Overcoming Language Barriers in Academia: Machine Translation Tools and a Vision for a Multilingual Future

EMMA STEIGERWALD, VALERIA RAMÍREZ-CASTAÑEDA, DÉBORA Y. C. BRANDT, ANDRÁS BÁLDI, JULIE TERESA SHAPIRO, LYNNE BOWKER, AND REBECCA D. TARVIN

Having a central scientific language remains crucial for advancing and globally sharing science. Nevertheless, maintaining one dominant language also creates barriers to accessing scientific careers and knowledge. From an interdisciplinary perspective, we describe how, when, and why to make scientific literature more readily available in multiple languages through the practice of translation. We broadly review the advantages and limitations of neural machine translation systems and propose that translation can serve as both a short- and a long-term solution for making science more resilient, accessible, globally representative, and impactful beyond the academy. We outline actions that individuals and institutions can take to support multilingual science and scientists, including structural changes that encourage and value translating scientific literature. In the long term, improvements to machine translation technologies and collective efforts to change academic norms can transform a monolingual scientific hub into a multilingual scientific network. Translations are available in the supplemental material.

Keywords: artificial intelligence, neural networks, scientific communication, multilingualism, plain language

The language in which science is primarily communicated has varied through time and space, cycling through Chinese, Sumerian, Egyptian, Persian, Greek, Latin, Arabic, German, and French, to name a few (von Gizycki 1973, Montgomery and Crystal 2013). The use of English as the scientific *lingua franca* began only 400 years ago, alongside Great Britain's growing colonial empire. After the World Wars, it continued to expand with the increasing military, economic, and technological clout of the United States (Canagarajah 2002, Gordin 2015). Since then, English dominance has extended across the entire globe, as no language has previously done. Today, 98% of peer-reviewed scientific publishing is in English (Ammon 2012, Liu 2017), and English is the official language of most scientific events and international and indexed academic journals.

Having a common language benefits science by facilitating international scientific communication and creating a monolingual repository for publications and data (Montgomery and Crystal 2013). The maintenance of a common scientific language is also useful for the dissemination and recognition of research performed by scientists whose primary language is not widely spoken, as well as for facilitating communication between such scientists and the wider scientific community. Having a shared scientific language also facilitates international mobility and limits the number of additional languages required for international collaboration. However, despite the benefits of a common language, maintaining a single universal scientific language creates barriers by requiring the majority of researchers in the world to become proficient in an additional language prior to engaging with the global academic community. Through its "Recommendation on Open Science," UNESCO has called on scientific institutions to foster global, multilingual, and cross-disciplinary research programs in order to provide more equitable access to scientific knowledge and careers (UNESCO 2021).

In the present article, we summarize the costs of a single universal language in science and provide a set of practical approaches that individuals, academic societies, and institutions can take to help break down language barriers, focusing on machine translation tools for written sources and structural change that would better support a multilingual

academy. Although the suggestions contained in the present article are built from and sometimes particularly pertinent to our research experiences in ecology, evolution, and conservation, these ideas may be useful to a broader scientific audience.

The costs of a single universal language in science. Although maintaining a central language has its benefits (see above), it also stymies the advancement of science, creates barriers within academia, and complicates applying scientific evidence to decision-making outside of academia. For example, because academic knowledge is mostly communicated in English, scientists and other members of society often overlook knowledge generated in other languages. One concrete manifestation of that is using keywords exclusively in English during literature searches (Pabón Escobar and da Costa 2006, Kirchik et al. 2012, Liang et al. 2013, Neimann Rasmussen and Montgomery 2018, Amano et al. 2021a). This effect can be amplified by language biases in search engines (Rovira et al. 2021). Overlooking non-English studies can result in large gaps within global databases, which affects policy, management, and decision-making (Amano and Sutherland 2013, Amano et al. 2016, 2021a, Konno et al. 2020, Angulo et al. 2021, Kirpotin et al. 2021). For example, the exclusion of the many studies on conservation interventions published in languages other than English can reduce the evidence being considered during decision-making processes and lead to less-optimal natural resource management (Amano et al. 2021a). In addition, non-English-language literature could expand both the geographical and the taxonomic coverage in biodiversity studies (Khelifa and Mahdjoub 2022). Biases in who contributes to science and makes these management decisions also reduce the credibility and global buy-in to these management practices (Baldi and Palotas 2021).

English proficiency also influences who participates in science at a global scale, which is detrimental to science, because a diversity of perspectives bolsters the construction of robust and innovative scientific knowledge (Bennett 2013, AlShebli et al. 2018, Hofstra et al. 2020). Proficiency in English is often a requirement for professional advancement, such as publishing in high-impact journals, receiving international grants, and participating in international conferences (Hwang 2005, Clavero 2010, Huttner-Koros and Perera 2016, Ramírez-Castañeda 2020). Non-Anglophones are therefore under constant pressure to improve their English language skills (Tardy 2004, Lindsey and Crusan 2011, Corcoran 2015, Suzina 2021), which can be a source of anxiety and an emotional burden (Ramírez-Castañeda 2020, Amano et al. 2021b). Moreover, this challenge is not experienced equally across English learners but, rather, weighs particularly heavily on learners whose dominant language is highly divergent from English and on learners from regions in which English-language instruction or media are not widely available, two issues that are not mutually exclusive. Language barriers can impose a severe financial burden on individuals, who may pay for English classes, proofreading, and translation services, reinforcing socioeconomic inequity in science, especially because these burdens are experienced to a greater extent by those in countries with a lower gross domestic product (Schofield and Mamuna 2003, Kieffer 2010, González Mellado et al. 2020, Ramírez-Castañeda 2020, Khelifa and Mahdjoub 2022). Biases during peer review may lead non-Anglophones to publish in lesser-known journals or in regional journals that publish in other languages, making their research less discoverable (Mur Dueñas 2012). These burdens intensify the dependence of many non-Anglophone scientists on scientists with high English proficiency (Ordóñez-Matamoros et al. 2011). Ultimately, these barriers can impede non-Anglophones from obtaining jobs, tenure, or promotion (Moreno 2010).

Constraining diverse points of view to fit within the structure and vocabulary of a single language impoverishes scholarly discourse and observations of nature. For instance, language shapes how we perceive color (Siok et al. 2009), our understanding and memory of events (Fausey et al. 2010), and our ability to gauge the awareness or knowledge of others (Jara-Ettinger and Rubio-Fernandez 2021). When we write only in English, we limit our way of describing the relationships between ideas—a type of loss that has been analogized to the creation of an epistemological monoculture (Martin 2009, Bennett 2013, Aguilar Gil 2020). Moreover, constraining global scientific discussions to a single language can limit who builds, has access to, and communicates scientific knowledge to the broader public (Canagarajah 2002, Tardy 2004, Huttner-Koros and Perera 2016, O'Neil 2018), profoundly affecting the relationship between science and society. Scientific monolingualism may reduce the dissemination of science to non–English-speaking institutions and communities, which can leave new knowledge inaccessible to the people for whom it is most relevant, such as those living near study sites, local public media, and regional policymakers (Márquez and Porras 2020). This is likely particularly impactful for people in countries with low English proficiency, who have reduced access to knowledge communicated exclusively in English (Amano et al. 2016, Saha et al. 2019), sometimes even to studies that feature these regions (Barath 2019). Although the disconnection between science and society is unfortunate for any scientific field, the cost is particularly high for applied sciences and crisis disciplines such as climate science, epidemiology, and conservation (Meadow et al. 2015, Saha et al. 2019, Amano et al. 2021a), where the rapid dissemination of new results makes a material difference to urgent decisions that must be made despite incomplete evidence.

The existence of a single universal language of science may currently serve to share new knowledge broadly and practically. However, those who bear the costs of a single language also tend to face additional barriers—for example, those associated with colonialism, because the language that an individual speaks is tied to the history of their country and culture. Therefore, maintaining a single language in

science without providing adequate support to people who do not speak that language will continue to perpetuate historical imbalances. Attempts to create a more accessible centralized language (e.g., Esperanto) have not gained traction (Tonkin 1987), and although English may present some linguistic advantages (e.g., relatively simple and genderless grammar), it is not the only language with these attributes, and its dominance can be attributed to the historical factors mentioned above. Therefore, we propose that science would benefit from integrating multiple languages. Multilingual science will also benefit our community by creating support systems that can facilitate potential future transitions, because although it may feel unlikely, history has shown that dominant languages are likely to continue changing over time.

Short-term actions: Translation and the promotion of multilingual science

Science benefits from diverse viewpoints, and language is one of many axes of diversity (AlShebli et al. 2018, Hofstra et al. 2020). However, little structural support exists in the present to help non-Anglophones publish and advance professionally in English. Recently, Amano and collaborators (2021b) highlighted some practical tips to overcome language barriers, such as promoting multilingual activities, being empathetic with those who face language barriers, providing an English proofing network for preprints (Khelifa et al. 2022), and translating scientific literature (Amano et al. 2016, 2021b, Márquez and Porras 2020, Ramírez-Castañeda 2020). Multilingual publishing is another mechanism that actively promotes and places value on contributions in different languages. Machine translation tools can help scientists take concrete steps toward publishing in multiple languages, including in English. In the present article, we largely focus on machine translation tools for written sources; however, we highlight that similar efforts can be extended to the spoken word.

An overview of machine translation tools and how to improve them for scientific literature. The earliest approaches to machine translation used painstakingly programmed linguistic rules and very large dictionaries, but they had limited success because language is full of ambiguity and computers had no access to the type of real-world knowledge and social interactions that people use to interpret language (Way 2020). Following the introduction of the Internet and the increasing trend of producing texts in digital form, machine translation researchers moved away from linguistic approaches and toward data-driven machine translation, which capitalized on the strengths of computers (e.g., pattern matching, rapid calculations). Around the turn of the millennium, statistical machine translation systems began to appear, including early free online tools, such as Google Translate. In statistical machine translation, the developers fed the computer with vast quantities of previously translated texts, and the system used these examples to calculate the probability that a given

phrase should be translated in a certain way in a future text (Way 2020). Statistical machine translation tools produced better-quality output than linguistic approaches, but there was still considerable room for improvement. Another data-driven approach, known as *neural machine translation*, appeared in late 2016, and it has presented another leap forward in terms of translation quality. Today, the majority of machine translation tools use artificial neural networks in combination with artificial intelligence–based techniques such as machine learning (Forcada 2017). These techniques require developers to provide the machine translation system with many training examples of original source texts and their translations for the system to learn. Therefore, translation tools are more easily tuned to widely used languages or languages with more of these examples. Although they are not perfect, neural machine translation systems provide a more viable starting point than older machine translation systems, which relied on linguistic or statistical approaches. The results of neural machine translation systems can be used for basic knowledge acquisition or as a first draft that can then be improved (e.g., for academic writing; Parra Escartín and Goulet 2020). Increasing numbers of people are using neural machine translation tools because of their ease of use and free online availability (e.g., DeepL and Google Translate; Bowker 2021).

However, using machine translation tools still requires good judgment, which is why there is a need for machine translation literacy (Bowker and Ciro 2019, Bowker 2021). Machine-learning technologies are very sensitive to the quantity and quality of their training data. To work well, machine translation systems need access not only to enormous quantities of previously translated texts and their corresponding original texts but also to good quality texts that are relevant to the focal topic (Chu and Wang 2020). For example, the language used in specialized fields contains many technical terms and constructions that are not part of everyday language. Therefore, for a machine translation system to accurately translate texts in the field of biology, it would need to be provided with millions of examples of previously translated texts specifically from this domain. Moreover, these examples would need to cover all the desired language combinations (e.g., English and French, Chinese and Hindi, English and Chinese). In some cases, when a particular language pair has relatively few translated texts available, the lack of training data can be overcome by using a widely spoken language as a pivot language (e.g., translating from Spanish to Chinese using English as an intermediary), although this approach may propagate errors (Kim et al. 2019). Similarly for spoken communication, the recent COVID-19 pandemic rapidly increased the need for and use of online communication platforms that provide closed-captioning in multiple languages. However, piping two imperfect technologies (machine translation and speech recognition) together can compound translation errors (Sulubacak et al. 2020), similar to problems arising from the use of pivot languages.

Do you meet one or more of these conditions:

1. My work is conducted in a region or country where the primary language is not English
2. My primary language or that of one or more of my collaborators is not English
3. My work directly or indirectly affects a group of people whose primary language is not English

No → Yes

Is it under one of these copyrights: CC BY, CC BY-SA, CC BY-NC, or CC BY-NC-SA? ← Yes — Is the paper already published? No

Consider translating your research for a lay audience (e.g., a plain language summary, blog post, comic strip, podcast, etc.)

Yes

Great! You can do the translation without requesting copyright permission

Be sure to include your translation in the supplementary materials to avoid future copyright fees

No

Can you obtain copyright permission from the publisher for the translation?

Ok! — Ok!

Yes, for free!

Do you or a collaborator possess the skills to do the translation?

Yes, for a fee

Yes

Can your institution or another organization provide funding?

No — Yes

Translate using machine translators and provide a disclaimer with the translation that explains it was not corrected

Translate by hand or using machine translators and then correct for accuracy

No

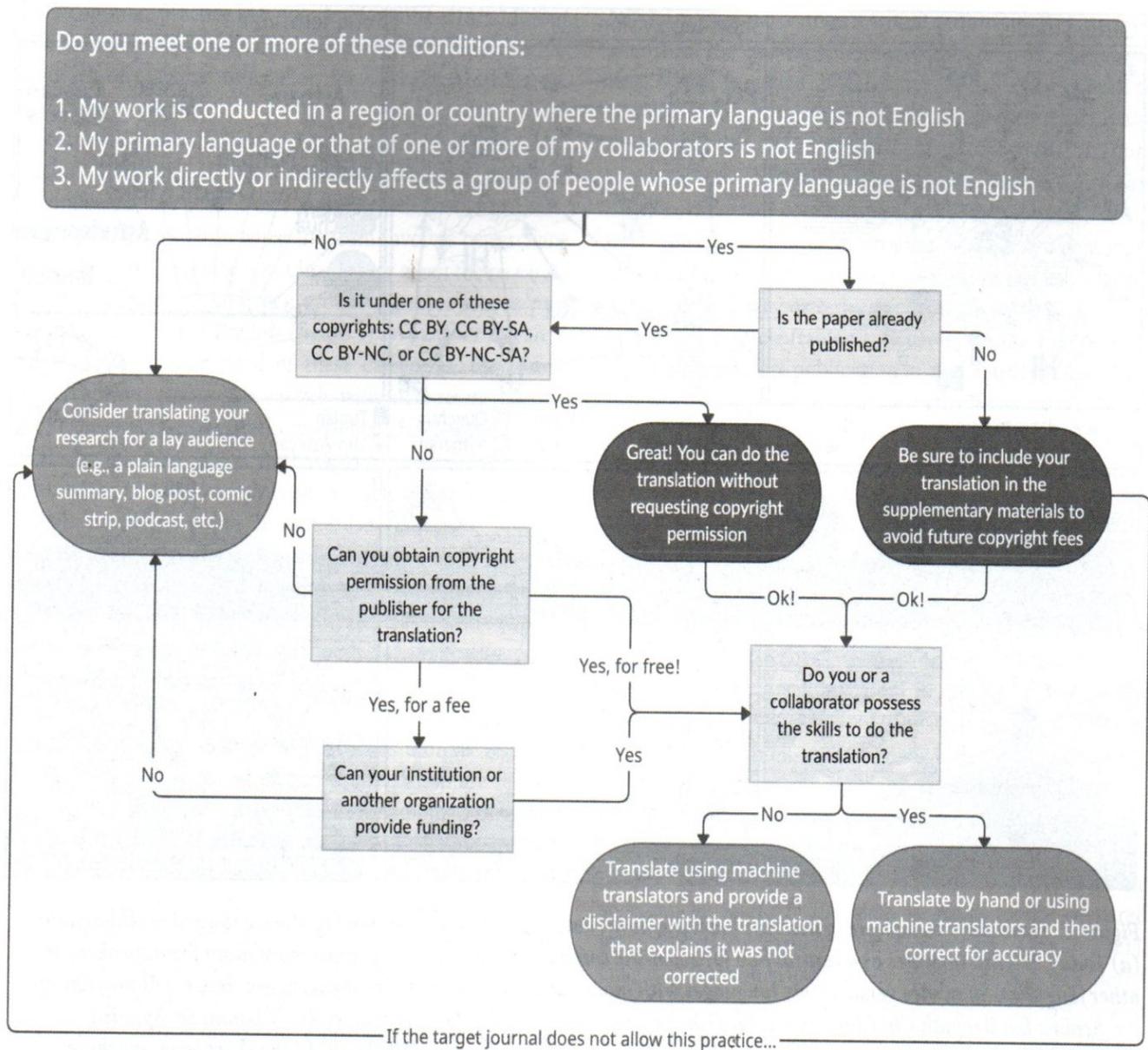— If the target journal does not allow this practice… —

Figure 1. An example decision tree that authors can use to decide when and how to translate their research output. Scientists whose research programs meet one or more of the listed circumstances may consider translating into languages relevant to those circumstances. Understanding that researchers are often limited by resources and time, we provide this diagram as a suggestion of when to prioritize translation, because translations may be useful in additional circumstances.

There are clear steps that scientists and machine translation tool developers can take to improve the implementation of technologies in scientific translation. A concerted effort toward providing open-access, human-verified, and high-quality translations of abstracts in scientific journals would significantly contribute toward generating the data necessary for training machine translation systems. At the moment, free online translation tools are trained mainly on general language data rather than on scientific jargon or specialized language. Researchers and tool developers could collaborate on open-access tools that train machine translation systems for specialized fields of research. Simultaneously, we could encourage scientists to develop or contribute to multilingual glossaries of specialized terminology, in part to help keep up with the constant generation of new scientific jargon (Nkomo and Madiba 2012, Wild 2021). For instance, Wikipedia is an excellent open-access platform for finding multilingual translations of technical and scientific topics. However, it is currently underused by several scientific disciplines, and several languages with large numbers of speakers (such as Hindi and Turkish) are underrepresented (Kincaid et al. 2020, Roy et al. 2021).

When, why, and how scientific literature can be translated. With the aid of translation tools, contributing translations of abstracts, keywords, and entire articles could become the norm for research programs that cross languages (figure 1; Amano et al. 2021b). Indeed, translating scientific abstracts is already a common practice for some journals in bilingual or (primarily) non-Anglophone countries (e.g., the
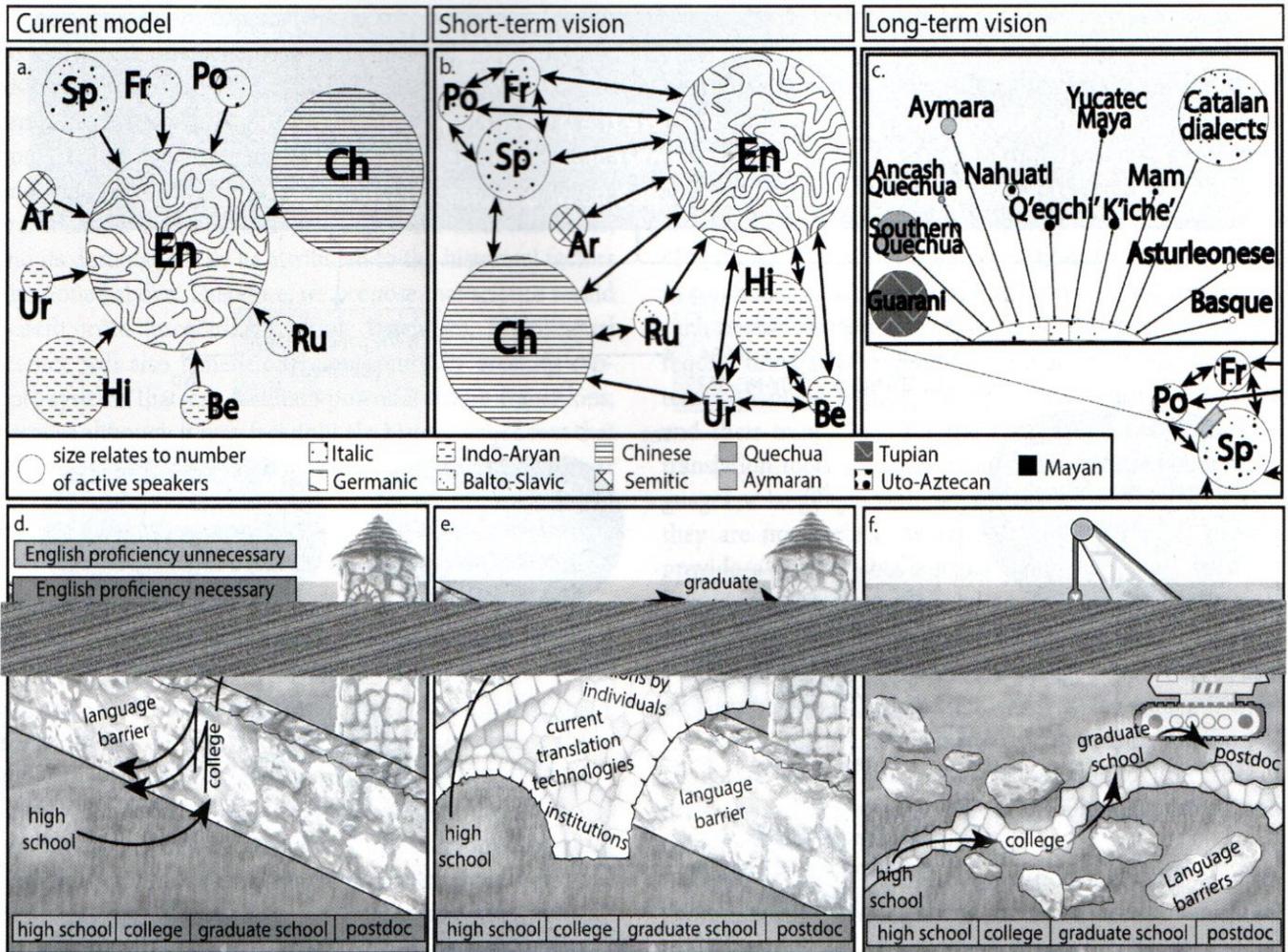
Figure 2. *Two visual metaphors to describe breaking down language barriers and moving science toward multilingualism.
(a) Today, English operates as a central hub for scientific communication, receiving much more input from speakers of
other languages than vice versa. (Only languages with more than 230 million active speakers are shown.) Abbreviations:
Ar, Arabic; Be, Bengali; Ch, Chinese; En, English; Fr, French; Hi, Hindi; Po, Portuguese; Ru, Russian; Sp, Spanish;
Ur, Urdu. The numbers were estimated according to Eberhard and colleagues (2022). (b) In the short term, machine
translation tools and efforts by scientific communities can help form secondary language hubs (see the main text) that
create and disseminate scientific knowledge among all languages within each language family. For instance, Hindi may
serve as a connector language; science translated into Hindi can then be more easily translated from Hindi into other
Indo-Aryan languages. (c) As machine translation technologies improve, greater exchange across language families will
indirectly benefit the speakers of languages with smaller numbers of active speakers (inset), who, owing to geography
or history, often must learn a second language from one of these major families. For instance, the greater availability of
texts translated to Italic languages will facilitate translation into languages historically and geographically associated
with Spanish (i.e., indigenous languages of Iberia, South America, and Central America). (d) Currently, students must
become proficient in English during or prior to their graduate studies if they wish to pursue science as a career, presenting
a language barrier that may intersect with associated barriers. (e) In the short term, structural changes by institutions,
actions by individuals, and machine translation tools can help students bridge the barrier. (f) In the long term, advanced
translation technologies and a more multilingual scientific academy will help demolish language barriers. Under this more
accessible paradigm, scientists may be able to advance their careers and their English proficiency in parallel, rather than
needing English proficiency as a prerequisite for a career. Ultimately, a more multilingual scientific community will make
science more accessible to the multilingual public.*

Canadian Journal of Forest Research, the *Brazilian Journal
of Biology*). Normalizing the practice of translation will
increase access to scientific research for scientists, students,
teachers, policymakers, journalists, and members of society
at large. It could also shift the work of translation to be
more equally shared between native English speakers and
those who speak English as an additional language, because
translations would not only happen from other languages to
English (figure 2a) but also from English to other languages
(figure 2b). When native English speakers cannot directly
translate to a language they do not speak, they will need
to find and pay for (or reciprocate) translation services,

as is common practice for non-Anglophones. In addition, translating abstracts will help to substantially improve the accuracy of machine translation for scientific texts, as we described above.

We recommend that researchers consider translating articles (or, minimally, abstracts and keywords) when the work is conducted in a region or country in which the primary language is not English, when the team involves researchers whose primary language is not English, or when the research directly or indirectly affects a group of people whose primary language is not English (figure 1). Whenever research fits into the first or third of these categories, the teams should always work collaboratively with local researchers to avoid the practice of parachute science (Haelewaters et al. 2021, de Vos et al. 2022). Authors who speak the language selected for translation may wish to create a first draft using DeepL Translator, Baidu Translate, Naver Pagago, Yandex. Translate, Google Translate, or a similar tool, which can then be manually edited. Authors who do not speak the additional language can work with journals to find reciprocal translation partners or other modes of support (supplemental table S1; Amano et al, 2021b), or they can search for reciprocal translation opportunities through forums such as ResearchGate or preprint servers such as bioRxiv (Khelifa et al. 2022). Because some aspects of translation are subjective (e.g., specific vocabulary choices or idiomatic translations), it is critical to reference the person or software that was used and whether the machine translation (if one was done) was verified by a human (Croft 2021; see the translation note in the acknowledgments). Importantly, creating a byline for translation or language editing will normalize the acknowledgment of these critical services, provide scientists with an alternative option to exchanging authorship for editing assistance, and provide a language contact to whom translation questions can be directed.

Translations of previously published scientific articles (even by the authors themselves) often cannot be posted publicly because of copyright restrictions. Therefore, researchers may wish to include a translation of a full manuscript as part of the supplemental material when submitting their work for publication or as part of material stored on accessible web platforms such as the Open Science Framework (osf.io) or GitHub (GitHub.com), which can be updated at a later time point to include additional translations. In the future, we suggest that journals with copyright restrictions could implement fee waivers for translated versions (see table S1). Note that *open access* itself does not imply anything about copyright, but some journals publish open access articles with Creative Commons licenses such as CC BY, CC BY-SA, CC BY-NC, or CC BY-NC-SA, which allow translation without copyright infringement (BY means "Credit must be given to the creator," SA means "Adaptations must be shared under the same terms," and NC means "Only noncommercial uses of the work are permitted"; see creativecommons.org for more information). Creative Commons licenses with the ND term ("no derivatives") would require written permission from the copyright holders to publicly post a translation, which is a type of derivative. If the authors wish to conduct a translation once a paper has been published and it is not published under one of these Creative Commons licenses, they do have a few options, including paying the copyright fee, obtaining a fee waiver (not easy, in our experience), requesting an erratum to append a document to the supplemental files, and choosing to publish a plain-language summary or reflection instead, perhaps as a blog or magazine article (table S1, figure 1). In the case of posting preprints before publication in a peer-reviewed journal, some servers permit authors to share information in languages additional to English (e.g., EcoEvoRxiv), although this is not the case for all (e.g., bioRxiv).

A contribution that all researchers and journals can make, regardless of their native language, is to prepare a plain language summary that is both reader friendly and translation friendly (Bowker and Ciro 2019). A text that is less structurally complex can still be rich in meaning, but it will be easier for readers to digest and for machine translation tools or human translators to translate. Because the goal of plain writing is simply to write as clearly as possible, the technique can be applied to any language. However, the specific approaches for reducing structural complexity or linguistic ambiguity may differ from one language to the next (see table 1 for examples that apply to English). More detailed information on how to write in an easy to translate style can be found in the plain language toolkit prepared for scientists by Evidence for Democracy (Qaiser 2021). One way that journals can help make papers better suited for machine translation and more accessible to readers with lower English proficiency is to soften word limits, because the methods to shorten sentence structure tend to introduce grammatical complexity and ambiguity. The advent of online-only journals has provided a great opportunity for journals to soften word limits without incurring production fees (table S1).

The role of academic institutions in promoting translation efforts

Journals and academic societies have the power to change norms, because they are important forums in which scientists engage with each other and are recognized for their work. Journals can actively contribute to addressing language barriers and supporting multilingual science by providing clear guidelines regarding when authors are expected to translate articles or abstracts (e.g., see figure 1), how translations can be included in published articles, how research in other languages should be cited (Fung 2008), and how to search for journal content in languages other than English (see table S1 for additional suggestions and supplemental table S2 for examples). Including multilingual scientists on boards and committees can help drive and facilitate these efforts (table S1). Societies or journals could also provide free translation services or promote

Table 1. Plain language writing tips to reduce structural complexity and linguistic ambiguity in English, including ideas from Bowker and Ciro (2019).

Action	Explanation	Example
Use shorter sentences	The longer the sentence, the more challenging it is to identify the relationships between the different elements.	Try to keep sentences under 25 words.
Use the active voice	It is easier to identify the agent in the sentence and to understand its relation to the other elements.	*We report the findings* instead of *The findings are reported*
Avoid long strings of modifiers	When connecting words (e.g., prepositions) are eliminated, readers and machine translation tools must infer the relations between the words.	*liquid oxygen tank*, a tank for liquid oxygen, versus *red oxygen tank*, a red tank for oxygen
Include optional relative pronouns (*that, which*)	Relative pronouns (*that, which*) help readers to understand how different elements are related. Even though it is possible to omit them in some cases, it is better to include them to clarify the relationships.	MCPyV, as well as Epstein-Barr virus, normally connected with humans under the form of subclinical infection, versus MCPyV, as well as Epstein-Barr virus, *which are* normally connected with humans under the form of subclinical infection
Define and use terminology consistently	All languages have synonyms, but it may be challenging to recognize that different words can refer to the same concept. Using terms consistently (and defining them, if possible) reduces confusion for readers and machine translation tools.	Instead of alternating between *amyotrophic lateral sclerosis* and *motor neuron disease*, choose one term and use it consistently.
Minimize the use of abbreviated forms	Abbreviated forms are challenging for machine translation tools, which may try to recognize them as words. They may also be difficult for speakers of other languages. Use sparingly.	MS could be a short form for *multiple sclerosis, master of science, manuscript,* or even a polite term of address for a woman, and a machine translation tool may choose incorrectly.

Note: Recommended, free online tools that can suggest how to accomplish these goals for a given piece of writing can be found at sites such as https://hemingwayapp.com and https://datayze.com/?category=writing.

mentorship within academic societies to provide English proofing (e.g., translatE 2020, Khelifa et al. 2022). In addition, several recent papers have highlighted how individual scientists can reduce bias and improve the peer-review process for nonnative English speakers when acting as reviewers or editors (Romero-Olivares 2019, Mavrogenis et al. 2020, Amano et al. 2021b).

The translation of titles, abstracts, keywords and full texts—which can greatly improve machine translation tools—could be facilitated if journals create a streamlined process for authors to add translations during or after publication and provide a clear statement of copyright policy regarding whether adding translations is subject to the same copyright restrictions as other use cases. Some journals already have systems in place for abstract translations (see table S2 for examples). Multilingual scientific literature and conference booklets would permit researchers and other members of society to use their primary language when scanning the literature or conference abstract books to find relevant articles and talks. Finally, author guidelines that encourage the inclusion of multilingual graphical abstracts (e.g., figure 1 in Chu et al. 2021) also increase accessibility, and plain-language abstracts or highlights have the additional benefit of being machine translation–friendly (see our long-term vision below; see Shailes 2017 for examples of plain-language summaries produced by journals, societies, and other organizations).

The actions cataloged above could incur at least two types of burden on journal staff and conference organizers: the financial burden of providing free translation services and the time needed to review translated texts. If a journal or conference cannot presently afford to freely translate their contributors' science, they (or a consortium of journals) might consider creating a forum on their website or via existing preprint servers (Khelifa et al. 2022) where contributors can identify reciprocal or volunteer language editing and translation partners. For example, Cochrane, a United Kingdom–based charitable organization, has a network of volunteers that translate their systematic reviews of medical literature from English to various languages (www.cochrane.org/join-cochrane/translate). Although these solutions relieve the burden on the staff of journals and conferences, they demand free labor from researchers. To compensate, journals and conferences could create systems of incentive such as discounts in publication or registration fees. One alternative to overcome the time needed to review translations is to require authors to label translations with standardized disclaimers, such as "manually translated by a fluent speaker," "machine translated," or "machine translated and manually edited for accuracy." Journals could simply note that these translations have not undergone peer review, as is already the case for most supplemental material (e.g., see the *Molecular Ecology* journal guidelines for abstract translation in table S2). Other actions outlined in table S1

might also come with additional cost (e.g., providing closed captioning in conference talks), but we see this cost as an important and worthwhile investment toward a more inclusive academic environment, which will benefit science, the individuals that participate in it, and society at large.

Universities can promote efforts to overcome language barriers through both their educational role and their role in shaping research program priorities. For example, they could emphasize or recognize the importance of publications in (non-English) national and regional journals for tenure and promotion files, contract renewals, or degree requirements. Faculty and students often feel pressure to publish in English-language journals, because this boosts the rankings and impact factor of their institution, but national and regional publications play an important role in disseminating knowledge (Moreno 2010, Vaidyanathan 2019), which closely aligns with many university missions.

Importantly, because machine translated texts are imperfect, machine translation literacy is essential (Bowker 2021). Universities can develop cross-disciplinary courses to teach and enact the practice of scientific translation, which is itself part of a vast field of study (Munday 2016). Universities can make machine translation literacy training part of a standard STEM (science, technology, engineering, and mathematics) curriculum, so that new researchers are acquainted with the strengths and limitations of translation technologies (Bowker 2021). Students are already widely using these technologies, but perhaps without an appreciation for how to work around their limitations (Bowker 2021). In addition, students in the sciences could be encouraged to study foreign languages, as is common in the humanities (Kellsey and Knievel 2004), especially if conducting research in non-Anglophone regions.

Many other institutions can do their part in improving scientific standards, making science more accessible, and, therefore, ultimately more globally impactful (table S1). Public databases, such as GenBank or Online Mendelian Inheritance in Man, are critically important resources, and a multilingual approach to their online platforms, as is demonstrated by the International Union for the Conservation of Nature or Fonoteca Zoológica, would permit broader engagement with these resources. Funding agencies can include clauses that encourage or require researchers working internationally to include local scientists in their research and encourage budget items to support translating results and outreach that engages with local communities in their local languages. In addition, international funding agencies could permit the submission of grant or scholarship proposals in several languages, especially if these funds are focused on communities or students who do not necessarily speak English.

Long-term vision: From a language hub to a language network

At present, scientific publishing is largely centered on the English language, with relatively few languages receiving substantial input from the English hub (figure 2a; the estimated number of active speakers for each language displayed is from Eberhard et al. 2022). For this reason, nonnative English speakers must generally acquire English proficiency before or during their graduate career or else forgo participation in the international scientific community (figure 2d). We envision that multilingualism is the outcome of a prolonged process of inclusion of languages brought about by improved translation technologies and changes in community norms. A first step that can be taken toward multilingual science is the creation of temporary secondary language hubs that can act as networking communities and knowledge centers for non-Anglophones (figure 2b), supporting these scientists throughout the launch of their early careers (figure 2e). For example, hubs for Mandarin, Hindi, or Spanish would establish information streams between sets of languages with many speakers, additional languages pertaining to each language family, and the central English node. Efforts to facilitate the creation of these secondary hubs in science are already happening through multilingual conference activities, bilingual journals, and regional academic societies (table S2; Márquez and Porras 2020, Amano et al. 2021b). In the future, tertiary hubs could be established until greater multilingualism is achieved (figure 2c) and English proficiency is no longer requisite for participation in the international scientific community (figure 2f). Geographic proximity, political history, and language origin can be some of the strategies used to define a tertiary hub.

Most of Western society has accepted that a universal language is integral to the scientific enterprise (Aguilar Gil 2020); therefore, we acknowledge how unreachable or unnecessary a multilingual future may appear. However, a multilingual vision encompasses more than academia; it also aligns with multidisciplinary and plurinational efforts to preserve languages, culture, and knowledge (UNESCO 2021; Endangered Languages Project 2022). To reach such a long-term goal, we envision that accurate and readily available translation technologies, as well as collective efforts supporting and integrating multilingual science, will play important roles. The ideas presented in this article are a starting point only and will require further discussion. They are not exclusive, universal, or definitive, and the community will require other changes to make scientific fora more inclusive. We encourage the creation of discussion groups on this topic to generate new and innovative ideas to help solve language barriers.

of Vertebrate Zoology community, who have been pushing forward several initiatives to improve equity and inclusion in our field, and Tatsuya Amano, whose published articles and resources on the value of multilingual science have guided many of our discussions. This work was supported in part by the Zuckerman STEM Leadership Program (to JTS); by the Julius H Freitag Memorial Award, the Hannah M. and Frank Schwabacher Fund, and the David and Marvalee Wake Fund (to ES); by the National Research, Development, and Innovation Office of Hungary (contract no. ED_18–1–2018–0003 to AB); by UC Berkeley start-up funding (to RDT); and by the Social Sciences and Humanities Research Council of Canada (grant no. SSHRC 435–2020–0089 to LB). Finally, we want to acknowledge the unseen effort of all the people that work protecting endangered languages and researchers' efforts to make science available in different languages and formats.

Translation acknowledgements: The Spanish, Portuguese, and Hungarian translations were conducted using DeepL Translator (www.deepl.com/translator) and were proofread and edited by VRC, DYCB, and Brigitta Palotás, respectively. The French translation was conducted by Thibault de Poyferré who manually corrected translations that were initially generated with Google Translate and DeepL Translator. Additional translations are acknowledged in the repository linked in the supplemental material.

This article was the product of an interdisciplinary collaboration between scholars of evolution, ecology, and conservation (DYCB, VRC, RDT, ES, JTS, and AB) and a scholar of translation studies (LB). DYCB and VRC learned English as a second language while pursuing basic and higher education in Brazil and Colombia. They moved to the United States for graduate school, where they are constantly confronted with language barriers and recognize the additional burden that English dominance causes to their colleagues in South America and other regions. ES and RDT grew up speaking English; learned Spanish in high school and while conducting fieldwork, respectively; they acknowledge that advancing their research goals has been facilitated both by having English as a first language and by having collaborators who speak English. LB grew up speaking English, learned French and Spanish as part of her training to become a translator, and recognizes the advantages of being able to work in a native language in comparison to an additional language. JTS grew up as a native English speaker in the United States. She has learned multiple languages, which has facilitated working in countries in which English is not the primary language. However, she has found that being a native English speaker has provided many opportunities she otherwise would not have had. AB learned English as a second language during higher education in communist Hungary, where little importance was placed on the English language. When he started his career, he was handicapped by the lack of an English-speaking environment, in addition to the difficulties inherent in the fact that English is not closely related to Hungarian.

Supplemental material
Supplemental data are available at *BIOSCI* online.

References cited

Aguilar Gil YE. 2020. Aguilar-Guevara A, Bravo Varela J, Ogarrio Bodillo G, Quaresma Rodríguez V, eds. Ää: Manifiestos sobre la Diversidad Lingüística. Almadía.

AlShebli BK, Rahwan T, Woon WL. 2018. The preeminence of ethnic diversity in scientific collaboration. Nature Communications 9: 5163.

Amano T, Sutherland WJ. 2013. Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. Proceedings of the Royal Society B 280: 20122649.

Amano T, González-Varo JP, Sutherland WJ. 2016. Languages are still a major barrier to global science. PLOS Biology 14: e2000933.

Amano T, et al. 2021a. Tapping into non-English-language science for the conservation of global biodiversity. PLOS Biology 19: e3001296.

Amano T, C Rios Rojas, Boum II Y, Calvo M, Misra BB. 2021. Ten tips for overcoming language barriers in science. Nature Human Behaviour 5: 1119–1122.

Ammon U. 2012. Linguistic inequality and its effects on participation in scientific discourse and on global knowledge accumulation: With a closer look at the problems of the second-rank language communities. Applied Linguistics Review 3: 333–355.

Angulo E, et al. 2021. Non-English languages enrich scientific knowledge: The example of economic costs of biological invasions. Science of the Total Environment 775: 144441.

Baldi A, Palotas B. 2021. How to diminish the geographical bias in IPBES and related science? Conservation Letters 14: e12786.

Barath H. 2019. Indian initiatives aim to break science's language barrier. Nature 571: 289–290.

Bennett K. 2013. English as a *linguafranca* in academia: Combating epistemicide through translator training. Interpreter and Translator Trainer 7: 169–193.

Bowker L. 2021. Promoting linguistic diversity and inclusion: Incorporating machine translation literacy into information literacy instruction for undergraduate students. International Journal of Information, Diversity, and Inclusion 5: 127–151.

Bowker L, Ciro JB. 2019. Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community. Emerald.

Canagarajah AS. 2002. A Geopolitics of Academic Writing. University of Pittsburgh Press.

Chu C., Wang R. 2020. A survey of domain adaptation for machine translation. Journal of Information Processing. 28: 413–426.

Chu C, et al. 2021. Conceptual and theoretical dimensions of biodiversity research in China: Examples from plants. National Science Review 8: nwab060.

Clavero M. 2010. "Awkward wording. Rephrase": Linguistic injustice in ecological journals. Trends in Ecology and Evolution 25: 552–553.

Corcoran JN. 2015. English as the International Language of Science: A Case Study of Mexican Scientists' Writing for Publication. PhD dissertation. University of Toronto, Toronto, Canada.

Croft J. 2021. Why translators should be named on book covers. Guardian (10 September 21).

De Vos A. 2022. Stowing parachutes, strengthening science. Conservation Science and Practice 4: e12709.

Eberhard DM, Simon GF, Fennig CD. 2022. Ethnologue: Languages of the World 25th ed. Ethnologue. www.ethnologue.com.

Endangered Languages Project. 2022. Endangered Languages Project: Supporting and celebrating global linguistic diversity. Endangered Languages Project. www.endangeredlanguages.com.

Fausey CM, Long BL, Inamori A, Boroditsky L. 2010. Constructing agency: The role of language. Frontiers in Psychology 1: 162.