

“Well, parts of linguistics is open”: Insights into Linguists’ Diverse Understandings of Open Science

ELEN LE FOLL

Abstract: Broadly defined as the study of language, linguistics is a diverse field spanning many disciplines. Recent studies on the prevalence of Questionable Research Practices (QRPs) in linguistics (e.g., Isbell et al. 2022) suggest that it suffers from many of the same issues that triggered the replication crisis in psychology (see, e.g., Sönning and Werner 2021). While surveys have indicated that linguists are generally in favor of Open Science/Open Scholarship (OS), there appears to be a “a misalignment between the attitude to and the adoption of OS practices” (Liu and de Cat 2024, 64). The present study aims to gain insights into this misalignment by exploring linguists’ understanding of what constitutes OS and of the specificities of linguistic research that (can) affect its applicability to (subdisciplines of) linguistics. To this end, the study draws on the results of an anonymous, small-scale survey and the qualitative analysis of semi-structured interviews conducted with 26 linguists based in Northern Europe representing all career stages and a wide range of subdisciplines within linguistics.

The results reveal diverse understandings of OS among linguists. While some focus on the accessibility of research (for both academics and the wider public), others prioritize the sharing of data, materials, and code to promote transparency, reproducibility, and replicability. The latter group also emphasizes the importance of OS principles and values such as rigor, fairness, and collaboration. Linguists report learning about OS through conferences, workshops, library services, and social media but, most importantly, in personal interactions with other researchers, thus making much of this knowledge network dependent. The interviewees highlight several challenges and considerations that they believe need to be addressed when applying OS to linguistics. These include ethical and legal issues concerning data sharing, the high inter-person variability inherent to many linguistic studies, the need for (more) funding for open access monographs, and the need for training in data management and statistical methods.

Keywords: replication, reproducibility, open scholarship, open access linguistics, open research, open science

1 Introduction

Focusing on the multifaceted discipline of linguistics, this study seeks to gain insights into the apparent disconnect between Open Science/Open Scholarship (OS) practices and some of the epistemological and methodological traditions of the humanities (Knöchelmann 2019). While traditionally anchored in humanities faculties, linguistics—broadly defined here as the scientific study of language—is a particularly interesting case study in this endeavor as, decades after its alleged “quantitative turn” in the 1990s and early 2000s (Kortmann 2021; McGillivray and Jensen 2023), linguistic research continues to range from non-empirical research and studies based on introspective data to large-scale data-driven studies. This is in part due to the interdisciplinary nature of many subdisciplines of linguistics, which have been influenced to a greater or lesser degree by the practices of neighboring disciplines such as psychology in the case of psycholinguistics and computer science for computational linguistics. As such, linguistics can be considered to be at the crossroads between the humanities, the social sciences, and the “hard” biomedical sciences (Bochynska et al. 2023, 2).

While the Open Science movement has gained significant traction in psychology, it represents a relatively new development in linguistics (see, e.g., Casillas et al. 2023; Liu 2023; Plonsky 2024b; Sönning and Werner 2021). That said, the momentum for OS in linguistics has grown discernibly, as evidenced by a surge in publications, the formation of research networks, and the organization of symposia and conferences dedicated to OS (Liu 2023; Liu et al. 2023; Liu and de Cat 2024). However, the fact that much of the infrastructure and discourse surrounding OS originated in experimental psychology (see, e.g., Gelman 2016; Open Science Collaboration 2015) presents some challenges: The tools and resources are not always aligned with the priorities and methodologies of (applied) linguists (Liu 2023, 444).

A significant complicating factor is the diversity of epistemological and methodological traditions present in linguistics. For example, researchers employing constructivist or ethnographic approaches may find some OS practices, such as preregistration, less relevant or even incompatible with their work (Al-Hoorie et al. 2024, 14). Moreover, some linguists perceive OS to suffer from a “quantitative bias” (Liu and de Cat 2024, 94), a sentiment that appears to be shared by qualitative researchers in other disciplines (see, e.g., Prosser et al. 2024 for a recent survey of attitudes towards OS in qualitative management and organization studies). In fields such as language testing, privacy or copyright constraints can make data sharing difficult or unethical (Isbell and Kim 2023, 13; Chappelle and Ockey 2024), leading some to question the

universal applicability of OS practices to linguistics. These challenges highlight the need for ongoing dialogue within the discipline to contextualize OS in ways that respect its varied paradigms and research practices (Liu 2023, 448).

Different rationales are known to motivate the adoption of OS practices, including social justice, epistemic responsibility, inclusivity and diversity, and economic and personal gains. For example, in their article titled “(Why) Are Open Research Practices the Future for the Study of Language Learning?” Emma Marsden and Kara Morgan-Short (2023, 348) note that “[t]he respective roles and impacts of these different rationales themselves merit empirical scrutiny, perhaps through interviews and surveys to improve understanding of drivers and affective variables that might (better) underpin and shape the open research movement.”

The present study is an initial attempt at addressing this need. Its aims are two-fold. It seeks to gain insights first into linguists’ diverse understanding of what constitutes Open Science/Research/Scholarship (see section 2.1 on terminology) and second into the specificities of linguistics that (can) affect its applicability to (subdisciplines of) linguistics. To this end, a survey was sent out to the subscribers of the ReproducibiliTea in the HumaniTeas mailing list, most of whom are linguists (see section 3.1), and semi-structured interviews were conducted with 26 linguists representing a variety of subdisciplines and career stages (see section 3.2). The present study does not formulate any hypotheses. Instead, it addresses the following broad sets of research questions:

- RQ1.** What do linguists understand and encompass under the terms *Open Science*, *Open Research*, *Open Scholarship*, and *Open Education*?
- RQ2.** To what extent are linguists aware of OS practices? How did they reach this awareness? Where do linguists obtain knowledge about OS practices?
- RQ3.** What specificities of their (sub)discipline(s) do linguists feel need to be considered when applying and promoting OS practices in linguistics?

The terminological question was motivated by the literature (see section 2.1) and the call for papers for this special issue. The second set of questions was motivated by Scott Sterling’s (2024, 46) claim that “[i]t would not be much of a stretch to suggest that open science is being conducted by a small group of educated—and, to some extent, insulated—academics.” Finally, the third research question aims to bring answers to recurrent questions about the applicability of (all) OS practices to humanities and social science research (Ferguson et al. 2023; Knöchelmann 2019). It aims to make a contribution to Meng Liu and Cécile de Cat’s (2024, 90–91) call for “[o]pen discussions

bringing together [applied linguistics] researchers with diverse epistemological stances” to “collectively define optimal OS practices across research paradigms.”

2 Open Science in Linguistics: A Brief Literature Review

In the following, I provide a brief overview of previous literature that has attempted to shed light on OS practices in specific subdisciplines of linguistics, as well as more broadly across the field. I begin with a brief discussion of the diverse Open Science/Open Scholarship (OS) terminology used in linguistics.

2.1 Terminology

Some linguists, notably many involved in applied linguistics, prefer the term *Open Research* to *Open Science*, arguing that the latter “poses different challenges and affords different benefits for different types and approaches to research” (Marsden and Morgan-Short 2023, 374n1). Although their article is titled “Open Science: Considerations and Issues for TESOL Research,” Al-Hoorie et al. (2024) also criticize the term for being too narrow. However, rather than *Open Research*, they suggest *Open Scholarship* as a more inclusive term that covers a broader range of practices, including the creation and use of Open Educational Resources (OERs; see also Liu et al. 2023). This rebranding has been adopted by some research networks, such as Open Applied Linguistics (Al-Hoorie et al. 2024, 14). That said, a recent edited volume about open practices in applied linguistics also features *Open Science* in its title (Plonsky 2024b).

In sum, while linguists use different terms to refer to various principles and practices, it would appear that *Open Science* (with or without capitalization) remains the most widely used term. It has featured prominently in special issues (e.g., Sønning and Werner 2021; Kremmel and Isbell 2024) and across a wide variety of subdisciplines from historical corpus linguistics (Kesäniemi et al. 2018) to phonetics (Garellek et al. 2020) and computational linguistics (Rohatgi et al. 2023). Following Plonsky (2024b), I am adopting the acronym OS for Open Science/Open Scholarship in an attempt to be as inclusive as possible.

2.2 Sharing Data, Materials, and Code in Linguistics

In what may be the first substantial discussion of reproducibility in linguistics, Andrea Berez-Kroeker et al. (2018) emphasized the importance of data management and

sharing in language documentation and typological research. In this position paper, the 14 co-authors reported on the views of 41 linguists (mostly from North America, but representing diverse subfields of linguistics) who had convened to address “reproducibility as it applies to linguistic scientists, especially with regard to facilitating a culture of proper long-term care and citation of linguistic data sets” (Berez-Kroeker et al. 2018, 2).¹ They justify their focus on reproducibility by explaining how “true replicability is not possible to achieve” “in many fieldwork-based life and social sciences” (Berez-Kroeker et al. 2018, 5) and claim that reproducibility is thus a more realistic goal. The authors prioritize the transparency of both methods of data collection and analysis and the availability of the source data.

Two surveys of linguistic data citations suggest that, in both regards, much remains to be done in linguistics. Examining 100 descriptive grammars published between 2003 and 2012, Lauren Gawne et al. (2017) concluded that very few authors made their methods or data sources explicit. Berez-Kroeker et al. (2017) reached a similar conclusion in their analysis of 270 articles from nine prestigious linguistics journals. More recently, Agata Bochynska et al. (2023) examined the availability of materials, raw and processed data, and analysis scripts in two random samples of 250 linguistics journal articles published prior to the replication crisis (RC) being widely acknowledged (2008/2009) and post RC awareness (2018/2019). Figure 1 shows the proportion of articles with shared materials relative to the number of articles in which each characteristic was applicable for the two sampled periods. It clearly shows that, across linguistics, sharing data, code, and research materials remains very much the exception rather than the norm.

Bochynska et al. (2023, 25) acknowledge that practices vary across subdisciplines and journals, some of which have begun to adopt open data policies. That said, disappointingly low rates of reproducibility have been observed in some journals even post open data mandates (see, e.g., Laurinavichyute et al. 2022). Preliminary results from an ongoing survey of sharing practices in corpus linguistics research (Le Foll, forthcoming) suggests that, at least in some subdisciplines of linguistics, rates of materials, data, and code sharing have not dramatically increased since Bochynska et al.’s survey.

2.3 Publishing and Communicating Linguistics Research

A recent bibliometric study found that the humanities and social sciences, including linguistics, are dominated by journals operating hybrid models, with only a small percentage

1. Here, and throughout this article, *reproducibility* refers to the ability to obtain the same results using the original dataset(s) and method(s). This is in contrast to *replication*, which involves applying the same method(s) to novel data to answer the same research question(s) and/or test the same hypotheses (Turing Way Community 2022).

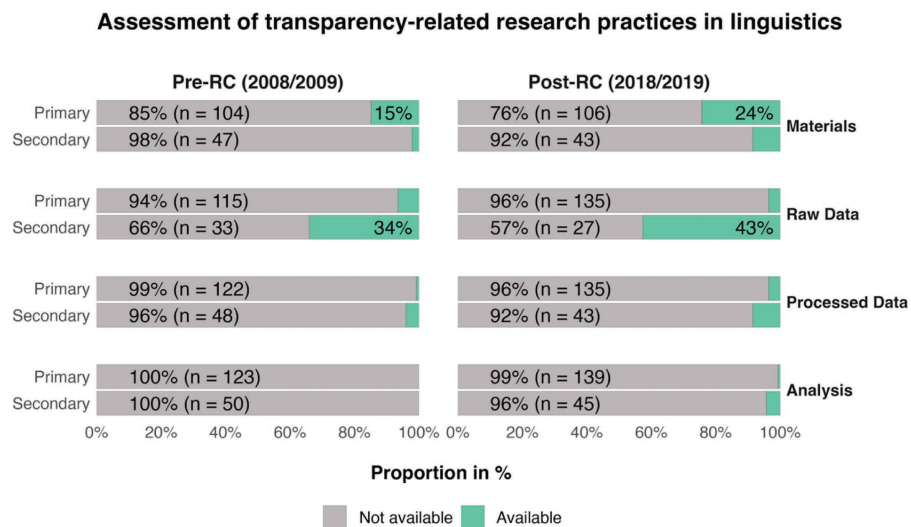


Figure 1. Percentages of the sampled empirical linguistics articles for which materials, raw data, processed data, and analysis scripts were found to be available, split into articles published for the pre-replication-crisis (left) and post-replication-crisis (right) (reproduction of figure 2 from Bochynska et al. 2023, 11, CC BY 4.0).

of articles published as gold (11.8%) or diamond (3.3%) open access (Butler et al. 2022). In the field of applied linguistics specifically, the rate of open access papers in hybrid journals was also found to be low (18% according to Alferink 2022, cited in Andringa et al. 2024, 5). To move towards more equitable and ethical forms of open access publishing, some linguists have been advocating for the establishment of new, non-commercial diamond open access journals or the “flipping” of existing journals from a commercial model to a non-profit one (see, e.g., Andringa et al. 2024; Rooryck 2023).

Another form of open access publication consists in publishing pre- and postprints of research outputs (also known as green open access). However, in most non-computational subdisciplines, the linguistics community appears to engage very little with preprints. A recent, large-scale study comparing peer-reviewed outputs with preprint availability across all disciplines concluded that, in 2023, just 1.5% of peer-reviewed publications in the field of “language, communication and culture” had been pre-printed (Rzayeva et al. 2025, 19). To raise awareness and encourage linguists to post their Author Accepted Manuscripts (AAM), a group of applied linguists launched the Postprint Pledge initiative (Al-Hoorie and Hiver 2023). As of October 2025, 132 researchers had signed the pledge.

Several recent initiatives seek to widen the accessibility of applied linguistics research outputs to beyond academia. These include OASIS (Alferink and Marsden 2023), TESOLgraphics (Sato et al. 2024), the TBLT Language Learning Task Bank (Gurzynski-Weiss et al. 2024), and a host of popular linguistics podcasts (see, e.g., Gawne 2024).

2.4 *Surveys of Linguists' Attitudes Towards OS*

I am aware of three surveys that have attempted to capture (applied) linguists' attitudes towards OS. An unpublished survey showed that, out of 326 survey respondents (59% of which were from the United States and United Kingdom), 90% claimed that openly available research materials and data were "very beneficial" (see Marsden 2019). In another survey, 354 researchers involved in second language (L2) research (from 45 different countries) described their replication practices and attitudes towards replication (McManus 2022). Just over half of the respondents reported having attempted to replicate an empirical study in the past; 65% of those who had not done so claimed that they wished to in the future. The results suggest that L2 researchers generally have positive attitudes towards replication, especially regarding its relevance and value (McManus 2022).

The third and to date most recent survey was conducted in early 2021 and attracted responses from 157 applied linguists across different career stages (Liu and de Cat 2024). It concluded that most respondents had positive attitudes towards reproducibility, with PhD students being the most positive. However, a significant portion of participants reported neutral or mixed attitudes towards data sharing. Despite this, all groups reported a high level of willingness to adopt more open practices, such as sharing data and code. PhD students also reported the highest self-efficacy in open practices. Notably, there was a significant gap between PhD students and more senior researchers in terms of code sharing, with the latter having extremely low rates of "(almost) always sharing R code" (Liu and de Cat 2024, 76–77). Less than half of the participants had an Open Science Framework (OSF) account, and a significant proportion of researchers in all groups were unaware of what OSF was, which the authors cautiously interpreted as a lack of awareness of existing OS infrastructures.

In all three surveys, self-selection bias is likely to have impacted the results. Liu and de Cat (2024, 64) explicitly acknowledge this: "We do not claim that the sample is representative of the AL [applied linguistics] community as a whole. Rather, our sample could be more appropriately characterised as composed of those who were sufficiently interested in OS at the time of the survey (e.g., with a certain level of prior knowledge and interest in OS) to take the time to respond."

In particular, it is worth noting that, in this most recent large-scale survey, none of the surveyed linguists reported using exclusively qualitative data and only a handful reported using qualitative data at all (Liu and de Cat 2024, 68). Thus, the views and perspectives of linguists working primarily with qualitative methods have not yet been adequately examined, even though insights from other disciplines suggest that they are likely to differ from those of researchers mostly invested in quantitative approaches (Gowie et al. 2024; Pownall 2025; Prosser et al. 2024; Salet et al. 2025).

Crucially, these surveys should be interpreted in light of (applied) linguists’ admitted involvement in research fraud and Questionable Research Practices (QRPs). In a large-scale survey with 351 respondents (Isbell et al. 2022; Plonsky et al. 2024), 17% of applied linguists working with quantitative methods admitted to one or more forms of fraud and 94% to one or more QRP (see also Farangi and Nejadghanbar 2024 on the prevalence and perceived severity of QRPs among Iranian applied linguists and Larsson et al. 2023 among US and Swedish humanities researchers). Taken together, these findings point to a potential “misalignment between the attitude to and the adoption of OS practices” (Liu and de Cat 2024, 64). This potential misalignment warrants further examination and motivated the present study.

3 Methods and Data

The present study draws on two sources of data. The first was collected via an online survey circulated among recipients of the newsletter of *ReproducibiliTea in the HumaniTeas*, an OS initiative that I co-organize with two other linguists at the University of Cologne, Germany. The second is the outcome of content-based qualitative analyses of semi-structured interviews that I conducted with 26 linguists between February and April 2025.

3.1 Survey Data

ReproducibiliTea is a global grassroots journal club initiative that provides forums “to discuss diverse issues, papers and ideas about improving science, reproducibility and the Open Science movement” (*ReproducibiliTea*, n.d.; cf. FitzGibbon et al. 2020). *ReproducibiliTea in the HumaniTeas* was launched in December 2023 as a dedicated forum to discuss these issues in the context of humanities and social science research specifically. We meet six to seven times per semester, both on-site at the University of Cologne and online via Zoom. Invited guest speakers lead discussions on topics as varied as reproducible workflows, FAIR (Findable, Accessible, Interoperable, and Reusable) data sharing, and research ethics and integrity. In addition, we organize workshops on tools such as Git, Quarto, and Docker. Most attendees are early career researchers (ECRs) from linguistics and neighboring disciplines. Participation is roughly equally split between on-site and online.

The survey was conducted online using LimeSurvey in February 2025. It was circulated via the *ReproducibiliTea in the HumaniTeas* mailing list, which, at the time, had around 120 subscribers. With just 38 participants (not all of whom completed all

questions), the response rate was relatively low. Just over half of respondents (55%) reported a current or previous affiliation with the University of Cologne. The remaining respondents were all affiliated with European institutions except one who was affiliated with a non-European institution and five who did not report an affiliation.

The online survey consisted of 18 questions (see Le Foll 2025b). The relevant questions for the present study concern respondents' reported awareness of and involvement with different OS practices (Q07), where they learned about OS outside of ReproducibiliTea sessions (Q08), and what they felt could be done in their (sub)discipline to increase the uptake of OS practices (Q09).

About a third of respondents were students (at undergraduate to graduate to post-doctoral levels). Five lecturers and three professors also completed the survey. The largest group of respondents were postdoctoral researchers ($n = 10$), who are typically also the most active participants at our meetings. Most respondents (70%) chose "linguistics" as their main field of research/studies and an additional 16% identified with neighboring disciplines such as "Romance languages," "literature," and "digital humanities." A wide range of subdisciplines was reported (see Le Foll 2025a).

3.2 Interview Data

A call to participate in the interview project was sent via the mailing list of ReproducibiliTea in the HumaniTeas on February 24, 2025. However, only four individuals volunteered as a result of this call. I subsequently sent personalized emails to a further 27 linguists of different career statuses and from a broad range of subdisciplines (six of which were also subscribers of the mailing list at the time of writing) with the aim of interviewing some 20 linguists. The positive response rate to these personalized emails was higher than expected and led to the recruitment of 22 additional interviewees.

I drafted interview questions (see Le Foll 2025b) to obtain more in-depth answers to the three research questions. However, the interviews were only semi-structured; the exact wording, order, and number of questions varied depending on the interviewees' responses. I conducted all 26 interviews online over Zoom from February to April 2025. The interviewees consented to the interviews being recorded and to anonymized transcriptions being published on an open repository. Most interviews lasted around half an hour (range: 13–52 min.). All were conducted in English except one (A01 in German).

The interviews were first transcribed with the help of a locally run instance of Whisper (Radford et al. 2022) and subsequently checked, corrected, and anonymized. To protect the identity of the interviewees, mentions of all concrete projects were anonymized as PROJECT. Other aspects that were anonymized include mentions of specific institutions (INSTITUTION), cities (CITY), countries/regions/languages (COUNTRY),

Table 1. Academic status of the 26 interviewees

Position	<i>n</i>
Postdoc	7
Associate professor	4
Doctoral researcher	3
Full professor	4
Lecturer	2
MA student	2
PhD no longer in academia	2
BA student	1
Librarian	1

and colleagues and supervisors (PERSON). I interviewed 15 linguists who identify as female and 11 as male. To protect the identity of the interviewees, I use the singular *they* to refer to all interviewees. Table 1 shows the distribution of the interviewees' roles in academia at the time of the interviews.

The two interviewees who are no longer in academia had very recently completed a PhD in linguistics and had decided to pursue a job outside of academia. Most interviewees are affiliated with a German institution or were up until they left academia ($n = 21$); two with a British university; and one each based in Belgium, Norway, and Sweden. Of the interviewees with a German affiliation, eight are affiliated with the University of Cologne. Nine other German institutions are represented. Three of the interviewees are currently affiliated with European institutions but come from or have spent many years teaching in countries of the so-called Global South. These experiences have shaped their relationship with OS, and this is reflected in the interviews.

As I expected some disciplinary differences, my first question asked the interviewees to situate themselves and their research within linguistics. Almost all mentioned several subdisciplines. In the transcripts, highly specialized subdisciplines and unusual combinations of subdisciplines that could potentially lead to the identification of the interviewees have been anonymized. Across all 26 interviews, the following subdisciplines were mentioned more than once: corpus linguistics (11), phonetics (7), applied linguistics (5), sociolinguistics (5), discourse analysis (4), language teaching (4), phonology (4), second language acquisition (4), language learning (3), psycholinguistics (3), theoretical linguistics (3), English linguistics (2), World Englishes (2), cognitive linguistics (2), computational linguistics (2), language documentation (2), pragmatics (2), teacher training (2), typology (2), and variational linguistics (2) (for full list, see Le Foll 2025a).

I acknowledge that the recruitment methods employed in this study inevitably come with some self-selection and researcher biases. That said, personally inviting individual linguists allowed me to reach out to interviewees of different career statuses (in particular, students) and to encourage linguists who felt that they did not know enough about OS to participate. Hearing the voices of those who do not (yet) personally associate

with the OS “movement” was important for this project. These individuals are unlikely to respond to an open call that mentions OS in the title. I also targeted a wide range of subdisciplines of linguistics. However, that the most frequently mentioned subdiscipline is corpus linguistics may be due to a disciplinary bias in my personal networks.

3.3 Analysis Methods

The survey responses were downloaded as a CSV file (see Le Foll 2025b), and descriptive statistics and data visualization were computed in R. To analyze the interview transcripts, a qualitative content analysis (QCA) approach was adopted. One of the strengths of QCA is the possibility to integrate prior knowledge into the analysis process (Kuckartz and Rädiker 2023, 31): The interview guide generated a set of a priori categories, which were refined and expanded through inductive category development based on the interview transcripts. This iterative process was conducted using OpenQDA (Belli et al. 2025). It involved multiple coding cycles during which the categories were refined.

The coding process aimed to structure, systematize, and interpret the data, rather than quantify it. Many passages were assigned to multiple categories. The final taxonomy of codes comprises 39 categories, of which nine correspond to the leading questions of the interview structure (see Le Foll 2025a), and 10 to OS practices (of which the most frequently mentioned were—in descending order of frequency—open data, open access, open code/methods, open education, and preregistration). The remaining codes are the result of more interpretative analyses; they concern both motivations to do OS and reasons not to get involved or resist OS, as well as challenges and concrete suggestions to improve the status quo.

The process of category development and coding was informed by my personal understanding of the topic, itself based on my experiences as an ECR linguist, linguistics lecturer, and OS advocate based in Germany. The aggregated metadata for all interviewees, anonymized, and a CSV file with all the coded passages and their corresponding categories are available on Zenodo (see Le Foll 2025a). While acknowledging that my qualitative analyses are not reproducible, it is hoped that this transparency will allow others to engage with the data from their own perspective(s).

4 Results and Discussion

The data collected as part of this study are extremely rich. In the following, however, I focus on the aspects relevant to the research questions formulated in the introduction. Wherever possible, I draw on both the survey and interview data.

4.1 On Terminology

One of the questions that I asked the interviewees concerned the use of the term *Open Science* for linguistics. I explained that the term *Open Research* was preferred by some linguists, whereas others preferred *Open Scholarship*, interpreted by some as being more inclusive and encompassing *Open Education* too (see section 2.1 and Le Foll 2025a). All interviewees reported not having given the terminology much thought, if any, prior to the interview. On reflection, however, the vast majority felt that *Open Science* was entirely suitable for linguistics—although some recognized the potential divide between what *is* and *is not* considered a science:

So it doesn't surprise me that that linguistics is not considered science for some people. That's something I hear a lot or have heard a lot. But I think it's not wrong to add linguistics or language research or language science into the field of science.

(A07)

Several native speakers of languages other than English mentioned that the distinction between the humanities and the sciences did not come naturally to them because, in their first language, both the terms for linguistics and humanities include the word *science* (e.g., in German, *Geisteswissenschaften* and *Sprachwissenschaft* include the word *Wissenschaft*). Thus, it appears that the terminological preoccupation summarized in section 2.1 may be largely an Anglo-Saxon one. In fact, some interviewees were somewhat offended at the thought of linguistics not being considered a science:

ihr [Linguistik] den Titel Science abzusprechen, finde ich ziemlich irrsinnig. Und nicht besonders fair [to deny it [linguistics] the title of science, I find quite absurd. And not particularly fair].

(A01)

While the majority *did* consider linguistics a humanity and anchored their own research in the humanities, others reported that linguistics, if anything, is more of a social science to them. Moreover, several linguists made clear that they did not consider such distinctions to be very meaningful:

I think linguistics is kind of unique, especially for somebody doing phonetics and phonology, because we really kind of have like one leg in the more scientific type of world and another one in the more humanities-like type of world. So I feel okay with either [Open Science and Open Research] personally.

(A12)

To be honest, I haven't given it much thought. I mean, for me, if we do want to use this distinction between humanities, the hard or natural sciences and perhaps social sciences. I mean, I do see linguistics kind of as language science. So I think it's a suitable term. I personally haven't come across open scholarship or open research. But I guess that sounds pretty interchangeable to me.

(A09)

Many considered *Open Research* and *Open Science* to be “synonyms” (A06), although they themselves used *Open Science*. Some interviewees justified this choice by saying that *Open Science* is simply more common, and one interviewee also mentioned that *Open Science* features in the name of the Open Science Foundation (OSF), a platform that was mentioned by many of the linguists who engage in OS practices. Although the term *Open Research* was largely acknowledged as suitable for linguistics, *Open Scholarship* was met with much more skepticism. Very few had come across the term before, but many of the interviewees spontaneously associated it with more narrow definitions than Open Science that focus on the researchers rather than the research:

Open scholarship sounds a bit too individualistic.

(A02)

[Open Scholarship] also entails maybe that you just really put yourself out there so that you do have something like a very detailed personal website.

(A04)

So scholarship for me is always something financial. So where you get something. So I'd be misled probably a little bit by this term.

(A25)

Others felt that the term was old-fashioned:

But for me, if I read open scholarship, it sounds even more, honestly, it sounds like a humanities thing [. . .] because being a scholar sounds a little bit antiquated. [. . .] To me, “I'm a scholar” that sounds like I would sit in libraries a lot and read and think really hard about the stuff that I do.

(A19)

Only few linguists were convinced by the idea of using *Open Scholarship* as an umbrella term that would include *Open Education*. The latter was also understood differently. Somewhat surprisingly, Open Education was occasionally exclusively associated with the ability to access educational materials in pre-tertiary education. However, some

interviewees primarily associated Open Education with research dissemination and transfer:

So I feel like open research, open science is more to do with research practice while open education also has to do with the teaching that we do informed by our own research practice. Although then the teaching that we do, is could be considered part of the dissemination activity. That is the endpoint of research in any case. So they're very closely related.

(A13)

One linguist also associated the term *Open Scholarship* with social justice and activism in a way that attempts to address global inequalities (A03).

4.2 On Linguists' Understanding of Open Science

Although the interviews revealed a broad consensus for the applicability of the term *Open Science* to linguistics, the interviewed linguists differed much more in what they understood as OS. During the analysis of their associations, two groups emerged: One set of linguists (primarily or exclusively) associates OS with the accessibility of research outputs, whereas the other focuses on sharing data, materials, and code for transparency, reproducibility, and/or replicability. To many in the first group, accessibility is associated not only with open access publication but also with science communication and transfer.

I would say the accessibility of results and, well, science to the general public. So to everyone who might think that this might be of use to their work, profession, or maybe also just out of interest.

(A14)

The second group of linguists associated OS not only with specific practices such as sharing (raw) data, analysis code, research materials, and preregistration but also frequently with principles and values such as transparency, sharing, rigor, collaboration, honesty, fairness, and democracy. In answer to my question about what they associated with OS, their responses included:

So open science kind of makes me think of *collaborative* research within a community of practice. So being able to *collaborate* with other colleagues who are willing to *share* what they're doing and the tools that they are using or the data sets that they are using. So that basically within this community of practice, well, knowledge can be built in a sort of a *collaborative* way, building on what has been done previously. And so

learning from previous, perhaps mistakes or difficulties and kind of helping each other to improve practice and in the end, knowledge.

(A13; emphases added)

[Open Science] is also the idea that science should be *transparent* and aim at being *reproducible* to as large an extent as possible. I mean, you can't, nobody will be able to interview me in 100 years' time, but they could do an interview with similar people. And it's also a *democratic* question, question of *democracy*, not only to pay back for the ones of us who work at tax-funded universities, but also to help anyone in the world to be able to take, well, read and analyze the data we produced. *Sharing* our tax-funded employments or similar.

(A02; emphases added)

While most interviewees whose associations belonged to the first group tended to be less involved in OS practices, this is by no means a perfect correlation. Some interviewees were aware of sharing practices and of some of the principles of OS without practicing them themselves (yet). There were also no obvious disciplinary associations, apart from the fact that preregistration was almost exclusively mentioned by researchers in psycholinguistics, neurolinguistics, and phonetics (A06, A09, A11, A17).

It goes without saying that the interviewees' personal associations with OS are shaped by the OS practices that they are (1) aware of and (2) involved in. The data concerning these two aspects are analyzed in the following section.

4.3 On Linguists' Awareness and Experience of OS Practices

Both the 38 survey respondents and the 26 interviewed linguists reported very different degrees of knowledge and experience of OS practices. Most of the attendees of ReproducibiliTea in the HumaniTeas had at least heard of practices such as preregistration, sharing and reusing FAIR data, and creating and reusing OERs—even if, for most practices, only a minority had actively engaged in these practices at the time of the survey (see fig. 2).

The results presented in figure 2 are in line with the target audience of ReproducibiliTea in the HumaniTeas, which aims to bring together humanities students and researchers with an interest in OS, often seeking to learn more about how to implement these principles in their own research. It is also likely that some of the respondents, especially undergraduate and graduate students, first heard of these practices at one of the meetings.

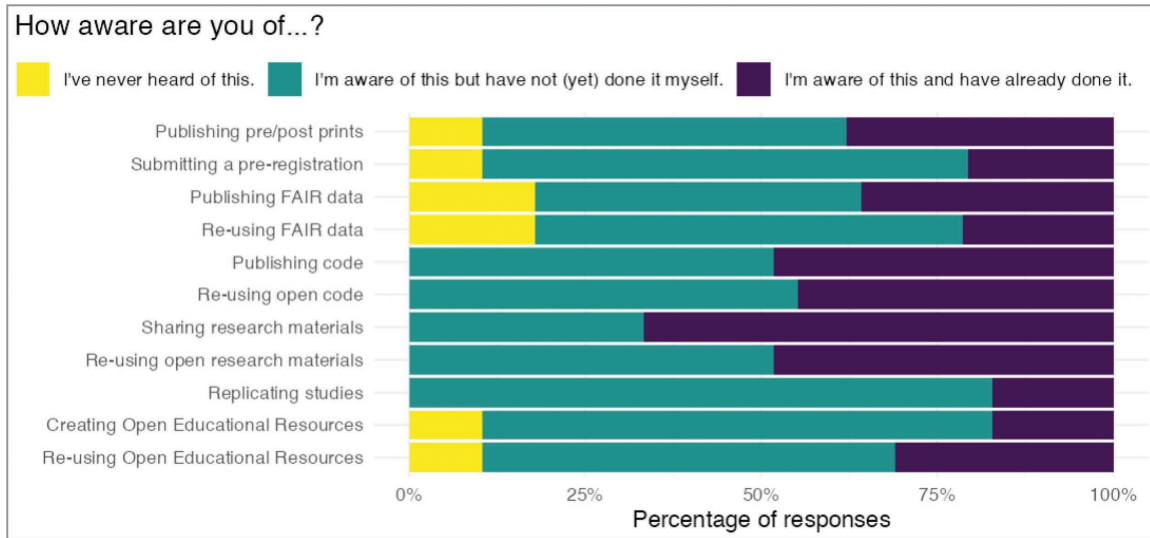


Figure 2. Respondents’ awareness of OS practices ($N_{\max} = 38$)

Similarly, the linguists interviewed as part of this study demonstrated varying degrees of personal involvement in OS. One interviewee (A14) described themselves solely as a “consumer” of OS because they had not yet published anything open access but had benefited from open access publications and science communication resources. At the other end of the spectrum, A12 enthusiastically described sharing research data, code, and materials openly, stressing that “as far as I can have any control over something, I’ll share everything” (A12). Others reported partial involvement:

I have taken part in research projects that kind of embraced open science principles, more or less, because I think it’s not a yes or no thing. It’s more of a continuum.

(A13)

Some felt that they could do more and reported an intention to implement more aspects of OS in their upcoming projects. For example, A09 said that they practiced OS “a little bit,” adding that they probably hadn’t “been at the forefront like other people have.”

Among the interviewed linguists, the most practiced aspects of OS were publishing journal articles in open access and sharing data and analysis code. Many were familiar with general repositories such as the OSF; far fewer mentioned linguistics-specific repositories such as TROLLING² (two mentions) or IRIS³ (one mention). Interestingly, some considered that sharing research materials, data, and/or code for peer review (without publishing them afterward) or exclusively with colleagues to—at least partially—be a

2. <https://dataverse.no/dataverse/trolling>.

3. <https://iris-database.org/>.

contribution to OS, thus suggesting a continuous rather than binary understanding of degrees of openness. Very few interviewees mentioned reusing open data, conducting secondary studies, or meta-analyses. Many were not aware of the terms *preprints* and *postprints*. When I explained them, most said that they had made use of postprints via commercial platforms such as ResearchGate and Academia.edu, but those who were not familiar with the term had not published any themselves. Finally, very few interviewed linguists outside of phonetics and psycho-/neurolinguistics were familiar with the concepts of preregistrations and registered reports.

4.4 On Linguists' Sources of OS Knowledge

One of the questions of the online survey was: “Apart from attending ReproducibiliTea in the HumaniTeas, how have you learned about Open Science practices?” Twenty-five participants responded to this question, most with multiple mentions that included reading about the topic; attending conferences, workshops, university seminars, and lectures; taking advantage of library services; speaking to colleagues; and social media.

Social media was typically mentioned as a source of OS knowledge without further specification, though sometimes explicitly in the plural form—that is, “social media platforms” (A09). Two interviewees specifically mentioned Twitter, even though it had already been rebranded as X by the time the interviews were conducted; one notably used the past tense to indicate that they no longer consider the platform as valuable as they used to (A09). One interviewee also referred to ResearchGate and Academia.edu as “academic social media platforms” (A16), and another reported learning about aspects of linguistics and OS “that are not necessarily part of the program and things that I could not afford at the time” on YouTube, eventually leading them to produce their own YouTube videos on using open source software for linguistics research (A20).

Of the 26 interviewed linguists, two mentioned that they were (strongly) encouraged to preregister their studies and make their data and code publicly available by their supervisors:

So in the beginning, it was because my supervisor and the postdoc working at the project at that time encouraged me to do this because I thought, okay, it's like just the usual practice that you do.

(A06)

I was a bit reluctant in the beginning. I was like, oh my God, I cannot make my data openly available. Everyone will see what's wrong with it or something. But yeah, I didn't, I mean, [my supervisor] convinced me with the advantages. So I think it's

good to make research reproducible. And yeah, so basically he was the main person who encouraged me.

(A17)

However, many of the other ECRs involved in OS explicitly reported not having any such encouragement or support:

It was, it's fully my own motivation there I mean and it's not because my supervisors did not want it or, or but they, they were not aware of that.

(A23)

Several mentioned online resources such as webinars, video recordings, open handbooks, and social media as sources of knowledge and inspiration. This process of self-motivated learning-by-doing was summarized by A24 as learning “[f]rom doing and failing, I guess. And trying again.” In addition, many interviewees reported learning directly from “specific persons who are very passionate about the topic” (A18). Several mentioned one role model colleague, who often worked at their current or previous institution as their primary source of inspiration (mentions of specific people have been anonymized in the transcripts, but see Le Foll 2025a for a full list). All in all, this implies that much of linguists’ awareness and knowledge of OS depends on their personal networks, as highlighted by A19:

I talked to people. So here's the network aspect again. I looked at how other people did it. I saw other papers where they said, ah, data and code are available on OSF. Or this is available on my GitHub. And I thought, well, okay, this is obviously how you do it. So I checked it out. I talked to them. And then I tried it myself.

(A19)

In the interviews, some of the participants recalled—in the words of one interviewee—“revelations” (A23) that triggered their interest/involvement in OS. One such revelation concerned a sense of unfairness with regard to the for-profit academic publishing industry: These interviewees became OS practitioners as a result of struggling with paywalls when studying or researching/teaching in the Global South. For A20, access to learning materials in their native country was also an issue. They recalled how learning about corpus linguistics via a MOOC and YouTube videos inspired them:

So I thought, hey, the same way he [a linguist who posted recordings of their lectures on YouTube] helped me out when I couldn't have access to other things, I could help other people.

(A20)

Another interviewee (from the Global North) also came to OS from Open Education:

So I published OERs. And then when I published my first papers, I thought, well, now you do all this open education stuff and are very transparent about your teaching. Why aren't you doing it in your research as well? Sounds kind of inconsequential. So I started doing it. And I was informed a lot about CC licenses in the OER context. Anyway, that was basically where I started out with my research materials. And then I started publishing my code and data where possible. And yeah, documentation for the annotation. And then it kind of grew from there.

(A19)

Finally, two interviewees—from two different generations, an undergraduate student and a professor—explained how their interest in OS came from a long-standing interest in open source software. As one noted:

I just kind of heard the idea [of Open Science] and I thought it was very adjacent to, very similar to the philosophy of free software and open-source software. And so I found it quite appealing in that regard.

(A22)

4.5 *On the Specificities of Open Linguistics*

When asked, “What are the specificities of linguistics that ought to be taken into consideration when trying to apply and promote Open Science in linguistics?” interviewees’ initial response was often one of bemusement. Many did not feel capable of giving a meaningful response until I added that they could focus on their own subdiscipline(s). One linguist explained the dilemma as follows:

I'm not sure if you can even put it like that, open science practices in linguistics, because linguistics deals with so many different things and types of data and is adjacent to so many other disciplines. So when I do psycholinguistics, I have to adhere to the psychology traditions. When I do sociolinguistics maybe I'm more to the sociology traditions, when I do historical linguistics and more like and so on, right, and the data is very varied. And so going through a whole open science cycle in linguistics, I don't think is a thing, maybe. Maybe it's subdiscipline specific.

(A19)

Some interviewees felt that while some aspects of OS are generally applicable to all linguistics subdisciplines (e.g., preprints), others are not (e.g., sharing data and code).

Here, the aforementioned tensions between quantitative and qualitative researchers’ stances on OS surfaced. These are discussed in the following section.

4.5.1 *Open Methods*

Related to differences across subdisciplines, a common theme that emerged from the QCA was interviewees’ tendency to distinguish between the applicability of OS to quantitative versus qualitative or non-empirical research paradigms:

I think the quantitative people are probably a lot more open to open science practices than qualitative people.

(A04)

This finding is in line with Liu and de Cat’s (2024, 94) observation that some linguists perceive OS to suffer from a “quantitative bias.” Indeed, across disciplines, the OS movement has been criticized as promoting “a view that valuable knowledge is derived only from empirical, observable evidence” and therefore “incompatible standards” for much qualitative research (Pownall 2025, 556).

Among the interviewees, only one reported working exclusively with qualitative methods. Even though they were not (yet) involved in many of the aforementioned OS practices, they were very enthusiastic about them. Interestingly, they commented on aspects of transparency that the others did not mention such as researcher positionality:

So there is this transparency link that is, to my knowledge, or maybe my modest knowledge, is absolutely not to be found most of the times [in OS discussions], it sounds like data are falling from the sky, right? But nothing is neutral. I mean, we have this obsession with objectivity in working with data, but I do believe that we should be more honest with ourselves and clearly say that somehow the way we collect data, especially the way we select them, inevitably is based on decisions and intuitions that we have as researchers in a certain position and having only some material available rather than other material available. (A25)

Also addressing the need to expand the maxim for transparency to beyond “just” sharing data and analysis code, one corpus linguist called for greater transparency at the data pre-processing stage, citing tokenization as an example:

We do tend to share also scripts and codes. But what I’m thinking of is linguistic tools and software that sometimes is available through either an online platform or a desktop application, which doesn’t allow you to actually understand what happens behind the

scenes because there's no clear explanation of what happens, what processing the data goes through, and the actual code that is used by the software is not shared. So, that is something that perhaps might also be included more in discussions about open science.

(A13)

These excerpts exemplify the fact that *open methods* is understood differently across different research paradigms within linguistics. Finally, two interviewees—both very involved in OS—commented on a specificity of linguistics that is more closely associated with linguistics as a social science than a humanity: the high inter-person variability that we can expect across different language users. This led A12 to cite Jack Grieve's (2021) article on replication failure in linguistics and state that “when it comes to the replication crisis and what we need to expect, I think linguistics is slightly unique.” In a similar vein, one interviewee expressed some doubts as to how meaningful it is to preregister the hypotheses of linguistics studies:

because maybe language is so open and creative and you can't really always know what to expect. Or also the participants differ so much. So you have to take, maybe you don't know what you actually have to take into account. So this was, yeah, again, thinking of preregistration. Maybe to, I don't know, maybe you have to be a little bit more open to the data than in other fields. I don't know, because there's so much variability.

(A11)

4.5.2 *Open Data*

Regarding sharing data, two recurring subdiscipline-specific themes were identified in the interviews: data privacy and copyright. There appears to be a need for further training and greater guidance on these issues as both were also mentioned when participants of ReproducibiliTea in the HumaniTeas were asked what they would like to learn (more) about in future sessions. In the interviews, the first aspect was typically mentioned by linguists doing experiments or fieldwork with human participants, especially those working with vulnerable populations such as (migrant) children:

Bei Kindern sind wir nochmal extra sicher. Da könnten wir zum Beispiel, anders als man das vielleicht in anderer Forschung macht, auf keinen Fall Videos veröffentlichen. Und selbst bei Transkripten, wenn man die veröffentlichen wollte, deswegen findet man im Grunde auch nichts im Netz. Es ist so wahnsinnig schwierig. Man muss diese Daten komplett bereinigen, weil die Kinder sehr viel von sich preisgeben. Man müsste ganze Teile rausschneiden, um sie überhaupt in einer öffentlich zugänglichen Datenbank zu verwenden. [. . .] Das ist zum Teil hochgradig persönlich und eventuell

rückverfolgbar, auch wenn ich sie anonymisiere. Und deswegen ist für mich dann Datenschutz die absolute Maxime weit über Open Science, in was auch immer für einer Form. [With children, we are extra safe. For example, unlike in other research, we cannot publish videos under any circumstances. And even with transcripts, if you wanted to publish them, that’s basically why you can’t find anything online. It’s so incredibly difficult. You have to completely clean up this data because the children reveal a lot about themselves. You’d have to cut out whole sections in order to even make them available in a publicly accessible database. [. . .] Some of it is highly personal and possibly traceable, even if I anonymize it. And that’s why, for me, data protection is the absolute maxim far above open science, whatever form it takes.]

(A01)

By contrast, copyright issues were mostly mentioned by linguists working with corpus data:

This is maybe another issue that I had when I started thinking about this whole open access, about this whole open research or open science idea which I really like, but one thing that kind of troubles me a little bit is the question of you know copyright and all sorts of legal questions that I find incredibly difficult to answer or to find answers to particularly [. . .] with the whole corpus issue. So I’m really not sure whether I’m allowed or entitled to actually share this [corpus] with, with the wider public.

(A25)

The legal repercussions of publishing certain types of data were a concern to many interviewees. Some mentioned these concerns as a justification for not being (very) involved in OS practices, whereas those already invested in OS tended to mention them as a reason to initiate (systemic) change:

We have a big problem here and I think we didn’t fight this legally so far. Or well enough. [. . .] Everybody is very afraid and rightfully so. And when they go to their legal advisors in the university, you know, they prefer to err on the side of caution. So no, no, no, don’t do that. We have to contest that eventually.

(A12)

Finally, a third, subdiscipline-specific factor concerning open data emerged: the relative effort required to compile a dataset. One interviewee, an experimental linguist who reported systematically sharing their data on open repositories, reflected on the concerns of their typology colleagues who invest a lot of time, effort, and money to painstakingly collect data over weeks or months:

But I feel like with like fieldwork, for example, [. . .] you really, really do a lot of, you put a lot of effort into like collecting the data, then transcribing the data, then translating the data that you might not want others to also just use it, you know, so that you feel like this is mine. And I want to take everything, like I want to, really work with the data and get everything out of it and then, if I, if I'm done with it, then other people can have it. Maybe this is like the kind of thinking that's there because like, with me, like with reaction time data, I mean, yeah, it is a pain to collect data to find the students, but in the end, I mean, it's like a 20-minute experiment. So if other people can like benefit from these data, sure go ahead, use it, use it as a reference, whatever but it's not that, I mean, I did put effort into this but I guess it's nothing compared to a field work effort.

(A06)

4.5.3 *Academic Cultures*

Some interviewees focused on subdiscipline specificities related to the cultures of research dissemination and science communication—primarily differentiating between theoretical and applied linguistics:

I think for certain areas of linguistics, it's very important that they kind of become more open to the public, especially when it comes to language education. So I think many language teachers will be interested in reading publications about language teaching, for example. And I think that should be taken into consideration so that educators have the option to read books, works for free.

(A14)

Three further factors related to the varied academic cultures of linguistics were identified as areas requiring consideration when applying OS to linguistics. The first concerns the publication culture of subdisciplines closely associated with the humanities, where monographs and edited volumes remain an important means of disseminating research. Several interviewees reported that, although they would like to publish these in open access, this is difficult due to extremely high article processing charges (APCs) for gold open-access monographs and funding being mostly reserved for journal articles (only one interviewee mentioned the diamond open-access book publisher Language Science Press). Second, some interviewees cited the lack of incentives for (large and/or interdisciplinary) collaborations in linguistics as a factor potentially contributing to the (s)low uptake of OS practices in some subdisciplines. Third, many mentioned a lack of (adequate) training in research data management (RDM), statistics, and/or OS practices in linguistics as compared to psychology or STEM study programs.

5 Concluding Thoughts

Both the small-scale survey of ReproducibiliTea in the HumaniTeas attendees and the 26 semi-structured interviews contributed answers to the three sets of research questions formulated at the beginning of this article.

It transpired that the answers to RQ1 (concerning linguists’ understanding of OS) and RQ2 (concerning their awareness of OS practices) are partially correlated. In addition to the disparities anticipated on account of disciplinary and/or methodological differences, the analysis demonstrated that the varying degrees of awareness of OS among linguists also account for a substantial proportion of the differences in their conceptualizations of OS. Thus, a minority of linguists continue to equate OS largely with open access publishing. However, some of these individuals are keen to expand this narrow definition to encompass the accessibility of *all* research outputs, including those intended for non-academic stakeholders. Those with greater awareness of OS practices are more likely to relate OS to the sharing of data, materials, and code, in order to promote transparency, reproducibility, and replicability in linguistics. This group also frequently associates OS with principles and values such as collaboration, democracy, fairness, honesty, and rigor. In answer to the second part of RQ2, personal contacts, often within their immediate team, research group, or institute, appear to be the most important source of information about OS for linguists. Additionally, some linguists reported gaining OS knowledge at conferences, courses and workshops, and/or via their university library services. Several linguists highlighted the importance of role models who inspired them to become (more) involved in OS practices. Social media was also mentioned as a source of OS knowledge and inspiration.

Turning to RQ3, the interviewees outlined several challenges and considerations that they believe need to be addressed when applying OS to linguistics. For some, securing funding for open access publishing—particularly for monographs—was a major concern. Others were primarily concerned with the ethical and legal ramifications of sharing their data. Two interviewees expressed some doubts about the applicability of preregistration to linguistics due to high inter-person variability. Another perceived challenge was the lack of training in research data management and statistical methods in many linguistics programs.

The study’s findings are necessarily constrained by the small sample of respondents who, at the time of data collection, were all based in Northern Europe. In addition to the limitations of the recruitment procedures mentioned above, I also acknowledge the potential impact of social desirability bias. One interviewee made this explicit by mentioning that they had read Le Foll (2024) prior to the interview to “be more prepared” (A25). While this interview was not excluded from the analysis, it was interpreted with this in mind. Although using personal networks to recruit interviewees

enabled me to reach out to linguists who would not usually respond to an open call to participate in an interview about a topic with which they have little to no experience, all of the respondents showed a genuine interest in the topic. Hence, I agree with A24 who, when asked about what could be done to increase the uptake of OS practices in linguistics, retorted that I “should be interviewing some sceptics.” There undoubtedly remains much to be done to understand the dynamics of the awareness, acceptance, and uptake of OS in an interdisciplinary field as varied as linguistics. In a follow-up study, I will be analyzing linguists’ perceived barriers and challenges to applying OS in linguistics and examining their suggestions for increasing the uptake of OS in linguistics.

In sum, while significant progress has been made in raising awareness and establishing platforms for OS in at least some subdisciplines of linguistics, the challenges of implementation, epistemological diversity, and broad-based participation remain. Much of the progress appears to be localized, and many of these concerns are subdiscipline specific, as explained by one OS enthusiast:

Well, parts of linguistics is open. Certain journals, certain publishers, and certain people and groups advocate open science in linguistics, but I would say that still a large part of linguistics is not fully open as I understand it.

(A16)

Ultimately, the findings of this study underscore the need for continued efforts to promote the awareness, understanding, and the necessary knowledge and skills for the adoption of OS practices in linguistics, while also acknowledging the complexities and challenges that arise from the field’s diverse subdisciplines and epistemological traditions.

Open Peer Review Reports

Open peer review reports for this article are available at the following location: <https://doi.org/10.17613/tfj13-egc54>

Declaration of Competing Interests

I am the initiator and co-organizer of ReproducibiliTea in the HumaniTeas. Two of the students interviewed had attended one of my courses in the past, though neither course was specifically about OS.

Acknowledgments

I would like to thank my colleagues and co-organizers of ReproducibiliTea in the HumaniTeas, Gabriele Schwiertz and Denis Arnold, for their valuable input on the survey questions and Gabriele for their implementation on LimeSurvey. Many thanks to Vishar Kavehamoli and Julia Weinberger who corrected the automatic transcriptions and anonymized most of the interview transcripts. Last but certainly not least, I am deeply indebted to the 26 students and colleagues who participated in the interviews, provided such insightful answers to my questions, and agreed to anonymized transcripts of the interviews being shared with the research community.

Author Biography

Elen Le Foll is a post-doctoral researcher and lecturer in linguistics at the Department of Romance Studies at the University of Cologne. She has a strong interest in quantitative corpus linguistics methods and applications of corpus research to language teaching and learning. She is co-project investigator of a project on the role of gender as a prominence feature within the Collaborative Research Center “Prominence in Language”. As a keen educator, she enjoys teaching about quantitative methods, R, statistics, data visualization, critical data literacy, and Open Science practices for to (future) linguists and language teachers.

References

- Alferink, Inge. 2022. “Using OASIS Summaries: Reconciling Direct Access to Research Findings with a Need for Information Brokering: Open Scholarship in Applied Linguistics Symposium.” Presentation at Open Scholarship in Applied Linguistics Symposium.
- Alferink, Inge, and Emma Marsden. 2023. “OASIS: One Resource to Widen the Reach of Research in Language Studies.” *Innovation in Language Learning and Teaching* 17 (5): 946–52. <https://doi.org/10.1080/17501229.2023.2204100>.
- Al-Hoorie, Ali H., Carlo Cinaglia, Phil Hiver, et al. 2024. “Open Science: Considerations and Issues for TESOL Research.” *TESOL Quarterly* 58 (1): 537–56. <https://doi.org/10.1002/tesq.3304>.
- Al-Hoorie, Ali H., and Phil Hiver. 2023. “The Postprint Pledge—Toward a Culture of Researcher-Driven Initiatives: A Commentary on ‘(Why) Are Open Research Practices the Future for the Study of Language Learning?’” *Language Learning* 73 (S2): 388–91. <https://doi.org/10.1111/lang.12577>.
- Andringa, Sible, Maria Mos, Catherine van Beuningen, Paz González, Jos Hornikx, and Rasmus Steinkrauss. 2024. “Diamond Is a Scientist’s Best Friend: Counteracting Systemic Inequality in Open Access Publishing.” *Dutch Journal of Applied Linguistics* 13. <https://doi.org/10.51751/dujal18802>.

- Belli, Alessandro, Jan Küster, Florian Hohmann, et al. 2025. OpenQDA. Version 1.0.1. Zenodo, released March 14. <https://doi.org/10.5281/zenodo.15024779>.
- Berez-Kroeker, Andrea L., Lauren Gawne, Barbara F. Kelly, and Tyler Heston. 2017. "Survey of Reproducibility in Linguistics Journals, 2003–2012." <https://sites.google.com/a/hawaii.edu/data-citation/survey>.
- Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, et al. 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>.
- Bochynska, Agata, Liam Keeble, Caitlin Halfacre, et al. 2023. "Reproducible Research Practices and Transparency Across Linguistics." *Glossa Psycholinguistics* 2 (1). <https://doi.org/10.5070/G6011239>.
- Butler, Leigh-Ann, Lisa Matthias, Marc-André Simard, Philippe Mongeon, and Stefanie Haustein. 2022. "The Oligopoly's Shift to Open Access Publishing: How For-Profit Publishers Benefit from Gold and Hybrid Article Processing Charges." 26th International Conference on Science and Technology Indicators (STI 2022). Zenodo, September 7. <https://doi.org/10.5281/zenodo.6951572>.
- Casillas, Joseph V., Gabriela Constantin-Dureci, Iván Andreu Rascón, et al. 2023. "Opening Open Science to All: Demystifying Reproducibility and Transparency Practices in Linguistic Research." Preprint, PsyArXiv, December 22. <https://doi.org/10.31234/osf.io/spz4w>.
- Chapelle, Carol A., and Gary J. Ockey. 2024. "Open Science in Language Assessment Research Contexts: A Reply to Winke." *Language Testing* 41 (4): 882–85. <https://doi.org/10.1177/02655322241239377>.
- Farangi, Mohamad Reza, and Hassan Nejadghanbar. 2024. "Investigating Questionable Research Practices Among Iranian Applied Linguists: Prevalence, Severity, and the Role of Artificial Intelligence Tools." *System* 125 (October): 103427. <https://doi.org/10.1016/j.system.2024.103427>.
- Ferguson, Joel, Rebecca Littman, Garret Christensen, et al. 2023. "Survey of Open Science Practices and Attitudes in the Social Sciences." *Nature Communications* 14 (1): 5401. <https://doi.org/10.1038/s41467-023-41111-1>.
- FitzGibbon, Lily, Daniel Brady, Anthony Haffey, et al. 2020. *Brewing up a Storm: Developing Open Research Culture through Reproducible Tea*. Open Research Case Studies, University of Reading. <https://doi.org/10.17864/1926.92781>.
- Garellek, Marc, Adrian Simpson, Timo B. Roettger, et al. 2020. "Letter to the Editor: Toward Open Data Policies in Phonetics: What We Can Gain and How We Can Avoid Pitfalls." *Journal of Speech Sciences* 9 (September): 3–16. <https://doi.org/10.20396/joss.v9i00.14955>.
- Gawne, Lauren. 2024. "Linguistics and Language Podcasts." *Superlinguo* (blog), December 18. <https://www.superlinguo.com/post/770166813430546432/linguistics-and-language-podcasts>.
- Gawne, Lauren, Barbara F. Kelly, Andrea L. Berez-Kroeker, and Tyler Heston. 2017. "Putting Practice into Words: The State of Data and Methods Transparency in Grammatical Descriptions." *Language Documentation & Conservation* 11:157–89.
- Gelman, Andrew. 2016. "Why Is the Scientific Replication Crisis Centered on Psychology?" *Statistical Modeling, Causal Inference, and Social Science Search* (blog), September 22. <https://statmodeling.stat.columbia.edu/2016/09/22/why-is-the-scientific-replication-crisis-centered-on-psychology/>.
- Gowie, Evangeline, Anna Tsakalaki, and Etienne B. Roesch. 2024. "Making Interview Transcripts Open: Preliminary Results from a Scoping Review." Preprint, OSF Preprints, December 12. <https://doi.org/10.31219/osf.io/pzyvt>.
- Grieve, Jack. 2021. "Observation, Experimentation, and Replication in Linguistics." *Linguistics* 59 (5): 1343–56. <https://doi.org/10.1515/ling-2021-0094>.
- Gurzynski-Weiss, Laura, Lara Bryfonski, and Derek Reagan. 2024. "Teacher IDs and Task Adaptations: Making Use of the TBLT Language Learning Task Bank." In *Individual Differences and Task-Based*

- Language Teaching*, edited by Shaofeng Li. John Benjamins. <https://www.degruyterbrill.com/document/doi/10.1075/tblt.16.11gur/html>.
- Isbell, Daniel R., Dan Brown, Meishan Chen, et al. 2022. “Misconduct and Questionable Research Practices: The Ethics of Quantitative Data Handling and Reporting in Applied Linguistics.” *Modern Language Journal* 106 (1): 172–95. <https://doi.org/10.1111/modl.12760>.
- Isbell, Daniel R., and Jieun Kim. 2023. “Developer Involvement and COI Disclosure in High-Stakes English Proficiency Test Validation Research: A Systematic Review.” *Research Methods in Applied Linguistics* 2 (3): 100060. <https://doi.org/10.1016/j.rmal.2023.100060>.
- Kesäniemi, Joonas, Turo Vartiainen, Tanja Säily, and Terttu Nevalainen. 2018. “Open Science for English Historical Corpus Linguistics: Introducing the Language Change Database.” In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference*, edited by Eetu Mäkelä, Mikko Tolonen, and Jouni Tuominen. Helsinki. <https://ceur-ws.org/Vol-2084/paper4.pdf>.
- Knöchelmann, Marcel. 2019. “Open Science in the Humanities, or: Open Humanities?” *Publications* 7 (4): 65. <https://doi.org/10.3390/publications7040065>.
- Kortmann, Bernd. 2021. “Reflecting on the Quantitative Turn in Linguistics.” *Linguistics* 59 (5): 1207–26. <https://doi.org/10.1515/ling-2019-0046>.
- Kremmel, Benjamin, and Daniel R. Isbell. 2024. “Open Science Practices in Language Assessment: Introducing the Special Issue.” *Language Testing* 41 (4): 697–702. <https://doi.org/10.1177/02655322241264092>.
- Kuckartz, Udo, and Stefan Rädiker. 2023. *Qualitative Content Analysis: Methods, Practice and Software*. 2nd ed. Sage Publications.
- Larsson, Tove, Luke Plonsky, Scott Sterling, Merja Kytö, Katherine Yaw, and Margaret Wood. 2023. “On the Frequency, Prevalence, and Perceived Severity of Questionable Research Practices.” *Research Methods in Applied Linguistics* 2 (3): 100064. <https://doi.org/10.1016/j.rmal.2023.100064>.
- Laurinavichyute, Anna, Himanshu Yadav, and Shravan Vasishth. 2022. “Share the Code, Not Just the Data: A Case Study of the Reproducibility of Articles Published in the *Journal of Memory and Language* Under the Open Data Policy.” *Journal of Memory and Language* 125: 104332. <https://doi.org/10.1016/j.jml.2022.104332>.
- Le Foll, Elen. 2024. “Why We Need Open Science and Open Education to Bridge the Corpus Research–Practice Gap.” In *Corpora for Language Learning: Bridging the Research–Practice Divide*, edited by Peter Crosthwaite. Routledge.
- Le Foll, Elen. 2025a. “Semi-Structured Interviews with Linguists About Open Science.” Aggregated interviewee metadata, consent form, anonymized interview transcripts, anonymized annotated passages as semi-colon separated CSV file. Version 3. Zenodo, November 17. <https://doi.org/10.5281/zenodo.17630137>.
- Le Foll, Elen. 2025b. “Survey About Reproducibility in the Humanities.” Questionnaire in PDF and XML formats, raw data as CSV file. Version 1. Zenodo, November 11. <https://doi.org/10.5281/zenodo.17583920>.
- Le Foll, Elen. Forthcoming. *Sharing Is Caring? A Scoping Review of Open Science Practices in Corpus Linguistics Research*.
- Liu, Meng. 2023. “Whose Open Science Are We Talking About? From Open Science in Psychology to Open Science in Applied Linguistics.” *Language Teaching* 56 (4): 443–50. <https://doi.org/10.1017/S0261444823000307>.
- Liu, Meng, and Cécile de Cat. 2024. “Open Science in Applied Linguistics: A Preliminary Survey.” In *Open Science in Applied Linguistics*, edited by Luke Plonsky. Applied Linguistics Press. https://www.appliedlinguisticspress.org/home/catalog/plonsky_2024.

- Liu, Meng, Sin Wang Chong, Emma Marsden, et al. 2023. "Open Scholarship in Applied Linguistics: What, Why, and How." *Language Teaching* 56 (3): 432–37. <https://doi.org/10.1017/S0261444822000349>.
- Marsden, Emma. 2019. "Open Science and Applied Linguistics: Where Are We and Where Are We Heading?" Plenary, American Association of Applied Linguistics, Georgia, Atlanta, March 11. <https://osf.io/wbkj6/>.
- Marsden, Emma, and Kara Morgan-Short. 2023. "(Why) Are Open Research Practices the Future for the Study of Language Learning?" *Language Learning* 73 (S2): 344–87. <https://doi.org/10.1111/lang.12568>.
- McGillivray, Barbara, and Gard B. Jensen. 2023. "Quantifying the Quantitative (Re-)Turn in Historical Linguistics." *Humanities and Social Sciences Communications* 10 (1): 1–6. <https://doi.org/10.1057/s41599-023-01531-2>.
- McManus, Kevin. 2022. "Are Replication Studies Infrequent Because of Negative Attitudes? Insights from a Survey of Attitudes and Practices in Second Language Research." *Studies in Second Language Acquisition* 44 (5): 1410–23. <https://doi.org/10.1017/S0272263121000838>.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716. <https://doi.org/10.1126/science.aac4716>.
- Plonsky, Luke. 2024a. "The Era of Open Science Is Upon Us (Or, Why a More Open Science Is Also a Higher Quality Science)." In *Open Science in Applied Linguistics*, edited by Luke Plonsky. Applied Linguistics Press. https://www.appliedlinguisticspress.org/home/catalog/plonsky_2024.
- Plonsky, Luke, ed. 2024b. *Open Science in Applied Linguistics*. Applied Linguistics Press. https://www.appliedlinguisticspress.org/home/catalog/plonsky_2024.
- Plonsky, Luke, Dan Brown, Meishan Chen, et al. 2024. "'Significance Sells': Applied Linguists' Views on Questionable Research Practices." *Research Methods in Applied Linguistics* 3 (1): 100099. <https://doi.org/10.1016/j.rmal.2024.100099>.
- Pownall, Madeleine. 2025. "Bridging Qualitative Methods and Open Research." *Nature Reviews Psychology* 4 (9): 556–57. <https://doi.org/10.1038/s44159-025-00477-3>.
- Prosser, Annayah M. B., Olivia Brown, Grace Augustine, and David A. Ellis. 2024. "It's Time to Join the Conversation: Visions of the Future for Qualitative Transparency and Openness in Management and Organisation Studies." Preprint, SocArXiv, May 30. <https://doi.org/10.31235/osf.io/ntf73>.
- Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. "Robust Speech Recognition via Large-Scale Weak Supervision." arXiv, December 6. <https://doi.org/10.48550/ARXIV.2212.04356>.
- ReproducibiliTea. n.d. "ReproducibiliTea." Accessed November 11, 2025. <https://reproducibiliatea.org>.
- Rohatgi, Shaurya, Yanxia Qin, Benjamin Aw, Niranjana Unnithan, and Min-Yen Kan. 2023. "The ACL OCL Corpus: Advancing Open Science in Computational Linguistics." Preprint, arXiv, October 24. <https://doi.org/10.48550/arXiv.2305.14996>.
- Rooryck, Johan. 2023. "Lingua to Glossa." <https://www.rooryck.org/lingua-to-glossa>.
- Rzayeva, Narmin, Stephen Pinfield, and Ludo Waltman. 2025. "Adoption of Preprinting Across Scientific Disciplines and Geographical Regions (1991–2023)." Preprint, OSF, April 30. https://doi.org/10.31235/osf.io/xdwc4_v2.
- Salet, Xavier, John Gelissen, Guy Moors, and Jelte Wicherts. 2025. "Good, Bad, Different or Something Else? A Scoping Review of the Convictions, Conventions and Developments Around Quality in Qualitative Research." *Royal Society Open Science* 12 (6): 242001. <https://doi.org/10.1098/rsos.242001>.
- Sato, Masatoshi, Sin Wang Chong, Tasnima Aktar, Jennifer Cowell, Ming Sum Kong, and Mehdi Shaahdadi. 2024. "Creating and Sustaining a Platform for Researchers and Teachers to Communicate:

- An Example of *TESOLgraphics*.” *Innovation in Language Learning and Teaching*. <https://doi.org/10.1080/17501229.2024.2404613>.
- Sönning, Lukas, and Valentin Werner. 2021. “The Replication Crisis, Scientific Revolutions, and Linguistics.” *Linguistics* 59 (5): 1179–206. <https://doi.org/10.1515/ling-2019-0045>.
- Sterling, Scott. 2024. “Research Ethics in Open Science Within Applied Linguistics.” In *Open Science in Applied Linguistics*, edited by Luke Plonsky. Applied Linguistics Press. https://www.appliedlinguisticspress.org/home/catalog/plonsky_2024.
- Turing Way Community. 2022. “The Turing Way: A Handbook for Reproducible, Ethical and Collaborative Research.” Version 1.0.2. Zenodo, July 27. <https://doi.org/10.5281/zenodo.3233853>.