Preregistration

A preregistration stands as a meticulously timestamped document that delineates comprehensive details about a study, encompassing research questions, hypotheses, methodologies, and analytic strategies. As a pivotal facet of rigorous research, distinguishing itself from fellow open science methodologies like registered reports and preprints, a preregistration is written prior to data collection sans peer review. The depth of content detail within a preregistration spans a spectrum: in the simplest case, a preregistration can comprise merely a hypothesis or perhaps a brief description of the methods; on the other extreme, a detailed preregistration can include code, power analyses, participant exclusion criteria and beyond, supplementing the fundamental information. In this section will delve deeper into the various constituents adaptable to a preregistration, centering on its advantageous impact on linguistic research. Specifically, it will expound upon the actors involved (who), the rationale driving this practice(why), its content(what), and the procedural intricacies(how), elucidating the process of preregistering a study within the realm of linguistics.

Linguists engaged in empirical studies delve into data collection and observations to explore various dimensions of human language. These investigations span diverse areas such as corpus analysis, interview analysis, experimental analysis, and more. However, as highlighted by @roettger2021preregistration, researchers are human and humans have evolved to filter the world in irrational ways, which can lead to questionable research practices (QRPs) and other problems that may affect the replicability of published research. Preregistration emerges as a powerful instrument empowering linguists to bolster the trustworthiness and credibility of their inquiries through a systematic and predefined methodology. Moreover, beyond linguists, the practice of preregistration extends its benefits to a broader audience, encompassing students, professionals, and academics alike.

Researchers face vital decisions while designing and executing research, with an inherent flexibility involved in the process of designing and conducting the experiment, and in analyzing the results(@simmons2011false). This type of flexibility is termed as “researcher degree of freedom”. For instance, envision a phonetician intrigued by exploring lexical stress. In this scenario, the researcher might concentrate on distinct acoustic markers linked with stress, like pitch, duration, and/or intensity. Beyond selecting which acoustic markers to assess, they must pinpoint the specific realm for these measurements. Does it encompass the nucleus of stressed/unstressed syllables or the entirety of the syllable? Where exactly does the measurement occur - at the midpoint or as an average across the entire syllable? These choices wield significant influence on subsequent analyses. Preregistration serves the purpose of meticulously documenting these crucial facets of the study. Moreover, preregistration acts as a deterrent against questionable research practices (QRPs) like HARKing or p-hacking by mandating explicit declaration of decisions in the research trajectory before any data collection commences. Preregistering also acts as substantiation that researchers are conducting confirmatory rather than exploratory data analysis by delineating a clear chronology of decisions made prior to and post data collection. Additionally, an enhanced level of specificity in the preregistration compels researchers to consider facets of their study that might usually be deferred to a later stage, such as statistical analyses. This proactive approach demands more initial time investment from the researcher but significantly enhances the prospects of uncovering pivotal flaws in the study design.

The scope of preregistration extends to any facet of research deemed worthy of temporal documentation preceding the initiation of the study, and the essential components predominantly include Research Questions/Hypotheses, Methodological Framework, and Analytical Approaches.The specific elements that will encompass a preregistration document will appear to be considerably diverse dependent upon the specific domain within linguistics and the nuanced nature of the study in question.

An example illustrating this variability is as follows. Consider the scenario where a psycholinguist is poised to undertake an experiment utilizing the self-paced reading methodology. In this context, the focus of the preregistration would logically revolve around pivotal aspects, including explanations of research inquiries, formulation of hypotheses, and meticulous delineation of methodological intricacies. This encompasses characterization of participant demographics, recruitment strategies, sample size considerations, and analogous critical details. Additionally, the psycholinguist may find it imperative to explain the notable variables subject to manipulation within the experimental design, articulate planned data transformations, and analyze the chosen statistical modeling strategies for subsequent inferential analyses. Not all of the aforementioned components are equally prioritized in a preregistration.

It is important to acknowledge that incorporating the entirety of these components into a preregistration may be a formidable challenge, particularly if Open Science is a new endeavor for the researcher. In such instances, a judicious approach is permissible, with researchers encouraged to commence with elements they perceive as most valuable to their study. Over time and through iterative practice, researchers will gain a better discernment of relevance within the Open Science framework.

Many concerns regarding the implementation of Open Science or adopting preregistration revolve around the potential burden of ‘extra work.’ Conversely, preregistration is intended to streamline our workflow, fostering efficiency both in the short term and the long run. Moreover, it offers complete control over the level of detail one chooses to include. The depth of preregistration directly correlates with the effort invested; the more comprehensive the preregistration, the greater the initial workload, leading to reduced effort in subsequent stages. Presented here are select levels of detail as suggestions to help gauge the extent of information to incorporate into your preregistration (Figure 1). The shading denotes the degree of detail; darker shades signify more exhaustive information (thus requiring more work). It’s worth noting that flexibility allows for varying levels of detail in different sections.