Joint International Conference of SNLP-O-COCOSDA, pp. 60-67, 09-11 May 2002, Hua Hin, Thailand

Collection of a Chinese Spontaneous Telephone Speech Corpus and Proposal of Robust Rules for Robust Natural Language Parsing

Thomas Fang Zheng[1,2], Pengju Yan[1], Hui Sun[1], Mingxing Xu[1], and Wenuhu Wu [1]

[1] Center of Speech Technology, State Key Laboratory of Intelligent Technology and Systems,
Department of Computer Science & Technology, Tsinghua University,
Beijing, 100084, China
Tel./Fax: +86-10-6277-2001
[2] d-Ear Corporation, Rm. 134, 1/F Tech Center, 72 Tat Chee Avenue, Kowloon, Hong Kong
Tel.: +852-3106-2112, Fax: +852-8226-0603

e-mail : [fzheng, yanpj, sunh, xumx]@sp.cs.tsinghua.edu.cn, wuwh@tsinghua.edu.cn

Abstract

In this paper, a Chinese Spontaneous Telephone Speech Corpus in the flight enquiry and reservation domain (CSTSC-Flight) of 6 GB raw data containing about 50 hours' valid speech is introduced, including the collection and transcription principles and outline. Analysis on the spoken language phenomena contained in this corpus is then performed. Based on this, four types of grammatical are proposed so as to cover as many Chinese spoken language phenomena as possible for robust natural language parsing and understanding in spoken dialogue systems.

1 Introduction

For spoken dialogue systems, a high-quality spontaneous speech corpus is much important. Such a database normally has the following purposes: (1) it can be used to train a good acoustic model suitable for spontaneous speech recognition because it contains a lot of spontaneous speech phenomena; (2) it can be used to collect and analyze the spoken sentences (in text) because it contains many spoken language phenomena, which is no doubt useful for natural language parsing; (3) it can be used to extract the domain-specific knowledge, including domain-specific keywords, key phrases and so on, for domain-specific applications.

There are often two kinds of corpus collection methods. One, speakers are asked to utter a predefined sentence set in which phonemes are well balanced; two, a monitoring program is embedded in a spoken dialogue system (human-machine dialogue) or a real-world dialogue system (human-human dialogue) to record the spontaneous speech. These two methods have their own advantages and disadvantages. For real-world applications, the second one is preferred, because it is helpful not only for the training of the spontaneous acoustic model but also for the collection of domain-specific dialogue examples and rules.

On the other hand, a spoken dialogue system is designed for use in real-world human-computer applications where the utterances are spoken instead of written, spontaneous instead of canonical. To parse such sentences in which a lot of spoken language phenomena exist, the normal rules are not enough. Based on this analysis, there should be additional types of rules that can be used to parse spoken languages.

In this paper, the authors first give the details of the collection and transcription of a Chinese spontaneous telephone speech corpus in the flight enquiry and reservation domain (CSTSC-Flight), then analyze the sentences transcribed from the corpus, and finally describe four types of rules for spoken Chinese language parsing in this domain as well as their application to the *EasyFlight* system.

2 Corpus Collection

A spoken dialogue system mainly includes two parts, a spontaneous speech recognizer, and a robust spoken language parser.

Though there are techniques available to improve the performance of the speech recognizer for spontaneous speech (Zheng 2001), a spontaneous speech corpus is still the most fundamental and important thing to do. The domain of our application to be described later in this paper is the flight enquiry and reservation, therefore our corpus is also collected in this domain. The database is named as Chinese Spontaneous Telephone Speech Corpus on Flight Enquiry and Reservation (CSTSC-Flight), which is the first in our spontaneous corpus collection series.

The corpus is collected as follows.

(1) A monitoring program is developed and installed into a computer equipped with a multi-channel telephone speech card. The computer is inserted into the telephone system of a flight enquiry and reservation agency legally under consent. With aid of the multi-thread programming techniques (Yu 1999, and Sun 2000), this monitoring and recording system support multi-channel recording of several conversations simultaneously and is running and recording automatically.

(2) The sampling rate is 8 kHz, and the samples are in an 8-bit A-law/miu-law mono PCM format.

(3) All the data are human-human conversations on the topic of flight enquiry and reservation.

(4) All the dialogues are fully spontaneous because the customers are not aware of being monitored and recorded.

(5) Each wave file contains one and only one conversion between a customer and an operator.

CSTSC-Flight contains more than 6GB raw data. Because this corpus is being collected in a fully real-world manner, there are a lot of silence segments (for example when the operator is checking the reservation status from the computer) and some other topic-irrelevant sentences. By getting rid of these useless parts, we have the CSTSC-Flight with about 50 hours' topic-related pure speech data.

There are about 12,000 useful conversations (16,000 in total) in this corpus, each for one customer, so roughly there are 12,000 speakers, with different education, gender, age, accent, and so on, and CSTSC-Flight contains rich Chinese spontaneous language phenomena.

3 Corpus Transcription

The transcription should include three kinds of information, speaker information, speech information, and text information. In this section, we will present the transcription format, method, and platform.

3.1 Transcription format

We have two kinds of transcription files. One is the binary format file while the other is text. The binary format transcription file of a specific wave file contains all the following information:

(1) **Speaker information**. Including the speaker gender (male/female), and the speaker class (operator/customer).

(2) **Sentence segmentation information**. Each wave file contains a conversation between a customer and an operator, so there could be several sentences included in a single wave file. The sentence segmentation information contains the starting and ending sample points of each sentence in this conversation.

(3) **Pinyin [1] string**. The pinyin string as a pronunciation transcription of each sentence is given.

(4) **Syllable time-boundary information**. This information is often used to train an initial acoustic model. And,

(5) **Spontaneous acoustic phenomena information.** As defined in (Li 2000), several labels related to spontaneous phenomenon are used to independently annotate the spoken discourse phenomena, including lengthening, breathing, laughing, crying, coughing, noise, disfluency, murmur, modal, lip smacking, silence, non-Chinese, and uncertain segments. Such information can be used for garbage/filler modeling.

[1] In Chinese, "Pinyin" is a text representation of a syllable, and "Syllable" is a pronunciation representation of a Chinese character.

The text format transcription file of a specific wave file contains the following information:

(1) **Speaker information**.

(2) **Sentence segmentation information**.

(3) **Chinese character string**. The Chinese character string of each sentence is given with word segmentation information.

(4) **Pinyin string**.

(5) **Spontaneous acoustic phenomena information.**

An example of a text format transcription file of a wav e file is given as below.

Gender:	F
Speaker:	Customer
0:	(0:5788) 喂你好
Word Segments:	\喂\你好\
Pinyin String:	wei4 ni3 hao3

Gender:	F
Speaker:	Operator
1:	(5788:7645) 哎你好
Word Segments:	\哎\你好\
Pinyin String:	ai1 ni3 hao3

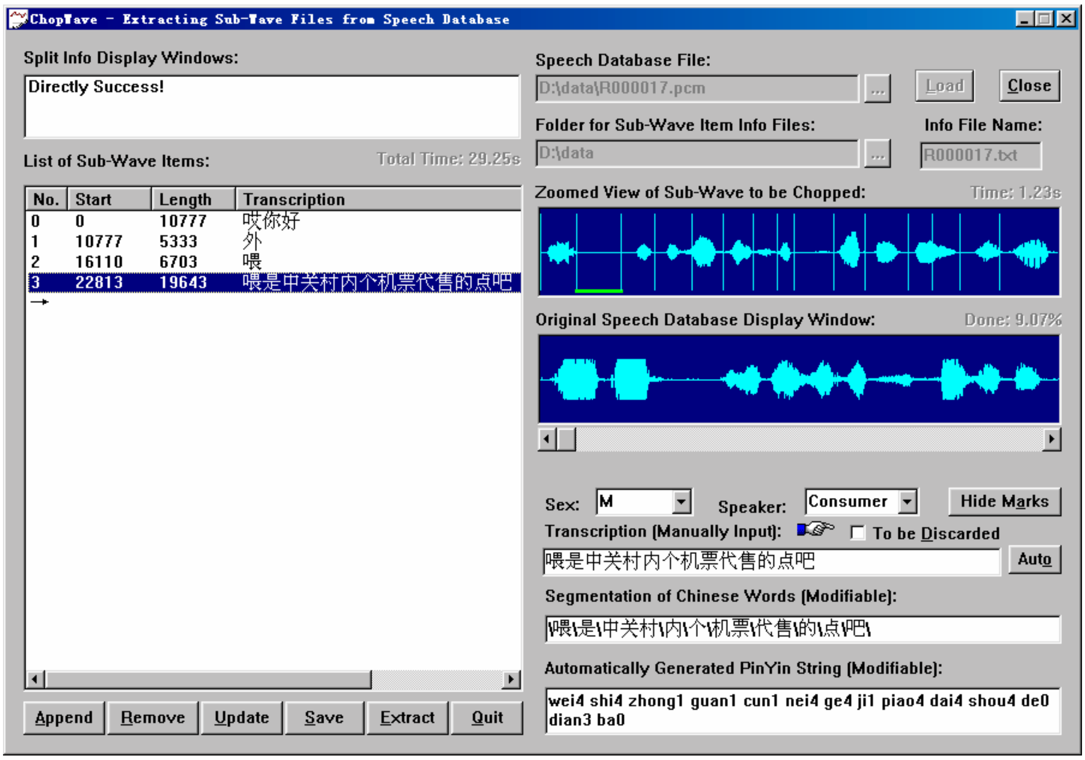
Gender:	F
Speaker:	Customer
2:	(7645:11360)售票处么
Word Segments:	售票\处\么\
Pinyin String:	shou4 piao4 chu4 me0

Gender:	F
Speaker:	Operator
3:	(11360:17640) <silence>对<noise>
Word Segments:	<silence>\对\<noise>\
Pinyin String:	<silence> dui4 <noise>

Gender:	F
Speaker:	Customer
4:	(17640:54364) 我 请 问 一 下 那 个 <lengthening>八月十号去上海的都有几个航班
Word Segments:	我\请\问\一下\那个\<lengthening>\八\月\十\号\去\上海\的\都\有\几\个\航班\
Pinyin String:	wo3 qing3 wen4 yi1 xia4 na4 ge0 <lengthening> ba1 yue4 shi2 hao4 qu4 shang4 hai3 de0 dou1 you3 ji3 ge4 hang2 ban1

3.2 Transcription tool and procedure

We design a tool to semi-automatically transcribe each conversation in the CSTSC-Flight corpus. The tool is friendly and easy-to-use. The most time-consuming parts (including the word segmentation and pinyin string transcription) can be done automatically with a little bit labor. Figure 1 shows the appearance of this tool.



Split Info Display Windows:

Directly Success!

List of Sub-Wave Items: Total Time: 29.25s

No.	Start	Length	Transcription
0	0	10777	哎你好
1	10777	5333	外
2	16110	6703	喂
3	22813	19643	喂是中关村内个机票代售的点吧

Speech Database File: D:\data\R000017.pcm

Folder for Sub-Wave Item Info Files: D:\data

Info File Name: R000017.txt

Zoomed View of Sub-Wave to be Chopped: Time: 1.23s

Original Speech Database Display Window: Done: 9.07%

Sex: M

Speaker: Consumer

Transcription (Manually Input): ☐ To be Discarded

喂是中关村内个机票代售的点吧

Segmentation of Chinese Words (Modifiable):

\喂\是\中关村\内\个\机票\代售\的\点\吧\

Automatically Generated PinYin String (Modifiable):

wei4 shi4 zhong1 guan1 cun1 nei4 ge4 ji1 piao4 dai4 shou4 de0 dian3 ba0

Figure 1. The CSTSC-Flight transcription tool.

The transcription procedure consists of the following steps.

(1) Load a wave file of a flight enquiry and reservation conversation to be transcribed. The waveform will be displayed in the "Original Speech Database Display Window" (Windows 2). As stated in Section 3.1, the transcription will be performed sentence by sentence, and each sentence is referred to as a "Sub-Wave Item" in this tool.

(2) Fill in speaker information boxes (sex and speaker).

(3) In Window 2, by dragging your mouse and listening to the selected waveform part (from the beginning to the current mouse position), select a valid sentence to form a new sub-wave item. This will be displayed in the "Zoomed View of Sub-Wave to be Chopped" (Window 1), and meanwhile it will be deleted from Window 2. That is to say, Window 2 is always displaying the un-transcribed part of the whole wave file. If this item is not a valid speech segment, it can be discarded.

(4) Transcribe the Chinese sentence (character string) of this item into the "Transcription (Manually Input)" box.

(5) Press the "Auto" Button to generate the Chinese word segmentation and pinyin string automatically (Zheng 1999) into the following two boxes. Any segmentation error or character-to-pinyin conversion error, can be manually corrected in the corresponding box. Because the segmentation and conversion accuracy is very high, the manual labor is much limited.

(6) The merging-based syllable detection automaton (Zhang 1999) is used to supply syllable time-boundary information (as vertical lines) automatically as shown in Window 1. A manual adjustment to these syllable boundaries is permitted simply by deleting, inserting, or moving the vertical separating line(s) using your mouse. Meanwhile, the spontaneous acoustic transcription can be performed manually.

(7) After all the above transcription is finished and perfect, append this item into the "List of Sub-Wave Item" list-box. By selecting any item in the list-box, you can also remove or update it later if necessary.

4 Corpus Analysis

4.1 Overall description

The CSTSC-Flight is a domain-specific corpus; therefore the topics are relatively concentrated. The topics in this corpus include: (1) enquiry on the flight departure and arrival time; (2) enquiry on available tickets; (3) enquiry on price; (4) enquiry on flight number, plane model, airline company, and/or airport; (5) enquiry on agency location; (6) enquiry on route to airport; and (7) telephone booking.

Analysis on the corpus transcription comes to the following summaries:

(1) Heavy background noises mainly at the operator's end, as well as telephone channel noises;

(2) Comparative low-volume or sometimes even unclear speech at the customer's end;

(3) Serious phoneme deletion and co-articulation; and

(4) So severe spontaneous linguistic phenomena that sometime a long sentence with several spoken language phenomena is very difficult even for the transcribers to understand.

4.2 Spoken language phenomena

Regarding the last point mentioned in Section 4.1, the detailed classifications as well as the corresponding examples are given as follows, where C: leading the customer's utterances and O: leading the operator's utterances.

- The courtesy items / sentences inessential for semantic analysis.

  C: 喂，你好，请问是中关村航空客运代理处么?
  Hi, hello, could you tell me ...?

- Simple repetitions because of the pondering or thinking when speaking.

  C: 我问一下那个四月三十呃四月三十号北京到...
  ... 30th April ... 30th April ...

- Semantic repetitions for emphasis.

  C: 请问那个周四就是四月三十号北京到...
  ... Thursday ... 30th April ...

- Speech corrections or repairs (Heeman 1994).

> C: 呃，那个什么星期三，呃，星期四的去南京的机票还有吗？
> ... Wednesday ... Thursday ...

- Ellipsis in the context.

> C: 我问一下那个四月三十呃四月三十号北京到福州的机票最后一班还有么？(Asking if there are tickets available for the last flight from the *departure city* to the *arrival city*, on the *departure date*.)
> O: 只有一班有。(“Only one flight with tickets available”)
> C: 那个那五月一号的下午三点有么？(How about the flight at a *departure time* on another *departure date*?)

- Constituents appearing in any order (as long as the sufficient information is given).

> C: ...五点二十五国航飞深圳的... (Time, airline code, location and some other items can appear in any order)

- Constituents appearing in reverse order.

> C: ...的机票 多少钱 得?
> How much cost
> (Normal order: “得” “多少钱”)

- Parol (verbal idioms) or unnecessary terms.

> C: 那，那个八点二十那个是去什么机场的呀？
> (“那”/”那个” is somewhat similar to “uhm”)

- And long sentences with all required information. Or additional explanation following the previous sentence.

> C: 哎，您好，这样那个我订一张(one)那个明天(tomorrow)下午(afternoon)五点(5 o’clock)四十五(45)去北京(from Beijing)到上海(to Shanghai)的那个机票(ticket)的。
>
> C: 您给我看一下有没有(is there)北京到湛江(from Beijing to Zhanjiang)的(ticket)？二十九(29th)、三十号(30th)?

All these general phenomena and especially the spontaneous speech phenomena are obviously great challenges to Chinese dialogue systems.

In addition to the above spoken language phenomena, there are some other phenomena specific to the flight domain, including:

- Generally the customer will begin the enquiry/reservation conversation with a long sentence containing all information needed to determine a flight.
- In the Q&A dialogue, there are more single sentences than compound sentences. And there are few modifiers in the single sentence.
- In almost all conversations, there are multiple confirmations between the customer and the operator regarding the enquiry conditions, contents and results.

It can also be seen that the question sentences are very common. Secondly, brief affirmations and negations are also very common sentence patterns. Others include the statement sentence, the imperative sentence, and the statement sentence followed by a short question sentence.

The design of the human-computer dialogue system should be based on such an analysis on the human-human dialogues, especially in the specific domain.

5 Robust Parsing Rules

The rule-based parsing is a prevalent method for the natural language understanding (NLU) and has been introduced in dialogue systems for spoken language processing (SLP). However, additional measures must be taken to cope with the severe spoken linguistic phenomena as described in last section. We present in this section a robust parsing scheme, which integrates the following methods. Keywords are used as terminal symbols; hence the symbol set of the grammar is purely within the semantical category. According to the analysis on the spoken language phenomena contained in CSTSC-Flight, the definition of the grammar is extended to accommodate four types of rules, called *up-tying*, *by-passing*, *up-messing*, and *over-crossing,* respectively, to cover most of the spoken language phenomena (Yan 2001).

5.1 Definition of grammar

In spoken dialogue systems, the traditional grammars where word-classes or part-of-speeches are taken as the terminal symbols, with which linguists are quite familiar, will not work efficiently because a great deal of daily spoken sentences will be rejected due to the narrow coverage of the grammars. At this point, we define a grammar in which the terminal/non-terminal symbols are all semantically meaningful constituents; therefore we call it a semantics-based grammar or semantic grammar in brief.

In Figure 2, the definition of the grammar is given formally in the context-free-grammar manner where "*rule_text*" denotes the start symbol, and the terminal symbols are the characters in quotes for the grammar transcription.

```
rule_text → rule_list
rule_list → rule | rule   rule_list
rule → symbol   [rule_type]   '→'   symbol_list
symbol_list → symbol | symbol   symbol_list
symbol → symbol_prefix | symbol_prefix   symbol_suffix
symbol_prefix → alphabetic
symbol_suffix → alphanumeric | alphanumeric   symbol_suffix
alphanumeric → alphabetic | numeric
alphabetic → '_'|'a'|'A'|'b'|'B'|...|'z'|'Z'
numeric → '0'|'1'|...|'9'
rule_type → '*'|'@'|'#'
```

Figure 2. Formalized definition of the grammar

We have four types of grammar rules: *up-tying* (*→), *by-passing* (→), *up-messing* (@→), and *over-crossing* (#→). An *up-tying* type rule is a conventional rule used in conventional grammars where the constituents are strictly tied up without any flexibility. By using a *by-passing* type rule constituents can be reduced by skipping irrelevant segments. An *up-messing* or an *over-crossing* type rule will be helpful to group constituents despite the order they occur. One difference between the last two types is that an *up-messing* rule does not contain any *over-crossing* sub-constituents. Another difference is that, the latter rule will help to reduce sub-constituents no matter whether their parsing occupations, where the occupation of a constituent is defined as the in-sentence-positions of all its leaf nodes, overlaps with each other or not, while the former one will not.

In some other methods the coverage of the grammar can be extended by means of skipping unnecessary speech segments. However in our method, the four types of rule are explicitly incorporated into the grammar as a whole, which results in a systematic way.

5.2 Transcription of semantic grammar

Though in some literature it was reported that semi-automatic approaches were used for grammar generation (e.g. Siu 1999), we generate the grammar manually because the transcription effort is greatly alleviated using our approach where the grammar is a semantic one. Based on sufficient analysis on the corpus, the system designer can employ a comparatively small lexicon to write the grammar easily because the semantic elements instead of part-of-speeches are used.

We present here some rule examples in the domain of flight enquiry and reservation to explain how to use the four types of rules in different situations.

5.2.1 *Up-tying* rules

The *up-tying* rules are needed in at least one case when the customer's ID card no. is to be parsed where the ID card no. is taken as a crucial piece of information forbidden to be inserted by or mixed with other terms. There are two versions of ID card no. in China, one is 15-digit long and the other 18-digit, and therefore three rules are needed.

```
sub_id_card_head *→ ato_0to9_yao  ato_0to9_yao ...
        ato_0to9_yao (15 identical terms)
id_card_no → sub_id_card_head
id_card_no *→ sub_id_card_head  ato_0to9_yao
        ato_0to9_yao  ato_0to9_yao
```

5.2.2 *By-passing* rules

A great deal of rules belong to the *by-passing* type, under the assumption that the input keyword string is full of recognized fillers/rejections, speech fragments or some other nonsense parts. For example, "星期啊三嗯星期" ("week *ah* three *en* week four"/Wedn-*ah*-esday and *en* Thursday) is admitted if the following *by-passing* rules exist.

```
sub_week_day → ato_week  ato_1to6
sub_week_day_list → sub_week_day
sub_week_day_list → sub_week_day  sub_week_day_list
```

5.2.3 *Up-messing* rules

The *up-messing* rules are required in case that some sub-constituents make up of a larger one without any restriction of the occurring order. In the flight enquiry and reservation domain, constituents of time, location, and plane type can be described by the *up-messing* rules.

```
timeloc_info_cond @→ info_date_time_cond
info_fromto
plane_info @→ mat_airline_code  mat_aircraft_type
flight_info_cond @→ timeloc_info_cond  plane_info
```

5.2.4 *Over-crossing* rules

Some concepts, which can be defined as the task-related minimal elements, may be derived from several different *by-passing* rules and can be used to form larger constituents. In this case, *over-crossing* rules are used to avoid the definition of many similar rules, and the runtime ambiguities can also be reduced. For example, "是不是 (be or not) *confirm_c*" , "*confirm_c*是不是", "*confirm_c*是吗 (be or not?)", and "是(be) *confirm_c*吗 (*question mark*)" can be described by a single *over-crossing* rule, where *confirm_c* denotes an item need to be confirmed.

```
mark_q_is → tag_is_or_not
mark_q_is → tag_is  tag_question_mark
mark_q_is → tag_is_q
confirm_request #→ mark_q_is  confirm_c
```

Totally about 200 rules are written in the semantic grammar for this domain, and most of them are *by-passing* rules. The coverage of the grammar is proven to be wide enough, and the semantic extractions can be performed directly because the concepts, e.g. *mark_q_is* in section 5.5.4, are formalized in the rules.

5.3 Application in *EasyFlight*

The rules are applied to *EasyFlight*, a flight enquiry and reservation system (Yan 2001), and some primary evaluation is made.

There are four modules in *EasyFlight*, as illustrated in Figure 2. The keyword spotter produces the n-best keyword strings. The marionette parser processes the keyword strings and output the resultant tree/forest. The semantic frame is used to represent the customer's semantics.

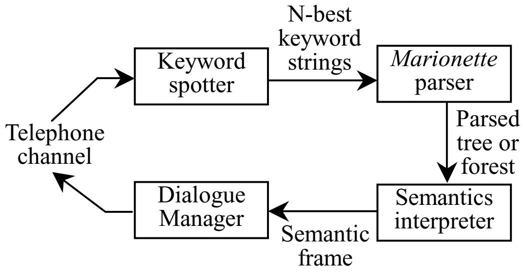


Figure 2. *EasyFlight* modules

A semantic function tree, which is isomorphic to the parsing-resultant tree, is used to analyze the semantics, where each constituent in this tree is associated with a function in the semantic function tree. The semantics interpretation is performed by means of calling the root functions and the middle/terminal functions are called recursively if necessary.

The applying of these robust rules as well as the marionette parsing scheme on *EasyFlight* achieves a satisfying robust parsing performance. The speech phenomena, such as acoustic garbage, linguistic garbage, repetition, ellipsis, word disordering, fragment and ill form, are overcome efficiently.

The following is an example showing how a user books tickets through the *EasyFlight* system. In this conversation, most spoken language phenomena exist, and the system can perform context-dependent parsing and enable the user to change topics freely.

User：你好
System：你好
U：我想订一张票
S：您想订从哪儿起飞的航班？
U：从北京到上海
S：您想订哪一天的?
U：明天
S：查到14个起飞时间，您要订大概什么时候的？
U：越早越好
S：您要6月9日07:20从北京出发，09:15到达上海的MU582次航班的机票吗？
U：请问是什么机型？
S：是M11型飞机
U：那我买二张
S：您要6月9日07:20从北京出发，09:15到达上海的MU582次航班的机票吗？
U：没错
S：您要2张票，是吗？
U：对，谢谢
S：不用谢

6 Summary

The collection, transcription and analysis of a Chinese spontaneous telephone speech corpus in the flight enquiry and reservation domain (CSTSC-Flight) are introduced, which is very useful to the training of the acoustic model and the definition of the robust parsing rules. Based on the analysis on this corpus, four types of robust rules are also presented. The application of these rules to a flight enquiry and reservation system *EasyFlight* shows a great performance, where user can talk with the system spontaneously as he/she is talking with a human.

For the time being, the data collected in CSTSC-Flight are from human-human conversations, which could be different acoustically and linguistically from human-computer conversation data. However, it is enough to provide acoustic training data and domain-specific linguistic knowledge so that a spoken dialogue system can be established. For data collection, the next step will be the collection of human-computer conversation data using the WoZ method (Eskenazi 1999) or through an established spoken dialogue system. For spoken language parsing, the next step will be to improve both the parser and the dialogue manager based on the analysis on the spoken language phenomena in CSTSC-Flight, for example the short Q&A sentence patterns.

References

[1] ESKENAZI Maxine, RUDNICKY Alexander, GREGORY Karin, *et al*, 1999. "Data collection and processing in the Carnegie Mellon Communicator," *EuroSpeech*, 2039-2042, 1999, Budapest, Hungary

[2] HEEMAN Peter, and ALLEN James, 1994. "Detecting and correcting speech repairs," In Proceedings of *the 32nd Annual Meeting of the Association for Computational Linguistics*, 295-302, June 1994, Las Cruces, New Mexico, USA

[3] LI Aijun, ZHENG Fang, BYRNE William, *et al*, 2000. "CASS: a phonetically transcribed corpus of Mandarin spontaneous speech." *International Conference on Spoken Language Processing*, I-485~488, Oct. 16-20, Beijing, China

[4] SIU K. C., and MENG H. M., 1999. "Semi-automatic acquisition of domain-specific semantic structures", *EuroSpeech*, 2039-2042, 1999, Budapest, Hungary

[5] SUN Hui, 2000. Research on acoustic preprocessing and Chinese spoken language phenomena in a flight enquiry system: Undergraduate Project. Tsinghua University: Beijing, China, May 2000 (in Chinese)

[6] YAN Pengju, ZHENG Fang, and XU Mingxing, 2001. "Robust parsing in spoken dialogue systems," *EuroSpeech*, 3:2149-2152, Sept. 3-7, 2001, Aalborg, Denmark

[7] YU Huayun, and CAI Lianhong, 1999. "The multi channel control techniques in the design of computer telephony integration systems," *Computer Engineering*, 25(4), April 1999 (in Chinese)

[8] ZHANG Jiyong, ZHENG Fang, DU Shu, *et al*, 1999. "Merging-based syllable detection automaton in continuous speech recognition," *J. of Software*, 10 (11): 1212-1215, Nov. 1999 (in Chinese)

[9] ZHENG Fang, 1999. "A syllable-synchronous network search algorithm for word decoding in Chinese speech recognition," *ICASSP,* II-601~604, March 15~19, 1999, Phoenix, USA

[10] ZHENG Fang, SONG Zhanjiang, FUNG Pascale, BYRNE William, 2001. "Modeling pronunciation variation using context-dependent weighting and B/S refined acoustic modeling," *EuroSpeech,* 1:57-60, Sept. 3-7, 2001, Aalborg, Denmark