Preregistration in Psycholinguistic Research

Enhancing Transparency and Reproducibility in Language Science

Job Schepens, Project S, SFB 1252

2025-07-16

▶ **What** is preregistration and why does it matter for linguistics?
▶ **Why** preregister psycholinguistic studies?
▶ **How** to preregister experiments, corpus studies, and fieldwork
▶ **Where** to register language research
▶ **Practical considerations** for the SFB 1252 community

Goal: Make your language research more credible and transparent

Preparatory Reading Recommendation

Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations.
Linguistics, 59(5), 1227–1249.

Why this paper?

▶ Addresses practical concerns with real examples
▶ Discusses corpus studies, eye-tracking, phonetics
▶ Recent (2021) and comprehensive

Key takeaways to focus on:

▶ Examples of researcher degrees of freedom in linguistics
▶ How to handle existing data and model convergence issues
▶ Balance between exploration and confirmation
▶ When preregistration does and doesn't apply in linguistics

What is Preregistration?

Definition

*Preregistration refers to posting a **timestamped outline**
of the research questions, hypotheses, method, and analysis plan
for a specific project **prior to data collection and/or analysis***

Key principle: Distinguish between:

▶ **Confirmatory research** (pre-planned)
▶ **Exploratory research** (data-driven discovery)

The Preregistration Spectrum

Three Levels of Preregistration:

1. **Simple:** Basic hypotheses & methods (Easy)
2. **Detailed:** Comprehensive analysis plans (Medium)
3. **Registered Reports:** Peer review before data collection (Difficult)

Preregistrations can vary from simple outlines to comprehensive analysis plans with pre-written code

Why Preregister?

Problem 1: Publication Bias

- **Null results rarely published**
- **Replication rate:** Only 1 in 400 studies
- **80% of tested hypotheses** reported as "confirmed" across 4,600 papers
- **Cross-linguistic variation** underreported

Result: Scientific record biased toward positive findings

Problem 2: Researcher Degrees of Freedom

▶ **Post-hoc acoustic measure selection**
▶ **Flexible participant exclusion criteria**
▶ **Multiple eye-tracking measures** available
▶ **Model specification** after seeing data

Consequence: False positives may mislead theory development

Why This Matters for Experimental Linguistics

Recent findings from our field:

- ▶ **Low replication rates:** Similar to psychology's "replication crisis"
- ▶ **McGurk effect replication failures:** Classic findings not always robust
- ▶ **Eye-tracking studies:** Different measures can yield different conclusions
- ▶ **Cross-linguistic assumptions:** English-based theories don't always generalize

Evidence: Roettger (2021) documents widespread analytical flexibility in linguistics

Reference: Roettger, T. B. (2021). Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics*, 59(5), 1227-1249.

Common Concern 1: "My data collection is unpredictable"

▶ **Preregister decision trees** for contingencies
▶ **Document changes** transparently
▶ **Example:** Children falling asleep during experiment

Common Concern 2: "I need exploratory analyses"

- ▶ **Preregistration only constrains** confirmatory part
- ▶ **Explore freely** after confirmatory tests
- ▶ **Just label findings** appropriately

Common Concern 3: "I'm working with existing corpora"

- ▶ **Can preregister analysis** of existing data
- ▶ **Example:** HCRC Map Task Corpus analysis
- ▶ **Reduces post-hoc** analytical flexibility

Common Concern 4: "Statistical models often fail to converge"

- ▶ **Preregister model simplification** procedures
- ▶ **Define convergence failure** handling
- ▶ **Plan for data transformation** needs

Common Concern 5: "I don't have concrete predictions yet"

- **Perfectly fine** for early-stage research
- **Explore first,** then confirm on new data
- **Frame exploratory studies** appropriately

Common Concern 6: "My field is observational, not experimental"

- **Preregistration mainly** for confirmatory research
- **Much linguistics is exploratory** by nature
- **Value different types** of inquiry equally

Key insight: Preregistration is flexible and adaptable to linguistic research

How to Preregister

What to Include: Key Questions (Roettger, 2021)

1. **Data collection:** Who, how many, where, when?
2. **Inclusion/exclusion:** Specific operational criteria
3. **Materials:** Stimulus selection and norming procedures
4. **Procedure:** Exact experimental protocol
5. **Variables:** How will constructs be measured?
6. **Statistical models:** Model formula, random effects structure
7. **Inference:** What constitutes support for your hypothesis?
8. **Contingencies:** What if models don't converge? Missing data?

Goal: Be specific enough that a skeptical reader is convinced you planned ahead

Templates Available

▶ **OSF Preregistration** - Comprehensive template
▶ **AsPredicted** - 9 simple questions, generates PDF
▶ **Secondary Data Analysis** - For existing corpora (Weston et al. 2019)
▶ **Replication Studies** - Specialized template
▶ **fMRI Preregistration** - Neuroimaging specific
▶ **Qualitative Research** - For qualitative methods
▶ **Clinical Trials** - Medical research specific

Resource: OSF Templates
Linguistics-specific: Secondary data template

Level 1: Simple Preregistration

Focus on the essentials:

- **Main research question**
- **Primary hypothesis**
- **Basic methodology**
- **Key analysis approach**

Good for: Beginners, exploratory studies, time constraints

Level 2: Detailed Preregistration

Include specifics:

- ▶ Handling missing data
- ▶ Multiple testing corrections
- ▶ Subgroup analyses
- ▶ Decision trees
- ▶ Pre-written analysis code

Good for: Confirmatory studies, complex designs

Level 3: Registered Reports

Two-stage process:

1. **Stage 1:** Submit intro, methods, analysis plan
2. **Review:** Peer review *before* data collection
3. **In-Principle Acceptance:** Publication guaranteed
4. **Stage 2:** Submit results, get published

Where to Preregister

Major Platforms

OSF (Open Science Framework) - Most comprehensive - Multiple templates - Integration with project management - Embargos up to 4 years

AsPredicted - Simple and quick - 8 basic questions - Good for beginners - Free to use

Platform Features Comparison

Feature	OSF	AsPredicted
Templates	Many	One (9 questions)
Output	Web page	PDF with URL
Embargo	4 years	Private option
Collaboration	Multi-author	Email approval
Cost	Free	Free
Integration	Project management	Standalone

Practical Implementation

Getting Started: Step by Step

1. **Choose your platform** (start with AsPredicted for simplicity)
2. **Select appropriate template**
3. **Draft your preregistration** (can save as draft)
4. **Discuss with advisors/collaborators**
5. **Finalize and register** (becomes timestamped)
6. **Conduct your study** as planned
7. **Report** confirmatory vs exploratory findings

Working with Advisors

- **Communicate early** about preregistration goals
- **Share resources** if they're unfamiliar with the process
- **Frame as written study design** (familiar concept)
- **Emphasize benefits** for the research quality
- **Start simple** if they're hesitant

Timeline Considerations

Typical Preregistration Timeline:

- ▶ **Planning Phase (4 weeks):**
  - ▶ Draft preregistration (2 weeks)
  - ▶ Advisor review (1 week)
  - ▶ Revisions (1 week)
- ▶ **Execution Phase (12+ weeks):**
  - ▶ Register study (1 day)
  - ▶ Data collection (8 weeks)
  - ▶ Analysis (4 weeks)

Managing Deviations

When things don't go as planned:

- ▶ **Document changes** transparently
- ▶ **Explain reasons** for deviations
- ▶ **Create new registration** if major changes needed
- ▶ **Distinguish** planned vs unplanned analyses in results

Remember: Transparency is the goal, not perfect adherence

What If Things Don't Go 'As Predicted'?

Standard language for reporting deviations:

▶ *"Contrary to expectations, we found that…"*
▶ *"Unexpectedly, we also found that…"*
▶ *"In addition to the analyses we pre-registered we also ran…"*
▶ *"We encountered an unexpected situation,*
    and followed our Standard Operating Procedure"

Key principle: Transparency, not perfection

Interactive Exercise: Issues That Arise

Scenario: You preregistered a study but encountered problems:

▶ **Lower response rate** than expected
▶ **Technical problem** with one measure
▶ **Discovered relevant covariate** during analysis
▶ **Found unexpected pattern** in data

Discussion: How would you handle each?

Examples and Practice

Example: Simple AsPredicted Registration

The 9 AsPredicted Questions:

1. **Data collection:** Have you already collected the data?
2. **Hypothesis:** What's the main question/hypothesis?
3. **Dependent variable:** What are you measuring?
4. **Conditions:** How many conditions?
5. **Analyses:** What statistical analysis?
6. **Outliers:** How will you handle outliers?
7. **Sample size:** How many observations?
8. **Other:** Anything else you would like to pre-register?
9. **Name:** Give a title to this AsPredicted pre-registration

Result: Time-stamped PDF with unique URL for verification

Example: Psycholinguistic Experiment

Research Question: How does prosodic prominence affect syntactic processing in German?

Hypothesis: Prominent words will show faster integration into syntactic structure

Participants: 40 German native speakers, 18-35 years, no language disorders

Materials: 120 sentences with prominence manipulation, normed for frequency/length

Procedure: Self-paced reading + comprehension questions

Analysis: Linear mixed-effects models with prominence as fixed factor

Exclusions: Accuracy <80% on comprehension, reading times >3 SDs

Example: Corpus Study with Existing Data

Research Question: Does word predictability
affect pronunciation in spontaneous speech?

Data: HCRC Map Task Corpus (Anderson et al., 1991)

Preregistered decisions: - **Predictability measure:** Trigram probability from Google
Books - **Acoustic measure:** Mean F0 of vowel nucleus - **Control variables:** Speaker
sex, utterance position, word frequency - **Exclusions:** Function words, words <3
phonemes - **Model:** Linear mixed-effects: F0 $\sim$ predictability + controls + (1|speaker)

Key insight: Even with existing data, many analytical choices remain

Key Takeaways

The Bottom Line for Linguists

- ▶ **Preregistration enhances** credibility of psycholinguistic research
- ▶ **Start simple** with basic hypotheses and methods
- ▶ **Language research is compatible** with preregistration principles
- ▶ **Exploratory linguistics** remains valuable (just label it clearly)
- ▶ **Not all linguistic subfields** need preregistration
  (observational research is different)
- ▶ **SFB 1252 can lead** the field in transparent language science
- ▶ **Individual benefits:** Better study design,
  protection from criticism, career advantages
- ▶ **Your theoretical contributions** become more impactful

Roettger's key insight: "Preregistration is not a panacea for all problems,
but it's a practice we can integrate into our work flow right away"

Next Steps for SFB 1252

1. **Explore platforms** (OSF recommended for complex linguistic studies)
2. **Try preregistering** your next experiment or corpus study
3. **Discuss with your project team** about adoption
4. **Consider joint preregistrations** for collaborative studies
5. **Share experiences** in future RDM workshops
6. **Advocate** for preregistration in linguistic journals

Immediate action: Choose one upcoming study to preregister

Resources for Further Learning

Essential websites:

- ▶ **Center for Open Science:** cos.io/prereg
- ▶ **OSF Preregistration:** osf.io/prereg
- ▶ **AsPredicted:** aspredicted.org
- ▶ **Templates:** osf.io/zab38
- ▶ **Registered Reports:** cos.io/rr

Reading recommendations:

- ▶ The Preregistration Revolution (Nosek et al., 2018)
- ▶ Research Preregistration 101 (APS)
- ▶ A manifesto for reproducible science (Munafò et al., 2017)

Hands-on Activity: Group Exercise

Small group task (10 minutes):

1. **Form groups of 3-4** people
2. **Choose a simple research scenario** from provided list
3. **Draft key preregistration elements** using AsPredicted format
4. **Present to class** (2 minutes per group)

Scenarios provided:

▶ Prosodic prominence and sentence processing
▶ Cross-linguistic comparison of word order effects
▶ Bilingual language switching patterns
▶ Corpus analysis of discourse markers

Common Questions (Part 1)

- ▶ *"How do I preregister when I don't know what acoustic measures to use?"*
- ▶ *"What if my linear mixed-effects models don't converge?"*
- ▶ *"Can I preregister corpus studies with existing data?"*

Common Questions (Part 2)

- ▶ *"How specific should my exclusion criteria be?"*
- ▶ *"What if children fall asleep during my experiment?"*
- ▶ *"How do I handle cross-linguistic variation I didn't anticipate?"*

Questions & Discussion

What challenges do you see for preregistering linguistic research?

How might preregistration help your current SFB 1252 project?

Who in your research area could be your accountability partner?

References

▶ **Roettger, T. B. (2021).** Preregistration in experimental linguistics: Applications, challenges, and limitations. *Linguistics*, 59(5), 1227–1249.

▶ Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49(3), 609–610.

▶ Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing Into Open Science: A Guide for Graduate Students and Their Advisors. *Collabra: Psychology*, 7(1), 18684.

▶ Lakens, D. (2019). The Value of Preregistration for Psychological Science: A Conceptual Analysis. *PsyArXiv*.

▶ Munafò, M. R., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021.

▶ Nosek, B. A., et al. (2018). The preregistration revolution. *PNAS*, 115(11), 2600–2606.

▶ Wagenmakers, E.-J., Dutilh, G., & Sarafoglou, A. (2018). The Creativity-Verification Cycle in Psychological Science. *Perspectives on Psychological Science*, 13(4), 418–427.