

Shanglun Li

2018/12/14

Referee Report on Sanchez et al. (2018)

Evaluation of the Research Question and Methodology

Unlike many other research projects where a research question is defined and answered, this paper would more accurately be described as a “problem-solution” research, where an existing issue is identified and a solution resolving said issue is introduced and explained. To this end, the researchers have done an adequate job addressing the purpose of their study. As stated in the beginning of the research, the “problem” of the study is found within the Child Language Data Exchange System (CHILDES), an existing system that has been playing an essential role in the research on child language development (Sanchez et al., 2018: p. 4). The issues embedded in the system include difficulties in mastering the complex functions and the inability to transfer data into Python or R, which fundamentally lead to inefficiencies, errors and inconsistencies, versioning problems, and most importantly, a lack of “computational reproducibility” (Sanchez et al., 2018: p. 5). To address these issues, this paper introduces childe-db, a system that “extends the functionality of CHILDES” (Sanchez et al., 2018: p.6) by improving the accessibility and usability of the current system, making it easier for casual explorers, students, teachers, as well as advanced researchers to engage in the study of language and cognitive development (Sanchez et al., 2018: p. 2).

The authors make a compelling case in demonstrating how childe-db solves the issues that come with the existing system and makes it easier for a wider range of researchers in the field. The organization of the paper goes as follows:

The researchers start by introducing the general background and methods used in the study of child language development. The basic functions of CHILDES and the challenges that

comes with it are sufficiently explained. Following that, the authors introduces childe-db as a system that could overcome the previously mentioned issues.

It starts by offering a description of the design and the technical approach of the infrastructure, demonstrating how childe-db would be more accessible and efficient by treating CHILDES as a set of “linked tables” instead of “CHAT files” (p.6) and how the users wouldn’t have to worry about the formatting details while operating with this system.

It goes on to introduce a series of other advantages that the system offers, including 1) a highly accessible open-source database (Sanchez et al., 2018: p.8), 2) accurate data guaranteed by effective data integrity checks (Sanchez et al., 2018: p.9), and 3) solved versioning problems by defaulting to the newest available database (Sanchez et al., 2018: p.9).

Next, the paper goes on to describe the interfaces for accessing childe-db, which are the childe-db web apps and the childeR R package. On one hand, the “easy-to-use” web application allows casual users to carry out frequency counts, derived measures, and population viewing without much technical prerequisites (Sanchez et al., 2018: p.11). On the other hand, the childeR package provides advanced researchers with a quick and easy way to manipulate large datasets (Sanchez et al., 2018: p.14).

To consolidate the claims above, the paper exhibits a number of worked examples to demonstrate the research applications of the web applications as well as childeR. Through practical case studies on color frequency, gender, and teaching with childe-db, the paper compellingly demonstrated how the learning, teaching and researching processes on the subject of child language development can be greatly enhanced by employing childe-db.

Finally, the paper points out some of the limitations that currently exist within the childe-db system, such as the lack of links to the underlying media files and the lack of a full range

of CHAT annotation tiers (Sanchez et al., 2018: p.26). The paper concludes by calling for feedback and suggestions in order for the system to grow and improve.

I deem the research paper to be well executed for the following reasons:

- Background information and context of the research is sufficiently established, which provides readers with a well-rounded perspective on the issue.
- The paper is organized in a clear and rational manner, with a logical flow that is easy to follow.
- The claims are trustworthy as they are backed by specific examples, solid evidence and detailed demonstrations of specific processes.
- Visual aids are well-used, which engages the readers and helps better understanding.
- The study not only mentions the advantages of the new system but also acknowledges the existing shortcomings, which adds to the sense of reliability of the paper.

Broader Literature and Citations

In fact, since CHILDES is such an essential and irreplaceable tool in the field of child language development, it makes this particular research paper quite innovative in the sense that it is quite difficult to discover scholarly sources that objectively critiques the accessibility and usability of the system. Unfortunately, this constitutes one of the weaknesses in the argument of the paper, since *The Emergence of an Abstract Grammatical Category in Children's Early Speech* (Meylan et al., 2017) is the only substantial source of reference in this study that points to the limitations of CHILDES.

Citations and references are well-applied in the most part; however, some minor issues still exist. On page 4, for example, the authors cite 7 scholar sources, including Marcus (1992), Snyder (2007), and Redington (1998), just to make the point that CHILDES has been a

widely-used tool, which appears redundant. Another unnecessary citation appears on page 6 line 130, since the study by Kline (2012) seems irrelevant to the case the paper is trying to make.

The paper also seems to have missed citations on two occasions. The first can be found on page 17 line 346, where it states, “because children learn most of their basic color words by the age of 5, the age range 1-5 years is a reasonable choice for Ages to include.” The second appears on page 18 line 363, where the study excludes gray, because “children hear [it] very rarely.” Both statements appear to be assumptions without valid references as evidence. *On the Development of Color Naming in Young Children: Data and Theory* (Bornstein, 1985) would be a good place to start, as the study provides a well-rounded examination of the subject in question.

Grammar, Spelling, and Style Errors

As a whole, the research paper is very well written regarding the structure of sentences and the organization of composition. However, a few minor errors were detected under scrutiny. For example, on page 16 line 326, a comma is missing after the term “However.” Furthermore, on page 26 line 484, the sentence would flow better if the semicolon is replaced with a period.

Extension

While the research paper adequately addresses the advantages of the new system, childes-db, with specific examples and detailed processes during research applications, the evidence still remain mostly one-sided since the research applications are executed by the researchers of the study themselves. On page 10, the paper mentions “community feedback,” which leads one to wonder if beta tests have already been implemented and what the responses indicate. As such, a possible extension of the current research question could

look into the actual user experience and whether the new system is genuinely capable of improving the research process on child language development. The research could further discover the limitations and shortcomings that currently exist in the system based on user feedback, and whether such challenges could be resolved in order to better serve the community's research and teaching needs.

References

- Bornstein, M. (1985). On the development of color naming in young children: data and theory. In *Brain and Language*
- Kline, M. (2012). CLANtoR. <http://github.com/mekline/CLANtoR/>; GitHub.
doi:[10.5281/zenodo.1196626](https://doi.org/10.5281/zenodo.1196626)
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, i–178.
- Meylan, S. C., Frank, M. C., Roy, B. C., & Levy, R. (2017). The emergence of an abstract grammatical category in children’s early speech. *Psychological Science*, 28(2), 181–192.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22 (4), 425–469.
- Sanchez, Alessandro, Stephan C. Meylan, Mika Braginsky, Kyle E. Mac-Donald, Daniel Yurovsky, and Michael C. Frank. (2018) “chilDES-db: A Flexible and Reproducible Interface to the Child Language Data Exchange System,” under review, *Communication and Learning Laboratory*, University of Chicago,
<https://callab.uchicago.edu/papers/smbmyf-brm-underreview.pdf>.
- Snyder, W. (2007). *Child language: The parametric approach*. Oxford University Press.