The Illusion of Diversity on Citizen Science Platforms: Why Linguistic Representation and Translation Matter for STEM[1]

by Renée Desjardins and Danielle Pahud

Renée Desjardins, Ph.D. <rdesjardins@ustboniface.ca> Associate Professor, School of Translation, Université de Saint-Boniface, Winnipeg, Manitoba

Danielle Pahud, Ph.D.c. <Danielle.Pahud@umanitoba.ca> Instructor I and Director of the Lockhart Planetarium, University of Manitoba, Winnipeg, Manitoba

The last decade has seen mobile technologies proliferate and the popularity of online participatory culture and online social media rise. In addition, crowdsourcing has become a default model for conducting large-scale projects in different contexts, including in academic research. The Search for ExtraTerrestrial Intelligence (SETI and SETI@home[2]) was a groundbreaking citizen science project and a well-known example of such a project within the citizen science and astronomy community. In many ways, opening research to the general public, particularly in online spaces, has led to a democratization of knowledge creation and dissemination, creating a sense that digital contexts are inherently more inclusive and diverse than traditional laboratories and classrooms. Indeed, the Tri-Agency statement on Equity, Diversity, and Inclusion (EDI) promotes increasing inclusivity in the research system[3]. While crowdsourcing research nominally meets this criterion, the implementation of projects and the degree to which these succeed in improving EDI, warrants nuanced reflection.

Early entrepreneurs of the social web-individuals like Facebook's Mark Zuckerberg or Reddit's Steve Huffman were optimistic that their platforms could remain neutral, and function without the gate-keeping mechanisms of 'traditional' media. However, many of these entrepreneurs had training in the areas of computer science or business. Few had formal training in disciplines in the social sciences and humanities (SSH), such as ethics, philosophy, or linguistics. Insights from these fields in the early days of social media, and arguably the early days of the Web, could have curtailed a number of issues we now face, such as: content moderation, technological surveillance, and techno-capitalism, which are often associated with contemporary social media giants. So, while the Web started as a 'uncolonized' space, over the years, intentionally or not, digital 'settlers' contributed to the moulding of this digital space. This has had an impact on access, accessibility, diversity, inclusion, and equity [1-3].

Our work is primarily concerned with investigating how scientific knowledge circulates in online digital spaces, using Translation Studies (TS) as one of our theoretical and methodological lenses. We contend that the web and most contemporary social media are inherently Anglocentric, which in turn has had some deleterious effects with regard to plurivocality [4] in the broadest sense. Bowker and Ciro [5] state that "English has emerged as the international language of scholarly communication–particularly in the domains of science and technology, despite the fact that only roughly 6% of the world's population speaks English as a native language." English is the language of the Internet and the *lingua franca* of the Web [6]. In TS, investigation of the asymmetrical 'flow' of knowledge creation and dissemination and the social and cultural implications this has is not an uncommon research strand; however, this body of research is seldom mobilized or even considered in the STEM fields. Succinctly: translation flow analyses (*i.e.,* the direction in which translated content/knowledge circulates) in our case studies point to the asymmetrical exchange of scientific knowledge and cultural capitals. Related research reveals that a proficiency in English generally facilitates increased access to scientific literature and a wider range of analytical, as well as technological, tools (*e.g.*, textbooks, software, apps) [7-10]. Our work started from a concerted interest in addressing what we considered disciplinary and epistemological silos between STEM and the SSH. More specifically, in our

Summary

This essay submission examines linguistic representation on citizen science online platforms, with specific attention given to Zooniverse and the Canadian Citizen Science Portal.

1. This research is supported in part by funding from the Social Sciences and Humanities Research Council of of Canada.
2. https://setiathome.berkeley.edu/.
3. https://www.nserc-crsng.gc.ca/NSERC-CRSNG/EDI-EDI/index_eng.asp.

SSHRC[4]-funded research, we sought to examine a particular instance of where social media, crowdsourcing, STEM, and linguistic (in)justice intersect: online citizen science. In doing so, we hope to inform best practices in online citizen science project development and deployment.

Citizen science is defined as "a form of research collaboration involving members of the public in scientific research problems to address real-world problems" [11]. Fundamentally, citizen science is premised upon all lay citizens being able to contribute meaningfully to the scientific process. As technology has evolved, many of the STEM disciplines have turned to the general public's interest in and willingness to participate in larger-scale scientific projects. In the research that our team collected and reviewed, it appeared that English citizen science literature implicitly assumed English proficiency among online users, failing to problematize this assumption explicitly: let us recall that only 6% of the world's population reports English to be their native language. Said differently, for 94% of the world's population, any interaction requiring English means varying levels of (self-, automated, and/or external) translation and interpretation. While digital and mobile technologies can help dismantle hierarchies and remove barriers to access, thus amplifying marginalized voices within citizen science work in STEM, we noted that explicit strategies to foster increased *linguistic* diversity related to citizen science initiatives remain limited. In short: if citizen science projects are conceptualized, programmed, and deployed using only English, does this truly mean all laypeople can contribute equally? And if not, what do we stand to lose from epistemological, empirical, and qualitative perspectives?

To address these questions, we focused our analysis on two citizen science social platforms (for more on how citizen science platforms are considered social and participatory media, see [10]): Zooniverse and Canada's Citizen Science Portal. We selected Zooniverse as it is ostensibly the most popular[5] citizen science platform, hosting both SSH and STEM projects, and Canada's Citizen Science Portal afforded readily accessible Canadian content. By examining more than one platform, we were also able to conduct comparative analyses.

We monitored Zooniverse on a monthly-basis over a two-year period (May 2018-July 2020), tracking active, paused, and finished citizen science projects across the disciplinary spectrum (Zooniverse hosts both STEM and SSH projects) and analyzing, when applicable, translation flows. We established a list of criteria that would help us identify translated projects and multilingual project features, including multilingual menus/buttons/tabs, embedded machine translation features, bilingual or multilingual prompts, research team profiles *(e.g.,* bilingual or multilingual project coordinators), external social accounts on platforms like Facebook or Twitter that published in more than one language, etc. Our research assistants, Racky Diallo (Université de Saint-Boniface) and Neil Doerksen (University of Manitoba) compiled preliminary data and used Python[6] to generate data visualizations. Our initial hypothesis was that inherent Anglocentrism underpinned Zooniverse despite some of the efforts it claimed to have put in place to promote linguistic diversity. While project builder templates enable translation, our data shows that only a limited number of projects were available multilingually over the two-year study period. Moreover, we noted that in a majority of cases, English was the point-of-entry language and default project ecosystem language throughout Zooniverse. However, data from 2020 (January to July) suggests that linguistic diversity and translated projects are on the rise: in January 2020, 3.8% (9/232) of Zooniverse's project catalog was available multilingually or had translation features. In July 2020, this percentage increased to 8.9% (24/269). Our data also showed that the number of languages into which projects were translated (which ranged between 10-15 different languages on a monthly basis throughout 2018-2020) tended to be languages from the Global North, including Spanish, French, Dutch, and Portuguese among others; *i.e.,* languages considered to be central (see [12] for terminology related to the positioning of languages according to a gravitational model). That said, our analyses indicate that July 2020 had an uptick in language diversity, with Kannada being one example of a language outside the Global North. Our Zooniverse analysis did not indicate the use of or translation into/out of any Indigenous languages from North America. This is worth pause: when we consider that some of the land on which citizen science projects are conducted and the histories some projects invoke, this lack of linguistic representation has symbolic importance, and, in some cases real-world effects.

One of the major differences between the projects that appear on Canada's Citizen Science Portal, in comparison with Zooniverse, is that a relative degree of translation is often a mandatory project and platform feature — mandatory from the standpoint of research funding, but also from the standpoint of public sector communication in Canada.[7] This means that at a minimum, all active project descriptions (28 as of July 2020) on the portal main page are available in the country's two official languages (English; French), even if some of the project ecosystems are in fact unilingual. This would indicate that an official language policy and an emphasis on translation/multilingual features does influence linguistic diversity in knowledge creation and dissemination, particularly compared to sites like Zooniverse that do not have such explicit policies in place. It is worth noting that it would be difficult to enforce a language policy given Zooniverse is a transnational platform, though amendments to community guidelines that more explicitly address linguistic

4. SSHRC refers to the Social Sciences and Humanities Research Council of Canada.
5 "The Zooniverse is the world's largest and most popular platform for people-powered research." https://www.zooniverse.org/about.
6. Python version 3.6. https://www.python.org/downloads/release/python-360/.
7. For more on Canada's Official Languages Act: https://laws-lois.justice.gc.ca/eng/acts/o-3.01/.

diversity and justice could be envisioned. As with Zooniverse, we note an absence of Indigenous language representation on the entirety of the Portal catalogue at the time of writing. Given Canada's commitment to reconciliation and the fact that these projects are carried out in traditional Indigenous territories, Indigenous language representation is essential in Canadian citizen science. This would likely encourage the inclusion of Indigenous epistemologies, frameworks, and methodologies and promote language revitalization[8]. In this vein, we do not wish to speak in place of Indigenous scientists, citizen scientists and researchers, but we align ourselves with Indigenous scholars who have made the call for decolonizing and Indigenizing education and academic research (see [13]) and suggest the relative absence of these languages should be further scrutinized. For instance, we may ask whether barriers to representation lay within backend programming (an argument Instagram used when it addressed the late programming of right-to-left languages, for example), or a lack of outreach and relational connections with specific communities.

The scope of this submission does not allow us to provide in-depth analysis of each aspect of our two-year project. However, we feel it is worth sharing with the physics, citizen science, and larger STEM communities that there is epistemological value in considering translation, multilingualism, and linguistic representation in the *conception*[9] and deployment of citizen science initiatives. Our work answers the call made within the citizen science community for implementing practices that promote and sustain EDI (see [14]). Our work shows that the discourse on the supposedly democratic nature of online citizen science rarely addresses the fact that to participate, however minimally, one must first have *access* to technology and possess baseline digital literacy (to say nothing of other relevant literacies). The lack of critical reflection in relation to the digital divide in citizen science is problematic and further exacerbated when we factor issues related to linguistic representation and justice. The fact that not all citizens can contribute equally to citizen science initiatives further enforces hierarchies in scientific inquiry, promoting exclusionary rather than inclusionary frameworks. We argue that if citizen science platforms start from a more humanistic approach, rather than focusing primarily on automation, expediency and scalability of scientific discovery, in addition to low-cost labour (it is worth recalling citizens are not necessarily financially or symbolically remunerated for their contributions), then citizen science platforms would likely engage individuals beyond traditional epistemologies, beyond dominant language hierarchies, and would address more issues related to digital access.

8. Given that our team comprises non-Indigenous members, we do not feel it is our place to suggest which Indigenous languages should be included or how Indigenous representation might be best conducted on these platforms. We can work in a collaborative fashion to suggest some of the translation tools and strategies that can be deployed to foster increased multilingualism, however, and hope such collaborative initiatives will take place in the future. Ultimately, our project's data point to a gap and a lack of representation, and our hope is that Indigenous consultation and input would be a necessary next step.

9. The emphasis on conception is to underscore the fact that translation and multilingualism are at times thought of in the latter stages of a project or simply to meet funding criteria. In the latter case, translation then becomes a sort of 'graft' instead of an integral part of the project's conceptualization from start to finish. Without citing every issue this causes, it can mean hastily produced translations or unwarranted recourse to machine translation (which can pose other challenges inherent to algorithms and training data).

REFERENCES

1. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Broadway Books, 2017.
2. Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: New York University Press, 2018.
3. D'Ignazio, C. and Klein, L.F. *Data Feminism.* Cambridge: MIT Press, 2020.
4. Nappi, C. "The Global and Beyond: Adventures in Local Historiographies of Science." *Isis*, **104**, 102-110 (2013).
5. Bowker L. and Ciro, J.B. *Machine translation and global research : towards improved machine translation literacy in the scholarly community.* United Kingdom: Emerald Publishing Limited, 2019.
6. Mahony, S. "Cultural Diversity and the Digital Humanities." *Fudan: Journal of the Humanities and Social Sciences,* **11**, 371-388 (2018).
7. Brisset, A. "Frontières de la traduction". *ÉcArts d'identité*, **2**, 30-37 (2008).
8. Buzelin, H. "Translating the American textbook". *Translation Studies*, **7**, 315-334 (2014).
9. McDonough, J. Dolmaya, Expanding the sum of all human knowledge: Wikipedia, translation and linguistic justice. *The Translator,* **23**, 143-157 (2017).
10. Desjardins, R. Are Citizen science 'Socials' Multilingual?: lessons in (non)translation from *Zooniverse,* in Desjardins, R., Larsonneur, C., Lacour, P. (Eds). *When Translation Goes Digital: Case Studies and Critical Reflections.* London: Palgrave Macmillan. (Forthcoming 2020).
11. Cohn, J.P. "Citizen science: Can Volunteers Do Real Research?" *BioScience,* **58**, 192-197 (2008).
12. Calvet, L.-J. *Pour une écologie des langues du monde.* Paris: Plon. 1999.
13. Cote-Meek S. and T. Moeke-Pickering, (Eds). *Decolonizing and Indigenizing Education in Canada.* Toronto: Canadian Scholars, 2020.
14. Sturm, U. *et al*. "Defining principles for mobile apps and platform development in citizen science." *Research Ideas and Outcomes,* **4**: e23394 (2018).