

Brian MacWhinney <macw@cmu.edu>: Mar 01 12:30AM

Dear Info-CHILDES,

We are happy to announce the release of a multimedia longitudinal child Mandarin corpus: The Tong Corpus. Given the importance of Mandarin Chinese as a language spoken by more than one billion people in China and overseas communities, this longitudinal corpus has far-reaching implications for the field. It contains naturalistic interactions between a Mandarin-speaking child Tong and his caregivers mainly in the home setting, from age 1;0 to 4;8.

For the first installment of this corpus, 22 hours of recordings of the child from the age of 1;7 to 3;4 have been deposited in CHILDES. The speech data have been transcribed in Chinese characters, and annotated with a morphosyntactic tier providing pinyin, parts of speech, and English translation. The transcripts are linked to audio or video files. Details of the corpus can be found at <http://cbrchk.org/the-tong> corpus/ <http://cbrchk.org/?page_id=1194&preview=true>.

This corpus goes beyond existing published corpora of child Mandarin in having more data for a single child, as well as media linking. It will contribute to related fields including Chinese linguistics, corpus linguistics, developmental psycholinguistics, education, and speech and language therapy.

With best wishes,

Xiangjun Deng

Virginia Yip

Bilingualism and Language Disorders Laboratory, Chinese University of Hong Kong (CUHK)-Shenzhen Research Institute

Childhood Bilingualism Research Centre, CUHK

CUHK-Peking University-University System of Taiwan Joint Research Centre for Language and Human

Complexity

dengxj98@gmail.com; vcymatthews@cuhk.edu.hk