Reproducible research practices and transparency across linguistics

Agata Bochynska, University of Oslo, NO, agata.bochynska@gmail.com

Liam Keeble, Newcastle University, UK, liamkeeble@gmail.com

Caitlin Halfacre, Newcastle University, UK, caitlin.halfacre@gmail.com

Joseph V. Casillas, Rutgers University, US, joseph.casillas@rutgers.edu

Irys-Amélie Champagne, University of Toronto Scarborough, CA, iryschampagne@outlook.com

Kaidi Chen, University of Connecticut, US; Columbia University, US, kaidi.chen@uconn.edu

Melanie Röthlisberger, University of Zurich, CH, melanie.roethlisberger@uzh.ch

Erin M. Buchanan, Harrisburg University of Science and Technology, US, ebuchanan@harrisburgu.edu

Timo B. Roettger, University of Oslo, NO, timo.roettger@iln.uio.no

Scientific studies of language span across many disciplines and provide evidence for social, cultural, cognitive, technological, and biomedical studies of human nature and behavior. As it becomes increasingly empirical and quantitative, linguistics has been facing challenges and limitations of the scientific practices that pose barriers to reproducibility and replicability. One of the proposed solutions to the widely acknowledged reproducibility and replicability crisis has been the implementation of transparency practices, e.g., open access publishing, preregistrations, sharing study materials, data, and analyses, performing study replications, and declaring conflicts of interest. Here, we have assessed the prevalence of these practices in 600 randomly sampled journal articles from linguistics across two time points. In line with similar studies in other disciplines, we found that 35% of the articles were published open access and the rates of sharing materials, data, and protocols were below 10%. None of the articles reported preregistrations, 1% reported replications, and 10% had conflict of interest statements. These rates have not increased noticeably between 2008/2009 and 2018/2019, pointing to remaining barriers and the slow adoption of open and reproducible research practices in linguistics. To facilitate adoption of these practices, we provide a range of recommendations and solutions for implementing transparency and improving reproducibility of research in linguistics.

1. Introduction

Linguistics, defined here broadly as scientific studies on language, lies at the intersection of the humanities and the social and biomedical sciences. It informs psychological and neural models of communication, categorization, and memory (Frermann & Lapata, 2021; McClelland et al., 2020); it guides methods for diagnosis and therapy of speech, development, and aging disorders (Bohn & Frank, 2019; Munsell et al., 2020); it informs methods for educational improvements and facilitates advancement in new technological solutions, such as speech recognition and speech synthesis (Malisz et al., 2020; Wang et al., 2019). Spanning across many subfields, linguistics is also a particularly variegated field when it comes to its methods and the nature of the empirical studies conducted, a field that – while historically observational (Grieve, 2021) – is increasingly shaped by quantitative data analysis. As such, linguistics, along with its neighboring fields, is undergoing a sea change in the way research is conducted and shared.

In recent years, the empirical and, in particular, quantitative sciences have experienced an unprecedented time of introspection and self-evaluation, with many scholars raising serious concerns about the credibility of scientific findings (Ioannidis, 2005). Recent meta-scientific efforts to evaluate the rigor and robustness of the quantitative findings in published literature discovered a disconcerting proportion of studies whose claims cannot be replicated using the same methods (Camerer et al., 2018; Errington et al., 2021; Open Science Collaboration, 2015), or whose analytical conclusions cannot be reproduced (Hardwicke et al., 2020, 2022). This is collectively referred to as the *replication crisis* or the *reproducibility crisis* (FORRT, 2021). Moreover, a growing body of evidence suggests that researchers' conclusions often vary even when they have access to the same data and answer the same research question (Breznau et al., 2022; Rotello et al., 2015; Silberzahn et al., 2018; Starns et al., 2019). This variability is often rooted in the inherent flexibility of data analysis, often referred to as researcher degrees of freedom, and in researchers' biases (Gelman & Loken, 2014; Simmons et al., 2011).

The realization that many published studies might be misleading, biased, or contain errors led to constructive dialogs and a number of reform efforts across disciplines (Kidwell et al., 2016; Klein et al., 2018; Zwaan et al., 2018). Central to these reform movements is encouraging scientists to share all relevant elements of the studies they conduct to increase transparency. In this context, sharing refers to making publicly available a version of the final manuscript (Moshontz et al., 2021), research materials, stimuli, and procedures, including a preregistered research plan (Nosek et al., 2018), raw and processed data, and analysis scripts (Gilmore et al., 2018; Laurinavichyute et al., 2022; Lindsay, 2017).

The guiding principle of these reforms and their recommendations is that both the scientific community and the public should be able to access relevant information in order to interpret and critically evaluate scientific claims (Munafò et al., 2017; Vazire, 2017). Such an open

stance can benefit and accelerate scientific activities by enabling the independent reproduction and replication of results, and by allowing one to synthesize evidence via, for example, meta-analyses (Nicenboim et al., 2018; Pigott & Polanin, 2020). Open access publications are also associated with higher citation rates and improved scholarly dissemination (Piwowar et al., 2018; Tennant et al., 2016). Open data and materials, such as experimental scripts and stimuli, can facilitate collaboration (Boland et al., 2017), increase efficiency and sustainability (Lowndes et al., 2017), make human error detectable (Nuijten et al., 2016), and are cited more often (Colavizza et al., 2020). Preregistration, i.e., registering research and analysis plans before the study is underway, complements transparent sharing of other materials (Nosek et al., 2018). Open preregistrations can boost a researcher's reputation (Stewart et al., 2020) and safeguard against post-hoc hypothesizing (i.e., HARKing – hypothesizing after the results are known) and post-hoc critique during peer-review (i.e., CARKing – critiquing after the results are known) (Hobson, 2019; Kerr, 1998).

Despite these benefits for both the public good at large and the individual researcher, open science has not yet been fully adopted in many disciplines. Recent assessments of biomedicine (Wallach et al., 2018), social science (Hardwicke et al., 2020), and psychological science (Hardwicke et al., 2022) suggest that rates of open access (25–65%), data and materials sharing (11–33%), and preregistrations (0–3%) can still be quite low. In addition, potential sources of bias, such as conflicts of interest, may not be disclosed (Cristea & Ioannidis, 2018; Hardwicke et al., 2022). Such discipline-specific meta-assessments are essential to both raise awareness of these important issues and discuss possible solutions. Indeed, it appears that, for example, the interest in the (lack of) reproducibility of psychological studies (Open Science Collaboration, 2015) and "psychology's renaissance" (Nelson et al., 2018) nudged psychology researchers to increase transparency in reporting analysis steps (Valentine et al., 2021).

As linguistics has become increasingly quantitative, researchers have articulated concerns that credibility-decreasing practices (e.g., low statistical power, lack of data sharing, selective reporting, etc.) are also prevalent in our field (Casillas, 2021; Kirby & Sonderegger, 2018; Laurinavichyute et al., 2022; Roettger, 2019; Vasishth et al., 2018). Recent meta-scientific assessments have investigated some aspects of research practices in language research, including the prevalence of direct replications (Kobrock & Roettger, 2023; Marsden, Morgan-Short, et al., 2018) and analytical flexibility (Coretta et al., 2023). Assessments of second language research (Marsden, Thompson, et al., 2018; Plonsky et al., 2015) reported limited sharing of materials (4–17%) and data (15%), and bilingualism researchers suggest poor availability of data, analysis, and materials in their subfield (Bolibaugh et al., 2021). An assessment of a language documentation and description subfield also concluded that methodology and data collection practices were not explicitly reported or shared in grammars and dissertations published between 2003 and 2012 (Gawne et al., 2017). A more recent assessment which targeted a specific psycholinguistic journal

reported 12–20% of articles sharing data and code before an open data mandate was introduced (Laurinavichyute et al., 2022). Even though this suggests that open science has not been fully adopted in some subfields of linguistics, a global assessment of transparency and reproducibility practices across the field has not been conducted yet. Thus, the present paper is a first systematic attempt to quantify practices related to transparency and reproducibility in a random sample of scientific articles across linguistics. We do this by sampling journal articles from before and after the so-called reproducibility crisis became widely acknowledged and comparing the number of transparency and reproducibility practices reported in the sample. This assessment aims at helping to track progress over time and calibrate future policies and training initiatives across the field. Moreover, it enables cross-disciplinary comparisons to further our understanding of possible challenges for open scholarship.

2. Methods

2.1 Design

Following Hardwicke et al. (2022), this study was a retrospective observational study with a cross-sectional design. Sampling units were individual articles. The study's design and data collection plan were preregistered on September 1st, 2021. The intention behind preregistering this non-confirmatory study was to restrict researcher degrees of freedom during data collection that could possibly bias the results. The preregistration can be accessed at https://doi.org/10.17605/OSF.IO/J2Q5P. Deviations from this protocol are explicitly acknowledged, including the prescreening procedure described in the preregistration (and with more details in the prescreening code and documentation: https://doi.org/10.17605/OSF.IO/GPXVA) and the update to the definition of *disagreement* for coding of articles described in 2.3. Other minor updates to the preregistration are noted as post-hoc below. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2012). All materials, data, and analysis scripts related to this study are publicly available on Open Science Framework: https://doi.org/10.17605/OSF.IO/ZX9KY. To facilitate reproducibility, the results can be re-run online in the stable Code Ocean container that captures the computational environment in which the study analyses were conducted: https://codeocean.com/capsule/9832712/tree/v2.

2.2 Sample

2.2.1 Identification of target articles

From Scopus (https://www.scopus.com/), we sampled all scientific articles with the subject terms "Language and Linguistics" and "Linguistics and Language" (All Science Journal Classifications = 1203, 3310) within the years of 2008/2009 (pre-replication-crisis, pre-RC, the early time

window before the so-called replication crisis became widely acknowledged) and 2018/2019 (post-replication-crisis, post-RC, the late time window after the so-called replication crisis became widely acknowledged[1]), respectively. We only sampled entries from academic journals (source type = journal (j)) that were articles or reviews (document type = article (ar) and review (re)) in English (language = English).[2]

This query amounted to 48,044 entries (retrieved on the 8th of April, 2021, i.e., sampling predated preregistration). Our target sample size for each time period was 250 codable articles (n = 500 in total). The chosen sample size corresponds to Hardwicke et al.'s (2022) sample size and was chosen to (a) offer a comparable dataset, and (b) match the authors' time resources to code these articles. To achieve the target sample size, we applied the following method: Out of the 48,044 articles, we randomly sampled 750 articles from the early time window (2008/2009) and late time window (2018/2019), respectively. The random sampling procedure was implemented to ensure that results obtained from the sample would approximate what would have been obtained if the entire population (here: all published linguistic articles) had been measured. Further, we randomly assigned 32 articles to a pilot in which the authors applied a preliminary coding scheme to sampled articles to identify possible coding issues. Subsequently, we randomly assigned 50 additional articles to a prescreening pilot (see 2.2.2) to gauge (a) the amount of papers that cannot be accessed, and (b) the amount of papers that are judged not to be part of linguistics, using a broad definition of *linguistics* from Oxford Languages: "the scientific study of language and its structure, including the study of morphology, syntax, phonetics, and semantics. Specific branches of linguistics include sociolinguistics, dialectology, psycholinguistics, computational linguistics, historical-comparative linguistics, and applied linguistics."[3]

Because we decided to use broad keywords to capture the breadth of linguistics, some articles that were related to language but were not a linguistic study (per above definition) might have been captured. Therefore, we excluded papers from adjacent fields found within our search. Non-exhaustive examples include literature studies, language-related biology not specifically covering language (audiology, vision), and law studies. The random sampling procedure was performed using *R* (R Core Team, 2022). The script for the sampling procedure can be found in the shared code: https://doi.org/10.17605/OSF.IO/GPXVA.

---

[1] Here we refer to the time after the replication crisis was acknowledged (e.g., via high-profile replication projects such as Open Science Collaboration 2015) and do not intend to suggest that it is over.

[2] The corresponding Scopus query is: SRCTYPE (j) AND (SUBJTERMS (1203) OR SUBJTERMS (3310)) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"re")) AND (LIMIT-TO (PUBYEAR,2008) OR LIMIT-TO (PUBYEAR,2009) OR LIMIT-TO (PUBYEAR,2018) OR LIMIT-TO (PUBYEAR,2019)) AND (LIMIT-TO (LANGUAGE,"English")).

[3] Originally sourced from https://www.lexico.com/en/definition/linguistics. This link no longer is active, but the current definition of linguistics found at https://www.dictionary.com/browse/linguistics conveys the same meaning.

2.2.2 Prescreening of articles

The pilot coding scheme indicated that some articles within the Scopus results might not fit the definition of linguistics applied in the study (see above). A pilot prescreening was performed to analyze the distribution of articles that would not be considered for the study (conducted by EB, IC, KC). Twenty-five articles from each time period were assessed by two coders for their potential inclusion in the sample. The prescreening document and the results of the prescreening analysis can be found here: https://doi.org/10.17605/OSF.IO/GPXVA. We excluded 32% of articles ($n = 8$) from the early time window (2008/2009) and 40% of articles ($n = 10$) from the late time window (2018/2019), because they were not linguistic studies. In addition, 10% ($n = 5$) were unavailable to code (either it was not possible to locate the full text or the article was not in English). Overall, this results in 36% exclusion (of studies that do not fall under the definition of linguistics), 54% inclusion, and 10% unavailability.

In the full prescreening, each article was marked by one coder as included, excluded, or unsure. If at least one coder marked the article for inclusion, it was included in the final coding scheme described below. The randomly sampled list was coded until there were 300 articles included for each time point (600 total), in order to account for disagreements on the inclusion coding to achieve the desired sample size of 250 for each time point. For example, prescreening coders might have decided that an article should be included in the final sample, but analysis coders could later disagree with this assessment. We originally planned to have a second coder mark all excluded or unsure articles; however, the number of included articles reached the desired minimum ($n = 600$), and we decided to only complete this step if the analysis coders did not meet their desired minimum ($n = 500$).

2.3 Procedure

From the prescreened sample ($n = 600$), each analysis coder (AB, JC, LK, MR, TR) assessed individual articles for each measured variable (according to **Table 1**, except for Open Access, see below), recording outcomes using a Google Form (see exported form in shared materials: https://doi.org/10.17605/OSF.IO/H8DFT). Coders who coded articles did not prescreen articles, and vice versa. Eighty percent of all articles were assigned to and coded by one coder only. The remaining 20% of articles were also assigned to a second independent coder. Any discrepancies between the first and the second coder were resolved through discussion, and if necessary, a third coder arbitrated in order to arrive at a consensus. We defined two types of discrepancies (post hoc): *overlapping response,* in which one coder's response was a superset of the second coder's response (e.g., coder 1 described the nature of the raw data as audio and text, and coder 2 described it as text only); and *disagreement,* in which coders described an item in non-overlapping terms (e.g., coder 1 described the nature of the raw data as audio, and coder

2 described it as text). We assessed the prevalence of disagreements for each preregistered coding category individually (according to **Table 1**). If a section was characterized by disagreements in more than 20% of articles, said section would be assigned to the same second independent coder for all articles of the corpus. In the current sample, no categories were over 20% disagreement. Overlapping discrepancies were combined (i.e., if one coder included two options, while another included one, we used the two options). Non-overlapping items were checked by a third coder and changed to the third coder's agreement with the original coder. Complete disagreement percentages can be found in the shared analysis results: https://doi. org/10.17605/OSF.IO/GPXVA.

For the assessment of open access, one individual coder (CH) followed the coder instruction for open access only, i.e., this coder did not code other aspects of the articles. To establish the open access status of each article, the coder used Unpaywall software (https://unpaywall.org/) while not being logged into a network that grants paid access to articles. If the Unpaywall button was green, that indicated an open access version was available; if it was gray, there was no open access version. If the Unpaywall button was not able to detect the paper via the DOI, the open access status was coded as unknown. According to Unpaywall's FAQ, this happens when the repository's data is not available to their API. All articles were coded for open access in December 2021.

2.3.1 Measured variables

Measured variables and their operationalizations are shown in **Table 1**. Data extraction was executed via a Google Form consisting of questions to the coder and response options (see https://doi.org/10.17605/OSF.IO/H8DFT for details).

2.3.2 Analysis

Following Hardwicke et al. (2022), the results are merely descriptive and report on the proportion of articles that fulfill relevant characteristics relative to the number of articles in which each characteristic is applicable. We did not preregister and apply inferential statistical procedures, for the following reasons: First, the current sample is small (relative to the entirety of linguistic articles), so inferential statistics would imply a false certainty in the generalizability of the results. We also lack knowledge about the effect magnitudes and variance components. Finally, using inferential statistics could shift the focus to binary interpretations (e.g., a difference being "significant" or not), losing sight of the magnitude of quantitative differences.

Table 1: Main measured variables (see the coding form table for the full operationalization and questions to coders: https://doi.org/10.17605/OSF.IO/H8DFT).

Coding Category	Coding Area
Article characteristics	Publication year
	Field
	Language investigated
	Journal impact factor at year of publication
	Year of journal impact factor
	Country (Corresponding author)
	Study type/design
	Type of empirical study
Preregistration	Preregistration
	Preregistration method
	Preregistration accessibility
	Preregistration content
Data sharing	Raw data type
	Raw data statement
	Raw data sharing method
	Raw data accessibility
	Raw data documentation
	Processed data statement
	Processed data sharing method
	Processed data accessibility
	Processed data documentation
Analysis script sharing	Analysis script availability
	Analysis script sharing method
	Analysis script accessibility
Materials / Methods sharing	Materials availability
	Materials sharing method
	Materials accessibility
	Replication
	Replication statement
Conflict of interest	Conflict of Interest statement
Open access	Open access status

3. Results

All analysis results are descriptive and focus on the number and proportion of articles that fulfill certain criteria. In calculating proportions, we used as a denominator the number of articles in which each criterion is applicable (see **Table 2**). Because linguistics is a multidisciplinary field with many different empirical traditions, our sample contained a wide variety of study types, which we categorized into three main types: empirical data, no empirical data, meta-analysis. From these categories, we attempted to bin empirical data into seven study types and designs. However, coding of these categories was often difficult, due to the lack of clear-cut distinctions between, e.g., observational study and experiment. In order to describe the observed patterns, we decided to instead divide the empirical study types into three post-hoc categories that are defined by the relationship between the data and the analysis: (1) *Primary data*, defined as data collected during the study (including experimental studies, observational studies, correlational studies, field studies or language description, case studies, surveys, and interviews), (2) *Secondary data*, defined as data from previous data collections or reanalysis (including corpus studies, descriptions of archives, discourse analyses, secondary data analyses from published data, typological studies), and (3) *Other*, a category distinct from the previous two for which most coded measures are irrelevant (modeling, simulations, introspections, formal linguistic analyses). Once categorized, we found very few Other category articles, but several articles that included multiple types of data. In order to ensure that count values were represented only once in the following descriptions, each was coded as Primary if it contained any primary data, Secondary if it contained secondary data, and Other if it contained other data types (reserved for studies with only other than primary or secondary data types; see **Table 2**).

3.1 Sample characteristics

Sample characteristics for all 600 articles are displayed in **Table 2** and **Figure 1**. Of the 600 articles, 519 were included in the final sample. Eighty-one articles out of 600 (13.5%) were excluded, due to either not being in English (0.2%), not being about language according to our definition above (11.2%), or not being accessible (2.2%).

Among the coded articles, we observed similar sample characteristics between articles published in the early time window (2008/2009) and the late time window (2018/2019), e.g., in terms of the number of journals represented and Journal Impact Factors as well as languages, authors' affiliations, and study and data types, showing a well-balanced sample between the two time points. Note that despite the well-balanced sample characteristics across the time points, the full sample shows an overall skew towards authors' affiliations in Western countries (see Figure 1, panel A) and Western languages, with the most frequent language under investigation being English (see **Figure 2, panel B**; see also Section 4).

Table 2: Sample study characteristics. Pre-RC refers to the early time window (2008/2009) and post-RC to the late time window (2018/2019). JIF refers to Journal Impact Factor metrics.

Metric	Total	Pre-RC	Post-RC
Coded Articles	600	300	300
Number of Journals	274	168	185
Number of Languages	84	58	53
JIF	1.22 (0.86) Mdn = 0.99	1.01 (0.81) Mdn = 0.80	1.44 (0.86) Mdn = 1.39
Number of Countries	61	45	47
Included Articles	519	264	255
Study Design – Empirical	363	175	188
Study Data – Primary	263	123	140
Study Data – Secondary	97	50	47
Study Data – Other	3	2	1
Study Design – Not Empirical	152	87	65
Study Design – Meta-Analysis	4	2	2

Note. All values are raw counts, except JIF, which represents the mean, standard deviation, and median. Number of Journals, Languages, JIF, and Countries represents the entire coded dataset. Included Articles represents data from the journal articles coded as language (linguistics) studies (according to the study definition), with agreement from the coders. Only Included Articles are reported in the Study Design and Study Data values. Articles reporting empirical studies (Study Design – Empirical) were categorized post hoc, after examination of the results, into studies with primary (Study Data – Primary) and secondary data (Study Data – Secondary).

Of the 519 articles included in the study sample, 360 reported study designs with primary or secondary empirical data, excluding the Other category ($n = 3$) as well as non-empirical ($n = 152$) and meta-analysis designs ($n = 4$). The availability of materials, raw and processed data, and analysis scripts as well as the presence of replications were assessed only for the articles reporting empirical study with primary or secondary data, due to these outputs and practices being primarily dependent on empirical designs. Preregistration and conflict of interest statements were assessed for the full included sample. Article availability was assessed for all three samples: articles reporting only empirical studies with primary and secondary data ($n = 360$), the full included sample ($n = 519$), and the initial coded sample ($n = 600$).

A: Author Affiliation

Country

Unknown
Other
Eastern Europe
Australia and New Zealand
Southern Europe
Western Europe
Northern Europe
Asia
Northern America

Pre-RC: 13.7%, 14.3%, 10.7%, 33%
Post-RC: 11%, 11.3%, 17.7%, 28%

B: Languages

Target Language(s)

Other
Chinese
German
Spanish
Cross Linguistic
Universal
English

Pre-RC: 32.6%, 9.9%, 44.1%
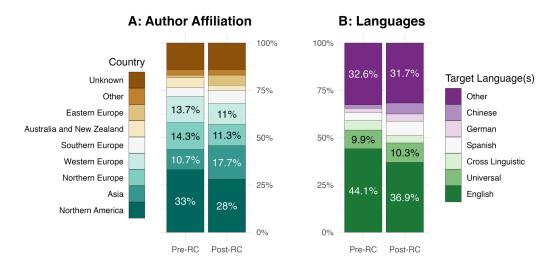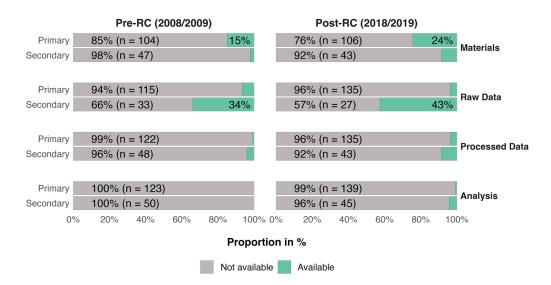Post-RC: 31.7%, 10.3%, 36.9%

Figure 1: Panel A: The proportion of first author affiliations by country binned into United Nations Subregions, collapsing all Asian subregions together and all other non-visualized regions as the Other category (Arel-Bundock et al., 2022). Panel B: The proportion of languages examined in these studies, with the 6 most frequent categories represented and the remaining languages collapsed together as Other. Cross Linguistic represents studies with five or more languages. Universal represents studies that made claims about all languages without basing their claims on a particular language. Pre-RC refers to the early time window (2008/2009) and post-RC to the late time window (2018/2019).

Assessment of transparency-related research practices in linguistics

Pre-RC (2008/2009)

Primary	85% (n = 104) — 15%
Secondary	98% (n = 47)

Primary	94% (n = 115)
Secondary	66% (n = 33) — 34%

Primary	99% (n = 122)
Secondary	96% (n = 48)

Primary	100% (n = 123)
Secondary	100% (n = 50)

Post-RC (2018/2019)

Primary	76% (n = 106) — 24%	**Materials**
Secondary	92% (n = 43)

Primary	96% (n = 135)	**Raw Data**
Secondary	57% (n = 27) — 43%

Primary	96% (n = 135)	**Processed Data**
Secondary	92% (n = 43)

Primary	99% (n = 139)	**Analysis**
Secondary	96% (n = 45)

Proportion in %

Not available   Available

Figure 2: Percentages of the available and not available materials, raw data, processed data, and analysis scripts for the pre-RC (left) and post-RC (right) time windows, displayed separately for primary data (Primary) and secondary data (Secondary), for the empirical study articles in the sample. The Other category was excluded.

3.2 Article availability (open access)

Of the 360 eligible empirical study articles (with primary and secondary data), we obtained a publicly available version (either the published version or pre-publication version from preprint repositories) for 129 (35.8%), whereas 208 (57.8%) were only accessible through a paywall. 23 articles (6.4%) were of unknown availability, i.e., it was unclear whether the article was publicly available or accessible only through a paywall. This rate was approximately the same when we examined the full initial dataset ($n = 600$) or the included dataset ($n = 519$). For the included dataset, 192 (37%) were available, 292 (56.3%) were paywalled, and 35 (6.7%) were not available. For the initial dataset, 211 (35.2%) were available, 342 (57%) were paywalled, and 47 (7.8%) were not available. The rate of the open access articles was lower in the early time window (28% of all sampled articles) compared with the late time window (42%).

3.3 Materials availability

Data for articles reporting empirical studies with primary or secondary data are visualized in **Figure 2**. Of those articles published in 2008/2009 and involving primary data ($n = 123$), 18 (14.6%) contained a statement or link regarding the availability of original research materials, such as survey instruments, software, or stimuli, but only 1 article ($n = 48$ overall, 2.1%) involving secondary data contained such a statement or link. Of the articles published in 2018/2019 ($n = 140$), 34 (24.3%) involving primary data and 4 ($n = 47$ overall, 8.5%) involving secondary data contained such a statement or link. For three additional articles, the Materials section was coded as *not applicable* due to the nature of the study design. Note that we marked materials as available also in cases when only partial materials were made available. For example, a full survey instrument was made available, but not tests for background measures or stimuli for an additional experiment.

Of the 57 articles overall for which materials were reportedly available, 4 did not have materials that were actually accessible, because of broken links.[4] Of the 53 articles for which we could access materials, the materials were made available in the article itself (e.g., in a table or appendix; $n = 10$), in a journal-hosted supplement ($n = 16$), on a personal or institutionally hosted (non-repository) webpage ($n = 3$), or in an online third-party repository ($n = 20$). All other materials locations were coded as Other ($n = 4$).

3.4 Data availability

3.4.1 Raw data

Raw data was defined as recorded information in its rawest, digital form, at the level of sampling units (e.g., participants, words, utterances, trials, etc.). Each article could include multiple data

---

[4] At the time of coding, one of the links redirected us to a website which hosted pornography.

types, and therefore 360 articles contained 434 unique data type combinations included in the article. These data types most frequently included text files ($n = 292$, 67.3%), followed by audio ($n = 95$, 21.9%), video ($n = 40$, 9.2%), and a few image files ($n = 7$, 1.6%).

Of those articles published in 2008/2009 and involving primary data ($n = 123$), 8 (6.5%) contained a statement or link regarding the availability of raw data such as audio recordings, transcriptions or data provided by experimental software. Of those articles involving secondary data ($n = 50$), 17 (34%) articles contained such a statement or link. Of those articles published in 2018/2019, 5 ($n = 139$, 3.6%) involving primary data and 20 ($n = 47$, 42.6%) involving secondary data contained such a statement or link.

Of the 50 articles for which raw data were reportedly available, 6 did not have raw data that was actually accessible, because of broken links or availability "upon request". Of the articles that we could access ($n = 44$), the raw data were made available in a journal-hosted supplement ($n = 4$), on a personal or institutionally hosted (non-repository) webpage ($n = 2$), or in an online third-party repository ($n = 38$). Of those articles, 6 data sources contained data documentation, such as metadata or data dictionaries.

3.4.2 Processed data

Processed data was defined as a derived form of the data that has undergone changes from its raw state (e.g., extraction of acoustic parameters via Praat, aggregates of responses, etc.). Of those articles published in 2008/2009 and involving primary data ($n = 125$), 1 (0.8%) contained a statement or link regarding the availability of processed data such as audio recordings, transcriptions or data provided by experimental software. Of those articles involving secondary data ($n = 50$), 2 (4%) articles contained such a statement or link. Of those articles published in 2018/2019, 5 ($n = 139$, 3.6%) involving primary data and 4 ($n = 47$, 8.5%) involving secondary data contained such a statement or link.

Of the 12 articles for which processed data were reportedly available, 3 did not have processed data that was actually accessible, because of broken links or availability "upon request". Of the articles that we could access ($n = 9$), the processed data were made available in a journal-hosted supplement ($n = 4$), on a personal or institutionally hosted (non-repository) webpage ($n = 1$), or in an online third-party repository ($n = 4$). Of those articles, 3 data sources contained data documentation. When examining raw data and processed data together, 44 studies included raw data but no processed data, 6 included processed data but no raw data, and 6 included both processed and raw data.

3.5 Analysis-script availability

For the articles published in 2008/2009 and involving primary or secondary data ($n = 175$), no analysis scripts were made available. For the articles published in 2018/2019 ($n = 188$), an

analysis script was shared for 3 articles (1.5%). The scripts were made available in a journal-hosted supplement ($n = 3$).

3.6 Preregistration

No articles in the included dataset ($n = 519$) reported a preregistered study.

3.7 Replication

Of the 360 articles that reported empirical studies, 4 articles (1.1%) were replications, according to the definition used in Kobrock and Roettger (2023).[5] Following Marsden, Morgan-Short, et al. (2018), observed replication studies were categorized post hoc by classifying them according to the number of changes made. As a result of this classification, all four were conceptual replications.

3.8 Conflict-of-interest statements

Of the 519 included articles, 52 included a statement about conflicts of interest (10%). All of these articles stated that there was no conflict of interest.

4. Discussion

4.1 Transparency practices are not widely adopted

Our assessment of transparency and reproducibility-related research practices in a random sample of 519 linguistic articles included in the study shows that approximately one third of the coded articles were publicly available, and important components of empirical research – including materials, raw data, and analysis scripts – were rarely made publicly available alongside them. We observed that sharing materials and raw data in articles with secondary data analyses was more common compared to raw data sharing in primary data studies and sharing of processed data and analysis scripts across categories. None of the articles in our sample contained a preregistration of hypothesis, data collection, or analysis, 10% contained a conflict of interest statement, and only 1% reported replication studies, none of which were direct replications. In the following, we will discuss the results in detail, compare them to recent assessments in other disciplines, and offer suggestions for ways forward. A summary overview of all recommendations is provided in **Table 3**.

---

[5] Following Kobrock and Roettger (2023), we first searched for the search string "replicat" and if there was a hit, we examined the title and the abstract of the paper, the text before and after occurrences of the search term "replicat", the paragraph before the Methods section as well as the first paragraph of the Discussion section. If the authors explicitly claimed that (one of) their research aim(s) was to replicate the result or methods of an initial study, this article was treated as a replication.

Table 3: Suggestions for the improvement of transparency and reproducibility practices in linguistics

	For researchers	For journals	For institutions and funders
Open Access	Use Open Access repositories for preprints and postprints, e.g., LingBuzz, EdArXiv, PsyArXiv, or find them via registries OpenDOAR and COAR for preprints	Be transparent about Open Access Policy; Allow self-archiving without restrictions	Have in place and make visible institutional policy on copyright or right of second publication; Make it required; Provide funding
Sharing Materials	Share materials on public general purpose repositories, e.g., Zenodo, OSF, Dataverse, or discipline-specific repositories, e.g., TROLLing or IRIS	Explicitly encourage it; Make it required where possible; Make it visible by assigning Open Materials badges to authors	Provide training; Explicitly encourage it; Make it required where possible
Sharing Data & Analyses	Share data and analyses scripts on public general purpose repositories, e.g., Zenodo, OSF, Dataverse, or discipline-specific repositories, e.g., TROLLing, IRIS, or CLARIN ERIC For personal/sensitive data: anonymize the data, ask for data sharing in informed consent, declare concerns why data cannot be shared, share at least the metadata	Explicitly encourage it; Make it required where possible; Make it visible by assigning Open Data and Open Analysis badges to authors	Provide training; Explicitly encourage it; Make it required where possible
Preregistration	Preregister the study on OSF, AsPredicted, or similar platforms; Submit preregistration as Registered Report to a journal	Allow for Registered Reports	Provide training; Explicitly encourage it; Provide examples and templates for preregistration if needed

(Contd.)

	For researchers	For journals	For institutions and funders
Conflict of interest	Declare conflict of interest even if not mandated and if none exist	Explicitly encourage it; Make it required; Include the COI information in published articles	Raise awareness of the importance of COI declaration
Replication	Establish resource-efficient ways to identify replication targets and conduct replication studies	Explicitly encourage it; Publish replication studies without barriers	Provide funding for replication studies; Reward replication studies in assessment criteria
Overall	Implement practices; Develop skills and keep yourself informed; Teach your students; Collaborate with others to increase awareness and develop standards	Implement guidelines and policies for transparent and reproducible studies and explicitly encourage them; Suggest standard phrases to facilitate meta-analyses	Increase awareness; Provide training, resources and incentives; Implement guidelines and policies for transparent and reproducible research

4.1.1 Open access

For 35.5% of sampled papers, we could access a publicly available version (open access). Although lower than the recent assessments in psychology (65%; Hardwicke et al., 2022) and social sciences (40%; Hardwicke et al., 2020), this is higher than similar assessments in biomedicine (25%; Wallach et al., 2018) and a recent cross-disciplinary assessment (28%; Piwowar et al., 2018). Note that if we compare the open access rate of the articles sampled only in the late time window (2018/2019) in the current study (42% of the post-RC sample), the rate becomes more similar to the recent assessments in social sciences.

Even though we have observed an increase in open access rates from 28% to 42% between the early and the late time windows in the current sample, still over half of articles from 2018/2019 did not have a publicly available version. Limited access to scientific articles decreases the value of academic papers and makes it difficult for other researchers, policymakers, and the general public to evaluate and reuse the research outcomes. The observed rate of open access articles suggests that there are still barriers to open publishing in our field. A possible lightweight solution is to share versions of the manuscript on preprint/postprint servers such as LingBuzz (https://lingbuzz.net/), EdArXiv (https://edarxiv.org/), PsyArXiv (https://psyarxiv.com/), and other discipline-specific repositories or institutional repositories at the researcher's own university. OpenDOAR (https://v2.sherpa.ac.uk/opendoar/) and COAR for preprints (https://doapr.coar-repositories.org/) are registries where one can find such discipline-specific open access

repositories for publications. These resources allow researchers to make manuscripts publicly available before or (even years) after publication, a practice that is usually explicitly allowed by journals. Once authors have transferred exclusive rights for publishing to the publisher, they will have to verify with the publisher's policy, however, which version of their manuscript they are still allowed to share and where. The Sherpa-Romeo database (https://v2.sherpa.ac.uk/romeo/) is a useful service in that regard.

4.1.2 Materials sharing

Only 16% of articles in our sample stated that research materials were available. There is a much higher sharing rate of materials for articles reporting primary data analyses (20%) than secondary data analyses (5%). This asymmetry might be due to the prevalence of corpus studies under Secondary data, which might not always have any materials to share (according to our definition). Primary data acquisition, on the other hand, with the majority being experiments or observational studies, usually requires instruments, stimuli, and/or software. For example, researchers collecting primary data might share experimental stimuli used in the study, such as images or videos shown to the participants, or observational instruments, such as surveys, interview guides, and questionnaires for others to reuse. In the current sample, around 20% of articles with primary data analyses reported available materials (pre-RC: 14.8%; post-RC: 24.3%), which is slightly higher than estimates from sociology (11%; Hardwicke et al., 2020) and psychology (14%; Hardwicke et al., 2022), but lower than estimates from biomedicine (33%; Wallach et al., 2018).

Sharing research materials is important for several reasons, but maybe most importantly it allows for direct and conceptual replication attempts, enabling other researchers to verify and generalize scientific claims (Zwaan et al., 2018). It also increases efficiency, because other researchers can reuse materials instead of having to recreate them (Chalmers & Glasziou, 2009). This sharing, in turn, can contribute to advancing subfields of linguistics more efficiently and in a collaborative effort (Gawne et al., 2017). Nowadays, publicly sharing research materials can be achieved by a few mouse clicks if these resources are not restricted for sharing, e.g., due to copyright. Linguists can share research resources on free-to-use third-party repositories that allow usage tracing and permanent links, such as the Open Science Framework (https://osf.io/) or Zenodo (https://zenodo.org/). Additionally, data curation websites, such as in Dataverse repositories (https://dataverse.org/), are also available for materials sharing. Instead of using general purpose repositories, linguists could also deposit their data with TROLLing, a discipline-specific instance of Dataverse for language research data (https://dataverse.no/dataverse/trolling), or IRIS for data from research in bilingualism and language teaching (https://www.iris-database.org/), or their institution's data repository (if existent). To facilitate materials sharing, journals can incentivize these practices by either making them mandatory at submission and/or

by offering publicly visible open science badges, a practice that has been shown to increase the sharing of materials and data (Hardwicke et al., 2018; Kidwell et al., 2016; Laurinavichyute et al., 2022).[6]

4.1.3 Raw data, processed data, and script sharing

Our study suggests that raw data were rarely shared for studies with primary data (4.9%), but more often shared for studies with secondary data (38.1%). The higher sharing rate for secondary data analysis can be explained by the predominance of corpus studies that used either publicly available or published corpora for their analysis. In these cases, availability of data (i.e., the corpus) is often indicated via a reference to previous literature or an existing online resource. The rate of data sharing for primary analyses in linguistics is in line with evidence that data underlying scientific claims are rarely shared (Hardwicke & Ioannidis, 2018; Iqbal et al., 2016).[7] Sharing raw data is nevertheless critical: it enables the evaluation and verification of underlying claims and allows for the evaluation of empirical, computational, and statistical reproducibility (LeBel et al., 2018). It allows for alternative analyses to establish analytic robustness (Steegen et al., 2016) and strengthens attempts to synthesize evidence via meta-analyses (Nicenboim et al., 2018).

The observed primary data sharing rates need to be interpreted in light of the fact that linguistic raw data are often audio and video recordings that can be difficult to share for technical but also legal reasons, as they are often tightly linked to the identity of the sample studied. This relates to commonly perceived barriers for sharing certain types of data, due to ethical, legal, or sometimes technical concerns in different fields of research (Cychosz et al., 2020; Gomes et al., 2022). For example, speech and video recordings are difficult, if not impossible, to anonymize, and the identity of participants in sociolinguistic survey data is difficult to mask, due to large amounts of demographic information that, combined, can lead to indirect identification of research subjects. Sharing raw data becomes impossible if one cannot anonymize the data sufficiently, or if participants in the study have not given their consent to share their personal data. Some countries also do not allow the anonymization of data without the explicit consent of the participants. On the other hand, while applicable to some studies, these concerns do not apply across the board (Meyer, 2018), especially if anonymization of raw or processed data is possible. Moreover, identifying information can be removed by techniques such as synthetic data creation (Quintana, 2020), and there are various tools available that allow for an automatic

---

[6] Although open science badges do not guarantee that the resources are actually accessible (Crüwell et al., 2022).

[7] Instead of sharing data, it is common practice for some authors to state that data is available "upon request" (Colavizza et al., 2020). However, several studies have shown that data requested is rarely made available (Tedersoo et al., 2021). For example, Hardwicke and Ioannidis (2018) were unable to obtain 68% of 111 highly cited psychology articles (see also Vanpaemel et al., 2015).

anonymization of both qualitative (e.g., QualiAnon) and quantitative (e.g., amnesia, sdcMicro) data. In the cases when neither anonymization nor synthetic data creation applies, authors should transparently declare these concerns in their manuscript to inform the reader why data is not shared (Morey et al., 2016), and share the metadata (i.e., information about the data) in order to increase the findability and transparency of their study. A recently published *Open Handbook of Linguistic Data Management* provides extensive information on different ways that linguistics data can be prepared for archiving and sharing (Berez-Kroeker et al., 2022).

Our assessment of sharing processed data and/or analysis scripts revealed only some occurrences of available resources (3.3% and 0.8%, respectively), suggesting that it is not yet a common practice in the field. This rate is in line with assessments from other fields (Hardwicke et al., 2020, 2022; Rowhani-Farid & Barnett, 2018; Wallach et al., 2018). A recent assessment of data and code availability in the *Journal of Memory and Language*, however, revealed a higher sharing rate for psycholinguistic studies published in this specific journal, where even prior to the journal mandating data and code sharing, data were shared in 20% of articles and code in 12% of articles, percentages that increased to 51% and 32%, respectively, after the mandate (Laurinavichyute et al., 2022). One possible explanation for more data and code sharing in psycholinguistics studies could be the close proximity to psychology, a discipline where the reproducibility crisis has been more broadly acknowledged and the need for increased transparency has been widely communicated (see, for example, Open Science Collaboration, 2015). It also showcases the influence a journal's policy can have on sharing practices in the community. See, however, the need for regular quality checking of data availability statements in journals, based on a recent assessment of open data sharing (Towse et al., 2021).

While sharing raw data might pose many challenges due to the presence of participants' identifying information, processed data might be easier to anonymize (e.g., by categorizing or binning variables such as participants' age), alleviating legal barriers to data sharing. In addition, a processed data table and a step-by-step description of the analysis (in form of a script or instructions for point-and-click software) are the minimal requirements for other researchers to reproduce the results and evaluate how researchers arrived at their conclusions based on the available data. This resource is important, because humans are error-prone (Nuijten et al., 2016), and methodological descriptions are often too vague and lack the detail to recreate the data analysis pipeline (Hardwicke et al., 2018; Laurinavichyute et al., 2022). Again, as with materials sharing, data and analysis protocols can be shared through free third-party repositories like the Open Science Framework, Dataverse, or discipline-specific repositories such as TROLLing, and language resources made available through CLARIN ERIC (https://www.clarin.eu/). Journals and funding agencies can facilitate this practice by mandating data availability (where data sharing is possible) or giving more visibility to these efforts, e.g., through open science badges.

4.1.4 Preregistration

None of the articles in the current sample reported a study preregistration. A preregistration implies a time-stamped document, describing hypotheses, data collection procedure, and analysis plan, stored on an independent repository before data collection or analysis commences. Preregistrations draw an explicit line between exploratory and confirmatory analyses (Wagenmakers et al., 2012), make questionable research practices such as selective reporting and HARKing (hypothesizing after results are known) detectable (John et al., 2012), and help reduce researcher degrees of freedom (Mertzen et al., 2021; Roettger, 2021; Simmons et al., 2011; Wicherts et al., 2016; but see also Szollosi et al., 2020, for other perspectives on the role of preregistration in research). The absence of preregistrations in our early sample (2008/2009) is to be expected, since the concept of preregistrations in behavioral sciences is relatively new (although it had been around for awhile in clinical research, see Dickersin & Rennie, 2012). Other meta-scientific assessments found that preregistrations were as rare as 0–1% in social sciences (Hardwicke et al., 2020) and biomedical sciences (Wallach et al., 2018), and more recent assessments found preregistration rates of around 3% in psychology (Hardwicke et al., 2022). However, the complete absence of any preregistered study in the late time window (2018/2019) in the current study suggests that efforts to advocate for preregistration have not yet been successful in our field. Note, however, that because there is a time lag between preregistering a study and its publication, it is possible that the time frame investigated was slightly too early to see the first adopters in linguistics.

Preregistrations can be logged using templates provided on websites such as the Open Science Framework Registry (https://osf.io/registries) and AsPredicted (https://aspredicted.org/). Additionally, more and more journals offer specific article types called Registered Reports, which are peer-reviewed preregistrations (Nosek & Lakens, 2014). If the study design is approved by reviewers and editors, publication of the study is, in principle, accepted prior to seeing the results (see also Roettger, 2021, for a discussion of this publishing format specifically in the context of linguistics). Several journals in the field have already implemented this new article type (e.g., *Laboratory Phonology, Language & Cognition, Cognitive Linguistics, Journal of Memory and Language, Glossa Psycholinguistics,* see https://cos.io/rr for an exhaustive list).

4.1.5 Conflict of interest

Our investigation suggests that linguistic articles were less likely to include conflict-of-interest (COI) statements (10%) than the social sciences (15%; 12), psychology (39%; 13), and the biomedical sciences (65%; 42). One explanation for the observed rate of COIs might be that mandating COIs is not a common practice in linguistics journals compared to other fields. Alternatively, journals might ask about COIs during submission, but not include this information in the final published article. Disclosing potential conflicts of interest in research articles informs readers about possible risks of bias (Bekelman et al., 2003; Cristea & Ioannidis, 2018). Without

journal mandates, authors may falsely assume that such statements are not relevant to them (Chivers, 2019). Articles should ideally always include a conflict of interest statement, even if it is only to explicitly declare that there were no funding sources and no potential conflicts of interest in the first place. Journals can facilitate adoption of COIs by mandating them and including them in the final published articles. This inclusion is particularly relevant in a field like linguistics, which has a wide reach in applied practice, possibly affecting clinical, educational, and technological applications.

4.1.6 Replications

Of the experimental articles we examined, 1% claimed to be a replication study, a rate that is in line with a recent assessment of the prevalence of replications in experimental linguistics (2.5%, Kobrock & Roettger, 2023), as well as previous estimates in psychology (5%, Hardwicke et al., 2022; 1.6%, Makel et al., 2012), social science (1%, Hardwicke et al., 2020), biomedicine (5%, Wallach et al., 2018), educational science (0.1%, Makel & Plucker, 2014), and economics (0.1%, Mueller-Langer et al., 2019). Because evidence provided by a single study is limited (Amrhein et al., 2019) and some theories are built on a rapidly growing body of experimental evidence, it is of critical importance to evaluate and substantiate existing findings in the literature. Scientists are trained to ensure the reliability and generalizability of scientific findings by conducting direct replication studies, i.e., studies that aim to arrive at the same scientific conclusions as an initial study by collecting new data and completing new analyses but using the same methodology. Thus, it is important to quantify this practice.

However, while a rate of 1% might appear low, it remains unclear what a desired replication rate actually looks like. An answer to this question is likely complex and weighs the possible replication value of the published studies against the certainty of their underlying claims (Isager et al., 2021). To facilitate higher replication rates, funding agencies, journals, but also editors and reviewers, need to start valuing direct replication attempts as much as they value novel findings, also in language research. For example, journals could dedicate space to direct replications (e.g., as their own article type or journal section). Researchers, on the other hand, need to establish resource-efficient ways to identify replication targets (Isager et al., 2021) and conduct them (Frank & Saxe, 2012; Open Science Collaboration, 2015). Finally, even though desired rates of replications would be dependent on the assessment of replication value of the studies in the field, each research study should aim for maximum transparency and reproducibility, where possible (respecting the legal and ethical limitations to sharing research outputs).

4.2 Transparency practices are only slowly adopted

The observed rates of transparency and reproducibility practices suggest that open science is still not the norm in linguistics, similar to assessments and conclusions in other disciplines. However,

one of the main aims of the current study was not only to assess transparency practices in linguistics in general, but also to measure how the adoption of these practices has changed over time. Thus, we selected two main time windows for this comparison, one before the so-called reproducibility crisis was widely acknowledged (2008/2009) and one after (2018/2019). We observed that some transparency practices, such as preregistration, have not been adopted in linguistics in either the early time window or the late time window ($n = 0$ preregistered studies in the study sample). Other practices, such as data, materials, and analysis sharing, showed an increase in the last decade. The availability of the study materials increased from 11.2% to 20.3% from the pre-RC to the post-RC time window. Raw data sharing changed from 34% to 42.6% for studies with secondary data; however, primary data sharing dropped from 6.5% to 3.6%. Availability of processed data increased for both primary data (from 0.8% to 3.6%) and secondary data (from 4% to 8.5%). No analysis scripts were available in the studied sample for the pre-RC time window ($n = 0$), but a small number of analyses was made openly available in the post-RC sample ($n = 3$, 1.6% of the collected sample).

Overall, the observed rates suggest that linguistics did not widely adopt transparency practices between the years 2008/2009 and 2018/2019. One possibility is that adoption of these practices has started later in linguistics compared with many disciplines within the social or natural sciences. In this case, the slow uptake of transparency practices observed in the present study between the selected time points would reflect late acknowledgement of the reproducibility and replication crisis in linguistics, with its effects becoming visible only later. That is, the chosen late time window (2018/2019) might not have picked up the practices that are being currently adopted in the field. For example, only more recently have researchers representing many different subfields of linguistics started developing data citation standards for the field, in order to increase the transparency and reproducibility of linguistics research (Berez-Kroeker et al., 2018; Tromsø Recommendations for Citation of Research Data in Linguistics, 2019). Another possibility is that the observed slow uptake reflects the complexity and diversity of the field. For example, only a few areas of linguistics overlap with fields that have a large amount of meta-scientific research on reproducibility and transparency practices, high awareness of the reproducibility and replication crisis, and a high rate of journal requirements for open data and materials, such as psychology (e.g., psycholinguistics) and neuroscience (e.g., neurolinguistics). This close proximity could facilitate the adoption of transparency practices in these few subfields through greater training opportunities and greater availability of specific open science practices, infrastructures, and other resources that make the uptake of open science easier. Many areas of linguistics, however, are still lacking necessary open science infrastructure and resources, making transparency and reproducibility difficult to implement. Many areas are also lacking journals that would implement open science policies or incentivize transparency practices. In many subfields of linguistics, it is also still common practice to incentivize article publications

far more than the sharing and publishing of datasets and analysis scripts (e.g., in hiring and promotion decisions; for an example, see Promotion & Tenure Guidelines – American Association For Applied Linguistics, n.d.), subsequently discouraging researchers from dedicating time and resources to preparing materials, data, and analyses for sharing. This incentive alignment could lead to significantly fewer transparency and reproducibility practices in many subfields, and consequently affect the observed rate of open science adoption in language research more generally.

Regardless of the reasons for the slow uptake, we believe the current scientific landscape presents an excellent opportunity for linguistics to keep moving forward with the implementation of transparent and reproducible research practices. The adoption of these practices requires a multipronged approach: from educators to train the next generation of scientists, from researchers to implement these practices within their workflow, from departments to support with education, resources, and incentives (such as changes in the research assessment systems), and from journals and funders to implement new guidelines (see, for example, Transparency and Openness Promotion Guidelines (TOP), https://www.cos.io/initiatives/top-guidelines).

4.3 Linguistics is still WEIRD

Besides limited transparency and reproducibility practices across linguistics, we also observed that language research shows a lack of openness to diverse languages and populations. The data on Authors and Languages in the current sample (see **Figure 1**) suggests that the field is biased toward WEIRD (Western, Educated, Industrialized, Rich, and Democratic) research. Linguistics, as a whole, might be unique in that it suffers from biases at two levels: when sampling from populations and when sampling from languages.

Here we add evidence that published linguistic research overrepresents work on a few languages, most notably Indo-Germanic languages (60.4%), with a focus on English (40.5%), as calculated across all languages in all sampled articles ($n = 600$). The focus on English has not changed much throughout the last decade, with only a small decrease of studies on English language from 2008/2009 (44%) to 2018/2019 (37%) in the current sample. This lack of diversity leads to generalizations over human language based on an arbitrary and historically biased language sample. Non-Indo-European languages and language families are either under-investigated or not investigated at all. This trend has been recently more closely examined in language acquisition research, showing that the languages under investigation were highly skewed towards English and other Indo-European languages, and represented only approximately 1.5% of the world's languages, based on data from four journals (Kidd & Garcia, 2022a, 2022b). The largest international corpus of child language, CHILDES, also shows an overrepresentation of English and other Indo-European languages (MacWhinney, 2007). Recent work has also shown an over-reliance on English in language and cognition research or cognitive science more broadly

(Blasi et al., 2022). This focus is problematic, because it leads to bias: researchers' prior beliefs as to how the next language manifests a certain communicative phenomenon might be biased by their preconceptions about the languages (Majid & Levinson, 2010) and cultures (Henrich, 2020) with which they are most familiar.

Looking at the characteristics of languages investigated in linguistics research (rather than the characteristics of people sampled), some authors argued that a focus on Western languages, and English in particular, has meant that much of linguistic theory is underpinned by Western/ English-centric concepts which may not apply generally to all languages (Evans & Levinson, 2009; Majid & Levinson, 2010). If other researchers are basing the phenomena that they study on this body of theory, naturally they will be exploring concepts that are relevant primarily to English and other Western languages, neglecting to do research on phenomena and concepts which may have particular relevance to non-WEIRD languages and language groups. This focus is an often discussed issue in and beyond linguistics (Gil, 2001; Goddard & Wierzbicka, 2014; Levisen, 2018; Wierzbicka, 2009), but despite some observations and calls for change made in a few subfields of linguistics (Andringa & Godfroid, 2020; Bergs, 2021; Blasi et al., 2022; Joshi et al., 2020; Kidd & Garcia, 2022a), to our knowledge, the present study is the first quantitative assessment of this bias based on a random sample across linguistics.

In addition to a bias toward certain target languages, our study also quantified the bias toward institutions from Western countries, in particular, Northern America (31%) and Europe (34%), and most noticeably Anglo-Saxon countries. Out of all sampled articles ($n = 600$), 211 had as their corresponding author a researcher located at an institution from either the United States of America (27%) or the United Kingdom (8%), making up over one third of the sample. A similar high skew towards author affiliations in Western countries was observed in a recent survey on language acquisition research, with 49% of papers with authors from North America and 38% from Europe (Kidd & Garcia, 2022a).

It is important to acknowledge here that this assessment has limits. Our sampling procedure itself introduced parts of the observed biases: First, we used a particular database to sample from (Scopus) which might come with its own selection biases as to what research is listed and indexed and how articles are categorized into relevant disciplines (i.e., more English-based research from the Global North may be indexed, see Tennant, 2020). Second, we restricted papers to those written in English, biasing our sample towards Anglo-Saxon authors or authors that have had the privilege of training in academic writing in English.

Despite this caveat, we believe that opening up linguistic research to more diverse populations and language samples as well as increased open publishing can facilitate removing barriers between academia and the wider society through increased access to research, and provide more representative and generalizable evidence in language studies (Andringa & Godfroid, 2020).

4.4 Meta-scientific assessment in linguistics is challenging

While it is the case that, based on the current results, linguistics seems to have been slowly adopting open science practices, we believe this is, in part, driven by the very nature of the field, which is broad, cross-disciplinary, and touts diversity of both data and study designs. In some research institutions, linguistics is included as part of the humanities, while in others it is often classified as a social science, and, for some, it represents a hypothesis-driven experimental science housed with other "hard" sciences. Linguistics is a methodologically diverse field with a tradition of exploratory and introspective research (Grieve, 2021). Studies can range from introspections, to vast qualitative and quantitative explorations of spoken and written corpora, up to highly controlled experiments. Much of the data collected in the distinct subfields of linguistics necessarily deal with identifiable audio/video recordings and/or transcripts, and the analyses conducted on these data can be qualitative, quantitative, or somewhere in between. Only very recently has there been an effort to standardize data citation and attribution practices to increase transparency about data sources across various subfields of linguistics and the reproducibility of linguistic research more generally (Berez-Kroeker et al., 2018). This diversity implies that many of the research practitioners in linguistics have vastly distinct training, and we believe this heterogeneity might at least partially explain the overall slow adoption of open science practices we observed. This diversity also came with its practical challenges during our coding of articles. That is, our sample reflects a wide range of subdisciplines and, consequently, certain assessments were either difficult to compare across subdisciplines or not applicable at all. The subset of our sample that is possibly most comparable to previous work is the subset of empirical linguistic studies, which is by no means representative of linguistics at large.

During our coding of articles, extracting information was also challenging due to a lack of standardized practices as to how to report on sharing (or not sharing) materials. This issue makes any meta-scientific assessment difficult and drastically reduces discoverability and reuse of resources. Our assessment of transparency practices relied on at least two assumptions. First, we assumed that the authors explicitly disclosed if they shared resources and if so, how. Second, we assumed that if transparent practices were disclosed, the coders were able to extract and interpret this information. Neither of these assumptions must hold; thus, any rates that are generated here are necessarily only proxies of the true rates of transparency practices. Not being able to detect, find, or reuse sources that are actually shared obviously reduces their value for the field beyond meta-scientific assessments.

In order to facilitate sharing practices and the discoverability of shared resources in the future, solutions have to be implemented at the journal, institutional, and funder policy levels. For example, the introduction of machine-readable, highly formalized method sections could help to ensure that readers can find, scrutinize, and reuse shared resources (Lakens & DeBruine,

2021). Adoption of explicit disclosure phrases, such as the 21-word solution proposed by Simmons et al. (2012), could help standardize method-relevant practices. Moreover, implementing TOP guidelines (see https://www.cos.io/initiatives/top-guidelines) could aid future meta-scientific assessments in order to track progress and guide policy making.

5. Conclusions

The study of language extends beyond the investigation of how we communicate and touches on important social, cultural, cognitive, technological, and biomedical aspects of the uniqueness of human nature and behavior. As the field has grown in scope and became increasingly empirical and quantitative linguistics is now faced with the challenges and limitations of the scientific practices that pose barriers to reproducibility and replicability.

One of the solutions proposed to the widely acknowledged reproducibility and replicability crisis has been the implementation of transparency practices, such as open access publishing, preregistering research plans before data collection or analysis, sharing study materials, data, protocols, and analysis scripts, performing study replications, as well as declaring possible conflicts of interest. In the present study, we have randomly sampled 600 journal articles from linguistics and assessed the prevalence of these practices in the sampled literature. In line with similar studies in other disciplines, we found that one third of the articles were published open access, and the rates of sharing materials, data, and analyses were under 10%. We also observed 1% of replications and 10% of conflict-of-interest reporting, along with no preregistrations in the studied sample. These rates have not increased noticeably between 2008/2009 and 2018/2019, pointing to remaining barriers to the adoption of open and reproducible research practices in linguistics. We conclude that, similar to other recently assessed fields, such as psychology, social sciences, and biomedicine (Hardwicke et al., 2020, 2022; Wallach et al., 2018), linguistics has not yet firmly established transparency and reproducibility as guiding principles in research.

The adoption of these principles and the subsequent implementation of transparency and reproducibility practices in the field can be facilitated by making it easy for individual researchers to register, share, and publish their outputs as well as apply relevant practices to make their research more reproducible. This implementation can be further facilitated by incentivizing transparent and reproducible practices by stakeholders (e.g., journals, funders, and institutions), in order to reward researchers for transparency. Finally, meta-scientific research should serve as an evidence base for guiding further implementation of these practices. In this spirit, the present work aims to help track progress over time in linguistics, enable cross-disciplinary comparisons of transparency practices, and aid in facilitating the adoption of open science in research more broadly.

Data accessibility statement

All materials, data and analysis scripts related to this study are publicly available on Open Science Framework: https://doi.org/10.17605/OSF.IO/ZX9KY. To facilitate reproducibility, the results can be re-run online in the stable Code Ocean container that captures the computational environment in which study analyses were conducted: https://codeocean.com/capsule/9832712/tree/v2.

Competing interests

Some authors of this paper have engaged with meta-scientific research before and have been actively involved in methodological debates about scientific practices, including transparency in and outside of linguistics. **AB** works at the Open Research Section at the University of Oslo Library and as such is responsible for teaching and implementing open and reproducible research practices across the university and is regularly invited and occasionally compensated for teaching on open and reproducible research at other higher education institutions. **JC** was compensated to teach invited workshops about topics related to reproducible and transparent research by higher education institutions, such as the Royal Holloway University of London, Penn State, and Texas Tech. **MR** is a member of the Open Science Services of the University Library at the University of Zurich and as such is responsible for supporting researchers in data management, open data and open access. **EB** has been compensated to teach open science topics at conferences, workshops, and on her YouTube channel. **TR** was compensated to teach invited workshops and summer schools about topics related to reproducible and transparent research by higher education institutions such as the University of Cologne, the University of Birmingham, and the Centre national de la recherche scientifique (CNRS). All other authors have no competing interests.

Authors' contributions

AB: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – original draft, Writing – review & editing. **LK**: Conceptualization, Data Curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **CH**: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **JC**: Conceptualization, Data Curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **KC**: Conceptualization, Investigation, Methodology, Writing – review & editing. **IAC**: Conceptualization, Investigation, Methodology, Writing – review & editing. **MR**: Conceptualization, Methodology, Writing – review & editing. **EB**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing – original draft, Writing – review

& editing. **TR**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Visualization, Writing – original draft, Writing – review & editing.

Erin M. Buchanan and Timo B. Roettger are shared last author.

References

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician, 73*(sup1), 262–270. DOI: https://doi.org/10.1080/00031305.2018.1543137

Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics, 40*, 134–142. DOI: https://doi.org/10.1017/S0267190520000033

Arel-Bundock, V., Yetman, C. J., Enevoldsen, N., & Meichtry, S. (2022). *countrycode: Convert country names and country codes* (1.4.0) [Computer software]. https://CRAN.R-project.org/package=countrycode

Bekelman, J. E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *JAMA, 289*(4), 454–465. DOI: https://doi.org/10.1001/jama.289.4.454

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., Dubinsky, S., Meier, R. P., Thieberger, N., Rice, K., & Woodbury, A. C. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics, 56*(1), 1–18. DOI: https://doi.org/10.1515/ling-2017-0032

Berez-Kroeker, A. L., McDonnell, B., Koller, E., & Collister, L. B. (Eds.). (2022). The Open Handbook of Linguistic Data Management. The MIT Press. DOI: https://doi.org/10.7551/mitpress/12200.001.0001

Bergs, A. (2021). The problem of universalism in (diachronic) cognitive linguistics. *Yearbook of the German Cognitive Linguistics Association, 9*(1), 177–188. DOI: https://doi.org/10.1515/gcla-2021-0009

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences, 26*(12), 1153–1170. DOI: https://doi.org/10.1016/j.tics.2022.09.015

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology, 1*, 223–249. DOI: https://doi.org/10.1146/annurev-devpsych-121318-085037

Boland, M. R., Karczewski, K. J., & Tatonetti, N. P. (2017). Ten simple rules to enable multi-site collaborations through data sharing. *PLOS Computational Biology, 13*(1), e1005278. DOI: https://doi.org/10.1371/journal.pcbi.1005278

Bolibaugh, C., Vanek, N., & Marsden, E. J. (2021). Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research. *Bilingualism: Language and Cognition*, *24*(5), 801–806. DOI: https://doi.org/10.1017/S1366728921000535

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P. C., Baumann, M., Baute, S., Benoit, V., Bernauer, J., Berning, C., … Żółtak, T. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences*, *119*(44), e2203150119. DOI: https://doi.org/10.1073/pnas.2203150119

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., & others. (2018). Evaluating the replicability of social science experiments in Nature and *Science* between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. DOI: https://doi.org/10.1038/s41562-018-0399-z

Casillas, J. V. (2021). Interlingual interactions elicit performance mismatches not "Compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. *Languages*, *6*(1). DOI: https://doi.org/10.3390/languages6010009

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. DOI: https://doi.org/10.1016/S0140-6736(09)60329-9

Chivers, T. (2019). Does psychology have a conflict-of-interest problem? *Nature*, *571*(7763), 20–24. DOI: https://doi.org/10.1038/d41586-019-02041-5

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, *15*(4), e0230416. DOI: https://doi.org/10.1371/journal.pone.0230416

Coretta, S., Casillas, J. V., Roessig, S., Franke, M., Ahn, B., Al-Hoorie, A. H., Al-Tamimi, J., Alotaibi, N. E., AlShakhori, M. K., Altmiller, R. M., Arantes, P., Athanasopoulou, A., Baese-Berk, M. M., Bailey, G., Sangma, C. B. A., Beier, E. J., Benavides, G. M., Benker, N., BensonMeyer, E. P., … Roettger, T. B. (2023). Multidimensional signals and analytic flexibility: Estimating degrees of freedom in human-speech analyses. *Advances in Methods and Practices in Psychological Science*, *6*(3), 25152459231162570. DOI: https://doi.org/10.1177/25152459231162567

Cristea, I.-A., & Ioannidis, J. P. A. (2018). Improving disclosure of financial conflicts of interest for research on psychosocial interventions. *JAMA Psychiatry*, *75*(6), 541–542. DOI: https://doi.org/10.1001/jamapsychiatry.2018.0382

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. L. (2022). *What's in a Badge? A computational reproducibility investigation of the Open Data Badge Policy in one issue of Psychological Science*. PsyArXiv. DOI: https://doi.org/10.31234/osf.io/729qt

Cychosz, M., Romeo, R., Soderstrom, M., Scaff, C., Ganek, H., Cristia, A., Casillas, M., de Barbaro, K., Bang, J. Y., & Weisleder, A. (2020). Longform recordings of everyday life: Ethics for best practices. *Behavior Research Methods*, *52*(5), 1951–1969. DOI: https://doi.org/10.3758/s13428-020-01365-9

Dickersin, K., & Rennie, D. (2012). The evolution of trial registries and their use to assess the clinical trial enterprise. *JAMA, 307*(17), 1861–1864. DOI: https://doi.org/10.1001/jama.2012.4230

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife, 10*, e71601. DOI: https://doi.org/10.7554/eLife.71601

Evans, N., & Levinson, N. (2009). The myth of language universals and the myth of universal grammar. *Behavioral and Brain Sciences, 32*(5), 452–453. DOI: https://doi.org/10.1017/S0140525X09990641

FORRT. (2021). *Reproducibility crisis (a.k.a. Replicability or replication crisis)*. FORRT – Framework for Open and Reproducible Research Training. https://forrt.org/glossary/reproducibility-crisis-aka-replicab/

Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science.* DOI: https://doi.org/10.1177/1745691612460686

Frermann, L., & Lapata, M. (2021). Categorization in the wild: Category and feature learning across languages. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(43). https://escholarship.org/uc/item/55v1x643

Gawne, L., Kelly, B. F., Berez-Kroeker, A. L., & Heston, T. (2017). *Putting practice into words: The state of data and methods transparency in grammatical descriptions.* http://hdl.handle.net/10125/24731

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis–a "garden of forking paths"–explains why many statistically significant comparisons don't hold up. *American Scientist, 102*(6), 460–466. DOI: https://doi.org/10.1511/2014.111.460

Gil, L. V. (2001). The European focus of the curriculum in the educational reforms in Spain at the end of the twentieth century. *Encounters in Theory and History of Education, 2.* DOI: https://doi.org/10.24908/eoe-ese-rse.v2i0.1734

Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science, 1*(1), 121–130. DOI: https://doi.org/10.1177/2515245917746500

Goddard, C., & Wierzbicka, A. (2014). Semantic fieldwork and lexical universals. *Studies in Language. International Journal Sponsored by the Foundation "Foundations of Language," 38*(1), 80–127. DOI: https://doi.org/10.1075/sl.38.1.03god

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences, 289*(1987), 20221113. DOI: https://doi.org/10.1098/rspb.2022.1113

Grieve, J. (2021). Observation, experimentation, and replication in linguistics. *Linguistics, 59*(5), 1343–1356. DOI: https://doi.org/10.1515/ling-2021-0094

Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLOS ONE, 13*(8), e0201856. DOI: https://doi.org/10.1371/journal.pone.0201856

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, *5*(8), 180448. DOI: https://doi.org/10.1098/rsos.180448

Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. A. (2022). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspectives on Psychological Science*, *17*(1), 239–251. DOI: https://doi.org/10.1177/1745691620979806

Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. A. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science*, *7*(2), 190806. DOI: https://doi.org/10.1098/rsos.190806

Henrich, J. (2020). *The WEIRDest people in the world*. Macmillan. https://us.macmillan.com/books/9780374710453/theweirdestpeopleintheworld

Hobson, H. (2019). Registered reports are an ally to early career researchers. *Nature Human Behaviour*, *3*(10), Article 10. DOI: https://doi.org/10.1038/s41562-019-0701-8

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. DOI: https://doi.org/10.1371/journal.pmed.0020124

Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., & Ioannidis, J. P. A. (2016). Reproducible research practices and transparency across the biomedical literature. *PLOS Biology*, *14*(1), e1002333. DOI: https://doi.org/10.1371/journal.pbio.1002333

Isager, P. M., van 't Veer, A. E., & Lakens, D. (2021). *Replication value as a function of citation impact and sample size* [Preprint]. MetaArXiv. DOI: https://doi.org/10.31222/osf.io/knjea

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. DOI: https://doi.org/10.1177/0956797611430953

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293. DOI: https://doi.org/10.18653/v1/2020.acl-main.560

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. DOI: https://doi.org/10.1207/s15327957pspr0203_4

Kidd, E., & Garcia, R. (2022a). How diverse is child language acquisition research? *First Language*, 01427237211066405. DOI: https://doi.org/10.1177/01427237211066405

Kidd, E., & Garcia, R. (2022b). Where to from here? Increasing language coverage while building a more diverse discipline. *First Language*, 01427237221121190. DOI: https://doi.org/10.1177/01427237221121190

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for

increasing transparency. *PLoS Biology*, *14*(5), e1002456. DOI: https://doi.org/10.1371/journal.pbio.1002456

Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, *70*, 70–85. DOI: https://doi.org/10.1016/j.wocn.2018.05.005

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W., Frank, M. C., & others. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1). DOI: https://doi.org/10.1525/collabra.158

Kobrock, K., & Roettger, T. B. (2023). Assessing the replication landscape in experimental linguistics. *Glossa Psycholinguistics*, *2*(1). DOI: https://doi.org/10.5070/G6011135

Lakens, D., & DeBruine, L. M. (2021). Improving transparency, falsifiability, and rigor by making hypothesis tests machine-readable. *Advances in Methods and Practices in Psychological Science*, *4*(2), 251524592097049. DOI: https://doi.org/10.1177/2515245920970949

Laurinavichyute, A., Yadav, H., & Vasishth, S. (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. DOI: https://doi.org/10.1016/j.jml.2022.104332

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389–402. DOI: https://doi.org/10.1177/2515245918787489

Levisen, C. (2018). Dark, but Danish: Ethnopragmatic perspectives on black humor. *Intercultural Pragmatics*, *15*(4), 515–531. DOI: https://doi.org/10.1515/ip-2018-0018

Lindsay, D. S. (2017). Sharing data and materials in *Psychological Science*. *Psychological Science*, *28*(6), 699–702. DOI: https://doi.org/10.1177/0956797617704015

Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., Jiang, N., & Halpern, B. S. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, *1*(6), Article 6. DOI: https://doi.org/10.1038/s41559-017-0160

MacWhinney, B. (2007). The Talkbank Project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora: Volume 1: Synchronic databases* (pp. 163–180). Palgrave Macmillan UK. DOI: https://doi.org/10.1057/9780230223936_7

Majid, A., & Levinson, S. C. (2010). WEIRD languages have misled us, too. *Behavioral and Brain Sciences*, *33*(2–3), 103–103. DOI: https://doi.org/10.1017/S0140525X1000018X

Makel, M. C., & Plucker, J. A. (2014). Facts Are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316. DOI: https://doi.org/10.3102/0013189X14545513

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542. DOI: https://doi.org/10.1177/1745691612460688

Malisz, Z., Henter, G. E., Valentini-Botinhao, C., Watts, O., Beskow, J., & Gustafson, J. (2020). *Modern speech synthesis for phonetic sciences: A discussion and an evaluation* [Preprint]. PsyArXiv. DOI: https://doi.org/10.31234/osf.io/dxvhc

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning, 68*(2), 321–391. DOI: https://doi.org/10.1111/lang.12286

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics, 39*(5), 861–904. DOI: https://doi.org/10.1017/S0142716418000036

McClelland, J. L., Hill, F., Rudolph, M., Baldridge, J., & Schütze, H. (2020). Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proceedings of the National Academy of Sciences, 117*(42), 25966–25974. DOI: https://doi.org/10.1073/pnas.1910416117

Mertzen, D., Lago, S., & Vasishth, S. (2021). The benefits of preregistration for hypothesis-driven bilingualism research. *Bilingualism: Language and Cognition, 24*(5), 807–812. DOI: https://doi.org/10.1017/S1366728921000031

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science, 1*(1), 131–144. DOI: https://doi.org/10.1177/2515245917747656

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C. C., Newman, D. P., Schönbrodt, F. D., Vanpaemel, W., Wagenmakers, E.-J., & Zwaan, R. A. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science, 3*(1), 150547. DOI: https://doi.org/10.1098/rsos.150547

Moshontz, H., Binion, G., Walton, H., Brown, B. T., & Syed, M. (2021). A guide to posting and managing preprints. *Advances in Methods and Practices in Psychological Science, 4*(2), 25152459211019948.

Mueller-Langer, F., Fecher, B., Harhoff, D., & Wagner, G. G. (2019). Replication studies in economics – How many and which papers are chosen for replication, and why? *Research Policy, 48*(1), 62–83. DOI: https://doi.org/10.1016/j.respol.2018.07.019

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 1–9. DOI: https://doi.org/10.1038/s41562-016-0021

Munsell, M., Oliveira, E. D., Saxena, S., Godlove, J., & Kiran, S. (2020). Closing the digital divide in speech, language, and cognitive therapy: Cohort study of the factors associated with technology usage for rehabilitation. *Journal of Medical Internet Research, 22*(2), e16286. DOI: https://doi.org/10.2196/16286

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review of Psychology, 69*(1), 511–534. DOI: https://doi.org/10.1146/annurev-psych-122216-011836

Nicenboim, B., Roettger, T. B., & Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German. *Journal of Phonetics, 70,* 39–55. DOI: https://doi.org/10.1016/j.wocn.2018.06.001

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. DOI: https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., & Lakens, D. (2014). A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. DOI: https://doi.org/10.1027/1864-9335/a000192

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226. DOI: https://doi.org/10.3758/s13428-015-0664-2

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). DOI: https://doi.org/10.1126/science.aac4716

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research*, *90*(1), 24–46. DOI: https://doi.org/10.3102/0034654319877153

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375. DOI: https://doi.org/10.7717/peerj.4375

Plonsky, L., Egbert, J., & Laflair, G. T. (2015). Bootstrapping in applied linguistics: Assessing its potential using shared data. *Applied Linguistics*, *36*(5), 591–610. DOI: https://doi.org/10.1093/applin/amu001

Promotion & tenure guidelines – American Association for Applied Linguistics. (n.d.). Retrieved June 30, 2023, from https://www.aaal.org/promotion-and-tenure-guidelines

Quintana, D. S. (2020). A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *ELife*, *9*, e53275. DOI: https://doi.org/10.7554/eLife.53275

R Core Team. (2022). *R: A language and environment for statistical computing* [Manual]. https://www.R-project.org/

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1). DOI: https://doi.org/10.5334/labphon.147

Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics*, *59*(5), 1227–1249. DOI: https://doi.org/10.1515/ling-2019-0048

Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review*, *22*(4), 944–954. DOI: https://doi.org/10.3758/s13423-014-0759-2

Rowhani-Farid, A., & Barnett, A. G. (2018). Badges for sharing data and code at Biostatistics: An observational study. *F1000Research*, *7*, 90. DOI: https://doi.org/10.12688/f1000research.13477.2

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., & others. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. DOI: https://doi.org/10.1177/2515245917747646

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. DOI: https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 word solution* (SSRN Scholarly Paper 2160588). DOI: https://doi.org/10.2139/ssrn.2160588

Starns, J. J., Cataldo, A. M., Rotello, C. M., Annis, J., Aschenbrenner, A., Bröder, A., Cox, G., Criss, A., Curl, R. A., Dobbins, I. G., & others. (2019). Assessing theoretical conclusions with blinded inference to investigate a potential inference crisis. *Advances in Methods and Practices in Psychological Science*, *2*(4), 335–349. DOI: https://doi.org/10.1177/2515245919869583

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. DOI: https://doi.org/10.1177/1745691616658637

Stewart, S., Rinke, E. M., McGarrigle, R., Lynott, D., Lunny, C., Lautarescu, A., Galizzi, M. M., Farran, E. K., & Crook, Z. (2020). *Pre-registration and registered reports: A primer from UKRN* [Preprint]. Open Science Framework. DOI: https://doi.org/10.31219/osf.io/8v2n7

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., Rooij, I. van, Zandt, T. V., & Donkin, C. (2020). Is preregistration worthwhile? *Trends in Cognitive Sciences*, *24*(2), 94–95. DOI: https://doi.org/10.1016/j.tics.2019.11.009

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, *8*(1), Article 1. DOI: https://doi.org/10.1038/s41597-021-00981-0

Tennant, J. P. (2020). Web of Science and Scopus are not global databases of knowledge. *European Science Editing*, *46*, e51987. DOI: https://doi.org/10.3897/ese.2020.e51987

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, Chris. H. J. (2016). The academic, economic and societal impacts of Open Access: An evidence-based review. *F1000Research*, *5*, 632. DOI: https://doi.org/10.12688/f1000research.8460.3

Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora's Box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. *Behavior Research Methods*, *53*(4), 1455–1468. DOI: https://doi.org/10.3758/s13428-020-01486-1

Tromsø recommendations for citation of research data in linguistics. (2019, December 3). *RDA*. DOI: https://doi.org/10.15497/RDA00040

Valentine, K. D., Buchanan, E. M., Cunningham, A., Hopke, T., Wikowsky, A., & Wilson, H. (2021). Have psychologists increased reporting of outliers in response to the reproducibility crisis? *Social and Personality Psychology Compass*, *15*(5), e12591. DOI: https://doi.org/10.1111/spc3.12591

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra*, *1*(1), 3. DOI: https://doi.org/10.1525/collabra.13

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language, 103*, 151–175. DOI: https://doi.org/10.1016/j.jml.2018.07.004

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology, 3*(1), 1. DOI: https://doi.org/10.1525/collabra.74

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632–638. DOI: https://doi.org/10.1177/1745691612463078

Wallach, J. D., Boyack, K. W., & Ioannidis, J. P. A. (2018). Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLOS Biology, 16*(11), e2006930. DOI: https://doi.org/10.1371/journal.pbio.2006930

Wang, J., Pan, C., Jin, H., Singh, V., Jain, Y., Hong, J. I., Majidi, C., & Kumar, S. (2019). RFID Tattoo: A wireless platform for speech recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(4), 155:1–155:24. DOI: https://doi.org/10.1145/3369812

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-Hacking. *Frontiers in Psychology, 7*. DOI: https://doi.org/10.3389/fpsyg.2016.01832

Wierzbicka, A. (2009). Overcoming Anglocentrism in emotion research. *Emotion Review, 1*(1), 21–23. DOI: https://doi.org/10.1177/1754073908097179

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*, e120. DOI: https://doi.org/10.1017/S0140525X17001972