Dialogue-Act Analysis with a Conversational Telephone Speech Corpus Recorded in Real Scenarios

Keyan Zhou[1], Aijun Li[2], Chengqing Zong[1]
[1]*NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190*
[2]*Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, 100732*
E-mail: [1]{kyzhou, cqzong}@nlpr.ia.ac.cn, [2]liaj@cass.org.cn

Abstract

CASIA-CASSIL is a large-scale corpus of Chinese spontaneous telephone conversations in tourism domain underdevelopment. This paper gives some statistics of linguistic characteristics based on the Dialogue-Act (DA) annotation in CASIA-CASSIL. Distributions of DA are presented and compared in different domains. And also, we describe and discuss two kinds of Question sentences in detail, which are Yes-or-No question and Wh-Question. In Yes-or-No question sentences, a large part of them can be called intonational question realized by intonation cues rather than any question markers. We believe the details on linguistic and paralinguistic information will help to study the prosodic analysis pertinent to DAs.

Key words: Dialogue-Act, question sentence, conversational telephone corpus, speech annotation

1. Introduction

Spoken language processing, including spoken language translation, speech recognition, spoken dialog system, and summarization is one of the most important research area in Man-Machine interactive system. Currently, data driven or machine learning is state-of-the-art technology in spoken language processing for mining and utilizing complex discourse phenomena. The technology will be benefited from the large-scale conversation corpus with rich phonetic, linguistic and paralinguistic annotation. In the last few decades, several English conversation corpora have been published, such as Switchboard-DAMSL (Jurafsky *et al*., 1997) of telephone conversations, the ICSI Meeting Corpus (Janin *et al*., 2003) and the AMI Meeting Corpus (Carletta *et al*., 2006) of natural meetings. Meanwhile, few researches on spoken Chinese discourse have been reported, implying that such an annotated corpus of Chinese dialogs is unavailable yet.

One of the studies that we are focusing on is to use the prosodic information, such as intimation, stress or prosodic boundary, to help the speech recognizer give corrective dialogue acts judgment.

In the corpus, we found a lot of question sentences called as intonational questions, i.e., the questions are realized by intonation cues rather than syntactic information such as question markers. How to recognize this kind of question is a big challenge for present system.

One example picked from our corpus is given in Figure 1, where the two counterparts have same texts but different intonations.

Q: "tai4 yue4 yuan2 er4 haor4 lou2 san1 ling2 jiu3?"(太月园二号楼三零九?)

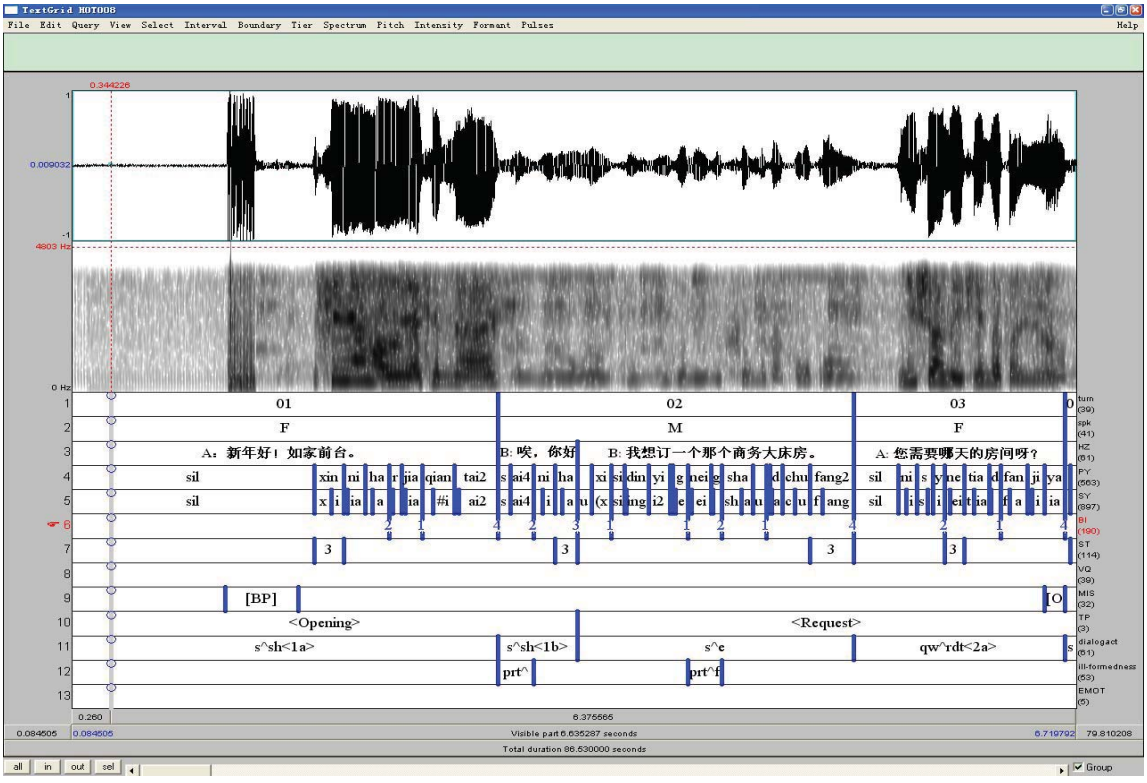
S: "tai4 yue4 yuan2 er4 haor4 lou2 san1 ling2 jiu3." (太月园二号楼三零九.)

Sentence *Q* expresses a question while *S* is a statement. For these two utterances, the automatic speech recognition results will be the same. Such questions like *Q*, which contains no question markers, are called intonational questions, which widely exist in Chinese spoken language.

These two utterances share the same prosodic structure. Besides, the sentence stresses of *Q* and *S* are both in character "er4". Then, how can we distinguish the sentence type of *Q* and *S* from prosodic information? Boundary tone is considered as the information carrier about intonational question and statement (Lin, 2006a; 2006b; Sun, 2006). The *Q* has a high boundary tone H%, while *S* has a low one L%. As in Figure 1, the boundary tone of syllable "jiu3" (a low dipping tone) shows a high rising part for *Q* while a low contour for *S*. Additionally, the final prosodic word has greater F0 range for *Q* than that for *S*.

Figure 1. F0 contours of a Question and Statement counterpart

Figure 2. The Annotation Interface of a Dialogue using Praat
(Three panels from top to bottom are for waveform, spectrogram, and 13 annotation layers.)



We suppose that this kind of phonetic knowledge will improve the performance of the speech recognition. Therefore, the detailed linguistic information and the dialogue acts must be surveyed first to give a clear map for the further acoustic analysis.

Remainder of this paper is organized as follows. Section 2 gives a brief introduction to CASIA-CASSIL corpus. Section 3 shows some statistics based on the annotation. Section 4 describes and discusses two kinds of question sentences in detail, which are Yes-or-no questions and Wh-Questions. Finally, we give concluding remarks in Section 5.

Res			Air			Hotel		
Label	**count**	**%**	**Label**	**count**	**%**	**Label**	**count**	**%**
qy^raf	392	14.1	qy^raf	853	20.5	qy^raf	897	15.4
s^aa	331	12.0	s^e	769	18.4	s^b	491	8.5
s^e	294	10.6	s^aa	476	11.4	s^aa	487	8.4
s^b	224	8.1	s^b	376	9.0	s^e	480	8.3
s^sh	170	6.1	qw^rdt	338	8.1	s^na	343	5.9
qw^rdt	153	5.5	s^sh	249	6.0	s^sh	305	5.2
s^cc	140	5.0	s^bye	115	2.8	qw^rdt	273	4.7
is^co	132	4.8	s^na	113	2.7	s^cc	226	3.9
s^f	96	3.5	s^cs	103	2.5	is^co	190	3.3
s^m	87	3.1	s^ft	92	2.2	s^df	148	2.5

Table 2. Top Ten DA Labels and Their Percentages (%) in Three Domains

2. CASIA-CASSIL Corpus

CASIA-CASSIL, a large-scale corpus of Chinese spontaneous telephone conversations in tourism domain, is now being built as a fundamental corpus for study on spoken Chinese phenomena. To develop the first edition of CASIA-CASSIL, we have collected a large number of spontaneous telephone recordings up to the present. After a strict selection, only a minority of dialogs remains, which are with good voice-quality, enough turns and strictly belong to required domains.

The annotation is designed as a multi-leveled framework based on previous annotation systems (Li *et al.*, 2000; Li *et al.*, 2001; Li, 2002; Li and Zu, 2006). Each level is time-aligned to the audio data. The annotation guideline is described in detail in (Zhou *et al.*, 2010).

These selected dialogs are then transcribed and now being annotated including Turns, Speaker Gender, Orthographic Transcription, Chinese Syllable, Chinese Phonetic Transcription, Prosodic Boundary, The Stress of the Sentence, Non-Speech Sounds, Voice Quality, Topic, Dialogue-Act (DA) and Adjacency Pairs (AP), Ill-formedness, and Expressive Emotion.

Up to now, we have annotated 350 dialogues in three domains out of 1036 selected dialogues in total. The general information is shown in Table 1. Praat[1] is used for annotation. Figure 2 gives one sample of annotation.

[1] http://www.praat.org

Domains	Dialogues	Turns	DA labels
Air	121	3,405	4,169
Hotel	158	4,348	5,810
Res	71	2,210	2,773
Total	350	9,963	12,752

Table1. Brief Statistics of Annotated Data

3. Statistics of Labels

3.1. Statistics of Turn and DA Labels

The statistics of 350 annotated data is shown in Table 1. A turn contains one or more intonation phrase. Each intonation phrase has a particular DA label. There are two levels of DA tags: general tags (9 labels) which represent the basic form of an utterance (e.g., statement, question etc.), and appended specific tags (36 labels) which represent the function or characteristics of an utterance. Specially, considering the integrality of utterance when turn changes, we propose a tag set called interruption, which contains 3 tags (abandoned, interrupted, and indecipherable). The general tag and specific tags are separated by symbol '^', while interruption tag follows with dot '.' (Zhou *et al.*, 2010).

3.2. Distributions of DA Labels

Table 2 gives the top ten DA labels which occur most frequently in the corpus and the statistics of their percentages in different domains. Since all the recordings belong to tourism service, it is reasonable that "qy^raf" (standing for "request affirmation") is the most common DA label. The distribution of other 9 DA labels vary a lot in three domains, yet, the overall ten types are similar. The common labels are qy^raf, s^aa, s^e, s^b, s^sh, and qw^rdt.

4. Question Sentences

In general DA tag level, “s”, “qy” and “qw” are the three tags which occur most frequently. Question sentences including “qy” and “qw” are one of the most complex and common phenomenon in Chinese spoken language, because there are intonational questions and a large number of question markers.

In this section, we will give the statistical anlaysis of “qy” and “qw” according to their question markers or intonation in detail.

4.1. Yes-or-No Questions

“Qy”, which stands for Yes-or-No question, is one of the most common grammatical categories in Chinese spoken language. “Qy” is often used to confirm a specific fact or an event in conversation.

Table 3 presents 16 types of “Qy” in our annotated corpus. Majority of them have Qy-Question markers in sentence finals. Yet, intonational question “Fxqy1” takes a partition of about 19% in total.

Most of the intonational questions behave as echo question. In conversations, especially in telephone conversations, speakers always have to repeat portion or the whole of the previous utterance, thereby he/she can make an understanding check or confirmation.

Tag	Hotel	Air	Res	Description
Fxqy1	209	190	90	intonational question
Fxqy2	23	31	5	X “bu4”(不) X？ X ”mei2” (没)X？
Fxqy3	518	247	274	“ma5?” (吗?)
Fxqy4	329	235	80	“ba5?”(吧?)
Fxqy5	59	41	10	“a5?”(啊?)
Fxqy6	9	14	2	“la5?”(啦?)
Fxqy7	9	22	1	“ne2?”(呢?)
Fxqy8	14	34	7	“ya5?”(呀?)
Fxqy9	5	4	0	“na5?”(那?)
Fxqy10	16	12	1	“le5?”(了?)
Fxqy11	23	19	1	“ha5?”(哈?)
Fxqy12	12	4	0	“mei2”/”mei2 you3?” (没?/没有?)
Fxqy13	5	4	0	“hai2?”(还?)
Fxqy14	9	1	3	X ”bu5?”(X 不?)
Fxqy15	4	0	0	“bei5?” (呗?)
Fxqy16	2	0	0	“wa5?”(哇?)
Total	1246	858	474	

Table 3. Sixteen Types of Yes-or-No Question according to Question Markers. *

Type	Hotel	Air	Res	Description
Fxqw1	218	153	67	With Qy-Question markers “ne5”, “ya5”, “de5”, “le5”, (呢, 呀, 的, 了)etc. in the final boundary.
Other	271	143	81	No markers in the boundary.

Table 4. Two Basic Types of Wh-Question

Type	Hotel	Air	Res	Description
Fxqw2	38	43	17	“na3”(哪)
Fxqw3	67	19	19	“shen2 me5” (什么)
Fxqw4	16	17	1	“zen3 me5”(怎么)
Fxqw5	34	33	1	“ji3”(几)
Fxqw6	29	29	3	“duo1 shao3” (多少)
Fxqw7	1	0	0	“wei4 shen2 me5” (为什么)
Fxqw8	3	0	0	“shei3”(谁)
Fxqw1 (*)	30	12	26	Only Qy-Question markers

Table 5. A Further Division of Fxqw1

Type	Hotel	Air	Res	Description
Fxqw1 (*)	30	12	26	Only Qy-Question markers
Fxqw2	93	81	33	“na3”
Fxqw3	130	42	50	“shen2 me5”
Fxqw4	25	35	3	“zen3 me5”
Fxqw5	75	62	7	“ji3”
Fxqw6	81	61	17	“duo1 shao3”
Fxqw7	1	1	0	“wei4 shen2 me5”
Fxqw8	7	0	0	“shei3”
Fxqw9	46	2	9	Intonational Question
Fxqw10	0	0	3	“sha3”(啥)
Total	489	296	148	

Table 6. Ten Types of Qw-Question according to Question Markers

4.2. Wh-Qustion

“Qw”, Wh-Question, is another familiar question type

in conversations which contain wh-question makers such as 'what(什么), where(哪里/哪儿) who(谁), why(为什么,怎么), or which(哪个)'. Wh-Question is always used to request for details about specific matters.

In analysis, we find "Qw" of Chinese dialogues can be divided into two types as shown in Table 4. Type "Fxqw1" is the one with Qy-Question markers in the final boundary, such as "ne5"(呢), "ya5"(呀), "de5"(的), "le5"(了), and so on as always used in Yes-or-No questions shown in Table 3. Fxqw1 takes almost 50% in the Wh-Question. The utterance like "您是要坐哪趟航班啊？"(Which flight do you like to take?) belongs to this type.

"Fxqw1"can be further divided into the following types described in Table 5. "Fxqw2" to "Fxqw8" are those with Wh-Question markers as well. Fxqw1(*) are those Wh-Questions with only Qy-Question markers.

Table 6 presents the entire 10 types of the Wh-Question appeared in annotated data. The top three types in each domain are listed as follows: Fxqw2, Fxqw3, Fxqw6 in Hotel, Fxqw2, Fxqw5, Fxqw6 in Air, and Fxqw3, Fxqw2, Fxqw1(*) in Res.

5. Conclusion

Based on the annotated conversation corpus, this paper makes a statistics analysis on DA. The distribution of DA labels varies in different domain, yet there are common labels of the top 10. They are qy^raf, s^aa, s^e, s^b, s^sh, and qw^rdt. Question sentences including Yes-or-No question and Wh-Question are one of the most complex and common phenomenon in spoken language.

Yes-or-No question sentence and Wh-Question sentence are classified and analyzed based on their question markers in detail, which will be the basis for future work on phonological analysis.

Although a majority of Yes-or-No questions have Qy-Question markers in utterance boundary, intonational question "Fxqy1" takes a partition of about 19% in total. Echo-question is also noticed in Yes-or-No question for its frequency and peculiar usage. It will be a research focus in our future work as well.

Most of the Wh-Questions have Wh-Question markers, however, in order to strengthen the interrogative mood, speakers always use Qy-Question markers in the end of utterance as well. Thereby, it will be more difficulties to recognize Wh-Question from Yes-or-No question in dialogues.

6. Acknowledgements

The research work described in this paper has been partially funded by the Natural Science Foundation of China under Grant No. 60975053 and 90820303, and also supported by the National Key Technology R&D Program under Grant No. 2006BAH03B02, and CASS Key Lab project.

7. References

[1] Carletta, J., S. Ashby, S. Bourban, *et al*. *The AMI Meeting Corpus: A Pre-Announcement*. In Steve Renals and Samy Bengio, editors. 2006.

[2] Janin, A., D. Baron, J. Edwards, *et al*. The ICSI Meeting Corpus. In *Proceedings of the 28st International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong. 2003.

[3] Jurafsky, D., L. Shriberg, and D. Biasca. *Switchboard SWBD-DAMSL Labeling Project Coder's Manual*, Draft 13. Technical Report 97-02, University of Colorado Institute of Cognitive Science. 1997.

[4] Li, A., F. Zheng, W. Byrne, *et al*. Cass: A Phonetically Transcribed Corpus of Mandarin Spontaneous, *ICSLP'2000*. 2000.

[5] Li, A., B. Xu, C. Aong, *et al.* A Spontaneous Conversation Corpus CADCC, *Oriental COCOCSDA'2001*, Korea. 2001.

[6] Li, A. Chinese Prosody and Prosodic Labeling of Spontaneous Speech. In B. Bel and I. Marlin (eds), *Proceedings of the Speech Prosody 2002 Conference.* Aix-en-Provence, France, pages: 39-46. 2002.

[7] Li, A., Y. Zu. Corpus Design and Annotation for Speech Synthesis and Recognition, as a chapter in Advances in *Chinese Spoken Language Processing*, edited by Chin-Hui Lee, Haizhou Li, Lin-shan Lee, Ren-Hua Wang, Qiang Huo, World Scientific Publishing Co. Pte. Ltd., Singapore. pp. 243-268. 2006.

[8] Lin, M. C. Zhao' viewpoint of Chinese intonation and boundary tone. *Report of Phonetic Research 2006*. 2006a.

[9] Lin, M. C. Interrogative Mood and Boundary Tone in Chinese. *Report of Phonetic Research 2006*. 2006b.

[10] Sun, N.H., F0 feature of Tones and boundary tones v.s. their phonetic rules. PhD theis, CASS. 2006.

[11] Zhou, K.Y., A.J. Li, Z. G. Yin, C. Q. Zong, CASIA-CASSIL: a Chinese Telephone Conversation Corpus in Real Scenarios with Multi-leveled Annotation, *in Proceedings of LREC 2010*, Malta, May 2010.

[Published in Proc. of O-COCOSDA 2010, November 24-25, in Kathmandu, Nepal]