



Oct. 19, 2013

Prof. Brian MacWhinney
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

Dear Brian,

I am writing to express my sincere gratitude to you for your creation of the CHILDES database. In the past two decades I have used the CHILDES database extensively (since my graduate school years), and found that it has benefited my research in tremendous ways. Looking back, I cannot imagine how I would have got my research done if I had not had access to this important database. I believe that many researchers in the field have exactly the same feeling as I do.

There are roughly three ways in which the CHILDES database has benefited my research.

First, because CHILDES contains electronic databases not only for English, but also for other languages (especially Chinese and Cantonese in my case), it is extremely helpful to researchers like me who have a strong interest in cross-linguistic studies of language acquisition. In my research, I have also made use of the Bilingual Chinese-English Corpus contributed by Steven Matthew and Virginia Yip in connectionist modeling of bilingual language processing (e.g., Li & Farkas, 2002; Zhao & Li, 2007, 2010, 2013). My students and I have also used the Tardif corpus and other Mandarin Chinese corpora for corpus-based analyses in my research (see Hao, Shu, & Li, 2008; Liu, Zhao, & Li, 2008; Zhao & Li, 2008).

Second, because CHILDES includes not only transcribed speech of children, but also parental speech directed to children, it is extremely important to researchers like me who are interested in identifying the role of parental input in determining the lexical-morphological patterns of children's speech. For example, we have published an extensive analysis of the parental speech in CHILDES in order to identify the role of frequency in the input (see Goodman, Dale, and Li, 2008), and have made use of the CHILDES parental corpus in connectionist modeling. The CHILDES parental corpus has been useful to many researchers in developmental psycholinguistics (see Goodman, et al. 2002, 2003), and hence has not just benefitted my own research.

Third, because CHILDES contains massive amount of empirical child language data, covering all aspects of children's lexical, morphological, sentential, and discourse acquisition, it is very important to researchers like me who would like to use or verify computational modeling results with empirical data from children. CHILDES aids me in at least three ways in this regard. (1) I have evaluated my modeling results against data available in CHILDES. (2) I have derived a comprehensive list of the early nouns and verbs that appear in the CHILDES English corpus, with information on the co-occurring morphological markers and frequency of use. This information serves as a basis for us to construct a realistic learning schedule for simulation. (3) I have analyzed both children's and adults' speech in CHILDES in order to derive semantic representations that correspond more closely to the child's developing lexical representations at different stages of development. Moreover, this co-occurrence based method has been incorporated into our connectionist model, DevLex, which extracts semantic representations from CHILDES parental speech and projects these representations onto a self-organizing map (see Li, Farkas, & MacWhinney, 2004; Li, Zhao, & MacWhinney, 2007).

Although all of these aspects are concerned with research only, CHILDES has proven to be an important educational tool. In my cognitive science and psycholinguistics classes, I have been teaching my students how to use CHILDES as a tool to study language acquisition. Given the highly electronic nature of CHILDES (e.g., being available on the world-wide-web 24 hours a day), students can access the database, retrieve the CLAN program, analyze individual children's speech, and write short reports. Several students have also gone beyond this level, and have conducted independent research on the database.

Over the years your CHILDES has become part of my research, and you have also been my closest collaborator, especially with regard to computational modeling using information from CHILDES. I want to go back to my earlier comment: researchers like me cannot imagine how it would have been had you not created the CHILDES database. I want to thank you and congratulate you on making such an important contribution to the field.

Sincerely

Ping Li, PhD

A handwritten signature in red ink, appearing to be 'Ping Li', written in a cursive style.

Professor of Psychology, Linguistics, and Information Sciences and Technology
Co-Chair, Inter-College Graduate Program in Neuroscience
Co-Director, Center for Brain, Behavior, and Cognition
Pennsylvania State University
University Park, PA 16802, USA

Appendix: Published articles or book (since 2003) that have used CHILDES:

Book:

Li, P., Tan, L.-H., Bates, E., & Tzeng, O. (2006). *The Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge University Press.

Articles:

Goodman, J.C., Dale, P.S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35, 1-17.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581-612.

Li, P. (2007). What's in a lexical system? The emergence of semantic structure in lexical acquisition. In V. Gathercole (Ed.), *Language Acquisition: A Festschrift to Melissa Bowerman*. Mahwah, NJ: Lawrence Erlbaum. (in press)

Li, P., & Zhao, X. (2007). Computational modeling of the expression of time. *The expression of time in language*. Berlin & New York: Mouton de Gruyter. (in press)

Zhao, X., & Li, P. (2007). The acquisition of lexical and grammatical aspect in a developmental lexicon model. *Linguistics: An Interdisciplinary Journal of the Language Sciences*. (in press)

Zhao, X., & Li, P. (2007). Bilingual lexical representation in a self-organizing neural network. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum. (*Computational Modeling Prize 1st Place, Language*).

Li, P. (2006). In search of meaning: The acquisition of semantic structure and morphological systems. In J. Luchjenbroers (Ed.) *Cognitive linguistics investigations across languages, fields, and philosophical boundaries* (pp.109-137). Amsterdam, Holland: John Benjamins, Inc.

Li, P. (2006). Modeling language acquisition and processing in connectionist networks. In Li, P., Tan, L.-H., Bates, E., & Tzeng, O. (Eds.), *The Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). (pp.320-329). Cambridge, UK: Cambridge University Press.

Li, P., Tan, L., Bates, E., & Tzeng, O. (2006). New frontiers in Chinese psycholinguistics: An introduction. In Li, P., Tan, L., Bates, E., & Tzeng, O. (Eds.), *The Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge, UK: Cambridge University Press (pp. 1-9).

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9, 220-225.

Zhao, X., & Li, P. (2005). A self-organizing connectionist model of early word production. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 2434-2439). Mahwah, NJ: Lawrence Erlbaum.

Li, P., Farkas, I., & MacWhinney (2004). Early lexical acquisition in a self-organizing neural network. *Neural Networks*, 17, 1345-1362.

Xing, H., Shu, H., & Li, P. (2004). The acquisition of Chinese characters: Corpus analyses and connectionist simulations. *Journal of Cognitive Science*, 5, 1-49.

Li, P. (2003). Language acquisition in a self-organising neural network model. In P. Quinlan (Ed.) *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 115-149). Hove & New York: Psychology Press.