Received: 6 May 2025 | Revised: 9 January 2026 | Accepted: 12 January 2026

DOI: 10.1002/asi.70055

RESEARCH ARTICLE

JASIST WILEY

How multilingual is scholarly communication? Mapping the global distribution of languages in publications and citations

Carolina Pradier[1] | **Lucía Céspedes**[1,2] | **Vincent Larivière**[1,2,3,4]

[1]École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, Québec, Canada

[2]Consortium Érudit, Montréal, Québec, Canada

[3]Department of Science and Innovation-National Research Foundation Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, Western Cape, South Africa

[4]Observatoire des Sciences et des Technologies, Centre interuniversitaire de recherche sur la science et la technologie, Université du Québec à Montréal, Montréal, Quebec, Canada

Correspondence
Vincent Larivière, École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, Québec, Canada.
Email: vincent.lariviere@umontreal.ca

Funding information
Fonds de Recherche du Québec-Société et Culture, Grant/Award Numbers: 338828, 362184; Social Sciences and Humanities Research Council of Canada, Grant/Award Number: 1007-2023-0001

Abstract

Language is a major source of systemic inequities in science, particularly among scholars whose first language is not English. Studies have examined scientists' linguistic practices in specific contexts; few, however, have provided a global analysis of multilingualism in science. Using two major bibliometric databases (OpenAlex and Dimensions), we provide a large-scale analysis of linguistic diversity in science, considering both the language of publications ($N = 87{,}577{,}942$) and of cited references ($N = 1{,}480{,}570{,}087$). For the 1990–2023 period, we find that only Indonesian, Portuguese, and Spanish have expanded at a faster pace than English. Country-level analyses show that this trend is due to the growing strength of the Latin American and Indonesian academic circuits. Our results also confirm the same-language preference phenomenon (particularly for languages other than English), the strong connection between multilingualism and bibliodiversity, and that social sciences and humanities are the least English-dominated fields. Our findings suggest that policies recognizing the value of both national-language and English-language publications have had a concrete impact on the distribution of languages in the global field of scholarly communication.

1 | INTRODUCTION

Scholarly writing is often said to be an inherently polyphonic, dialogic exercise (Bakhtin, 2010). Indeed, one of the skills every scholar must become increasingly proficient at is the incorporation of others' ideas and voices (Hyland & Jiang, 2019). Whether to reinforce an argument or to provide a counterpoint, the presence of others' words, craftily interwoven with one's own, is the defining feature of academic discourse. If science is built on cumulative knowledge, then the cited references present in publications are the traces of this process.

Citations thus constitute one of the most relevant markers of symbolic capital within the scientific field (Desrochers et al., 2018; Ennser-Kananen, 2019; Hyland & Jiang, 2019; Selvi, 2024). Citations are often

considered—despite limitations—as an indicator of impact, influence, or quality, and are used in research assessment systems, either at the institutional, national, or international level (Davis, 2009; Vasen et al., 2023). Since the importance of being cited cannot be overstated, it is not surprising that many studies have looked into citations as potential indicators of systemic inequities in science (Kozlowski et al., 2022; Pradier, Kozlowski, Shokida, & Larivière, 2025; Smith et al., 2021; Tóth et al., 2023).

This paper focuses on one source of such inequities: language (Flowerdew, 2019; Politzer-Ahles et al., 2016; Swales, 1997). The use of English as an academic lingua franca has obvious benefits, as it fosters international communication, collaboration, and mobility (Gordin, 2015; Kuteeva, 2023; Montgomery, 2004; Politzer-Ahles et al., 2016; Steigerwald et al., 2022). However, English is an *asymmetric* lingua franca (Gordin, 2015), as it is "the native tongue and the national language of the most influential segment of the global scientific community, but a foreign language for the rest of the world." (Ammon, 2012, p. 335). While some scholars have tried to downplay the role of language, and especially of the dominance of English, as a source of inequalities (Hyland, 2016), there is a growing body of evidence showing the enormous obstacles faced by non-native English speakers in science.

1.1 | Linguistic inequalities in science

Non-native English speakers spend more time reading and writing articles, and preparing to present their work at international conferences—which they are also less likely to attend (Ramírez-Castañeda, 2020)—and their work is more likely to get rejected by journals on the basis of language (Amano et al., 2023; Beigel & Bringel, 2023). Evidence shows that language-related disadvantages are less pronounced in late-career researchers, but "this could be due to survivorship bias; only those non-native English speakers who have managed to conduct science in English as efficiently as native English speakers may have remained in a research career" (Amano et al., 2023, pp. 7–8). However, it is not just a matter of time. Given that it is harder to express ideas in a non-native language (Bortolus, 2012; Flowerdew, 2019; Gordin, 2015; Hanauer et al., 2019), non-native English speakers across linguistic communities systematically show greater dissatisfaction with their work in English compared to their first language, and also experience more anxiety (Hanauer et al., 2019; Ramírez-Castañeda, 2020). Nevertheless, the cost of learning English is not the same for all researchers—it is one of the hardest languages to learn for native speakers of linguistically dissimilar languages such as Korean or Japanese (Ammon, 2012, p. 340). Moreover, linguistic inequalities intersect with economic (Soler, 2020) and gender (Amano et al., 2025) inequalities. Well-resourced researchers may be able to afford professional translators or copyediting services (Amano et al., 2023; Ramírez-Castañeda, 2020), but others must resort to other strategies, such as relying on literacy brokers to support their publishing (Curry & Lillis, 2014) or even trading co-authorship positions for linguistic skills (Bortolus, 2012).

Given that diverse research groups are more likely to make novel scientific contributions (Hofstra et al., 2020), all these obstacles are detrimental to the advancement of knowledge (Amano et al., 2023; Steigerwald et al., 2022). Within the scientific field, overlooking scientific contributions in languages other than English may lead to the systematic neglect of a substantial body of valuable scientific work (Amano et al., 2016; Ammon, 2012; Curry & Lillis, 2014; Hunter et al., 2021; Pradier, Kozlowski, Mazoni, & Larivière, 2025; Steigerwald et al., 2022). Furthermore, language barriers can limit the accessibility of knowledge for some user groups—especially outside the scientific field—who, in many cases and contexts, may not have a proper understanding of the language. More broadly, addressing the general audience and policymakers requires translation in non-Anglophone contexts (Amano et al., 2016; Bortolus, 2012; Chavarro et al., 2017; Curry & Lillis, 2024; Montgomery, 2004).

Existing scholarship outlines several policy recommendations aimed at mitigating the harmful effects of monolingualism in science. To favor the discoverability of publications in other languages, authors have suggested elaborating periodical comprehensive reports in English on publications in other languages (Ammon, 2012), alerting authors and editors to cutting-edge research carried out in non-Anglophone contexts (Kuteeva, 2023), and considering other languages when conducting a systematic review (Amano et al., 2016). Conversely, to reach diverse linguistic communities, publishing English-language articles accompanied by abstracts in languages other than English can be a successful strategy (Kuteeva, 2023).

1.2 | Multilingualism in scholarly communication

Although a complex series of historical transformations led to cement the position of English as the current hegemonic language of science (Ammon, 2012; Gordin, 2015), the same challenges will persist as long as science remains largely monolingual: subordinated linguistic communities will continue to face obstacles to fully contributing to—and benefiting from—scientific knowledge,

and valuable insights will be systematically overlooked. As a result, there is a growing consensus on the importance of maintaining a balanced multilingualism that considers all the communication purposes of science, as well as all the languages needed to fulfill these purposes (Sivertsen, 2018, p. 4). Supporters of multilingualism in scholarly communications advocate for the dissemination of research findings in all languages, arguing for equal access and equal valorization regardless of the language in which they are written (Helsinki Initiative on Multilingualism in Scholarly Communication, 2019).

This article is aligned with such efforts and seeks to answer the following research question: **How much linguistic diversity is there in science?** Previous research on linguistic diversity in science has mostly focused on specific regions and languages (Beigel, 2022; Gong et al., 2019; Hanafi & Arvanitis, 2014; Kim, 2018; Larivière, 2018; Waast et al., 2010). Taken together, these studies suggest that across regions, research evaluation systems and institutional demands are the main factors shaping the incentives to increasingly publish in English and cite English-language literature. When surveyed, scientists also mention other motivations to communicate in English, such as the desire to reach a global audience (Kim, 2018). However, this motivation might simply reflect researchers' internalization of research evaluation criteria—as publishing in English (Larivière, 2018) and citing English-language literature (Gong et al., 2019) are associated with higher citation counts.

Previous studies consistently show that the social sciences and humanities remain the most linguistically diverse domains—mostly because of the localized nature of their research objects and their engagement with local public debates. Latin America (Beigel, 2022) and the Arab world (Hanafi & Arvanitis, 2014) are interesting cases of coexistence of English and local languages, as authors find language-segmented circuits of knowledge production and diffusion. The literature also highlights the complex relationship between language and the contents of science. Analyses of the French, German, Quebec (Larivière, 2018), and North African (Waast et al., 2010) contexts suggest that favoring publication in English-language journals might unintentionally discourage research focusing on local topics.

To the best of our knowledge, although this lack of evidence has been identified as a fundamental gap in the literature (Ammon, 2012), a single previous study by Garfield and Welljams-Dorof (1990) conducted a global analysis of publication and citation languages across countries. However, such a study is based on Web of Science data, which has well-documented linguistic biases (Asubiaro et al., 2024; Mongeon & Paul-Hus, 2016). Using inclusive data sources, which do not select the journals they index on the citations they receive, we study the presence of languages across academic circuits of publications, as evidenced by the incorporation of multiple languages into scientific outputs (journals and publications) and inputs (cited literature). Building on this framework, this paper attempts to provide a better understanding of how linguistic inequalities shape the circulation of knowledge.

2 | MATERIALS AND METHODS

Data for this article were retrieved from the Dimensions and OpenAlex databases (Herzog et al., 2020; Priem et al., 2022), as both sources strive to provide a comprehensive coverage of scientific outputs (Khanna et al., 2022; Singh et al., 2021; Visser et al., 2021). However, our main argument for using Dimensions rather than OpenAlex as our main data source is the greater quality and completeness of its metadata, specifically in terms of cited references (Alperin et al., 2024; Culbert et al., 2025). We examine all articles and conference proceedings indexed in Dimensions and published between 1990 and 2023. Language information—confirmed using manual validation of a sample (Céspedes et al., 2025)—was retrieved from OpenAlex, and matching between both bibliometric databases was based on Digital Object Identifier (DOI) matching. Throughout our results section, all languages below the threshold of 0.5% publications each are grouped into the category labeled as "Other" (see Figures B1-B7 for alternative threshold computations). Russian was also included in this group despite exceeding the 0.5% threshold due to potential indexing issues at the sources of data, which rendered the results unreliable.

Our data consist of 87,577,942 distinct articles and conference proceedings with language information, 54,822,930 of which also included information on their references, resulting in a total of 1,480,570,087 citation links. Our results should be interpreted with the following caveat: publications with information on their references are not randomly distributed across languages; they are much more frequently found in English documents than in other languages (see Supporting Information, Figure A1). Therefore, all of our estimations of the degree of multilingualism of references should be considered as lower bounds.

Dimension's disciplinary classification is based on the Fields of Research (FoR) classification of the Australian and New Zealand Standard Research Classification (ANZSRC) system. This classification system includes 22 research disciplines, which were regrouped into three broad research areas (Table 1).

TABLE 1 Reclassification of Australian and New Zealand Standard Research Classification into broad research areas.

Broad research area	ANZSRC fields of research
Biomedical and Health Sciences (MED)	Biological Sciences, Biomedical and Clinical Sciences, Health Sciences, and Psychology
Natural Sciences and Engineering (NSE)	Agricultural, Veterinary and Food Sciences, Earth Sciences, Environmental Sciences, Information and Computing Sciences, Mathematical Sciences, Built Environment and Design, Chemical Sciences, Engineering, and Physical Sciences
Social Sciences and Humanities (SSH)	Commerce, Management, Tourism and Services, Creative Arts and Writing, Economics, Education, History, Heritage and Archaeology, Human Society, Language, Communication and Culture, Law and Legal Studies, Philosophy and Religious Studies

Articles in our corpus were published in 96,971 distinct journals. Journal language was defined in terms of the most frequent language of publication. In cases where the most frequent language accounted for less than 90% of publications, the journal was considered multilingual (see Figures C1–C6 for alternative threshold computations). Finally, country-level authorship is computed using fractional counting. Each publication is divided by its number of authors, and these fractions are then assigned to each country according to each author's first institutional affiliation (in cases of multiple affiliations). These fractions are later aggregated to determine the proportion of the articles authored by each country (the sum of all fractions equals the number of publications in our dataset). While our country-level analysis cannot provide a detailed report for every single country, readers can explore an interactive version of our results at https://vlab.ebsi.umontreal.ca/languages_app/.

3 | RESULTS

3.1 | To what extent is English the language of science?

Considering all research areas, the proportion of English publications in Dimensions goes down from 93.86% in 1990 to 85.52% in 2023, mainly due to the growing percentage of Indonesian, Portuguese, and Spanish documents (Figure 1a). Interestingly, the share of publications in other languages—those below the 0.5% threshold—has also increased significantly, suggesting growing linguistic diversity within the corpus (for a detailed account of this group's composition, see Supporting Information, Section B). While the number of papers published in English has consistently increased from 877,493 in 1990 to 4,989,675 in 2023, the gap separating it from all other languages of publication seems to be closing (see Supporting Information, Figure A2). Besides the growth in publications in other languages, this relative decrease of English can also be explained by an increase in indexing of non-English language journals, that is to say, the growing visibility of these scientific linguistic communities.

The breakthrough of Indonesian is worth highlighting, from almost complete invisibility to 2.69% of publications in 2023 (Figure 1b). In 2014, Indonesian universities were mandated by government decree to make all scientific publications openly available in order to make knowledge available to a general audience (Huskisson, 2025), a landmark that is reflected in our data. On the other hand, the two main Latin American languages show a gradual but sustained increase in their share of publications, reaching similar proportions by the end of the period considered and surpassing, by far, central European languages with international language communities in decline (Ammon, 2006, 2012; Gordin, 2015; Melliti, 2019). French arguably experienced the most significant fall in relative participation, going from being the second largest language of scientific publication in 1990 to a mere 1.06% by 2023. German, after experiencing some growth in the early 2000s, has also experienced a relative decline in the last decade.

Although these results depict an increasingly multilingual portrait of science—or, at least, an improvement in bibliographic databases' capacity to provide a comprehensive portrait of science—English remains the overwhelmingly dominant language: 98.89% of all cited references in our corpus point to documents in English. However, our results confirm the close links between multilingualism and bibliodiversity (Beigel, 2022; Cruz Romero et al., 2024), as the percentage of English-language references varies across document types (Figure 1e). While citations to conference proceedings, preprints, articles and even book chapters in English are well above 90% of all references to these document types, references to books and monographs refer to documents in English in 90.91% and 87.35% of cases, respectively.

Our findings are also coherent with the phenomenon known as own-language preference (Figure 1c,d), that is, "the degree to which researchers in a certain field draw upon the literature published in their own language" (Bookstein & Yitzhaki, 1999, p. 338). A consideration is in order: Bookstein and Yitzhaki regard own-language preference as a phenomenon of language speakers. In

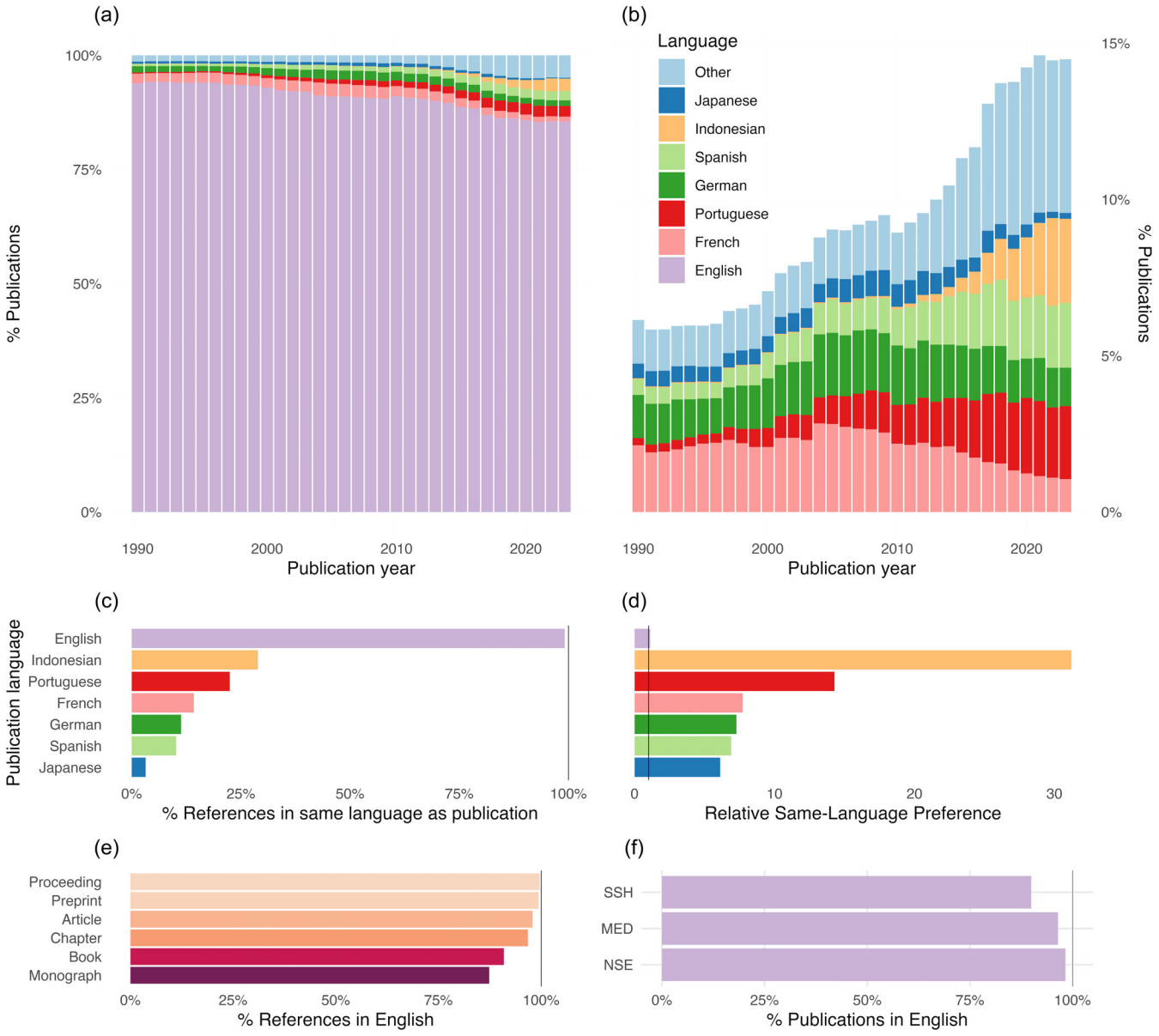

FIGURE 1 Multilingualism at the article level (1990–2023). (a) Percentage of articles indexed in Dimensions by language. (b) Percentage of articles indexed in Dimensions by language, excluding English. (c) Percentage of cited references in the same language as the citing publication. (d) Relative Same-Language Preference (ratio between the observed same language citation rate and the expected same language citation rate). (e) Percentage of English-language references by document type (as defined in Dimensions). (f) Percentage of publications in English, by broad research area (MED, Biomedical and Health Sciences; NSE, Natural Sciences and Engineering; SSH, Social Sciences and Humanities).

our interpretation, the coincidence between language of citing and cited documents may not involve multilingual authors' *own* language in the sense of their first language, but rather, that in which they regularly work, conduct research and seek to publish (Curry & Lillis, 2014). Consequently, to better capture the writing, citing and publishing practices of multilingual scholars, it would be more precise to speak of *same*-language preference (thus linking the language in which the citing and the cited documents are written rather than national official or co-official languages, or individuals' first or preferred languages).

However, these results must take into account the language composition of citable documents. The Relative Same-Language Preference (RSLP) indicator represents the ratio between the *observed* same language citation rate and the *expected* same language citation rate—the language's share of publications (Bookstein & Yitzhaki, 1999, p. 339).

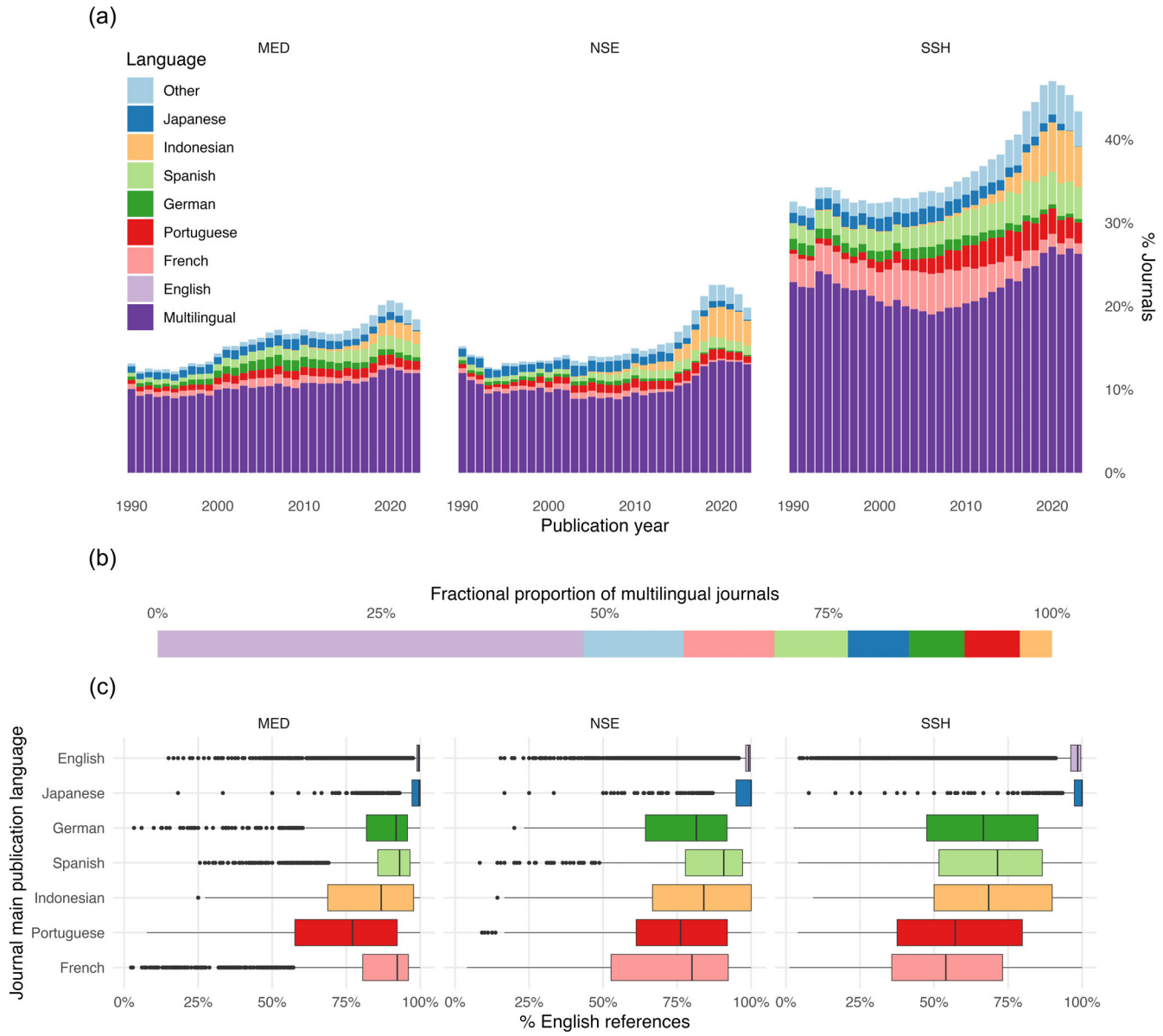

FIGURE 2 Multilingualism at the journal level (1990–2023). (a) Percentage of journals indexed in Dimensions by language and broad research area, excluding English (in cases where the most frequent language accounted for less than 90% of publications, the journal was considered multilingual). (b) Language composition of multilingual journals. (c) Distribution of English references by broad research area and journal publication language.

For instance, if 50% of language *L*'s references are to publications written in language *L*, but only 10% of the publications are written in *L*, we would obtain a ratio of 5.

English-language articles mainly cite literature in the same language (Figure 1c). However, this phenomenon is not due to a disproportionate preference to cite other English-language literature, but rather due to the overwhelming majority of citable documents in that language (Figure 1d). As for Indonesian, more than a quarter of the references found in articles written in this language are to other Indonesian publications, and this case does reflect an aggressive preference for articles written in Indonesian. In Latin America, both Portuguese and Spanish constitute their own circuits of regional scholarly publishing and knowledge circulation as well (Beigel, 2022, 2023; Salatino, 2019, 2022), but same-language preference seems to be less pronounced for hispanophone researchers than for their lusophone counterparts.

The role of languages also varies across scientific domains due to differences in research cultures, as well as the nature of certain objects of study across academia. Figure 1f shows the proportion of publications in English for each of the three broad categories defined in the Methods section. In line with studies showing the

markedly Anglophone monolingual circuits of mainstream science publishing and indexing (Mongeon & Paul-Hus, 2016; Salatino, 2022; Vera-Baceta et al., 2019), in NSE almost all documents are found to be written in English; this proportion is somewhat lower in MED. It is in SSH where we find the comparably lowest share of English-language publications; however, with an 89.91%, the supremacy of English in SSH publications can hardly be questioned.

3.2 | The role of journals as venues for non-English conversations

Given their role as venues of circulation of knowledge, journals are key to structuring academic fields (Bourdieu, 2001). In agreement with existing evidence, multilingual journals and journals published in languages other than English currently represent little more than 20% in MED and NSE, while this proportion is closer to 40% in SSH (Figure 2a). French and German still retain a minimal presence among MED and NSE journals, but it is in SSH where their decline has been more pronounced, particularly for French. As discussed by Larivière (2018), this trend is consistent with the decision of French, German, and Quebec-based local journals to accept a growing number of submissions in English. Japanese journals maintained a modest, if diminishing, proportion of journals in all three fields for most of the considered period, but sharply declined in the last 2 years.

Both Spanish and Portuguese journals have increased their shares in all three disciplinary fields. This growth can be attributed to the development of region-wide indexes and infrastructures for scholarly publishing and scientific information—such as Latindex (1995), Scielo (1998), Redalyc (2003) or LaReferencia (2012)—which have provided Latin American journals with the necessary support and tools to expand their activities (Beigel, 2022). These platforms represent a fundamental system of conservation and visibility of Latin American scientific output across research areas (Packer, 2020). The trajectory of Indonesian journals is somewhat different, since its explosive growth is not built on pre-existing academic circuits, but is driven by the massive adoption of the Open Journal System (OJS) as an infrastructure for digital publication over the last two decades, making the country the current world leader in open access journals editing (Huskisson, 2023).

While the proportion of multilingual journals has remained fairly stable for most of the period in MED and NSE, and even began to experience some growth in the last 5 years (more markedly in the case of NSE), these journals went down in SSH from 24.18% to 19% between 1993 and 2006. The tendency has been reversed, reaching a higher percentage in 2023 than that of 30 years ago. This speaks of a shifting trend in the policies of SSH journals, increasingly accepting submissions and publishing articles in at least two languages.

In order to verify whether this multilingual turn is real or just a formal declaration of principles by journals, we examined the language distribution of the papers in multilingual journals (Figure 2b). As we can see, almost half of the documents published in journals that formally accept submissions in different languages are actually in English. French comes in second, with 10.17% of articles published in multilingual journals. The scarce 3.6% of articles in Indonesian published in multilingual journals may indicate a preference for publication in Indonesian-only journals when the articles are written in this language.

Unsurprisingly, almost all references found in English journals are to English documents (Figure 2c). This applies to Japanese journals as well, revealing—at least in terms of the available data—a linguistic community that has been completely absorbed by the English mainstream. For all other languages, we observe a greater dispersion in the language composition of citations, indicating varying levels of integration into the mainstream English circuit within each linguistic community.

We find the lowest rates of citations to English-language literature in articles in SSH journals, especially for French, where median citations to English documents barely surpass 50% of the total. There certainly are many prominent French SSH journals, but their sphere of influence is limited to their national and linguistic fields, while they remain invisible to the USA (Heilbron & Gingras, 2018, p. 53; as cited in Kuteeva, 2023). This is indicative of a still closely-knit linguistic and academic community which reads and cites the knowledge that circulates in French, backed up by the symbolic capital held by some iconic French publications, editorials, and institutions, by digital infrastructures such as libraries and repositories, and by linguistic policies seeking to maintain (or restore) the relevance of the French language in the scholarly communications landscape.

Conversely, articles published in NSE, and especially in MED journals, are more likely to reference English-language literature. In this regard, Portuguese and Indonesian journals stand out, as they have built strong linguistic and academic circuits of publication that operate in all fields—not only in SSH, where multilingualism is widespread. In the English-dominated fields of NSE and MED, this is no small feat. As discussed by Beigel & Salatino (2015, p. 18), thinking in terms of circuits entails undoing the widespread identification between visibility

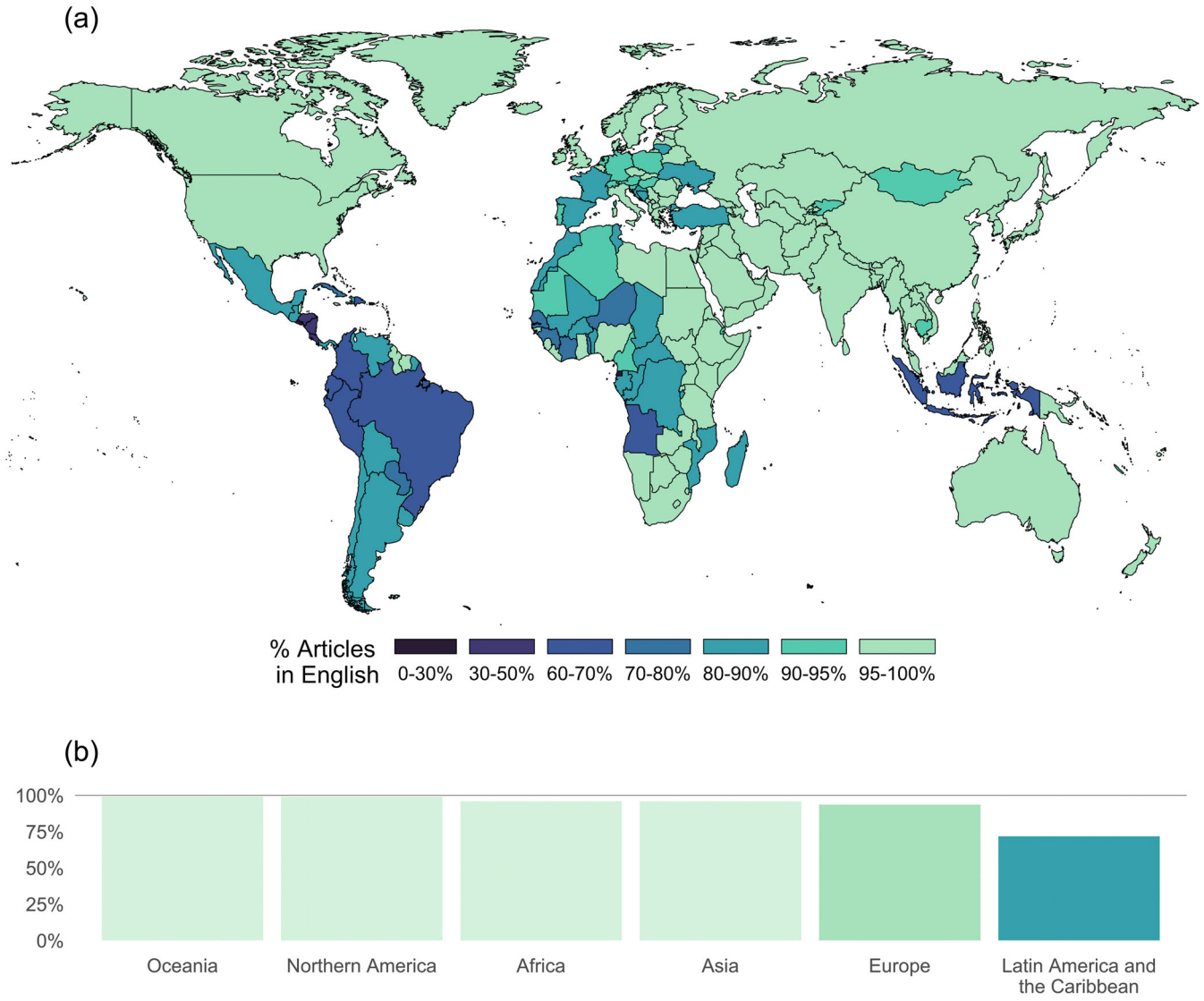

FIGURE 3 Multilingualism in publications across the world (1990–2023). Proportion of articles and conference proceedings in English by (a) country (countries with at least 30 publications) and (b) region. Country-level authorship is computed using fractional counting. Each publication is divided by its number of authors, and these fractions are then assigned to each country according to each author's first institutional affiliation (in cases of multiple affiliations).

and indexation, since the circulation of knowledge depends on its accessibility—and English-language publications do not necessarily lead to broader circulation in non-Anglophone contexts.

These results highlight the tension between the two main functions of scientific journals: being a means of communication and an instrument of consecration (Beigel & Salatino, 2015; Salatino, 2018). Even though evaluation systems encourage scientists to publish their work in high-impact—which is often synonymous with English-language—journals, non-mainstream journals in languages other than English provide essential venues for developing locally relevant scientific conversations and fostering the dissemination of scientific knowledge to the general public (Alperin, 2015; Chavarro et al., 2017; Pradier, Kozlowski, Shokida, & Larivière, 2025; Waast et al., 2010).

3.3 | Where is multilingualism coming from? National and regional dissemination circuits

At the article level, the distribution of English-language publications across countries (Figure 3) exhibits roughly the same pattern as their English-language references (Figure 4), but the bar is set much, much higher for references. Worldwide, even in the

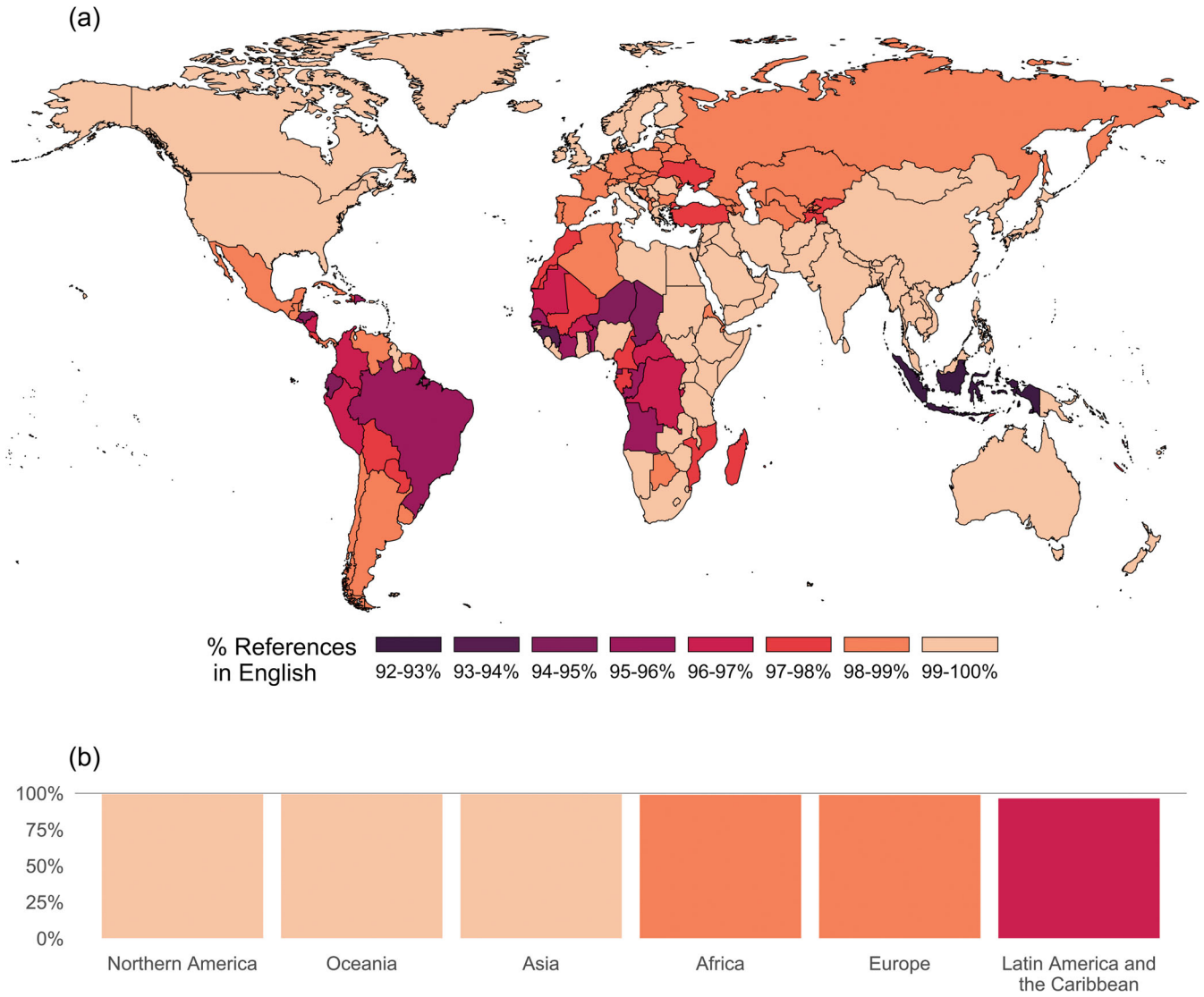

FIGURE 4 Multilingualism in cited references across the world (1990–2023). Proportion of references in English by (a) country (countries with at least 30 publications) and (b) region. Country-level authorship is computed using fractional counting. Each publication is divided by its number of authors, and these fractions are then assigned to each country according to each author's first institutional affiliation (in cases of multiple affiliations). Cited references are assigned to the same country (or countries) as the publication citing them.

countries with the lowest proportions of articles in English, such as Indonesia, Brazil, Ecuador, or Angola, references to English publications represent no less than 92%. While this result is partly a consequence of the overwhelming majority of citable documents in that language (Bookstein & Yitzhaki, 1999), it also illustrates how researchers in non-Anglophone communities must engage in dialogue with specialized debates and types of literature that do not necessarily coincide with those of their linguistic community (Beigel & Bringel, 2023).

This world distribution of English publications and references corresponds almost exactly with Kachru's (1992) model of concentric circles of world Englishes:

> These circles are defined with reference to the historical, sociolinguistic and literary contexts. (...) The Inner Circle represents the traditional bases of English, dominated by the 'mother tongue' varieties of the language. In the Outer Circle, English has been institutionalised as an additional language (...) The Expanding Circle includes the rest of the world where English is used as the primary foreign language, and the uses of English are unpredictably increasing. (p. 3)

Our results show that Inner Circle Anglophone countries such as the USA, the UK, Ireland, Australia, and

New Zealand practically do not publish in any other language than English. In the case of Canada, it can be inferred that the majority of English-speaking scholars' publication patterns drive the average of this country closer to an almost complete hegemony of English rather than towards a balanced bilingualism. As for references, these countries seem to be impervious to literature in languages other than English.

Outer Circle African and Asian countries where English is an official or co-official language or has any other kind of legal status in public administration or education, such as Kenya, Nigeria, South Africa, or India, also lean towards English publications and references almost exclusively. On the contrary, in countries with a history of invasion and colonization by the French, Portuguese, and Belgian empires—for example, Niger, Mozambique, or the Democratic Republic of Congo—or where Arabic plays an important role in scholarly communications—for example, Morocco, Algeria, or Tunisia (Melliti, 2019)—English does not appear as the sole language of scholarly communication. These patterns reveal the deeply entrenched colonial legacy in the linguistic heritage and academic cultures of former European colonies and highlight the political and epistemic efforts still ahead to decolonize the structures of production, circulation, and appropriation of scientific knowledge (Canagarajah, 2022, 2025; Khanna et al., 2022).

Meanwhile, in the Expanding Circle, central non-Anglophone countries with strong academic traditions exhibit slightly lesser proportions of publications and references in English (mainly France and Spain, and, to a lesser extent, Germany). The case of Italy is particularly striking, as the language of Dante seems to have been completely overridden by English in both publications and references. Interestingly, Portugal shows a higher anglicization than these countries and is in the antipodes of Brazil. This is a different situation than that of Spain, which leans closer to the behavior of some Spanish-speaking Latin American countries, both in terms of language of publications and references. Countries in Eastern Europe and Asia—except for Indonesia, and to some extent, Turkey, Ukraine, and Russia—have a near complete preference to cite English-language documents. However, we can hardly draw robust conclusions from our results for Russian, Korean and Chinese-speaking countries, as publications from these countries are less likely to be indexed, and OpenAlex performs particularly poorly in identifying publications in these languages (Céspedes et al., 2025; Zheng et al., 2025).

Finally, Latin American countries consistently exhibit low rates of publications and references in English. However, intra-regional nuances can be observed: Brazil, Colombia, Ecuador, and Peru publish in English markedly less than México, Venezuela, Chile, Argentina, or Uruguay. Guyana and Suriname, where the official languages are English and Dutch respectively, are particular cases due to their political and cultural ties with the Anglophone Caribbean and the Netherlands. El Salvador, Nicaragua, and Honduras, on the contrary, are the countries with the lowest proportion of English publications worldwide. These results are remarkable, since local journals in this region are frequently resource-constrained and therefore cannot afford to pay for DOIs in their publications (Authier, 2023; Beigel, 2024). As a result, our data source reflects only the subset of journals most integrated into the mainstream English-speaking publishing circuit (Pradier, Kozlowski, Mazoni, & Larivière, 2025). While this result supports the finding of the coexistence of multiple linguistic circuits in Latin America (Beigel & Salatino, 2015), it also points to a less favorable implication: researchers may bear the burden of cultivating two academic identities in order to engage in dialogues both in their first language and in English (Flowerdew, 2019).

Focusing on the regional level aggregations (Figures 3b and 4b), Latin America stands out as the only region that has successfully developed a research dissemination ecosystem that is relatively detached from the English-speaking mainstream. While the Indonesian case presents a promising example of successful national-level policies aimed at countering English-language hegemony, its linguistic isolation within the region poses challenges for the future growth of an Indonesian language community. Africa may seem paradoxical at first glance, since the continent has some of the countries with the highest and lowest percentages of English-language references. However, this division clearly corresponds to the linguistic distribution of the European colonization of Africa, where clusters of countries with low percentages of English-language publications and references are basically still under the influence of colonial language communities that are gradually losing their power (Ammon, 2006). Conversely, the characteristics of the colonization process in Latin America resulted in a landscape where almost all countries share a single official language, and the second most relevant language has common linguistic origins, resulting in low barriers to communication.

4 | DISCUSSION AND CONCLUSION

Despite the dominance of English, the production of scientific knowledge is a multilingual endeavor. Our study brings attention to the fact that, if English acts both as a bridge and a fence in international exchanges of scientific knowledge (Kuteeva, 2023), the dialogues happening under the bridge and around the fence should not be

overlooked. Drawing on comprehensive bibliometric data sources, we show that—partly thanks to a better indexing—there has been a change in the position of languages within the scientific field in the last two decades. The XXth Century's Anglo hegemony (Gordin, 2015) is gradually giving way to an increasingly South American- and South Asian-leaning landscape of Indonesian, Portuguese, and Spanish publication circuits. Meanwhile, French and German—once contenders for hegemonic positions in science (Ammon, 2006, 2012; Gordin, 2015; Melliti, 2019)—are in relative decline. These findings have clear policy implications, as they offer new evidence to understand both successful and unsuccessful cases of language policy.

Important limitations to these results have to be acknowledged. They represent lower-bound estimates of the linguistic distribution of publications and citations, as our data sources, despite being inclusive, still underrepresent non-English content. This is due to a number of factors. First, OpenAlex often misclassifies non-English texts as if they were in English (Céspedes et al., 2025), which leads to an overestimation of English in detriment to other languages. This means that, in reality, the research dissemination ecosystem is more multilingual than our results suggest. Second, citation links and discipline data are more limited for non-English publications. Third, bibliographic databases (even those with a broad coverage such as Dimensions or OpenAlex) do not provide complete coverage of non-mainstream circuits, where many publications often do not even have a DOI. This leaves out of the picture a very relevant part of scientific outputs from regions or institutions that cannot afford to pay for DOIs in their publications (Beigel, 2024). For example, as surveyed by Authier (2023), 54% of Latin American journals indexed in the Directory of Open Access Journals lack any kind of persistent identifiers. As large-scale bibliometric databases such as OpenAlex and Dimensions rely on Crossref as its backbone, journals that use other DOI registration agencies[1] are less likely to be covered. As most of those agencies are in Asia, journals from that region may suffer from indexing issues, which results in inaccurate representations of these countries' languages in our maps. The Web of Science does have an Arabic, Chinese, and Russian Citation Database as part of their collections (in addition to integrating Scielo and the Korean Journal Database). But WoS has a more restrictive coverage than Dimensions and OpenAlex; therefore, these regional indexes might still be leaving out a sizable proportion of journals. In turn, those covered by wider-ranging databases may be lacking well-curated metadata for citations and languages.

The fact that English-language publications and references still account for the vast majority points to the dominant position of theoretical and methodological perspectives associated with the use of this language (Beigel & Bringel, 2023; Bennett, 2015; Hunter et al., 2021; Kuteeva, 2023), and the potential loss of diversity of cultural thought patterns embodied in different rhetorical practices (Swales, 1997). Nevertheless, for non-Anglophone authors, publications or translations into English may also be a strategy to reach visibility in other national fields (Gingras & Khelfaoui, 2025; Melliti, 2019), thus potentially enabling South–South collaborations and dialogues facilitated by English as a vehicular or contact language. In MED and NSE, the overwhelming majority of English-language publications and references is a warning sign of an expanding epistemological monoculture (Bennett, 2015). On the other hand, the so-called parochialism or provincialism of SSH may be saving epistemic and linguistic diversity in that field.

English monolingualism can be understood in terms of what Law and Lin (2017) call "market-oriented parochialism", this is, the demands from commercial publishers—particularly those based in North America and Europe—that publications be made more "international", where "international" means attuned to dominant research agendas. As Curry and Lillis (2014) have observed, in recent decades the signifier "international" has slid towards the signifier "English", and together, both terms constitute a naturalized—but by no means natural—indicator of scientific quality. As a result, evaluation systems worldwide pressure researchers to publish in English (Bortolus, 2012; Flowerdew, 2019)—particularly in high-impact "international" English-language journals based in the UK, the Netherlands, and the USA (Beigel & Salatino, 2015; Chavarro et al., 2018; Hanauer et al., 2019).

The linguistic communities we have identified in this study—mainly Indonesian or Brazilian scholars citing literature written in their own languages—emerge as spaces that question the assumption that publishing in one's own country is an endogamic and therefore discouraged practice (Beigel, 2023), valuing, instead of marginalizing, "the ways of knowing and being entwined in these languages and the people and communities who embody, care for, and practice this knowledge" (Shi, 2023, p. 8). Given that both countries are deeply engaged with the open-access movement (Beigel et al., 2024; Packer, 2020; Seo, 2018; Van Noorden, 2019), their scholarly publishing ecosystem operates with considerable autonomy from commercial publishers' demands. Open-access infrastructures in these regions truly embrace multilingualism, as they publish content in languages other than English, and also combine these contents with English translations (Khanna et al., 2022; Packer, 2024). In accordance with Kuteeva (2023), our findings illustrate how these parallel language policies have contributed to the development of academic circuits that improved the discoverability

of publications produced within these linguistic communities. An additional unintended positive outcome of these initiatives is their contribution to the progress of translation technologies. While translation tools have the potential of enormously reducing the burden on non-native English speakers, they are trained using human-made translations—and therefore benefit from the greater availability of high-quality translations of scientific texts (Steigerwald et al., 2022).

Thus, if Indonesian or Portuguese occupy a more prominent position today than they did 30 years ago, it is owing to the massive adoption of infrastructures such as the Open Journal Systems (an open-source, free-to-use journal publishing platform) or to the enactment of policies that value publications in the national language as well as those in English. Both are examples of science policy with a linguistic angle that have had a very concrete impact on the distribution of languages in the global field of scholarly communication, effectively contributing to multilingualism and bibliodiversity and, in the process, allowing other voices to be heard over the sometimes deafening and seemingly impregnable Anglophone wall of sound.

AUTHOR CONTRIBUTIONS

Carolina Pradier, Lucía Céspedes, and Vincent Larivière designed research; Carolina Pradier, Lucía Céspedes, and Vincent Larivière performed research; Carolina Pradier analyzed data; and Carolina Pradier, Lucía Céspedes, and Vincent Larivière wrote the paper.

ACKNOWLEDGMENTS

The authors acknowledge funding from the Social Science and Humanities Research Council of Canada Pan-Canadian Knowledge Access Initiative Grant (Grant Number 1007-2023-0001), and the Fonds de recherche du Québec—Société et Culture through the Programme d'appui aux Chaires UNESCO (Grant Number 338828). Carolina Pradier acknowledges funding from the Fonds de recherche du Québec (Grant Number 362184).

DATA AVAILABILITY STATEMENT

The scripts used for this study are publicly available under: https://github.com/caropradier/languages_science. Detailed results are published in: https://vlab.ebsi.umontreal.ca/languages_app/. OpenAlex data is openly available, but restrictions apply to Dimensions' proprietary bibliometric data. To obtain the bibliometric data, readers can fill in the following application: https://form.asana.com/?k=bnz2qnJNn3o-wMd2RJoYjA&d=1204915780607962.

ORCID

Carolina Pradier https://orcid.org/0009-0007-5058-6352
Lucía Céspedes https://orcid.org/0000-0001-5896-3377

ENDNOTE

[1] https://www.doi.org/the-community/existing-registration-agencies/.

REFERENCES

Alperin, J. P. (2015). *The public impact of Latin America's approach to open access* [Doctoral, Stanford University]. https://purl.stanford.edu/jr256tk1194

Alperin, J. P., Portenoy, J., Demes, K., Larivière, V., & Haustein, S. (2024). An analysis of the suitability of OpenAlex for bibliometric analyses (No. arXiv:2404.17663) arXiv. https://doi.org/10.48550/arXiv.2404.17663

Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages are still a major barrier to global science. *PLoS Biology*, *14*(12), e2000933. https://doi.org/10.1371/journal.pbio.2000933

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J. D., Montaño-Centellas, F., Paudel, K., White, R. L., & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biology*, *21*(7), e3002184. https://doi.org/10.1371/journal.pbio.3002184

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J. D., Montaño-Centellas, F., Paudel, K., White, R. L., & Veríssimo, D. (2025). Language, economic and gender disparities widen the scientific productivity gap. *PLoS Biology*, *23*(9), e3003372. https://doi.org/10.1371/journal.pbio.3003372

Ammon, U. (2006). Language planning for international scientific communication: An overview of questions and potential solutions. *Current Issues in Language Planning*, *7*(1), 1–30. https://doi.org/10.2167/cilp088.0

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*(2), 333–355. https://doi.org/10.1515/applirev-2012-0016

Asubiaro, T., Onaolapo, S., & Mills, D. (2024). Regional disparities in Web of Science and Scopus journal coverage. *Scientometrics*, *129*(3), 1469–1491. https://doi.org/10.1007/s11192-024-04948-x

Authier, C. N. (2023). Cobertura de los identificadores persistentes (PID) para artículos científicos en Latinoamérica. Zenodo. https://doi.org/10.5281/zenodo.8007325

Bakhtin, M. M. (2010). *The dialogic imagination: Four essays*. University of Texas Press. Original work published 1981.

Beigel, F. (2022). Multilingüismo y bibliodiversidad en América Latina. In D. Bentivegna, J. del Valle, M. Niro, & L. Villa (Eds.), *Anuario de Glotopolítica Núm. 5* (pp. 119–132). Tipográfica. https://glotopolitica.com/aglo5/beigel/

Beigel, F. (2023). Les revues argentines de sciences humaines et sociales, entre la circulation régionale et l'ancrage local. *Biens Symboliques/Symbolic Goods*, *12*, 12. https://doi.org/10.4000/bssg.1790

Beigel, F. (2024). Cartographies for an inclusive Open Science. *SciELO Preprints*. 40. https://doi.org/10.1590/SciELOPreprints.10286

Beigel, F., & Bringel, B. (2023). Situated circulation and publication languages of the academic elites of the Southern Cone. *Tempo Social*, *34*, 153–180. https://doi.org/10.11606/0103-2070.ts.2022.194375

Beigel, F., Packer, A. L., Gallardo, O., & Salatino, M. (2024). OLIVA: La Producción Científica Indexada en América Latina.

Diversidad Disciplinar, Colaboración Institucional y Multilingüismo en SciELO y Redalyc (1995-2018). *Dados*, *67*, e20210174. https://doi.org/10.1590/dados.2024.67.1.307

Beigel, F., & Salatino, M. (2015). Circuitos segmentados de consagración académica: Las revistas de Ciencias Sociales y Humanas en la Argentina. *Información, Cultura y Sociedad*, *32*, 11–36.

Bennett, K. (2015). Towards an epistemological monoculture: Mechanisms of epistemicide in European research publication. In R. Plo Alastrué & C. Pérez-Llantada (Eds.), *English as a scientific and research language: Debates and discourses* (pp. 9–35). De Gruyter Mouton. https://doi.org/10.1515/9781614516378-004

Bookstein, A., & Yitzhaki, M. (1999). Own-language preference: A new measure of "relative language self-citation". *Scientometrics*, *46*(2), 337–348. https://doi.org/10.1007/BF02464782

Bortolus, A. (2012). Running like Alice and Losing good ideas: On the quasi-compulsive use of English by non-native English speaking scientists. *Ambio*, *41*(7), 769–772. https://doi.org/10.1007/s13280-012-0339-5

Bourdieu, P. (2001). *Science de la science et réflexivité: Cours du Collège de France, 2000–2001*. Raisons d'agir.

Canagarajah, S. (2022). Language diversity in academic writing: Toward decolonizing scholarly publishing. *Journal of Multicultural Discourses*, *17*(2), 107–128. https://doi.org/10.1080/17447143.2022.2063873

Canagarajah, S. (2025). Diversifying "English" at the Decolonial Turn. *TESOL Quarterly*, *59*(1), 378–389. https://doi.org/10.1002/tesq.3306

Céspedes, L., Kozlowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., Poitras, C., Ninkov, A. B., Ebrahimy, S., Ayeni, P., Filali, S., Li, B., & Larivière, V. (2025). Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*, *76*(6), 884–895. https://doi.org/10.1002/asi.24979

Chavarro, D., Ràfols, I., & Tang, P. (2018). To what extent is inclusion in the web of science an indicator of journal 'quality'? *Research Evaluation*, *27*(2), 106–118. https://doi.org/10.1093/reseval/rvy001

Chavarro, D., Tang, P., & Ràfols, I. (2017). Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling. *Research Policy*, *46*(9), 1666–1680. https://doi.org/10.1016/j.respol.2017.08.002

Cruz Romero, R., Stephen, D., & Stahlschmidt, S. (2024). Assessing bibliodiversity through reference lists: A text analysis approach. Zenodo. https://doi.org/10.5281/zenodo.14045592

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, *130*(4), 2475–2492. https://doi.org/10.1007/s11192-025-05293-3

Curry, M. J., & Lillis, T. (2024). Multilingualism in academic writing for publication: Putting English in its place. *Language Teaching*, *57*(1), 87–100. https://doi.org/10.1017/S0261444822000040

Curry, M. J., & Lillis, T. M. (2014). Strategies and tactics in academic knowledge production by multilingual scholars. *Education Policy Analysis Archives*, *22*(32), 1–28. https://doi.org/10.14507/epaa.v22n32.2014

Davis, P. M. (2009). Reward or persuasion? The battle to define the meaning of a citation. *Learned Publishing*, *22*(1), 5–11. https://doi.org/10.1087/095315108X378712

Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T. D., Pecoskie, J., Tsou, A., & Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, *57*(2), 223–248. https://doi.org/10.1177/0539018417752089

Ennser-Kananen, J. (2019). Are we who we cite? On epistemological injustices, citing practices, and #metoo in academia. *Apples - Journal of Applied Language Studies*, *13*(2), 2. https://doi.org/10.17011/apples/urn.201905092524

Flowerdew, J. (2019). The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. *Language Teaching*, *52*(2), 249–260. https://doi.org/10.1017/S0261444819000041

Garfield, E., & Welljams-Dorof, A. (1990). Language use in international research: A citation analysis. *The Annals of the American Academy of Political and Social Science*, *511*(1), 10–24. https://doi.org/10.1177/0002716290511001002

Gingras, Y., & Khelfaoui, M. (2025). The relative (in)visibility of sociologists in the French, American, British, and German national fields (1970–2018). *Social Science Information*, *64*, 05390184251317451. https://doi.org/10.1177/05390184251317451

Gong, K., Xie, J., Cheng, Y., Larivière, V., & Sugimoto, C. R. (2019). The citation advantage of foreign language references for Chinese social science papers. *Scientometrics*, *120*(3), 1439–1460. https://doi.org/10.1007/s11192-019-03146-4

Gordin, M. D. (2015). *Scientific babel: How science was done before and after global English*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226000329.001.0001

Hanafi, S., & Arvanitis, R. (2014). The marginalization of the Arab language in social science: Structural constraints and dependency by choice. *Current Sociology*, *62*(5), 723–742. https://doi.org/10.1177/0011392114531504

Hanauer, D. I., Sheridan, C. L., & Englander, K. (2019). Linguistic injustice in the writing of research articles in English as a second language: Data from Taiwanese and Mexican researchers. *Written Communication*, *36*(1), 136–154. https://doi.org/10.1177/0741088318804821

Heilbron, J., & Gingras, Y. (2018). The globalization of European research in the social sciences and humanities (1980–2014): A bibliometric study. In J. Heilbron, G. Sorá, & T. Boncourt (Eds.), *The social and human sciences in global power relations* (pp. 29–58). Springer International Publishing. https://doi.org/10.1007/978-3-319-73299-2_2

Helsinki Initiative on Multilingualism in Scholarly Communication. (2019). Federation of Finnish Learned Societies; Committee for Public Information; Finnish Association for Scholarly Publishing; Universities Norway; European Network for Research Evaluation in the Social Sciences and the Humanities. https://doi.org/10.6084/M9.FIGSHARE.7887059

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, *1*(1), 387–395. https://doi.org/10.1162/qss_a_00020

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Hunter, N. B., North, M. A., & Slotow, R. (2021). The marginalisation of voice in the fight against climate change: The case of Lusophone Africa. *Environmental Science & Policy*, *120*, 213–221. https://doi.org/10.1016/j.envsci.2021.03.012

Huskisson, M. (2023). *Guest post – Scholarly publishing as a global endeavor: Leveraging open source software for Bibliodiversity*. The Scholarly Kitchen.

Huskisson, M. (2025, January 23). *Guest post: Reflections from the Munin conference part three – Measuring impact*. The Scholarly Kitchen.

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. https://doi.org/10.1016/j.jslw.2016.01.005

Hyland, K., & Jiang, F. (2019). Points of reference: Changing patterns of academic citation. *Applied Linguistics*, *40*(1), 64–85. https://doi.org/10.1093/applin/amx012

Kachru, B. B. (1992). World Englishes: Approaches, issues and resources. *Language Teaching*, *25*(1), 1–14. https://doi.org/10.1017/S0261444800006583

Khanna, S., Ball, J., Alperin, J. P., & Willinsky, J. (2022). Recalibrating the scope of scholarly publishing: A modest step in a vast decolonization process. *Quantitative Science Studies*, *3*(4), 912–930. https://doi.org/10.1162/qss_a_00228

Kim, E.-Y. J. (2018). Scholarly publishing in Korea: Language, perception, practice of Korean university faculty. *Publishing Research Quarterly*, *34*(4), 554–567. https://doi.org/10.1007/s12109-018-9602-3

Kozlowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences*, *119*(2), e2113067119. https://doi.org/10.1073/pnas.2113067119

Kuteeva, M. (2023). Knowledge flows and languages of publication: English as a bridge and a fence in international knowledge exchanges. *Journal of English for Research Publication Purposes*, *4*(1), 80–93. https://doi.org/10.1075/jerpp.22008.kut

Larivière, V. (2018). Le français, langue seconde ? De l'évolution des lieux et langues de publication des chercheurs au Québec, en France et en Allemagne. *Recherches Sociographiques*, *59*(3), 339–363. https://doi.org/10.7202/1058718ar

Law, J., & Lin, W. (2017). Provincializing Sts: Postcoloniality, symmetry, and method. *East Asian Science, Technology and Society: An International Journal*, *11*(2), 211–227. https://doi.org/10.1215/18752160-3823859

Melliti, I. (2019). Une sociologie tunisienne francophone fait-elle encore sens ? *SociologieS*. 1–4. https://doi.org/10.4000/sociologies.9713

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of web of science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Montgomery, S. (2004). Of towers, walls, and fields: Perspectives on language in science. *Science*, *303*(5662), 1333–1335. https://doi.org/10.1126/science.1095204

Packer, A. L. (2020). The pasts, presents, and futures of SciELO. In M. P. Eve & J. Gray (Eds.), *Reassembling scholarly communications: Histories, infrastructures, and global politics of open access* (pp. 297–313). MIT Press.

Packer, A. L. (2024). Multilingualism in scientific literature communicated by journals from the SciELO Brazil collection. *European Review*, *32*(S1), S124–S144. https://doi.org/10.1017/S1062798724000103

Politzer-Ahles, S., Holliday, J. J., Girolamo, T., Spychalska, M., & Berkson, K. H. (2016). Is linguistic injustice a myth? A response to Hyland (2016). *Journal of Second Language Writing*, *34*, 3–8. https://doi.org/10.1016/j.jslw.2016.09.003

Pradier, C., Kozlowski, D., Mazoni, A., & Larivière, V. (2025). *Occluded topics: The hidden half of Brazilian research*. OSF. https://doi.org/10.31235/osf.io/msbnu_v1

Pradier, C., Kozlowski, D., Shokida, N. S., & Larivière, V. (2025). Science for whom? The influence of the regional academic circuit on gender inequalities in Latin America. *Journal of the Association for Information Science and Technology*, *76*(5), 790–802. https://doi.org/10.1002/asi.24972

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (No. arXiv: 2205.01833) ArXiv. https://doi.org/10.48550/arXiv.2205.01833

Ramírez-Castañeda, V. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. *PLoS One*, *15*(9), e0238372. https://doi.org/10.1371/journal.pone.0238372

Salatino, M. (2018). Más Allá de la Indexación: Circuitos de Publicación de Ciencias Sociales en Argentina y Brasil. *Dados*, *61*, 255–287. https://doi.org/10.1590/001152582018152

Salatino, M. (2019). Circuitos locales en contextos globales de circulación. Una aproximación a las revistas científicas argentinas. *Palabra Clave (La Plata)*, *9*(1), e073. https://doi.org/10.24215/18539912e073

Salatino, M. (2022). Linguistic circuits of Latin American scientific production. *Tempo Social*, *34*(3), 3. https://doi.org/10.11606/0103-2070.ts.2022.201928

Selvi, A. F. (2024). The myopic focus on decoloniality in applied linguistics and English language education: Citations and stolen subjectivities. *Applied Linguistics Review*, *16*(1), 137–161. https://doi.org/10.1515/applirev-2024-0011

Seo, T.-S. (2018). Open access full-text databases in Asian countries. *Science Editing*, *5*(1), 26–31. https://doi.org/10.6087/kcse.114

Shi, J. (2023). Articulations of language and value(s) in scholarly publishing circuits. *Canadian Journal of Academic Librarianship*, *9*, 1–33. https://doi.org/10.33137/cjal-rcbu.v9.38148

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Sivertsen, G. (2018). Multilingüisme equilibrat en ciència. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, *40*. https://doi.org/10.1344/BiD2018.40.24

Smith, C. A., Williams, E. L., Wadud, I. A., Pirtle, W. N. L., & Collective, T. C. B. W. (2021). Cite black women: A critical praxis (a statement). *Feminist Anthropology*, *2*(1), 10–17. https://doi.org/10.1002/fea2.12040

Soler, J. (2020). Linguistic injustice and global English: Some notes from its role in academic publishing. *Nordic Journal of English Studies*, *19*(3), S3. https://doi.org/10.35360/njes.575

Steigerwald, E., Ramírez-Castañeda, V., Brandt, D. Y. C., Báldi, A., Shapiro, J. T., Bowker, L., & Tarvin, R. D. (2022). Overcoming language barriers in academia: Machine translation tools and a vision for a multilingual future. *Bioscience*, *72*(10), 988–998. https://doi.org/10.1093/biosci/biac062

Swales, J. M. (1997). English as tyrannosaurus rex. *World Englishes*, *16*(3), 373–382. https://doi.org/10.1111/1467-971X.00071

Tóth, J. J., Háló, G., & Goyanes, M. (2023). Beyond views, productivity, and citations: Measuring geopolitical differences of scientific impact in communication research. *Scientometrics*, *128*(10), 5705–5729. https://doi.org/10.1007/s11192-023-04801-7

Van Noorden, R. (2019). Indonesia tops open-access publishing charts. *Nature*. https://doi.org/10.1038/d41586-019-01536-5

Vasen, F., Sarthou, N. F., Romano, S. A., Gutiérrez, B. D., & Pintos, M. (2023). Turning academics into researchers: The development of National Researcher Categorization Systems in Latin America. *Research Evaluation*, *32*(2), 244–255. https://doi.org/10.1093/reseval/rvad021

Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of science and Scopus language coverage. *Scientometrics*, *121*(3), 1803–1813. https://doi.org/10.1007/s11192-019-03264-z

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Waast, R., Arvanitis, R., Richard-Waast, C., & Rossi, P. L. (2010). What do social sciences in North African countries focus on ? In *World social science report* (pp. 176–180). UNESCO.

Zheng, M., Miao, L., Bu, Y., & Larivière, V. (2025). Understanding discrepancies in the coverage of OpenAlex: The case of China. *Journal of the Association for Information Science and Technology*, *76*, 1591–1601. https://doi.org/10.1002/asi.70013

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pradier, C., Céspedes, L., & Larivière, V. (2026). How multilingual is scholarly communication? Mapping the global distribution of languages in publications and citations. *Journal of the Association for Information Science and Technology*, 1–15. https://doi.org/10.1002/asi.70055