Abstract

This paper investigates the structure of Mandarin spoken dialogues by analysing the distribution of words and turns used in dialogues. The results of an empirical qualitative study show that independent of speakers, there exists a kind of basic vocabulary for daily Mandarin conversations. It is proposed that this is the minimal set of a lexicon for the use of spoken Mandarin. Moreover, a number of words in the basic vocabulary were specifically used for marking various constituent boundaries in discourse, such as turn-initial and utterance-final. Discourse markers and disfluency are also taken into consideration for their highly frequent occurrences and their function of marking significant positions in spoken discourse. By means of lexical distribution and its interaction with turn taking, this paper demonstrates a new attempt to analyse the structure of Mandarin spoken dialogues.

1. Introduction

The importance of collecting and analysing real spoken materials has been approved by the growing interests shared by linguists and computer scientists. On the one hand, recently developed theoretical considerations on spoken language have mainly been done by identifying the related phenomena through corpus studies. On the other hand, application-directed corpora collections also immediately contribute to the advances on speech analysis and processing. The complexity and the rich variety of spoken language phenomena [7] and [12] are the most fascinating aspects which can only be observed in spontaneous spoken dialogues. Spoken dialogues usually consist of multiple interlocutors, various topics and contain a large number of spontaneous speech phenomena. In this paper, spontaneous speech phenomena will be analysed from the point of view of lexical distribution and turns.

Thus, the study conducted in this paper focuses on the structural characteristics of Mandarin spoken dialogues and aims to explore lexical use in dialogues at various levels, namely the speaker level, the type-token level, the turn-taking level and the discourse level. What and how spoken dialogues are composed of are two essential issues discussed in this paper. Word frequency indicates the preference of speakers on choosing words they like to use. Discourse markers and disfluency dominantly occur in spontaneous speech, seldom in written texts. In addition, their function and location in spoken discourse can provide useful information on local turn and sentential structures.

2. Distribution of Words across Dialogues

Empirical observations have been made by examining Mandarin spoken dialogues. Data used to obtain the results are a part of the Taiwanese Putonghua Corpus (LDC98S72). Three dialogues have been analysed for our purpose. Putonghua refers to Mandarin. All subjects were born in Taiwan and their first language is Taiwanese. The subjects were given the instruction to talk on any topic they wanted to. The spoken data obtained are largely spontaneous and natural. Each dialogue is about 20 minutes long. Major topics in the dialogues are family, work and study, but the communication partners in dialogues did not stick to a specific topic all the time.

We used the original transcripts of the speech data with syntactic annotations adopted by adjusting the tagging system designed for the Academia Sinica Balanced Corpus [1]. It should be noted that the majority of the Sinica Corpus data are written texts and the tags have been developed restrictively for well-formed written sentences. Hence, for the completeness of coverage, we added extra-tags for ill-formed sequences and spontaneous speech phenomena (details cf. [17]).

2.1. Individual Speakers

The first step to investigate lexical distribution of dialogues is the separation of data produced by individual speakers, as shown in Table 1. 1-A/1-B stand for subjects in the first dialogue, 2-A/2-B and 3-A/3-B for the second and third dialogues, respectively. Because there is no unified definition of the unit “word” for Mandarin, the system used here was adopted from Sinica Corpus.

Table 1: Total Number of Types and Tokens

<table>
<thead>
<tr>
<th>Subjects</th>
<th>1-A</th>
<th>1-B</th>
<th>2-A</th>
<th>2-B</th>
<th>3-A</th>
<th>3-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokens</td>
<td>1688</td>
<td>1937</td>
<td>1247</td>
<td>1375</td>
<td>1782</td>
<td>1096</td>
</tr>
<tr>
<td>Types</td>
<td>587</td>
<td>585</td>
<td>439</td>
<td>522</td>
<td>616</td>
<td>430</td>
</tr>
</tbody>
</table>

2.2. Types and Word Frequency

We obtained a cross-speaker lexical distribution by analysing the relationship between types and word frequency. Words of the same token frequency were listed for comparison. In Figure 1, a similar pattern of lexical distribution among the subjects is shown. Words, which were only produced once and twice in the whole dialogues make up more than 80% of the total number of types for each subject. Therefore, we further look at words, which were used more frequently than twice.
2.4. Basic Vocabulary in Mandarin Dialogues

The elementary lexicon coverage described above is valid for each subject. The next step is to analyse the syntactic categories of these words. In total, there are 22 words. They are grouped as listed below. First to note is that the so-called nà-zhè words consist of determiners and adverbials: nà, nàge, nàme, nàyàng, zhè, zhège, zhéme and zhèyàng.\(^1\) These words serve the pragmatic purposes of discourse particles. When used in conversation, their function is more or less like connectives so, therefore, or then. Among these 22 words, 9 can clearly be identified as discourse markers: 3 nà-zhè words, 4 discourse particles and 2 adverbs jiù, dài. As expected, verbs and pronouns build an essential part of the minimal spoken lexicon for dialogues. In addition, there are 2 temporal-related words: time and the aspect particle, shìhòu and le. For spatial expressions, the verb zài (is located) was most frequently used in the dialogues.

- 6 verbs: zài, shì, jiùshì, sìlàō, yào, yǒu, “is located”, “is”, “that is”, “say”, “want”, “have”
- 4 discourse particles: EN, AI, LA, A,
- 3 adverbials: jiù, dài, “then”, “all”, “right/correct”
- 3 nà-zhè words: zhèyàng, nàge, nà, “this way”, “that one”, “then”
- 3 pronouns: nǐ, wǒ, nǐ, “he”, “I”, “you”
- 1 grammatical particle: le, usually for marking the past tense
- 1 noun: shìhòu and “time/time point”
- 1 negation: méiyǒu, “have not/have no”

Some of these words were investigated in recent pragmatic studies. Though, until now, no corpus analysis has been conducted for this purpose, the pragmatic functions have been taken into consideration such as backchannels (EN, AI, A) and reactive tokens (shì, jiùshì, dài, zhèyàng) in Clancy et al. [3]. The importance of spoken dialogue elements has obviously been noted in the literature. It would be interesting, for the next step, to explore the distribution of these words in terms of turns in dialogues and to see if this basic vocabulary has any relationship to the structure spoken dialogues.

3. Turns and Discourse-Marking Elements

This section focuses on the interaction of turns and three discourse components, which to some extent contribute to the structure of spoken discourse [9]. These are 1) specific words marking turn-initial and other turn- or utterance-related positions, 2) discourse markers and 3) speech repairs.

3.1. Marking Turn-Initial Positions

By investigating the same data, Tseng [17] reported that about 30% of the overall turns were initiated by discourse particles EN, OÙ, HAI, AI, O, A and E. The confirmation words dài and shì\(^2\) and the nà-zhè words nà, zhèyàng\(^2\) and zhèyàng are 5% and 6.6%, respectively. That is to say that these 12 words mark more than 40% of the total turn-initial positions. Most of them are function words. Among them, the discourse particles even do not have lexicalised semantic meaning.

Next, we investigated the relationship between all tokens of these words and their positions in turns: the distribution of these words in turn-initial and non-turn-initial positions. We would like to make sure that not a thin minority of all tokens of these words were produced in turn-initial positions, but a meaningful part of them. The results show that more than

\(^1\) Nà, nàge, nàme, nàyàng, zhè, zhège, zhéme and zhèyàng mean “that”, “that one”, “that way, then”, “that way”, “this”, “this one”, “this way, so” and “this way, so”.

\(^2\) Dài and shì can also be used as adjective “correct” and verb “is”. When they appear in turn-initial position, they usually confirm previous utterances, meaning right and yes.

\(^3\) Zhèyàng and zhèyàng have similar meanings “this way”, or “so”.
50% of all tokens of discourse particles EN, HAI, AI, O and OU as well as bei and bei A were found in turn-initial position. Based on the results, the set of words marking turn-initial positions is restricted to these seven words.

### 3.2. Discourse Markers

Discourse markers are usually defined as elements marking particular pragmatic functions in discourse, although they are often considered as non-lexicalised items. However, they do have the function of marking important sentential positions in spoken discourse [4], [6] and [14]. In total, 23 different discourse particles were produced in the dialogues, with the average of one particle per turn. Among them, 10 particles were preferred by all subjects and were used very frequently. It seems that speakers of a common language tend to use the same particles for specific purposes in spoken discourse. There also exists a regular mapping between the types of discourse particles and their sentential positions. MA is normally used in utterance-final, whereas EN is more likely to be found in the utterance-initial position [17].

Calculating discourse particles and nà-zhé words together, we found 1.6 particle words per turn on average. Similar to the typical discourse particles, nà-zhé words are more likely to be found in certain sentential positions than in other positions. For instance, nà (determiner) as a connective appears mostly in utterance-initial position, while nàgê (determiner + classifier) seems to retain more characteristics of determiners and is often located before nouns in mid-utterance position. Nà-zhé words are a good piece of evidence supporting the notion that some words are used differently in written and spoken language. More concretely, nà-zhé words are often used in spoken dialogues as particle-like discourse markers instead of determiners.

### 3.3. Speech Repairs

Among a number of research results reported on speech repair and disfluencies [10], [11], [13] and [15]. Chui [2] has investigated the properties of Chinese speech repairs. However, the interaction between speech repairs and discourse structure has not explicitly been taken into consideration in her results. In our study, exclusive of simple hesitations and repetitions, 373 immediate speech repairs were identified. Astonishingly, only 27 speech repairs among them contain an editing term. These are AI, A, HE, E, OU, NE, HEI and LA. The “edit signal hypothesis” has been approved by Labov [8] and Hindle [5]. The hypothesis proposes that speech disfluency in spoken discourse (they call it non-fluency) is usually accompanied by editing terms. However, our type of Mandarin spoken data does not explicitly support this hypothesis.

It is also to be noted that some discourse particles were only found in particular sentential positions. To take the most frequent discourse particle found in the turn-initial position EN as example, it is not included in the set of editing terms found. We interpret this result in the following way. EN is not used to indicate self-monitoring of the speaker, but to signal the intention of speaker to be ready to take over the turn. Therefore, turn-initial position and editing term position are somehow complementary positions from the pragmatic point of view.

With regard to the number of words involved in Chinese speech repairs, more than 10% of the overall words were found in repairing sequences (detailed definitions of speech repairs cf. [18]). An interesting result similar to a study on German speech repairs [18] is that the most likely position for Chinese speech repairs to be initiated is the phrasal boundary. For Chinese speech repairs, the second most likely position to initiate a speech repair is the morpheme with the central semantic content of the problem word involved. This is a clear evidence for the hypothesis that syntactic features play a role for the speakers to process the organization of their speech. When they have to interrupt their utterances or to resume their speech, they prefer certain syntactic locations.

### 4. Components of Mandarin Dialogues

This section proposes a preliminary annotation system for describing the structure of Mandarin dialogues by making use of previous results and the derived features.

#### 4.1. Annotations

Based on the previous analyses, we propose a system for transcribing Mandarin spoken dialogues, which is applied in our corpus collection project. Four elementary tiers are proposed to construct a multi-modal structure description of Mandarin dialogues. These are units, boundaries, discourse-marking signals and temporal alignment. Units consist of lexical items, phrases and turns. The reason for excluding utterances is that utterance boundaries in spontaneous speech are difficult and ambiguous to determine [16]. It can be left for further detailed annotation systems. Instead of choosing utterances as coordinating units between words and turns, we put our emphasis on the intermediate unit: phrase.

<table>
<thead>
<tr>
<th>Tiers</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td>blanks between words</td>
</tr>
<tr>
<td></td>
<td>&lt;p&gt;phrase&lt;/p&gt; for phrase</td>
</tr>
<tr>
<td></td>
<td>&lt;s Speaker_Name&gt;turn&lt;/s Speaker_Name&gt;</td>
</tr>
<tr>
<td>Boundary</td>
<td>&lt;!&gt; for unknown boundary</td>
</tr>
<tr>
<td>Discourse-Marking Signal</td>
<td>[d]discourse marker/[d] for discourse marker</td>
</tr>
<tr>
<td></td>
<td>[r]repair/[r] for repairs including repetitions</td>
</tr>
<tr>
<td></td>
<td>[par]particular sequence/[par] for particular sequence</td>
</tr>
<tr>
<td>Time</td>
<td>Time is given in all the brackets in ms.</td>
</tr>
</tbody>
</table>

As shown earlier, words, phrases and turns are all related to specific discourse-marking elements and functions from different perspectives. Because boundaries are already contained in the annotation for units, our annotation system only takes care of the unknown boundaries, for instance boundaries between interrupted words or phrases. Discourse-marking signals mark the location of single words or word sequences with particular discourse-related pragmatic functions, for instance discourse markers, speech repairs and particular sequence. All these three tiers have to be connected to a time tier to have all these elements temporally aligned.

#### 4.2. An Example: Overlapping and Alignment

By applying this annotation system, the problem of annotating overlapped turn-takings can be solved. We focus on changes of speakers instead of changes of turns. Furthermore, we use beginning and ending brackets to mark all sequences, so that the time alignment is possible for later labelling. Example 1 is
extracted from the ongoing project of a large Mandarin information dialogue speech corpus (MISC) to demonstrate this annotation. After adding the proposed annotations, the turn produced by speaker B is in the middle of the sequences produced by speaker A and we can easily get the marked up overlapping sequence.

**Example 1: a simplified annotated example**

```xml
<s A><r>cong shili yao dao shili meishuguan dehua[/r] women jiushii keyi zhiyao ni (d)[EN]/(d) jushi cong cong women xuexiaoyi yiyang da ersanliu zui fangbian de (d)[A]/(d)<s A><s B><d>MHM</d>(d)<s B><s A><r>ranhou [r]da dao yeshi dao gongguan[/r]) huan danshuixian ranhou ni zhiyao zuo dao dagai [/r]yuanshanzhan[/r]<s A/>
```

In Example 1, it should be noted that time alignment can be put into each beginning and ending brackets to have the transcript aligned with the speech signals. Similarly, to investigate the distribution of phrases, we can also assign phrasal brackets to each phrase. Furthermore, discourse-marking sequences and their positions in turn and time can be easily illustrated. For instance, the repair `[r]cong shili yao dao shili meishuguan dehua[/r]` was produced directly in the turn-beginning position and the discourse particle `(d)[A]/(d)` in the turn-final position.

5. Discussion and Conclusion

This paper has explored the distribution of words used in spontaneous spoken dialogues as well as the discourse-related sequences such as discourse markers and speech disfluencies from the structural point of view. Especially, it focused on the turn-related positions. This shows that the complexity of spoken dialogue structure is worth further investigating; especially the role of turns should receive more attention. More works should be done for determining constituent units and boundaries. This can be words, phrasal patterns or prosodic representations. This paper also showed that investigations on dialogue structure also contributes to the development of more practical and goal-oriented speech annotation systems. Further work on annotating corpus data and analysing dialogue structure on a bigger scale is in progress.

6. Acknowledgment

The study presented in this paper was in part supported by National Science Council (NSC-89-2411-H-001-098).

7. References