

**Danjie Su, Ph.D.**

Associate Professor of Chinese and Applied Linguistics

University of Arkansas

July 22, 2023

**1. Current research projects related to conversational transcription:**

Currently, I am collaborating with Dr. Kevin Goslar on a research project that explores methods of using AI to automatically transcribe large-scale Chinese conversational data in a precise manner that retains most conversational features (e.g., repetitions, fillers, etc.)

**2. Related recent talks:**

- Su, Danjie and Kevin Goslar. Evaluating pragmatic competence of Generative AI GPT-4 with the Lens concept. 2023 Technology-Enhanced Language Learning International Conference (TELIC), virtual. July 14, 2023.
- Su, Danjie and Kevin Goslar. Evaluating pragmatic competence of Artificial Intelligence with the Lens concept: ChatGPT for Chinese second language teaching. 20<sup>th</sup> Technology for Second Language Learning (TSL) Conference, Iowa State University. October 20-21, 2023.

**3. Conversational corpus constructed:**

- Su, Danjie. 2019. *The M. Chinese Video Corpus (MCVC)*. UCLA, Los Angeles & University of Arkansas, Fayetteville.  
(One-million-word (1,129,437) transcripts and 300 videos of spontaneous talk show conversations in Mandarin Chinese, totaling 100 hours of video data)

**4. Software for editing audio/video files:**

- Adobe Premiere Pro (See Su 2016 for an example, in which I used Premiere to extract screenshot images of a video of an English conversation to study gestural reuse.)

**5. Software for preparing frame-grabs for publications:**

- Online converters that convert an image to a line drawing:
  - a) Sketcher 1 <https://www.befunky.com/create/photo-to-art/> (less realistic, see Su 2022a for an example)
  - b) Inkify 1 <https://www.befunky.com/create/inkify/> (slightly more realistic)
- Adobe Photoshop (for everything else regarding images, e.g., for creating two multimodal transcription symbols in Su 2016)

**6. Software for working with corpora and analyzing transcripts:**

- AntConc (see Su 2017b for an example)
- UAM CorpusTool (see Su 2017b for an example)
- Microsoft EXCEL (see Su 2022a&b, 2023 for examples)

## 7. Software for transcription:

- VLC video player (using shortcut keys for speed control, etc.)
- SharpKeys (a Registry key that allows Windows to remap one key to any other key. Use it to speed up the workflow when working with large-scale data)
- Microsoft WORD (see Su 2016, 2017a, 2022a, 2023 for examples)
- Transana (used it before but no longer use it due to concerns of potential data loss in the long term)

## 8. Software for analyzing prosodic features:

- Praat (see Su 2016 for an example)

## 9. Software for Chinese text processing:

- NLPir-ICTCLAS Chinese lexical analysis system: [www.nlpir.org/](http://www.nlpir.org/) (for Chinese word segmentation and Part-Of-Speech (POS) tagging. I used in for the construction of the MCVc corpus.)
- OpenCC (Open Chinese Convert, 開放中文轉換), an opensource project for conversions between Traditional Chinese, Simplified Chinese and Japanese Kanji (Shinjitai). <https://github.com/BYVoid/OpenCC>

## 10. Publications using the software mentioned above:

Su, Danjie. 2023 (online first). Speakers' subjective evaluation of adversity: Adversity lens constructed through *bei*-passive in Chinese conversation. *Chinese Language and Discourse*. <https://doi.org/10.1075/cld.22002.su>

Su, Danjie. 2022a. Factuality lens: Choosing the unmarked passive construction in Chinese conversation. *Chinese Language and Discourse* 13(2). 242–274. <https://doi.org/10.1075/cld.21005.su>

Su, Danjie. 2022b. Early evidence of the Pareto Principle in grammatical distribution: Causative situations in Chinese conversational discourse. *Journal of Chinese Linguistics* 50(2). 443–474. DOI: [10.1353/jcl.2017.0113](https://doi.org/10.1353/jcl.2017.0113)

Su, Danjie. 2017a. Significance as a lens: Understanding the Mandarin *ba* construction through Discourse Adjacent Alternation. *Journal of Pragmatics* 117. 204–230. <https://doi.org/10.1016/j.pragma.2017.06.019>

Su, Danjie. 2017b. Semantics and chunking in written and conversational discourses: A corpus study of two near-synonymous words in Mandarin. *Chinese Language and Discourse* 8(1). 51–94. <https://dx.doi.org/10.1075/cld.8.1.03su>

Su, Danjie. 2016. Grammar emerges through reuse and modification of prior utterances. *Discourse Studies* 18(3). 330–353. <https://doi.org/10.1177/1461445616634551>