Case studies

Linguistic equity as open access: Internationalizing the language of scholarly communication

Race MoChridhe

University of Wisconsin–Milwaukee, United States of America

A B S T R A C T

The open access movement has called attention to ways in which financial barriers to participation in scholarly discourse inhibit the growth of knowledge and perpetuate global inequities. The majority of the focus, however, has lain upon two kinds of direct financial barrier: subscription fees and article processing charges. This article proposes that the use of English as the *lingua franca* of modern scholarly communication constitutes a 'hidden paywall' that counterproductively inhibits the participation of scholars from many parts of the world and particularly from the global south. After a brief review of the costs of this 'hidden paywall' and of the tradition of 'great power' *linguae francae* in which English now stands, this article suggests that 19th and 20th century proposals for the use of a constructed auxiliary language as an aid to global scholarship now deserve to be revisited, with contemporary developments in publishing technology and machine translation rendering them feasible alternatives to the *status quo* in ways that they were not when first introduced.

Introduction

Last year's announcement of 'Plan S' to drive all European research into open access (OA) outlets has convinced many that OA's time has come. Though not without critics, the road map is widely seen as overdue, driven by the steep rise of journal subscription costs and dwindling of library budgets, the proliferation of interest in library publishing, the need for exchange of research data and results on faster timelines than traditional publishing models have allowed, and the increasing prevalence of research outputs in nontraditional forms that disrupt long-established practices of publication and peer review.

Part of the motivation has come also from increased awareness of the diversity of the global research community and concerns about the disproportionate barriers that the cost of scholarly communication poses for scholars in the developing world, at small institutions, or without institutional affiliations—categories that have grown markedly due to changes in academic hiring practices. As more scholars find themselves on the margins, they find themselves also frozen out of research that could inform their work and/or channels to disseminate it. The loss of their talents and energies to the world's research communities is increasingly recognized as unacceptable.

The best means of increasing their engagement are debated, as ongoing arguments about 'green' and 'gold' OA models testify. Both subscription costs and article processing charges (APCs), however, pale in comparison with the cost of acquiring the language skills needed to read or write academic literature in the first place.

The hidden paywall of English

Research on the SCOPUS database estimates that 80% of all articles it indexes are written entirely in English and that most Western countries now produce more articles in English than in their native languages, with the proportion of English to vernacular papers rising steadily (Van Weijen, 2012). This consolidation facilitates international access in some ways, but in others actively discourages it. A study in Saudi Arabia found that the second most common reason for researchers there to avoid 'deep web' databases was that 'most of the resources available through the invisible web is [sic] in English'—a statement with which 85% of respondents either 'agreed' or 'strongly agreed' (Alyamu & Assiri, 2018, p. 89). That study's recommendations prominently included 'increas[ing] and develop[ing] the Arabic content of the invisible web' (p. 90), but such efforts are suppressed by the snowball effect of existing impact metrics, which largely rely on tallying citations. With such a large imbalance between output in English and in other languages, it is virtually guaranteed that papers in non-English languages will be less cited, and this is reflected in rankings such as SCImago's, where 48 of the top 50 science journals in 2017 were based in the United States or the United Kingdom, and both of the outlying journals (published in the Netherlands and Austria) published solely in English. It is reflected also in the policies of some universities that

automatically award greater weight to English-language publications when evaluating faculty research outputs (Phillipson, 2015, p. 28). The result is that English proficiency is a *de facto* requirement to participate in nearly all scholarly discourse. The advantages this gives to native speakers have been considered at length elsewhere (*e.g.* Li, 2013, p. 65). What may contribute to the discussion here is to attempt some quantification of the financial barrier imposed.

To put a precise number on the cost of acquiring English is difficult, as different learners have different levels of linguistic aptitude, different life circumstances determining their exposure to the language outside formal instruction, and different national contexts determining the level of taxpayer-funded instruction they enjoy. The broad inadequacy for higher academic purposes of the preparation in English offered by standard primary and secondary education in most countries may be judged in part, however, from the fact that intensive English programs (IEPs), intended as remedial language training for incoming international students, became in this decade the fastest-growing field of study in the US (Choudaha et al., 2013, p. 5), with large numbers of students entering from developed countries with robust educational systems, such as South Korea. South Korea is also a leading sender of students for English study abroad in other programs, the average cost of which per student totaled over $27,000 in 2011—a full $3000 greater than the country's GDP *per capita*. That these investments are rewarded has been shown by studies correlating higher wages, greater employment opportunities, and other benefits with the higher-level English acquired by those able to afford such training (Choi, 2015). The scale at which such additional, post-secondary English training is considered necessary may be gleaned by the total value of receiving such students to the US economy, estimated at nearly four billion dollars (Choi, 2015, p. 11). As Linn (2015) observed, 'In the market of English there are winners and there are losers … academic publishers get rich on the monolingual norm of the industry, and private language teaching is itself big business' (p. vii).

If this is true of even relatively affluent countries, it is more so in the developing world, where attempts to extend access to English learning opportunities have yielded poor results in both public and private models (Cronquist & Fiszbein, 2017). In the latter case, the large sums paid by poor families into schools that do not adequately prepare their children in English actively reinforce cycles of poverty, rather than lifting families out of it (Endow, 2018). What is more, most studies on this point have been concerned primarily with business English, whereas the competency required for academic work is often much higher. In sub-Saharan Africa, for instance, Jaygbay (1998) has estimated that as much as half of the population is effectively excluded from academic discourse by the inability to read English, French, or Portuguese at a sufficient level. The mastery needed to author academic prose being even greater than that required to read it, Pearce (2003) found that even advanced second-language speakers of English suffer from grammatical and diction errors that negatively impact their rates of acceptance for publication and hence their academic careers. Anecdotally, even relatively high-level speakers of English who meet or exceed the normal standards for professional business contexts report difficulty in following oral academic presentations in detail, greatly increased amounts of time needed to read articles, and social inhibition in connecting and collaborating with other scholars (Huttner-Koros, 2015). All of these factors contribute to a phenomenon much easier to quantify than the cost of English-language education and more suited for direct comparison with other publishing costs (particularly the APC)—namely, literacy brokering.

Literacy brokering refers to a variety of services by which linguistic aptitude in written communication in a desired language is traded as a commodity, including editing, translation, and consultancy. Though the segment of such services aimed specifically at academics might be thought niche, in many parts of the world it has become big enough business to warrant attention in the general media (Sugiharto, 2017). Prices vary by provider, locality, and level of service, but within the US

an academic copy editor generally costs a minimum of $35/h, with $50 being a more usual baseline. The most experienced and those working on the most specialized material may range up to $85. At those rates, a typical 30-page academic article by even a native English speaker not infrequently costs around $2000 to have independently edited (Belcher, n.d.); those by writers requiring more intervention may range still higher. Comparing this to the current average APC of $1684 per article (Pieper, 2018), it is easy to see how the 'hidden paywall' is often the higher of the two. Many scholars, of course, end up paying both.

This estimate ignores costs involved in an author's efforts to attain English language skills in the first place, which have already been discussed. To it must be added also indirect costs, such as additional needs for research and clerical assistance, increased teaching release time needed to prepare manuscripts in English, *etc.* (Lillis & Curry, 2010, p. 282). Even in their most conservative form, the numbers tell a story of disenfranchisement often ignored in our broader conversation, in large part because the cost of language acquisition and brokering is so thoroughly outsourced to scholars themselves that it is largely invisible to institutional budgets.

This outsourcing can sometimes have an additional perverse effect. Lillis and Curry (2006) studied several 'textual histories' of academic articles in preparation, looking at the interventions of a wider range of brokers, including academic colleagues and staff editors of journals. They found that even scholars with many years of experience in academic English remained dependent on literacy brokering to publish their work (p. 17), but also that the interventions of brokers often went beyond simple edits to more substantively reshape, reframe, and refocus the content of research. In one case, staff editors from a journal based in the Anglosphere effectively transformed a psychology article intended to present new knowledge into a 'peripheral' case study using the author's cultural context merely to confirm existing work from the English-speaking 'centre' (p. 30). In this way, even when the work of researchers from outside the English-speaking world makes its way into English-language discourse, it can do so in distorted forms that impair, or even nullify, the contribution it is intended to make.

Alternatives to the *status quo*

The extent to which English's role can or should be understood as 'linguistic imperialism' or critiqued in the context of Anglo-American systems of economic, political, cultural, and military power projection is an interminable debate. Relevant here is only the stark fact that English has *de facto* assumed the role which, in the West, was formerly held by Greek and Latin, and in other regions by Arabic, Persian, Aramaic, Sanskrit, Chinese, and Nahuatl. All of these have, at one time or another, served as common currencies of scholarly communication across diverse cultures and vernaculars, simultaneously opening doors and closing them by creating transnational spaces for the pursuit of knowledge that functioned simultaneously as walled gardens for social elites capable of dedicating the time, money, and effort to acquire their parlance. English is far less daunting for most second language learners than its predecessors in this role and yet, to those whose careers require an unremitting command of its eccentricities, it is no less effective a gatekeeper. To ask a biologist in Cambodia to write in English as a requirement of publishing in top-tier journals, receiving major grants, or being appointed to a tenure-track position represents the same largely arbitrary handicap that would be readily apparent were we to still require a computer scientist in Chicago to present her work in Latin.

This point has implications well beyond academe itself. Phillipson (2015) quoted Einstein's conviction that 'it is of major importance that the general public is given the opportunity to be made aware of … scientific research … Limiting discoveries to a narrow circle kills the philosophical genius of a people and leads to intellectual impoverishment' (p. 30). Even if it were assumed (and it would be an unjustified assumption) that mastery of English is a reasonable expectation of every person educated enough to hold a terminal degree, there would

still be an incalculable opportunity cost in holding the fruits of their labours aloof from the broader populations they are meant to serve. The otherwise rapidly eroding barriers between institutional and amateur scholars in citizen science initiatives, digital humanities crowdsourcing, and similar interactions can only magnify our sense of loss in reflecting on those still excluded.

In considering alternatives to this state of affairs, it is assumed here that having a standard *lingua franca* of scholarship is desirable. Even were it to be regarded as undesirable, however, it might still be seen as unavoidable insofar as efforts to increase multilingualism within academic settings have been shown to reinforce, rather than challenge, the role of English as a mediator (Björkman, 2013; Mortenson, 2014). The pertinent question, then, is not *whether* a standard language will dominate global scholarship, but *which* language might best serve the interests of the world's scholarly community and broader publics in doing so.

This is not a new question. In the late 19th century, when the three leading languages of science—French, English, and German—had no clear leader, multiple proposals were advanced for invented "auxiliary" languages whose simplicity and regularity of grammar would render them easy for researchers everywhere to learn. The most successful of these today was Esperanto, spoken still by over a million people and used for the past thirty years as the primary working language of the International Academy of Sciences in San Marino. Starting with a 1922 report for the League of Nations (International Language Commission, 1995, p. 7), repeated studies have demonstrated that students acquire high-level proficiencies in Esperanto in a fifth or less of the time required for any major European language (Nixon, 1953; Ruckmick, 1924).

Despite the repeated advice of committees and task forces to the League of Nations, the United Nations, and many national governments, however, no systematic international program of Esperanto education has ever been put into effect, owing in large measure to the diplomatic efforts of first the French and then the British and American governments to preserve the privileges of their respective languages by blocking action on the proposals (Garvía, 2015, pp. 161–4). Absent a coordinated, multilateral program of implementation, Esperanto's main limitation became readily apparent. While the language is easy to learn, it is unintelligible to anyone who has not previously studied it, and relatively few were willing to put in even the small effort to learn it in absence of guarantees that others would do the same. The Japanese meteorologist Wasaburo Ooishi was the first to discover the jet stream, but recognition of his findings was delayed by decades because he was one of the few bold enough to commit to publishing in the 'international language' without such reciprocal assurances (Maksel, 2018).

With this in mind, the International Auxiliary Language Association (IALA) published a new language—Interlingua—in 1951, using a method that derived a simplified, regularized average from the Romance languages and English, in order that the result would be not only easy to learn, but easy to understand if one had not already learned it. The source languages were chosen for proximity to the Greco-Latin scientific and technical vocabulary that had already been effectively internationalized, as well as for the broad colonial distribution of those languages which, for better or worse, had made them widely recognized throughout the world. It was hoped that this would open access to Interlingua, and to the scholarship produced in it, to the widest possible audience. Abstracts for scientific papers were the first practical application; for more than thirty years two dozen medical journals issued Interlingua abstracts and eleven international medical conferences published Interlingua summaries (Esterhill, 2002, pp. 26–7).

Abstracts and summaries proved to indeed be quite readable for those with no prior training, but the effect was found to work in reverse also. In one study at Varberg High School in Sweden, students who had studied only Interlingua successfully translated 95% of an article from the Italian magazine *Oggi*—a rate of comprehension higher than that achieved by students working on an article from *Der Spiegel* after having

studied German (a language more closely related to Swedish) for more than five years (Breinstrup, 2008, p. 3). The present author has suggested that Interlingua could thus be useful in the professional development of librarians requiring a general reading knowledge of Romance languages (MoChridhe, 2018).

Such targeted uses hold promise and deserve further exploration, but the question remains why, given the potential utility of a universal language of scholarly exchange, neither Esperanto nor Interlingua established itself, and of whether renewed efforts for these or any future project could succeed. The factors involved are many and complex, but Garvía (2015), in a review of major auxiliary language projects, suggested that they can be understood as following the same patterns that govern adoption of new technical standards (p. 153ff.). In such cases, the speed and scale at which the new standard can be disseminated and "consumer" exposure intensified are crucial to its success. It is here that circumstances have changed most notably since the past experiments with both Esperanto and Interlingua were conducted, and here that librarians have the greatest opportunity to contribute to the implementation of a new 'linguistic standard' for scholarly data.

At the time that Esperanto was used in scholarly journals and that Interlingua was piloted for abstracts, all translation was done manually. Alexander Gode, the director of the IALA, personally produced several million words to support the Interlingua trials (Esterhill, 2000, p. 19), and his death in 1970 led to a drastic reduction in capacity, as did the deaths of other prominent Interlinguists thereafter, leading to a bottleneck in the availability of content that forced many of the established pilot projects to be discontinued. Modern computer-aided translation allows a small number of people to process vastly more material than was previously possible, enhancing the availability of translations and reducing the cost of their production severalfold.

Similarly, previous trials of auxiliary languages were conducted before the advent of desktop publishing and digital distribution, when the cost of each line of text was significantly higher. Waltham (2005) estimated that a learned society publishing a print journal now incurs costs of around $710 per page, while an online journal costs only about $128 per page. Even the present cost of a print page, however, is greatly reduced by digital publishing tools compared to its twentieth-century levels. Under current conditions, in which an auxiliary language has not yet been widely adopted, the additional costs of printing even an abstract are not a good value proposition for a publisher. Digital publishing, however, reduces the loss incurred and thus makes implementation more manageable during the necessary transitional period in which the academic community is being familiarized with internationalizing content in this way. As that process progresses and more scholars become familiar with the chosen language (even if only as passive readers), translations would come to be seen as a value added and would eventually become revenue-positive, much like indexing or similar services.

To the extent that prior initiatives in using Esperanto and Interlingua for scholarly communication—otherwise well-received—were hobbled by the limited volume of translation work possible to small coteries of committed advocates and by the high cost of including that work in publications, contemporary machine translation and publishing technology clear a path to broader auxiliary language deployment than was formerly possible. Additionally, each step of any new initiative in this domain would benefit from far larger-scale and more systematic data collection and analytics than were previously available, allowing efficacy and return on investment to be determined at each milestone with much greater levels of confidence than past trials were able to achieve.

Plan S is now an experiment to see whether consortial efforts on the part of funders, researchers, and their institutions can force short-term losses within the scholarly communication industry in order to secure long-term gains. To the extent that it succeeds, it could offer a framework by which that industry could be opened beyond the horizons of licensing and copyrights. Future OA standards could require content to

be made available in an auxiliary language that can be more quickly, cheaply, and thoroughly learned than English, and perhaps even more readily understood by those who have not yet learned it. Interlingua, designed for this purpose from its inception, would be an obvious choice, but Esperanto also could offer significant advantages. An entirely new project could be developed. The precise nature of an OA linguistic standard is less important than that such a standard emerge and that it lower the cost of access to the international language of scholarship—the 'hidden paywall' that presently keeps so many out of the walled garden of scholarly prestige and cutting-edge conversation.

A role for libraries

Such a language's introduction could proceed in modest phases, beginning with requirements only for abstracting and metadata. As the global research, publishing, and library communities became more familiar with the emerging standard, this could be expanded to a requirement to accept new research submissions in the language. Only after adoption and use in this way had become widespread would requirements to provide full-text translations of submissions in other languages have to be considered, and by that time the transition from monolingual English to bilingual journals would likely begin to occur under its own economic logic.

Academic libraries constitute the natural leadership for such programs. They are already taking up an increasing role in scholarly publishing as metadata creators, providers of open educational resources, managers of institutional repositories, hosts to OA journals, and even homes to OA monograph presses. It seems likely that the publications and datasets in which trials of an auxiliary language would be both beneficial and practicable will increasingly be library products. In addition, the consortial models common in library publishing (such as that used by by the Open Library of Humanities) and the culture that surrounds them are more likely to be receptive to, and willing to sustain the necessary loss-leading for, such an experimental proposal than profit-driven publishers will be.

Beyond this, however, is the fact that librarians themselves are likely to be among the professionals who would find the most day-to-day use for such a language, owing to the large scale of increasingly international content they are required to process. It has already been suggested that Interlingua could function as a short-cut to developing reading knowledge across the major Romance languages, and its study thereby promises benefits to librarians even absent any wider introduction of the language. As the primary handlers of scholarly communication by volume, however, librarians would rapidly develop the most exposure to any auxiliary language if trials were to be undertaken at significant scale, which would prime them as the readiest source of expertise for supporting new initiatives and enhancing further growth—efforts that would dovetail naturally with Code4Lib, Wikipedia 'editathons', and a wide variety of other library-driven endeavours that promote outreach through democratization of knowledge creation and access to information. During the initial phases of introduction, in which some degree of 'literacy brokering' would still be required as an aid to the transition, the presence of this knowledge base among librarians as a profession would also cement libraries' role as content producers rather than mere handlers. In short, the implementation would create a window of opportunity to disrupt the scholarly communication industry in such a way as to re-form it not only in a more open and equitable configuration, but one that will place libraries more firmly at its centre.

Conclusion

So long as English remains the *de facto* language of global research and scholarship, the high costs in time and money that it imposes on non-native learners will inhibit information exchange, handicap academic careers, and impede public participation in research, especially in those regions where there is the greatest need for the opportunities that engagement with global scholarship promises and where articulate voices are most acutely missing from broader global conversations. As digital communication technologies penetrate deeper into the most impoverished and undereducated parts of the planet—and as digital humanities crowdsourcing and citizen science recruit increasing numbers of extra-, non-, and alt-academics into the research process—interlinguistic solutions such as those considered in this column, used either as a substitute for or as a complement to existing language policies in scholarly publishing, can harness the energy and insights of millions who today are silenced. The (re)introduction of a standard language of global scholarship that requires markedly lower investments to master than does English promises to open worlds both for the scholars who will be empowered to share their work through it and for those who will hear such scholars for the first time not as broken, halting voices, but as peers of equal strength and clarity.

References

Alyamu, H. Y., & Assiri, E. A. (2018). Invisible web and academic research: A partnership for quality. *International Education Studies, 11*(4), 84–91.

Belcher, W.. How to hire and work with an academic copyeditor. Wendybelcher.com. (2013). https://wendybelcher.com/writing-advice/how-to-hire-copyeditor/ (n.d.).

Björkman, B. (2013). *English as an academic lingua franca.* Berlin/New York: De Gruyter Mouton.

Breinstrup, T. (2008). Communication without borders. *Language at Work–Bridging Theory and Practice. 3(4)*https://doi.org/10.7146/law.v3i4.6238.

Choi, Y. (2015). The effects of English training abroad on labor market outcomes in Korea. *Research in Social Stratification and Mobility, 41*, 11–24.

Choudaha, R., Chang, L., & Kono, Y. (2013). International student mobility trends 2013: Towards responsive recruitment strategies. *World education news + reviews. 26(2)*.

Cronquist, K., & Fiszbein, A. (2017). *English language Learning in Latin America.* Pearson.

Endow, T. (2018, 26 June). Inferior outcomes: Learning in low-cost English-medium private schools—a survey in Delhi and National Capital Region. *Indian Journal of Human Development.* https://doi.org/10.1177/0973703018779725.

Esterhill, F. J. (2000). *Interlingua institute: A brief history.* New York: Interlingua Institute.

Esterhill, F. J. (2002). Interlanguage and 20th century scientific communication. *Journal of Universal Language, 3*(1), 15–31.

Garvía, R. (2015). *Esperanto and its rivals: The struggle for an international language.* Philadelphia: University of Pennsylvania Press.

Huttner-Koros, A. (2015, 21 August). The hidden bias of science's universal language. The Atlantic. https://www.theatlantic.com/science/archive/2015/08/english-universal-language-science-research/400919/.

International Language (known as Esperanto) Commission (1995). Study prepared for the Italian Ministry of Public Instruction. *Official Bulletin of the Italian Ministry of Public Instruction. 21–2. Official Bulletin of the Italian Ministry of Public Instruction* (pp. 7–43).

Jaygbay, J. (1998). Self-censorship in African scholarship and scholarly publishing. *Journal of Scholarly Publishing, 29*(2), 112–117.

Li, D. C. S. (2013). Linguistic hegemony or linguistic capital? Internationalization and English medium instruction at the Chinese University of Hong Kong. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.). *English-medium instruction at universities: Global challenges* (pp. 65–83). Bristol/Buffalo/Toronto: Multilingual Matters.

Lillis, T., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication, 23*(1), 3–35.

Lillis, T., & Curry, M. J. (2010). Academic research networks: Accessing resources for English-medium publishing. *English for Specific Purposes, 29*, 281–295.

Linn, A. (2015). Series preface. In S. Dimova, A. K. Hultgren, & C. Jensen (Eds.). *English-medium instruction in European higher education* (pp. vii–viii). Boston/Berlin: De Gruyter Mouton.

Maksel, R. (2018, April). Why was the discovery of the jet stream mostly ignored? Air & space magazine. https://www.airspacemag.com/as-next/as-next-may-unbelievablebuttrue-180968355/.

MoChridhe, R. (2018, September 25). How to cheat on the language skills section of your library résumé. Hack library school. https://hacklibraryschool.com/2018/09/25/cheat-on-language-skills/.

Mortenson, J. (2014). Language policy from below: Language choice in student project groups in a multilingual university setting. *Journal of Multilingual and Multicultural Development, 35*(4), 425–442.

Nixon, V. C. (1953). Lastatempaj eksperimentoj pri Esperanto en lernejoj. *Esperanto, 2*(60), 56–58.

Pearce, C. (2003). *Editing an African scholarly journal. 16(1),* Learned Publishing54–60. https://doi.org/10.1087/095315103320995096.

Phillipson, R. (2015). English as threat or opportunity in European higher education. In S. Dimova, A. K. Hultgren, & C. Jensen (Eds.). *English-medium instruction in European higher education* (pp. 19–42). Boston/Berlin: De Gruyter Mouton.

Pieper, D. (2018, 25 October). OpenAPC: Transparent reporting on article processing charges reveals the relative costs of open access publishing. LSE impact blog. http://blogs.lse.ac.uk/impactofsocialsciences/2018/10/25/openapc-transparent-reporting-on-article-processing-charges-reveals-the-costs-of-open-access-publishing.

Ruckmick, C. A. (1924). The Wellesley College Danish-Esperanto experiment. *International language* (pp. 51–56). .

SCImago (2017). SCImago journal rankings. https://www.scimagojr.com/journalrank.php?year=2017.

Sugiharto, S. (2017, 30 October). The struggle for linguistic equity. The Jakarta Post. http://www.thejakartapost.com/life/2017/10/30/essay-the-struggle-for-linguistic-equity.html.

Van Weijen, D. (2012, November). The language of (future) scientific communication. Research trends 31. https://www.researchtrends.com/issue-31-november-2012/the-language-of-future-scientific-communication/.

Waltham, M. (2005). *Learned society open access business models.* Princeton, NJ: JISC. http://www.marywaltham.com/JISCReport.pdf.