

Assignment #9

MACSS 30000

Delores Tang

12.6

Review of the Paper - CHILDES-db: a flexible and reproducible interface to the Child Language

### Data Exchange System

Different from most psychological research papers, Sanchez et al. (2018) was not aimed to explain a psychological phenomenon but to resolve a technological difficulty and to improve methodologies in the field for the convenience of other researchers. Therefore, their research question should be “what is CHILDES-db and how does this system along with its application programming interface (API) is able to help researchers and teachers to better utilize CHILDES and to increase computational reproducibility” (Sanchez et al., 2018, p.6). The article focuses on the development and description of CHILDES-db and its API and provides instructions for its uses and working examples in the field (p.6-24). Thus, Sanchez and his colleagues’ main research objective is to identify a technical difficulty in the field, to offer a solution and to introduce its creation processes, central features, its advantages and limitations, its instructions for users, and its practical applications.

I think that the authors of this article well defined their research question. They first described the importance and limitations of the Child Language Data Exchange System (CHILDES; Sanchez et al., 2018, p.4). Studying the processes of language development in children is helpful for researchers to understand language development in general (p.4). However, it is time- and resource-consuming and to collect video and audio data over a long duration of time. CHILDES hence provides linguists a comprehensive system of such data of different languages and children of different ages, but its format for encoding (CHAT) makes it

difficult for researchers to process data and to reproduce it computationally (p.5). Thus, they well justified the necessity for the development of an auxiliary tool to help with these problems.

The author answered their research questions compellingly by closely describing the development of CHILDES-db and its API, and their main features and benefits. Specifically, the authors reported the strength of CHILDES-db from three aspects. Different from the original data analysis tool CLAN (a command-line tool, p.5), CHILDES-db uses linked tables that resemble more to Excel, R, and Python files (Sanchez et al., 2018, p.6). In addition, its format allows researchers to associate identifiers and the corresponding data more easily (p.8), and its technical implementation allows a better ability for data storage (p.10), version updating (p.11), and data classification and processing (p.11-13). Their most compelling argument for the strength of CHILDES-db is that Sanchez and his colleagues provided worked examples (p.15) to demonstrate how to apply it in the field of developmental psychology and linguistics. Although I do believe, in order to introduce it to researchers who are not already familiar with the CHILDES system, it could be even more helpful to provide a diagram that illustrates the data analysis processes without CHILDES-db and its interfaces compared with the process with CHILDES-db.

I think that this paper offers enough background information and citations. Because, as I mentioned, the aim of this article is not to explain a human behavior or social phenomenon, so it is sufficient for the researchers to refer to the previous literature and identify the limitations of the existing CHILDES system. Especially in the introduction (p.4), Sanchez and colleagues listed many influential past studies that have utilized CHILDES to highlight the value of CHILDES and hence the necessity for a more accessible tool. I do hope that the authors could provide more examples from previous research and include more citations to elaborate on the

problems of writing custom parsers (p.5) to help non-expert of CHILDES to better understand the CHAT format it uses and how CHILDES-db could be a more convenient tool for analysis and formatting. I did not find any specific grammatical mistakes or errors from the text.

Using the data in CHILDES and the CHILDES-db tools and interfaces, I could think of many other extensions in the field of researching language development. For example, one research question I could think of is that “how Chinese and English speaking children differ in ways they acquire verbs and nouns?” Many researchers have discovered that adults and children process verbs and nouns differently across languages (Haman et al., 2017). In addition, children also acquire nouns and verbs distinctively (Parisse & Rossi, 2017). However, not much has been done to compare such noun/verb distinctions across languages. Therefore, I want to look for appropriate data sources in CHILDE, possibly with identifiers relevant to objects (nouns) and movements (verbs). By using CHILDES-db and its API, I could not only get access to previously collected data across nations but also visually compare children’s noun/verb acquiring patterns across age and development stages (Sanchez et al., 2018, p.17, Figure 5).

#### Reference

- Haman et al. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, 11–12, 818. doi:10.1080/902699206.2017.1308553
- Parisse, Christophe, & Rossi, Caroline.(2017). Semantic discrimination of Noun/Verb categories in French children aged 1;6 to 2;11. *Lexical Polycategoriality Cross-linguistic, cross-theoretical and language acquisition approaches*, 9789027265951.