Transcribing Mandarin Chinese Conversation: Linguistic and Prosodic Issues

Jee Won Lee¹, Hongyin Tao², Ping Lu³

Abstract

This study addresses some of the issues in the transcription of Chinese spoken discourse, with a particular focus on transcription solutions for various linguistic and vocal features of Mandarin conversation. While there is a great deal of complexity in Chinese spoken discourse, current practices in transcribing spoken Chinese and in building spoken corpora contain essentially no standards. Therefore, this study selects the linguistic and vocal features of spoken Chinese that meet the criteria of 1) highly recurrent in spoken Chinese; and 2) posing the most trouble in transcription. The features that need urgent attention in transcription include discourse particles, non-lexical vocalizations, and repairs. Given the centrality of spoken discourse, and conversation in particular, in discourse research, it is critically important to discuss key issues in transcribing Mandarin Chinese and then propose some ways for corpus building in the field of Chinese linguistics, which we do in this paper.

Keywords: transcription, Mandarin conversation, spoken language, corpus construction, linguistic and vocal features

1. Introduction

As any researcher of spoken discourse analysis is aware, transcription is one of the first steps toward building an adequate understanding of how spoken interaction works; yet transcription is never a straightforward matter (Ochs, 1979)[1]. According to Egbert, Yufu, and Hirataka’s (2016) investigation, there are almost as many different types of transcripts of...
non-English data as there are research articles surveyed[2]. In the case of Mandarin Chinese, the issue is further complicated by the fact that most texts are represented in Chinese characters, which are not phonetic. Thus, many important features of spoken Chinese may or may not be well represented with the standard character system. Moreover, when researchers build large corpora of spoken Chinese, there are very few standardized practices that are consistent across different teams at different institutions. While some earlier studies have made specific proposals about spoken Chinese transcription (e.g. Tao, 2003), transcription principles and methods still need to address important speech details[3]. This study aims to give an overview of the difficulties in transcribing Mandarin conversation and build spoken corpora, with a particular focus on possible transcription solutions for various vocal features of Mandarin conversation.

Before further discussions, let us first examine an excerpt that shows the complexity of talk and interaction in Mandarin Chinese conversation. The data are transcribed using the CA transcription system with slight modifications (Atkinson& Heritage, 1984)[4].

(1) A talk between two friends concerning carsickness.

1 A: 在說她暈(h)車的[問題.'
   zai shuo ta YUN(h)CHE de [wenti.'
   in the middle of say she carsick POSS problem
   ‘(We) were talking about her problem with car(h)sickness.’

2 B: [哇宰，我跟你說她暈車的經歷哈.
   [wa zai wo gen ni shuo ta yunche de jingli ha
   INT I with you say she carsick POSS history PRT
   ‘Oh, I will tell you about her carsickness.’

This transcript includes several speech phenomena represented by the various notational symbols. ‘[’ marks the start of overlapping utterances, ‘(h)’ indicates laughter coinciding with a word, and capital letters denotes that utterances are spoken much louder than the surrounding talk. The transcript also includes the discourse particle wazai and ha which are exclusively used in spoken language. (A literal translation of the recording can be found in the third line of the transcript, and an idiomatic translation appears the fourth line.)

Given the complexity of spoken discourse and the varying purposes of transcriptions in linguistic research, many researchers have highlighted the intrinsic difficulties in transcription
and various biases. Some researchers, for example, point out that transcription may not be a reliably reproducible process (Kerswill & Wright, 1990)[5]. Others point out that when researchers focus on certain features of the talk at the expense of others it means that these transcripts may not be able to capture the dynamic nature of talk (Cook, 1990)[6]. Still others contend that all transcripts are inextricably tied to the context of their production and application within a specific research area (Mondada, 2007), so transcript variation is generally seen as normal (Bucholtz, 2007)[7][8].

Compared with the extensive discussions in English discourse literature, the issue of transcribing spoken discourse has received relatively little attention in Mandarin Chinese scholarship. Most corpora of spoken Chinese contain no systematic and explicit statements of transcript conventions. Chinese discourse transcripts are usually very broad, and they often fail to capture vital aspects of talk in interaction. Given the importance of spoken language, and conversation in particular, in discourse research, there is clearly a need to discuss the relevant issues in transcribing Mandarin spoken discourse and attempt some solutions to the most important challenges in working with spoken Chinese. This paper is geared toward such a goal, but with the understanding that it is not possible to have one single system to deal with all spoken language phenomena.

In what follows, we first discuss what we believe to be some of the most important features of spoken Chinese that need urgent attention in transcription. These features are selected based on these criteria: 1) highly recurrent in spoken Chinese; and 2) posing the most trouble in transcription. After a discussion of the issues involved, we will explore some possible ways to transcribe them—with available systems. Further challenges and potential solutions will be provided following this discussion.

2. Prominent features of spoken Chinese and current systems

In the case of spoken Chinese, numerous language-specific features need to be addressed by discourse researchers. We will begin by outlining some of the most important issues in Chinese discourse transcription, which, as stated earlier, meet the criteria of 1) highly recurrent in spoken Chinese; and 2) posing the most trouble in transcription, and in doing so, we also review some of the solutions provided by some of the widely used transcription systems. Prominent features that meet these criteria include: discourse particles; non-lexical vocalizations; and repairs.

It is noted that almost all of the current transcription systems are not designed for
Chinese per se but are nevertheless being applied to the transcription of the Chinese language. We would like to highlight Chinese spoken discourse features and examine how these systems have been applied to Chinese data and how effective they are in so doing, as represented by current Chinese scholarship. We note that four main transcription systems have been used in representing Chinese data, which include the Conversation Analysis (CA; see Jefferson, 2004) system, the UC Santa Barbara system (SBS; see Du Bois et al., 1993), the Halb interpretative Arbeits transkriptionen (HIAT; see Ehlich, 1993) system, and the Gesprächs analytisches Transkriptions system (GAT; see Selting et al., 2009) system[9-12]. We now discuss in detail the three prominent features of spoken Chinese identified above, evaluate the application of the existing systems to Chinese, and offer our solutions as appropriate.

2.1 Discourse particles

As with many other languages, discourse particles are extremely common in spoken Chinese. They are typically non-lexical units that serve various pragmatic functions, including the organization of conversational interaction (Norrick, 2009; Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1996; Tao, 2003; Tseng, 2001, 2004) [13][14][15][16][17][18]. Mandarin discourse particles tend to occur at both the turn-initial position (Han, Dong, & Xue, 2010; Tseng, 2001) and the turn-final position (Chappell, 1991; Tao, 1996)[19][16][20][21].

Representations of discourse particles vary a great deal among currently available corpora of spoken Chinese. These variations may be due to the inherent linguistic features of the Chinese language, or they may occur when transcribers-unintentionally or deliberately-focus on specific aspects of speech. The vast number of homophones in Mandarin Chinese creates substantial orthographic variation: Calculating all possible combinations of Mandarin Chinese syllables plus lexical tones yields about 400 different segmental syllables and 1,277 distinct tonal syllables for the entire language while multiple Chinese characters may correspond to the same syllable. This results in discourse particles with more than one choice of characters available: ‘a’('啊', '阿'), ‘ai’('诶', '欸', '哎', '唉'), ‘ai yao’('哎哟', '哎喲'), ‘o’('喔', '噢'), ‘hei’('嘿', '嗨'), for example, are just some of the commonly seen cases.

At the same time, each of these morphemes has its own lexical content when represented with a different orthographic shape. For example, with the same syllable ‘ai’, ‘哎’ has the semantic meaning of ‘surprise, dissatisfaction’, while ‘唉’ has the function of ‘response’, which may result in different interpretation of the transcription. However, such subtle differences are often ignored and are used interchangeably in transcription. Whereas a human reader can
easily spot both variants based on sound similarities, computer programs and search engines may treat them as completely different terms. Therefore, if a researcher searches the data for only one variant without considering other potential variants, the result might be partial and thus misleading.

Practices in transcribing discourse particles vary, which prompts Lu et al. (2014) to suggest that if orthographic variations exist, transcribers should use the most commonly used forms listed in the Xiandai Hanyu Cidian dictionary[22]. They provide some recommended spelling conventions. In current practice, however, there is little consistency. For example, the Linguistic Data Consortium (2008) provides standardized Romanization spellings used to transcribe interjections[23]. The BJKY and the CADCC corpora projects use Chinese characters to represent the discourse particles, with a few established spelling conventions.

Among those who opt for phonetic notations, Tseng & Liu (2001) and Tseng (2006) transcribe all discourse particles in capital Roman letters (‘A’, ‘AO’, ‘EI’ etc.) whether the discourse particle is lexicalized or not[24][25]. They argue that readability is improved by using upper-case Roman letters to reflect the actual pronunciation of a word. In addition, among the most active researchers who have transcribed spoken discourse in research and who have published works in English, Roman alphabetic notations are used more often than Chinese characters. Biq (1998), Clancy, Thompson, Suzuki, & Tao (1996), Chui (1996), and Wu (2006) transcribe discourse-related items in alphabetic notation or Pinyin without tone marks[26][27][28][29]. In the SINICA project, turn-initial exclamations, discourse particles, repairs, filled pauses etc. are all written in capitalized Roman letters. The NCCU corpus and Li (2014), on the other hand, uses both Chinese characters and alphabetic characters or pinyin[30].

When representing Mandarin discourse particles, researchers must decide how to present them orthographically. These representations usually involve either using standard Chinese characters or a Romanization-based spelling of the word, and preferably both. For this category of spoken elements, we recommend the solutions offered in Lu et al. (2014)[22]. This study suggests that character plus phonetic details should be combined in cases where critical phonetic details are useful. Furthermore, if orthographic variations exist, transcribers should use the most commonly used forms listed in the Xiandai Hanyu Cidian dictionary. For specific spelling conventions, the reader is referred to Lu et al. (2014)[22].

2.2 Non-lexical vocalizations

The second type of prominent spoken features in Mandarin Chinese is non-lexical
vocalization. While most speech sounds can be transcribed with Chinese characters, transcribers are often confronted with words and sounds for which there are no obvious spelling conventions or character representations. Their use in spontaneous conversation normally deviates from original semantic content and unified semantic use is not shared even among native speakers. Therefore, they cannot be mapped exactly to established orthographic character standards.


Many editing terms are non-lexical vocalic forms. Editing terms can be hesitation markers, filled pauses, truncated sounds, special lexical forms, and so forth. For example, some demonstratives and other lexical words are used as filled pauses, e.g. zhege ‘this’, na ‘that’, nage ‘that’, shenme ‘what’, jiushi ‘just’, jiu ‘just’, among others. Non-lexical forms such as ‘m/mm/mmm’, ‘mh/mhm’, ‘hum’, ‘han’, ‘hon’, ‘um’, ‘eh’, ‘uh’, ‘unh’, ‘hm’, ‘huh’ etc. can also be used as filled pauses and hesitation markers.

In the UC Santa Barbara transcription system, Du Bois et al. (1993) argue that spellings for these non-lexical vocalizations can be derived from an already existing informal spelling convention discernable in the practice of playwrights, novelists, and cartoonists[10]. In CA, however, the spelling of words is altered to reflect how the speaker pronounces the word (termed ‘eye-dialect’ in Edwards 1992)[31]. With no established spelling conventions in Chinese due to the dominance of the character system and the limited use of the Romanization system (Pinyin and similar systems), researchers transcribe what they hear in diverse ways. Lu et al. (2014) suggest a viable Chinese transcription system for these non-lexical vocal cues, which would fall into the intermediate transcription category[22]. They argue that if there is no proper lexical entry in the dictionary, researchers should use the Chinese character that is closest to it in terms of pronunciation. For the intermediate level transcription, what is suggested by Lu et al. (2014) is to encode the actual pronunciation of a word in Pinyin form within parenthesis[22]. The HIAT system uses literary transcription, an adaptation of
orthographic conventions with the goal of capturing acoustic-auditory variants without specialized phonetic training. International Phonetic Alphabet (IPA) can also be used for capturing pronunciation variants. IPA is more precise than literary transcription, and some researchers have proposed using it in transcription (International Phonetic Association, 1989)[32]. However, specific phonetic and phonological training are necessary to use IPA, which is not widespread in all disciplines of discourse analysis.

In some transcription systems, spelling is not altered to reflect how the speaker pronounces the word. Instead, they use an alphabet-based literary transcription in place of the traditional orthographic representation of the same word. In some transcription systems, transcribers use the appropriate or similar Chinese character which can be replaced with the standard spelling of items. The fact that the pragmatic functions of Chinese backchannels is often more important than lexical meaning makes it necessary to mark these editing terms in an explicit way.

Thus our recommendation is to first of all represent these form as faithful as possible. In terms of forms to be used, again we recommend the character plus phonetic symbol hybrid system where character alone would not be sufficient. We further suggest that researchers list the phonetic symbols where no characters are available (e.g. ‘erm’ where there is a nasal ending in the syllable).

2.3 Repairs

The final prominent category in spoken Chinese is repair. Repair is defined as “any instance in which an emerging utterance is stopped in some way, and is then aborted, recast, or redone” (Sacks et al., 1974, 1977; Fox & Jasperson, 1995)[14][33][34]. As with other languages, speech repairs are frequent in Mandarin Chinese conversation. While much work has been done on Chinese repair (Chui, 1996; Tseng, 2006; Zhang, 1998), little reflective work has been carried out on the transcription of Chinese repair[28][25][35]. In this study, we deal with four major repair patterns: disfluent repetition, restart, abandonment repair, and overt repair. Disfluent repetition is direct repetition which cannot be explained or justified by Mandarin grammatical rules. In this type of repair, there is no alteration of the original message and the previous words are repeated without change. Thus we cannot clearly identify what is to be corrected and what is the correction. Levelt (1983) characterizes the repeat of one or more lexical items as covert repair[36]. Another characteristic of disfluent repetition is that many cases involve pauses and/or editing terms before the repair and/or repetition of the word.

The NCCU project closely transcribes cases where words are repeated and, after the
interruption and editing phase, the utterance is continued. But it does not use any transcription conventions to denote the disfluent repetition. Example (2) from an NCCU conversation shows that the speech is stopped at the conjunction *yinwei* ‘because.’ There is one word within a syntactic pattern immediately repeated without any lexicalized elements in between.

(2)

.. 因為.. 因為我們是走斜坡下去

*yin1wei4 yin1wei4 wo3men shi4 zou3 xie2po1 xia4qu4*

because because 1PL COP walk slope down

‘because..because we were driving down the hill.’

Restarts are partially repeated word sequences. Speech interruption can occur within a word, as exemplified by *zhen* of the whole compound *zhenhong*, a personal name. The interrupted name is completed with the completion of the entire repair sequence.

(3)

... (1.3) 禎.. 禎宏已經翻那麼多年了

*zhen zhenhong yi3jing1 fan1 na4me duo1 nian2 le*

REPAIR Zhenhong already translate so many year PRF

‘Zhenghong has already been translating for so many years.’

These first two repair patterns suggest a clause-internal constraint on recycling. However, in the third repair pattern below, the speaker abandons the whole construction completely. In Example (4), the incomplete declarative construction *jiushixiang* ‘it is like’ in the first place is abandoned, replaced by a question construction *weishenme* ‘why’, which suggests a new message is forthcoming.

(4)

..就是像..為什麼我們說我們是..前三志願

*jiu4shi4xiang4wei4shen2me wo3men2shuo1 wo3men2shi4 qian2san1zhi4yuan4*

that-is like why 1PL say 1PL COP top three choice

‘That’s like..why should we say that our school is the top three choices?’

In the SBS system, plain angle brackets ‘<>’ are used to enclose words that are false starts.
To illustrate this type of repair pattern, Example (4) has been slightly modified and reproduced as Example (5).

(5)
..<就是像>.為什麼我們說我們是.前三志願
jiù·xiàng. wèishénme wǒmen shuō wǒmen shì qian·san zhìyuán
that·is like why 1PL say 1PL COP top three choice
‘That’s like…why should we say that our school is the top three choices?’

Finally, an overt repair contains both the “repaired segment” and the “repairing segment,” that is, “the portion of the utterance which is being repaired…[and that] which is accomplishing the repair, respectively” (Fox and Jasperson, 1995)[34]. In an overt repair, we can clearly identify what is to be corrected and what is the correction. In Levelt’s (1983) transcription, overt repair information is placed immediately beside the relevant syllables: reparandum (Rep), editing terms (Et), and alteration (Alt)[36]. The reparandum denotes the speech that needs to be repaired, and the alteration is the repair itself. Editing terms are sequences produced between the reparandum and the alteration. Editing terms such as discourse particles, pause fillers or lexicalized items are used to bridge the gap between the reparandum and the alteration. These are recorded using single parentheses. Example (6) from the NCCU corpus indicates that *keshi* ‘but’ is substituted by *piru* ‘for example.’ ‘uh’ is used between them as an editing term.

(6)
可是..uh譬如說他有三點的課..三點半的課
ke·shì uh pi·ru·lǐ·bǐ·shuō tā yǒu sān diǎn dì de kè..sān diǎn bàn de kè
but PF for·example say 3SG have three o’clock ASSC class three o’clock half ASSC class
‘but, for example, if he has a class at three…at three thirty’

To illustrate this type of repair pattern, Example (6) has been slightly modified and reproduced as Example (7).

(7)
可是(Rep).. uh(Et) 譬如說(Alt)他有三點的課..三點半的課
ke·shì uh pi·ru·lǐ·shuō tā yǒu sān diǎn dì de kè..sān diǎn bàn de kè
but PF for-example say 3SG have three o’clock ASSC class three o’clock half ASSC class
‘but, for example, if he has a class at three···at three thirty’

Transcribers often skip repaired words as they are deemed incomplete, ungrammatical, or simply uninteresting. In other cases, transcribers transcribe them randomly. Consequently, it is difficult to take a close look at repair words and their functions in conversation. We thus suggest that orthographic and phonological features related to repairs need to be embedded in transcription to serve for many research functions of constructing a corpus. Furthermore, as the reader of the transcription may not always have access to the interrupted lexical form without knowing the intended Chinese word, it is advisable for researchers to mark repairs in Chinese character form wherever applicable.

3. Concluding Remarks

It should come as no surprise that transcript variations exist. Transcriptions are by their very nature limited, selective, and biased. There can never be one standard method of putting spoken word to printed/visual form, and no single system of transcribing speech could ever accommodate the needs of all users or serve every research project (Du Bois, 1991)[37]. In addition to these limitations, most researchers do not have the time or resources to include every detail of talk and interaction? they only include the details needed for a particular research approach (Bucholtz, 2007; Cook, 1990; Ochs, 1979)[8][6][1]. Thus, no transcription can record every detail of discourse as it occurs in real time, and differing transcription systems are necessary to suit different research needs.

However, researchers should explicitly describe the transcription conventions used in order to avoid the confusion that may be caused by multivalent symbols. It is also important to avoid inconsistency in the use of transcriptional conventions. Having a clear and consistent means of transcribing Mandarin Chinese conversation helps to construct large corpora and enables data sharing between researchers across disciplines.

While a complete transcription system or a complete guideline for transcribing all major details of spoken Chinese may not be possible with a single project like ours, we do hope that the issues we raised here will get the attention of practitioners in the field of Chinese discourse analysis. The recommendations made by the present authors are meant to be geared toward the goal of consistency and cross-study compatibility, and it is our hope that some consensus can be built toward more efficient systems for data processing as well as sharing
based on the issues we discussed here and the suggestions made based on the discussion.

References


(1996), pp. 52-133.


