a long time—benefits English and non-English speakers alike.

"If your first language is not English, and you're just trying to read the English language version of the paper, it will feel much less ambiguous and much more readable when the writer has used plain language," Steigerwald said. "But also, very importantly, when you go to translate that piece of text, the machine learning tools will have a much easier time of translating something that is written in plain language. So, this is kind of future-proofing your writing, so that if someone wants to translate it into a million languages, they'll have a much easier time of it when it's written in that way."

Obstacles remain to widespread translation of scientific articles, including where to make them available and how to deal with copyrights. Most journals do not even accept articles that are not in English, and few explicitly allow co-publication of articles with a translation. Tarvin has found that few journals have any policies about translations, and as a result of general copyright restrictions, many publishers charge exorbitant fees to post a translation online after publication.

"It's pretty astounding how many journals don't allow you to freely publish translations after publication, and how few have platform support where you could have even just an abstract in a second or third language," Tarvin said. "I think a major barrier for this is the web

platforms; not just the publishing and copyright rules, but also the platform functionality."

With the Breaking Barriers seminar and now the *BioScience* paper, Tarvin and her colleagues hope to gradually change the norm in science to default to translating papers into other languages, especially the language of the country where the research was done and the languages of the co-authors.

And the more translations out there, the more material there is for training machine translation systems to do a better job, gradually ratcheting up the quality of scientific translation.

"In my lab, we're translating a lot of our research, and now people in Emma's lab are doing that, too," she said. "I think sharing our positive attitude towards this and how it can make a difference for people has influenced a small, but growing, group of people who are starting to incorporate translation into their scientific workflow."

Additional co-authors of the *BioScience* paper include doctoral students Valeria Ramírez-Castañeda and Débora Brandt of UC Berkeley; András Báldi of the Institute of Ecology and Botany at the Centre for Ecological Research in Vácrátót, Hungary; postdoctoral fellow Julie Teresa Shapiro of the Ben-Gurion University of the Negev in Be'er Sheva, Israel; and Lynne Bowker, professor of [translation](#) and interpretation at the University of Ottawa in Canada.

More information: Emma Steigerwald et al, Overcoming Language Barriers in Academia: Machine Translation Tools and a Vision for a Multilingual Future, *BioScience* (2022). [DOI: 10.1093/biosci/biac062](https://doi.org/10.1093/biosci/biac062)

Provided by University of California - Berkeley

Citation: Machine translation could make English-only science accessible to all (2022, August 17) retrieved 27 April 2024 from <https://phys.org/news/2022-08-machine-english-only-science-accessible.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.