

Toward a Mandarin-French Corpus of Interactional Data

Helen K.Y. Chen^{1,2}, Laurent Prévot¹, Roxane Bertrand¹, Béatrice PriegoValverde¹, Philippe Blache¹

¹LPL, CNRS & Aix-Marseille Université

²The HK Polytechnic University

Introduction

This paper provides a preliminary description of the construction of an audio-video speech corpus of face-to-face Mandarin interaction. The corpus consists of 5 hours of two-party, face-to-face, spontaneous Mandarin interaction. Following the construction of the *Corpus of Interactional Data* (CID), an audio-video corpus in French recorded and processed at the Laboratoire Parole & Langage (LPL), Aix-Marseille Universités (Bertrand et al., 2008), the current project proposes the construction of a Mandarin speech corpus that will be annotated, based on a *multimodal* perspective, at various linguistic levels including prosody, morphology, syntax, as well as discourse and non-verbal representations such as gestures.

Design of the corpus -- the protocol

•Basic design and subjects

5 males and 5 females who are native speakers of Taiwanese Mandarin are involved in the recordings of face-to-face interaction. Preliminarily, it is proposed that the corpus will consist of 5 segments of two-party, face-to-face Mandarin conversational recordings. Each recording lasts about 1 hour. Thus at the initial stage of the construction of the corpus, the total recording time will add up to 5 hours.

•The task

The participants are invited to join the experiment, in which they are instructed to “have a chat” with another speaker. One of the following 3 sets of guidelines were provided randomly to the participants prior to each recording:

1. the participants may be instructed to talk about the most difficult things they've encountered while living in France;
2. the participants may be instructed to talk about one of the unusual things s/he has encountered during a recent trip;
3. the participants may not be given any instruction about what to talk about and simply started the conversation on their own

The reason to provide a guideline is to enable speakers to engage in the interaction rather quickly as soon as the recording starts. Although the speakers may be provided with one of the 3 guidelines as the initial topic for interaction, there was no further instruction about how long the topic should last. The speakers are free to switch to other topics as the interaction carries on.

•The setting of the recordings

Following the original French CID setting (Bertrand et al., 2008), all recording session took place in a soundproof chamber at LPL. The two speakers in each session sat side-by-side and slightly tilted towards each other. During the recording, each subject wore a headset and the voice from each speaker was recorded onto a separated track. In addition to the specific setting for the sound recording, subjects were also filmed in long and fixed shots.

Characteristics of the corpus

The conversations recorded for the CID corpus is presented as closely simulating the data of naturally occurred, face-to-face interaction; the dialogues of the current corpus resemble daily Mandarin interactions and can serve as a speech resource with rich information on turn-taking and sequential organisations of conversation (Sacks et al., 1974).

Proposed levels of transcriptions and annotations

The following levels of transcriptions and annotations are proposed preliminarily:

•Pre-processing and transcription

- In order to facilitate the transcriptions, the data will be pre-processed by an automatic segmentation into blocks of sound stream by silent pauses of 200ms (cf. Blache et al. (2009)) prior to the transcription
- The segmented data will then be transcribed orthographically (by using standard Hanyu Pinyin).

•Prosodic level

- A higher level of *intonational phrases*
- A more detailed level of interactionally related prosodic cues: i.e. duration, cut-off, lengthening, special voice quality (such as laryngealized voice); the annotations unit will be tokens

•Interaction and discourse level

- The basic unit will be *turn consturctional unit* (TCU), which corresponds to an action in interaction (cf. Schegloff, 2007)
- A parallel annotation of *disfluencies* will be identified; the annotations may include the reparandum, reparans, and break interval (cf. Shriberg, 1994)

•Non-verbal cue level: gestures

- Following McNeill (1992) and Chui (2003), 5 types of gestures will be identified: *iconic, metaphoric, deictic, spatial* gestures and *beats*
- Each type of gesture will be further marked by its *preperation, stroke, and retraction* (cf. McNeill, 1992)

Contribution of the corpus

- As the conversations recorded resemble much daily Mandarin interaction, such characteristic of the corpus may contribute to the advance analysis of the interactions between Mandarin speakers from various range of perspectives: the correlations between conversational interaction and sound/prosodic realization, and/or the examination of interaction in terms of non-verbal cues such as gestures
- In addition to obvious benefits of a richly annotated corpus with multimodal informations, the perfect replication of the CID experimental setting will allow for systematic cross-linguistic studies, i.e comparative studies between Mandarin and the original French data.
- the current Mandarin CID corpus can be further compare to other Mandarin speech corpus, such as Mandarin Conversational Dialogue Corpuse (MCDC, see Tseng 2004), as the two corpora are different in terms of the experimental settings

* This research was supported by the Erasmus Mundus Action 2 program MULTI of the European Union, grant agreement number 2009-5259-5.

