Reproducible, expandable materials for a longitudinal study using artificial languages

Making research materials **F**indable, **A**ccessible, **I**nteroperable and **R**eusable

Pablo Bernabeu[1] **Gabriella Silva**[2] **My Ngoc Giang Hoang**[1]
Vincent DeLuca[1] **Iva Ivanova**[3] **Claudia Poch**[2] **Jason Rothman**[1,2]
Jorge González Alonso[2,1]

[1] *UiT The Arctic University of Norway*
[2] *Nebrija University*
[3] *The University of Texas at El Paso*

Introduction

ERRARE HUMANUM EST
PERSEVERARE
AUTEM DIABOLICUM

Mistakes are fine, but no need to persevere in them, either.
[modern translation]

- **Scientific methods and findings:** inextricable (Barsalou, 2019; Nosek et al., 2012). Therefore, methods are very important.
- **Reproducibility:** long under scrutiny (Nosek et al., 2012) and modest (Kobrock & Roettger, 2023; Open Science Collaboration, 2015).
- **Desirable methodology:** allowing our future selves and others to reproduce, test and expand our materials.
  - *The FAIR Guiding Principles for scientific data management and stewardship* (Wilkinson et al., 2016)
- Different levels of accessibility, reproducibility and expandability of materials from **artificial language studies**.
  - *High:* materials and workflow are reproducible, thanks to modular framework based on minimal components. Final stimuli are testable, modifiable, expandable.
  - *Medium:* some materials are accessible, but the workflow is not reproducible (Cross et al., 2021; Mitrofanova et al., 2023).
  - *Low:* materials are not directly accessible (González Alonso et al., 2020; Morgan-Short et al., 2012; Pereira Soares et al., 2022).

Extensive documentation

Participant-specific parameters in OpenSesame

Each participant was assigned certain parameters in advance, including the mini-language, the order of the resting-state parts, and the stimulus lists. The code that was used to create this assignment is available in the 'stimulus_preparation' folder.

```
participant_parameters =
  pd.read_csv(exp.get_file(
    '../parameters per participant/' + var.study_site +
      ' site, parameters per participant.csv'))

var.language = participant_parameters.loc[
  participant_parameters['participant'] ==
    var.subject_nr]['language'].iloc[0]
```

Modular framework allowing flexible reuse

Modularity will facilitate expanding the materials within the same languages or to other languages.

verb_ID	verb_type	verb	verb_contrast_ID
1	copula_be	is	
2	copula_be	are	
3	copula_look	looks	
4	copula_look	look	
5	transitive	remembered	A
6	transitive	forgot	A

Traceable and testable code scripts

Use of free, open-source, script-based software, such as R and OpenSesame, augments credibility and reliability of research.

```
└── stimulus_preparation
    ├── Norway site, base stimuli.csv
    ├── Spain site, base stimuli.csv
    ├── base_images.R
    ├── R_functions
    │   ├── Session2_Pretraining_vocabulary.R
    │   ├── Session2_Training_gender_agreement.R
    │   ├── Session2_Test_gender_agreement.R
    │   ├── Session2_Experiment_gender_agreement.R
    ...
    ├── compile_all_stimuli.R
```

Tests throughout the workflow

Certain stimuli and experimental conditions should appear equally often to prevent repetition effects. To ascertain this, check whether all elements in certain columns appear equally often. If they do not, show warnings. Please note that this basic check only helps prevent blatant disparities, but it does not verify all the controls applied.

```
columns_to_check = c('noun1_gender', 'number', 'person',
                     'verb', 'noun1', 'wrapup_noun')
for(i in seq_along(columns_to_check)) {
  column = columns_to_check[i]
  number_of_unique_frequencies =
    combinations %>%
    filter(complete.cases(get(column)), get(column) != '') %>%
    group_by(get(column)) %>% tally() %>% select(n) %>%
    n_distinct()
  if(number_of_unique_frequencies != 1) {
    warning(paste0('Some elements in the column `', column,
                   '` appear more often than others.'))
    } }
```

Event-related potentials in OpenSesame

Custom Python code was used in OpenSesame to time-lock electroencephalographic measurements to onset of specific stimuli, by sending triggers to the serial port.

```
# Open the first serial port available
serialport =
  serial.Serial(serial.tools.list_ports.comports()[0].device)
# Send triggers to the port
def send_trigger(trigger):
    serialport.write(trigger.to_bytes(
      length = 1, byteorder = 'big'))
    # 10 ms separation from next trigger
    time.sleep(0.01)
    # reset port
    serialport.write(int(0).to_bytes(
      length = 1, byteorder = 'big'))
    return;
```

Conclusion

Adhering to best practices in the creation of research materials facilitates researchers' work beyond the shortest term, and increases its reliability.

Acknowledgements

Thank you to Merete Anderssen, Gaute Berglund, Anders Gabrielsen, Mona Fossum, Tekabe Feleke, Björn Lundquist, Natalia Mitrofanova, Yulia Rodina, Jade Sandstedt, Toms Voits, Marit Westergaard.

References

Barsalou, L. W. (2019). Establishing generalizable mechanisms. *Psychological Inquiry*, *30*(4), 220–230.

Cross, Z. R., Zou-Williams, L., Wilkinson, E. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2021). Mini Pinyin: A modified miniature language for studying language learning and incremental sentence processing. *Behavior Research Methods*, *53*(3), 1218–1239.

González Alonso, J., Alemán Bañón, J., DeLuca, V., Miller, D., Pereira Soares, S. M., Puig-Mayenco, E., Slaats, S., & Rothman, J. (2020). Event related potentials at initial exposure in third language acquisition: Implications from an artificial mini-grammar study. *Journal of Neurolinguistics*, *56*, 100939.

Kobrock, K., & Roettger, T. B. (2023). Assessing the replication landscape in experimental linguistics. *Glossa Psycholinguistics*, *2*(1).

Mitrofanova, N., Leivada, E., & Westergaard, M. (2023). Crosslinguistic influence in L3 acquisition: Evidence from artificial language learning. *Linguistic Approaches to Bilingualism*, *13*(5), 717–742.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLOS ONE*, *7*(3), e32974.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II: Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Pereira Soares, S. M., Kupisch, T., & Rothman, J. (2022). Testing potential transfer effects in heritage and adult L2 bilinguals acquiring a mini grammar as an additional language: An ERP approach. *Brain Sciences*, *12*(5), 669.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018.

The Research Council of Norway

Materials
rb.gy/jrusi8

the PoLaR lab
AcqVA Aurora
LaBL Center for Language, Brain & Learning
The Arctic University of Norway · UiT ·