

The submission reviewed here is “Opening open science to all: Demystifying reproducibility and transparency practices in linguistic research,” submitted to the journal *Linguistics*. The article is both a plea and a set of recommended practices for open science practices in linguistics. The introduction is an overview of the concept of open science, the scandals in psychology in the 2010s, and the perceived lack of attention to open science in linguistics. This section also introduces eight steps that can be taken by linguists in the interest of promoting open research practices. These are: positionality statements, declarations of COI, open data, reproducible code, literate programming, preregistration, registered reports, and pre-prints. Each of these steps is a further section of the paper. In each of these sections, the author presents the idea with a bit of background, and describes how linguists might undertake each step. The paper concludes with a reference to an online series of tutorials known as FOSIL (for “foster open science in linguistics”), ostensibly a project developed by the authors of this paper.

Let me begin by first clarifying my own positionality with regard to the subject at hand: I am an enthusiastic supporter of open science practices in linguistics, and have been for many years. So in this sense, I am solidly in favor of more exhortations in print to our colleagues to improve their practices, and recognize that many linguists are not trained in these practices and left to more or less fend for themselves. This is a sad state of affairs for our field and others, and needs to be corrected through outreach and education.

This paper, however, seems to me to be quite weak on two fronts. The first is what seems to be a concerning lack of awareness of recent work in linguistics to promote open science, especially with regard to research data. Of particular note is the absence of two fairly recent references in linguistics. The first is a highly-cited, multi-authored position paper calling for reproducible linguistics research that was published *in this very journal* in 2018:

Berez-Kroeker, Andrea L., Lauren Gawne, Susan Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard P. Meier, Nicholas Thieberger, Keren Rice & Anthony Woodbury. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56(1): 1–18.  
<https://doi.org/10.1515/ling-2017-0032>.

The second is an open access handbook from the very prominent MIT Press Open, with 56 chapters on open data management from more than 100 international authors across many subfields (which I refer to later in this review as OHLDM):

Berez-Kroeker, Andrea L., Bradley McDonnell, Eve Koller & Lauren B. Collister (eds.). 2022. The open handbook of linguistic data management. Cambridge, MA: MIT Press Open.  
<https://doi.org/10.7551/mitpress/12200.001.0001>.

And then of course there is the 2021 statement from the Linguistic Society of America on the Scholarly Merit of Open Scholarship in Linguistics:

<https://www.linguisticsociety.org/content/statement-scholarly-merit-and-evaluation-open-scholarship-linguistics>.

It seems to me that the authors have not done their due diligence in terms of understanding the state of literature in the field, especially in light of their claims that there is no such literature.

That is the first major weakness of this paper. Misrepresenting the current state of the field makes it very easy for readers to dismiss the topic altogether. Papers urging the research culture of the entire field forward need to carefully and accurately describe the problem at hand, so that readers can recognize the problem for what it is. This in turn makes it harder for readers to ignore the problem, and encourages them to think carefully about how they can help solve it in their own work. My own review is a case in point: I don't recognize the world of linguistics as described in this paper -- that is, one in which no linguists have really thought seriously about this issue before today -- and I am writing a negative review despite the fact that I hold this topic dear.

The second weakness is that the eight steps outlined here are handled only superficially, and in some cases, in a way that can be potentially harmful. Each section could be a paper unto itself (and interesting special issue idea, perhaps?). Below I highlight a few of the major oversights in the sections.

*Positionality statements.* The section on positionality statements does not address the counterpoints that have been made regarding the extra burden that required statements put onto BIPOC authors. See, e.g.:

Massoud, Mark Fathi. 2022. The price of positionality: Assessing the benefits and burdens of self-identification in research methods. *Journal of Law & Society* 1-23.

Roberts, Steven O., et al. 2020. Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science* 15: <https://doi.org/10.1177/1745691620927709>.

*Open data.* The section on open data, while acknowledging that privacy concerns should be respected, does not mention any discussion of the concerns of Indigenous peoples regarding colonial practices in linguistics and especially around who may own, and therefore share, language data (see eg Holton, Leonard & Pulsifer, chapter 4 of OHLDM and that chapter's very helpful list of references).

*Preregistration and registered reports.* The section on registered reports does not take into consideration very real pressures of the systems of academic reward, which bias university researchers against any additional steps in the research process that increase the time to publication (that is, the authors overlook the need for systemic change in academia to encourage open science -- see Alperin et al, chapter 13 of OHLDM).

One final comment about the paper is the last-moment mention of the tutorials in the FOSIL repository. Its lone mention and its placement at the very end of the paper, in the final two sentences, is a bit awkward. It makes it seem like FOSIL is the only extant solution to the problems presented in the paper. Rather than a list of possible educational opportunities for readers to avail themselves of, I'm left wondering, are there simply no other tools to learn from? Yet I know that there are, all across the sciences and humanities, easily findable online or even through any university library.

On the other hand, it's also possible that the authors are using this paper as an introduction to and publicity for FOSIL (I'm guessing the authors of both are the same people), in which case, FOSIL should be mentioned at the outset of the paper, and the following sections of the paper should justify each of the tutorials in the repository, with particular reference to the structure and contents of each online tutorial. Then this paper might have a more understandable and unified rationale...although then I would suggest it should not be submitted to a journal, but rather included in the repository itself. Then the entire FOSIL project could be announced someplace like the LINGUIST List.