# A MULTIMEDIA CORPUS OF CHILD MANDARIN: THE TONG CORPUS
**Deng Xiangjun   Virginia Yip**
*The Chinese University of Hong Kong*

ABSTRACT

This article features a new multimedia corpus[i] with 22 hours of recordings of a Mandarin-speaking child from the age of 1;7 to 3;4. We review the state of the art in the use of corpora for first language acquisition of Mandarin, and highlight the importance of corpus studies in evaluating children's language developmental patterns vis-a-vis adult input. The transcripts in our new corpus are annotated with a morphological tier indicating parts of speech, and linked to audio or video files. This corpus goes beyond existing published corpora of child Mandarin in having more data for a single child, as well as media linking. It contributes to a number of fields including language acquisition, Chinese linguistics, corpus linguistics, developmental psycholinguistics, education, and speech and language therapy.

_____

**Deng Xiangjun** [dengxj@szu.edu.cn]; The Research Centre for Language and Cognition, School of Foreign Languages, Shenzhen University, 3688 Nanhai Ave., Nanshan District, Shenzhen, Guangdong Province 518060, P. R. China.  ⓘ  https://orcid.org/0000-0002-9693-5593.

**Virginia Yip** [vcymatthews@cuhk.edu.hk], Department of Linguistics and Modern Languages, G/F, Leung Kau Kui Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

   i. TalkBank/CHILDES [Index to Corpora: Chinese/Tong Corpus (DOI: 10.21415/T5PC7Q)]

*Yuyan Fazhan Yanjiu: Guoji Ertong Yuliaoku Yanjiu Fangfa de Yingyong yu Fazhan* 汉语儿童语言发展研究：国际儿童语料库研究方法的应用与发展 (Research on the language development of Chinese children: The application and development of the research method of international child language corpora). Beijing: Jiaoyu Kexue Chuban She.

ZHOU, Jiangling, Ziyin Mai, and Virginia Yip. 2015. Null objects in Cantonese-English bilingual children. Paper presented at the 23rd Annual Conference of the International Association of Chinese Linguistics (IACL-23), Seoul, Korea.

ZOU, Ke. 1994. Resultative V-V compounds in Chinese. *MIT Working Papers in Linguistics* 22:271-90.

ZSKYYCB. *See* Zhongguo Shehui Kexueyuan Yuyan Yanjiusuo Cidian Bianjishi.

# 一个多媒体汉语普通话儿童语料库：
## 同语料库（TONG CORPUS）

邓潇君  叶彩燕

香港中文大学

**提要**

本文发布一个新的多媒体语料库的首阶段成果。这部分内容记录了一名普通话儿童从 1 岁 7 个月到 3 岁 4 个月期间的语言发展，共录得 22 个小时的语料。借此机会，我们回顾了汉语普通话一语习得研究中语料库使用的最新情况，强调语料库研究在考察儿童语言发展和成人语言输入时的重要作用。在我们这个新的语料库中，文字转写材料添加了词类注释层，并已实现与多媒体材料的链接。这个语料库在单个普通话儿童数据量和音频视频链接上超越了现有已发表的语料库。它将为语言习得、汉语语言学、语料库语言学、发展心理语言学、教育以及言语治疗等诸领域做出贡献。

**关键词**

儿童语料库 汉语普通话 语言输入 多媒体链接 词类注释层