Chinese Multimodal Resources for DA Analysis

Alex Chengyu Fang, City University of Hong Kong, Hong Kong SAR
Yanjiao Li, Shandong University, PR China
Jing Cao, Huazhong University of Economics and Law, PR China
Harry Bunt, Tilburg University, the Netherlands

Abstract:
Interactive speech constitutes a large proportion of human communication and represents a full repertoire of communicative strategies and mechanisms, the understanding of which has proved to be key to the proper management of spoken resources. In the light of this, recently, spoken corpora and multimodal resources have offered an especially powerful instrument that enables the extraction of linguistic information to help deepen our understanding about human speech and language, especially in the interpretation and exploration of communicative strategies and dialogue mechanisms. Increasingly, attention has gone beyond the lexical and syntactic information and focused on semantic and pragmatic properties in dialogues. In particular, the international research community has tapped into dialogue acts (DAs) that primarily pinpoint speaker intentions in specific contexts for practical applications in dialogue systems. In this context, the compilation of corpora of interactive spoken Chinese and the exploration of dialogue acts based thereupon have become an emergent requirement that is yet to be fruitfully met with. This article is a survey of the Chinese multimodal resources in general and interactive spoken corpora in particular, focusing on the existing coding schemes of dialogue acts.

1. Introduction

Interactive communication features prominently in human language in that it is "the dominant component of general language both in terms of language reception and language production" (Crowdy 1993:259). Meanwhile, "[t]he rich variations in human speech can only be adequately analyzed and represented in properly recorded, annotated and processed speech data" (Liu et al. 2006: 724). In other words, the

understanding of conversational dialogues calls for the availability of appropriately handled spoken resources. Most importantly, proper annotation of such resources has become an important instrument to enable the retrieval of salient linguistic features characterizing the spectrum of communicative strategies and speaker intentions which are broadly referred to as dialogue acts (DAs; see section below). The insight thus afforded, while broadening our understanding of human language in general and spoken interaction in particular, has contributed immensely to the development of practical man-machine dialogue systems, which has attracted the attention from both linguists and speech engineers (Tseng 2005a).

In addition to the more common linguistic annotation of spoken resources at phonological, lexical and syntactic levels, recent attention has now been focused on more sophisticated semantic and pragmatic features of dialogues. While the analysis of such features is technically challenging, how to fully exploit the semantic and pragmatic knowledge in human language and apply it to the field of language information processing has become a crucial and urgent question. The need for a set of useful semantico-pragmatic properties with a clearly defined use scenario has generated and, indeed, propelled research in the area of DA analysis, which is concerned with the annotation of spoken utterances for a range of communicative functions as well as a specific set of semantic meanings. See Section 2 below for a more detailed description. This article aims at a full survey of the state of the art within this area in the Chinese context for a two-fold purpose: to provide a fair insight into the past achievements and, more importantly, to indicate directions for future research. It is organized as follows: Section 2 will first introduce the concept of DA and review the trends of DA research based on interactive spoken corpora in the international community. Section 3 will present a survey of Chinese spoken corpora compiled in different countries and regions. Section 4 will examine in detail two existing DA coding schemes which have been applied to Chinese interactive spoken corpora. Section 5 will conclude with discussions and suggestions.

2. Dialogue Acts in Interactive Speech

To gain a better understanding of interactive speech, the term dialogue act (DA) has been used to signal the speakers' intention under specific contexts. According to the newly developed ISO standard (24617-2 2012; Bunt et al. 2010 and 2012), DA refers to "communicative activity of a dialogue participant, interpreted as having a certain communicative function and semantic content" (ibid: 2). See Example (1) for a set of

questions in dialogues.

(1)

a) *"Does the meeting start at ten?"*

b) *"What time does the meeting start?"*

c) *"The meeting starts at ten, right?"*

d) *"Is the meeting going to start at ten or eleven?"*

(ibid: 65-66)

According to the ISO coding scheme, these questions are analysed differently with different dialogue acts, namely, *propositional question*, *set question*, *check question*, and *choice question*. Of course, the same utterances may be analysed differently according to other, different DA coding schemes but the essence is the same: the communicative functions performed by the speaker (*Sender*) in these examples are different while the semantic content of these utterances is (approximately) the same. That is to say, the DA information helps to make explicit the communicative functions of the utterances. In addition to the ISO DA coding standard officially released in 2012, there are a few other DA taxonomies such as TRAINS (Traum and Hinkelman 1992; Traum 1996), DAMSL (Allen and Core 1997, Core and Allen 1997), SWBD-DAMSL (Jurafsky et al. 1997), ICSI-MRDA (Dhillon et al. 2004), AMI (AMI Consortium 2005) and DIT++ (Bunt 2009 and 2010). They were invariably proposed and tested on actual corpus data for automatic DA annotation and recognition: TRAINS has been applied to the TRAINS Corpus, SWBD-DAMSL to the Switchboard Dialogue Act Corpus, ICSI-MRDA to the ICSI Meeting Recording Dialogue Act Corpus, AMI to AMI Corpus, DIT++ to the LIRICS corpus, and ISO-DA to SWBD-ISO (Fang et al 2012a and 2012b; Bunt et al. 2013; Fang et al. 2013) and DialogBank (Bunt et al. 2016)[1]. In return, these corpora annotated with DA information have been employed as training data to achieve better recognition performance according to an "annotate-train-test" paradigm that has been found to be quite successful (Ritter et al. 2010: 172). As a matter of fact, the success in automatic classification and recognition of DAs has led to eminently practical applications in the field of natural language processing (NLP), such as speech recognition (e.g. Stolcke et al. 2000), dialogue summarization (e.g. Murray et al. 2006), construction of dialogue systems (e.g. Allen et al. 2007), information extraction and retrieval (e.g. Armstrong et al. 2003), machine translation (e.g. Sridhar et al. 2008a, 2008b),

[1] Accessible at http://dialogbank.uvt.nl.

and flirtation detection (e.g. Ranganath et al. 2009).

As can be noted, current research on DAs indicates two noteworthy phenomena. First, an important global trend is the standardisation of various DA schemes which were originally motivated by project-specific needs and practical requirements. The second major phenomenon is that most of the DA studies have been intensively carried out in English dialogues and that languages such as Chinese have not received due attention, which could be a more productive direction.

3. Chinese Interactive Speech Resources

Before the analysis of the existing Chinese DA schemes, a general introduction to the Chinese interactive speech resources will be made as an overview of the current situation in the Chinese context. As noted by Adolphs and Carter (2013: 1), "[d]atabases for the study of written language run into millions of words. Yet there are still relatively few projects devoted to spoken corpus linguistics." This is also true with spoken Chinese resources given the large number of Chinese speakers. Up to date, fourteen Chinese interactive speech corpora have been established, and annotations concerning a variety of information have also been marked up to reflect the nature of speech. Section 3.1 introduces the existing corpora and Section 3.2 is devoted to the descriptions of the annotation.

3.1 Chinese Interactive Speech Corpora[2]

Six of the 14 corpora of interactive spoken Chinese were constructed in Mainland China. The first three are held at the Chinese Academy of Social Science (CASS). As early as 1998, Prof Yueguo Gu initiated the compilation of the Spoken Chinese Corpus of Situated Discourse in the Beijing area (SCCSD), which is still an on-going project. According to the project's official website (www.multimodal.cn), the corpus contains three sub-corpora: the Corpus of Social Activities (CSA), the Corpus of Baby-Adult Discourse (CBAD) and the Corpus for Educator Development (CED). Among them, the first two sub-corpora have been transcribed and proofread. In 2000-2002, the Chinese Annotated Dialogue and Conversation Corpus (CADCC) was

[2] Since this chapter is concerned with interactive speech only, corpora containing only monologues and/or read speech have been left out.

compiled by Li et al. (2002). Later in 2001, jointly with the Institute of Automation at the Chinese Academy of Sciences, CASS initiated a domain specific corpus named Telephone Speech Corpus 973 (TSC-973). Table 1 presents the basic information about the three corpora.

Table 1 Corpora held at CASS

Corpus	*Time*	*Size*	*Mode*	*Topic*
SCCSD	1998-ongoing	600-hour audio 250-hour video (till 2007)	Face-to-face conversation	Home & work
CADCC	2000-2002	14-hour audio	Face-to-face conversation	No limitation
TSC-973	2001- 2002	2-hour audio	Telephone conversation	Hotel reservation

As can be noted in Table 1, while CADCC and TSC-973 are restricted to audio recordings, SCCSD is truly multimodal with both audio and visual recordings and thus represents a collection of distinctive "multi-modal texts" (Gu 2006: 128) with the purpose to model "total saturated experience (TSE)" (Gu 2009: 436) in communication by capturing "the real-life dynamics of actual discourse" (Gu 2002: 4).

Tsinghua University created the fourth corpus in 2002, named the Chinese Spontaneous Telephone Speech Corpus on Flight Enquiry and Reservation (CSTSC-Flight; Zheng et al. 2002). This domain-specific corpus is designed for the development of a spoken dialogue system, *EasyFlight*. It contains 50 hours' worth of interactions about flight enquiries and reservations.

The fifth corpus, CASIA-CASSIL, was initiated in 2002 by the Institute of Automation at the Chinese Academy of Sciences (Xie et al. 2002). Managed by Prof Chengqing Zong, this project intends to compile a corpus of Chinese spontaneous telephone conversations in the domain of tourism (Zhou et al. 2010). It aims to provide resources for research on spoken Chinese to promote the development of speech-to-speech translation systems and man-machine dialogue systems. By 2010, the corpus has increased to consist of 7,639 conversations in the tourism domain covering hotels, restaurants, airports, and travel agencies.

The sixth corpus was launched in 2008 by the National Language Committee of China as a long-range project for the construction of an Audio Database of Chinese Linguistic Resources (Li 2010). In particular, the project is aimed at the protection of linguistic diversity in China and therefore focuses on two types of language variations: dialects and spoken Mandarin with regional accents. While the Mandarin data is collected through reading (two articles) and narrating and therefore falls in the category of read speech, the dialect data

is collected in three different ways:

- Read speech (reading of 1000 characters, 1200 words, and 50 sentences);
- Narration；and
- Spontaneous conversations.

All the recordings will be phonetically and orthographically transcribed. To date, the components that have been finished include Beijing, Shanghai, Jiangsu, Guangxi, Liaoning, Hubei, and Fujian provinces.

Taiwan is the second major contributor towards the construction of spoken Chinese resources and has witnessed the creation of two corpora. The Taiwan Mandarin Conversational Corpus (TMC) as part of the National Digital Archives Project (2002-2006) is initiated by Prof Tseng at the Institute of Linguistics, Academia Sinica (Tseng 2004). The TMC Corpus is composed of three sub-corpora: the Mandarin Conversational Dialogue Corpus (MCDC, Tseng 2005a), the Mandarin Topic-oriented Conversation Corpus (MTCC, Tseng 2005b), and the Mandarin Map Task Corpus (MMTC, Tseng 2008). See Table 2 for the detailed information on the corpus.

Table 2 Components of the TMC Corpus

Corpus	*Size*	*Topic*
MCDC	30 1-hour spontaneous conversations	Topic chosen by the speakers
MTCC	30 20-minute conversations	Discussions about the news or events in 2001
MMTC	26 10-minute conversations	Map tasks (following HCRC Map Tasks designed by Anderson et al. 1991)

In this way, “the TMC Corpus provides speech data of a variety of speaker groups communicating in different speaking styles and situations” (Tseng 2013:4). The other NCCU Corpus of Spoken Chinese was constructed at the National Chengchi University (NCCU) in Taiwan (Chui and Lai 2008). See Table 3 for the updated information of the NCCU Corpus of Spoken Chinese (available at http://140.119.172.200/).

Table 3 Components of the NCCU Corpus

Corpus	*Size*	*Topic*
Corpus of Spoken Mandarin	36 face-to-face conversations	No limitation
Corpus of Spoken Hakka	21 20-minute conversations 29 20-minute narratives	/
Corpus of Spoken Southern Min*	/	/

* No public information is available.

As can be noted in Table 3, the NCCU Corpus comprises three varieties of spoken Chinese in Taiwan, including two dialects (i.e. Hakka and Southern Min) as well as Mandarin.

Hong Kong is also a major player in the construction of spoken resources, where two corpora of Chinese interactive speech have been compiled. The Hong Kong Cantonese Adult Language Corpus (HKCAC) is developed by Hong Kong University (Leung and Law 2001), aiming to provide resources for research in the Cantonese dialect. It comprises the call-in programs and forums on the radio from 1998 to 2000 with a wide range of topics such as political and economic issues, and personal problems. The entire recordings last more than eight hours. The HKUST Mandarin Telephone Speech Corpus (HKUST/MTS) was compiled in 2004 by Hong Kong University of Science and Technology (Liu et al. 2006) as a resource for research in Mandarin speech recognition. The corpus design is similar to that of the Switchboard corpus, covering a wide spectrum of topics in daily life. To be more specific, the speech data has been collected over public telephone networks, and each conversation lasts ten minutes between two native Mandarin speakers. The HKUST Corpus of Mandarin Telephone Speech (Part 1), a collection of 897 telephone conversations, was released by LDC in 2005 (Fung et al. 2005).

The remaining four resources of spoken Chinese were created from outside Greater China. Table 4 lists the basic information about them.

Table 4 Corpora of spoken Chinese outside Greater China

Corpus	*Country*	*Language specifics*	*Size*	*Mode*
CALLHOME Speech (Canavan and Zipperlen 1996)	USA	Mandarin Chinese	120 30-minute conversations	telephone conversations
CALLFRIEND (Canavan et al. 1996a, 1996b)	USA	Mandarin Chinese (Mainland)	60 conversations (5-30 minutes each)	telephone conversations
		Mandarin Chinese (Taiwan)	60 conversations (5-30 minutes each)	telephone conversations
The Lancaster/ Los Angeles Spoken Chinese Corpus (Xiao and Tao 2006)*	UK & USA	Mandarin Chinese	1 million words	face-to-face conversations, telephone calls, plays/movies, TV talk shows, debates, oral narratives, edited oral narratives
Wenzhou Spoken Corpus (Newman et al. 2007)	Canada	Wenzhou dialect	about 150,000 words	face to face conversations, phone calls, internet chats, story-telling, news, and songs

*No public access

It should be pointed out that, amongst the four corpora, CALLHOME and CALLFRIEND have been collected specifically for practical applications such as speech recognition and language identification. The CALLHOME Mandarin Chinese Speech Corpus was released together with a selected transcript (Wheatley 1996) and a lexicon of 44,405 words (Huang et al. 1996) with phonological, morphological and frequential information. Later, an XML version of CALLHOME Mandarin Chinese transcripts was also released (McEnery and Xiao 2008).

In summary, the above-mentioned survey sketches out the general features of the Chinese interactive speech corpora.

- The Chinese interactive speech corpora consist of dialogues collected from a wide range of sources, such as face-to-face conversations, telephone conversations, debates and internet chats.
- In addition to Mandarin, the existing corpora have also covered different dialects in China and Mandarin with regional accents.
- While the majority of the corpora deal with the general topics in daily life, there are spoken Chinese corpora specific to some chosen domains such as hotel reservation and flight enquiry.
- While most of the spoken corpora are stored in the form of audio signals with transcriptions, the SCCSD includes both audio and visual data and therefore is defined as multimodal, which constitutes a valuable unique resource for linguistic studies and practical applications.
- A significant portion of spoken resources has been specifically designed for direct applications in the field of speech engineering (i.e. TSC-973, phonetic engineering; CASIA-CASSIL, dialogue system; HKUST/MTS and CALLHOME, speech recognition; CALLFRIEND, language identification).

3.2 Annotations of Spoken Chinese Corpora

An indispensable preliminary to the annotations of spoken corpora is to transcribe the acoustic data. In the case of Chinese corpora, the orthographic or phonetic transcriptions often present in two patterns: characters only and characters with Pinyin, as illustrated by Examples (2) - (4).

(2) 當然還是我還是覺得美國人把人命看得比較值錢

(Extracted from the MTCC, Tseng 2005b: 209)

(3) B: 你们现在紧不紧呀?.

B: ni3 men0 xian4 zai4 jin3 bu4 jin3 ya0?

(Extracted from the CADCC, Li et al. 2002)

(4) 我请问一下那个<lengthening>八月十号去上海的都有几个航班

Wo3 qing2 wen4 yi1 na4 ge0<lengthening> ba1 yue4 shi2 hao4 qu4 shang4 hai3 de0 dou1 you3 ji3 ge4 hang2 ban1

(Extracted from CSTSC-Flight, Zheng et al. 2002: 62)

The utterance in Example (2) is transcribed in traditional Chinese characters only. Examples (3) and (4) showcase the character-plus-pinyin form, where the Arabic numerals after the Pinyin of each character indicate the tones, where 1 refers to the high and level tone, 2 the rising tone, 3 the falling-rising tone, 4 the falling tone, and 0 means the neutral tone.

Different from the English language, the processing of the Chinese language very often requires the identification of words in the first place, which is often known as word segmentation. The existing spoken Chinese corpora once again fall into two categories: with and without segmentation. See Examples (5) - (7).

(5) A: 喂 您好 前台

B: 您好 我 想 在 你们 这里 预定 一 个 房间

(Extracted from the TSC-973, Li et al. 2002)

(6) B: 你们现在紧不紧呀?.

\B:\你们\现在\紧\不\紧\呀\?\

(Extracted from the CADCC, Li et al. 2002)

(7) F1: 小五..跟我上次…上次..教到的那个小五

F2: 对对那个那个

(Extracted from the NCCU corpus, http://140.119.172.200/)

As can be seen in Examples (5) and (6), the word segmentation may come in different forms. The simplest way is to use space as the delimiter, as shown in Example (5). Another common way is to use slash as the delimiter and, in the case of Example (6), a backward slash. Example (7) is a case without any segmentation.

In addition, a special feature in the transcription of recorded spoken data is to transcribe the non-speech features such as pause, overlapping, breathing, and laughing (Gu 2009). See Examples (8) – (10).

(8) 你知道(short break)當然還是我還是覺得美國人把人命看得比較值錢

(Extracted from the MTCC, Tseng 2005b: 209)

(9) A: [LA]大哥还挺神.

\A: [LA]\大哥\还\挺\神\.\

A: [LA] da4 ge1 hai2 ting3 shen2 .

(Extracted from the CADCC, Li et al. 2002)

(10) F1: 小五..跟我上次…上次..教到的[那個小五]

F2: [對對那個]那個

(Extracted from the NCCU corpus, http://140.119.172.200/)

As can be seen in Example (8), a 'short break' is being identified. In Example (9), *LA* stands for laughing. In Example (11), the dots between words indicate the pause, where two dots (..) refer to short pause and three dots (…) mean medium pause. Meanwhile, the sections within the square brackets ([]) are the overlapped part of the utterance. In other words, when the speaker *F1* says "那個小五" , *F2* is saying "對對那個" simultaneously.

Apart from these conventional features of the transcription of spoken data, Gu (2009) proposed a multimodal structure to capture the various aspects in spoken data that otherwise may be lost in the transcription. In particular, an agent-oriented model (AOM) is constructed in the expectation of capturing the relevant information embedded in real-life activities at four levels:

- Occurrence Modeling (Level 0)
- Type Modeling (Level 1)
- XML and RDF Modeling (Level 2)
- Corpus Modeling (Level 3)

The uniqueness of such a model is to integrate the transcription with multi-level annotations. See Example (11) for a Level 0 (Occurrence Modeling, or O-Agent) coding.

(11) Orthographic transcript: 姥爷：怎么啦？

sTurn	*Agents*	
	sTurnAgent	姥爷
	sChunk	怎么啦
	uIntoType	呼唤调

According to Gu (2009:457-458), the whole transcript is a time-bound speaking turn (sTurn), and therefore

the orthographic transcript in Example (11) is labelled as a *sTurn*. As for the agent coding, there are three agents or attributes in this case, namely, *sTurnAgent*, *sChunk* and *uIntoType*. The speaker "姥爷" is the person who performs the turn, and is defined as *sTurnAgent* and what the speaker said (i.e. 怎么啦) is labeled as a speech chunk (*sChunk*). A third agent is the intonation type (i.e. 呼唤调), annotated as utterance intonation type (*uIntoType*), which very often reveals information that is not expressed by words.

In addition to intonation, the spoken Chinese corpora are also annotated with various linguistic features. For example, in the MTCC, phonetic feature of "assimilation" is annotated, as shown in Example (12).

(12) 其它的<b clear throat>@</b clear throat><b exhale>@</b exhale> <b assimilation>三分</b assimilation>之二是<b inhale>@</b inhale>警察局自己<b pause>@</b pause>

(Extracted from the MTCC, Tseng 2005b: 209)

Example (13) presents the top 10 words with part-of-speech tags sampled from the Taiwan Mandarin Conversational (TMC).

(13)

	Word	*POS*	*Meaning*
1	的	DE	structural particles
2	是	SHI	copula
3	一	Neu	numeral determinatives
4	在	P	preposition
5	有	V_2	you3
6	個	Nf	Measure
7	我	Nh	Pronoun
8	不	D	Adverb
9	這	Nep	Demonstrative determinatives
10	了	Di	aspectual adverb

(Tseng 2013: 17-18)

Annotations at semantic and pragmatic levels have been performed in some of the corpus projects on spoken Chinese. Two projects have referred to their coding with examples. For example, according to Gu (2009:457), utterances like "Good morning" and "How are you" have the illocutionary force of greeting and therefore should be annotated as such. The other corpus, TSC-973, has listed the five sentence functions that are marked up, namely, statement (S), interrogative (Q), imperative (I), exclamation (E), and bitty utterance (B) (http://www.chineseldc.org/doc/CLDC-SPC-2005-016/label.htm). See Example (14) for the annotation of a question (Q) and a statement (S).

(14)

Transcript		*Function*
啊	a2	Q
可以呀	ke3 yi3 ya0	S

The other two corpora (i.e. the CASIA-CASSIL and the MTCC) have been annotated with a set of dialogue acts and thus will be discussed in detail in Section 4.

4. DA Annotation in Chinese Interactive Speech Corpora

Dialogue acts, as semantic/pragmatic information, are used to indicate communicative functions performed by particular utterances in interactive speech. Of the various spoken Chinese corpora discussed in the previous section, two of them have been manually coded with dialogue act information：the CASIA-CASSIL (Institute of Automation at the Chinese Academy of Sciences 2002) and the Mandarin Topic-oriented Conversation Corpus (MTCC) (Academia Sinica 2002). Section 4.1 below will take a closer look at the two coding systems.

4.1 CASIA-CASSIL Coding Scheme

Zhou 周可艳 proposed a coding scheme for DA annotation based on the interactive speech data collected in CASIA-CASSIL (Zhou 周可艳 2010a), which is designed with reference to three DA coding schemes in English interactive speech, namely, SWBD-DAMSL, ICSI-MRDA, and AMI. In all, 48 tags have been applied to 350 conversations selected from CASIA-CASSIL (Zhou et al. 2010b). They can be classified into three categories: general tags (9 tags), disrupted (3 tags), and specific (36 tags). See Table 5 for a comparison between the DA tagset of CASIA-CASSIL and the three reference coding schemes. The table is constructed on the basis of an earlier mapping between the SWBD-DAMSL and ICSI-MRDA schemes by Dhillon et al. (2004: 3-5).

Table 5 CASIA-CASSIL Coding Scheme vs. References

CASIA-CASSIL			*SWBD-DAMSL*	*ICSI-MRDA*	*AMI*
General tags	Statement	s	sd, sv	s	INFORM
	Y/N Question	qy	qy	qy	ELICIT-INFORM
	Wh-Question	qw	qw	qw	
	Or Question	qr	qr	qr	
	Or Clause After Y/N Question	qrr	qrr	qrr	
	Open-end Question	qo	qo	qo	

	Rhetorical Question			qh	qh	qh	INFORM
	Imperative Sentence			is	ad	co	/
	Exclamatory Sentence			es	fe	fe	/
Disrupted tags	Interrupted			%-	/	%-	FRAGMENT
	Abandoned			%--	%-	%--	
	Indecipherable			%	%	%	
Specific tags	Responses	Positive	Accept	aa	aa, ny	aa	ASSESS; INFORM
			Partial accept	aap	aap	aap	
			Affirmative answer	na	na	na	
		Negative	Reject	ar	ar, nn	ar	
			Partial Reject	arp	arp	arp	
			Dis-preferred answer	nd	nd	nd	
			Negative answer	ng	ng	ng	
		Uncertain	Maybe	am	am	am	
			No knowledge	no	no	no	
	Action Motivators		Command	co	ad	co	/
			Suggestion	cs	co	cs	ELICIT-ASSESSMENT; SUGGEST
			Commitment	cc	cc	cc	/
			Soliloquy	so	t1	t1	/
	Checks		“Follow Me”	f	/	f	/
			Repetition request	br	br	br	ELICIT-INFROM
			Understanding check	bu	/	bu	ELICIT_COMMENT-ABOUT-UNDERSTANDING
			Backchannel	b	b	b	BACKCHANNEL
	Restated Information	Repetition	Repeat	r	/	r	/
			Mimic	m	m	m	/
			Summary	bs	bf	bs	/
		Correction	Correct Misspeaking	bc	bc	bc	/
			Self-Correct misspeaking	bsc	/	bsc	/
			Self-affirm	bsa	/	/	/
	Supportive Functions		Defending/ explanation	df	/	df	INFORM
			Elaboration	e	e	e	INFORM
			Collaborative completion	2	2	2	INFORM
	Politeness Mechanisms		Downplayer	bd	bd	bd	BE-POSITIVE
			Sympathy	by	by	by	
			Apology	fa	fa	fa	
			Thanks	ft	ft	ft	
			Welcome	fw	fw	fw	
			Say Hello	sh	fp	/	
			Bye	bye	fc	/	

	Request	Request affirmation	raf	/	/	ELICIT-INFORM; ELICIT-OFFER-OR-SUGGESTION
		Request details	rdt	/	/	
		Request suggestion	rsg	/	/	

The following features can be observed in Table 5.

1) It is interesting to note that the CASIA-CASSIL DA scheme has a closer affinity with ICSI-MRDA than with SWBD-DAMSL. First, the overall three-category structure of CASIA-CASSIL shows a close similarity to ICSI-MRDA. Second, CASIA-CASSIL shares more identical DAs with the ICSI-MRDA (41) than with SWBD-DAMSL (37). In addition, where a DA does occur in all of the three coding schemes, CASIA-CASSIL tends to follow the tag coding of ICSI-MRDA, as in the cases of *s*, *co*, *cs*, and *bs*. The unusual aspect is that content-wise CASIA-CASSIL is closer to SWBD-DAMSL for its transcriptions of telephone conversations while ICSI-MRDA contains transcriptions of multi-party meetings. A possible explanation is that SWBD-DAMSL is in the general domain and that ICSI-MRDA is specific to certain topics.

2) It is apparent that CASIA-CASSIL does not share much similarity with AMI. The two schemes use different DA tags, such as *s* vs. INFORM. Moreover, due to the limited types of DAs in the AMI scheme, one AMI DA very often corresponds to more than one CASIA-CASSIL DA. Therefore, the reference to AMI seems to be mainly in the confirmation of the fundamental communicative functions in the interactive speech such as statement, questions and answers.

3) It is worth pointing out that two communicative functions are shared only by CASIA-CASSIL and SWBD-DAMSL albeit different tags are used. They are “Say Hello” and “Bye”, which are characteristic of telephone conversations.

4) Out of 48 tags, there is one DA type unique to CASIA-CASSIL, which is self-affirm (or auto-feedback in ISO terms), coded as *bsa*. As mentioned earlier, the domain of CASIA-CASSIL is tourism, where the information is mainly exchanged between a customer and the clerk at the front desk of hotels, restaurants, airport services and travel agencies. Therefore, the conversations sampled in the corpus involve various types of request and confirmation (information-seeking and -giving) where

self-confirmation is a frequent phenomenon when the customer double-checks the given information. The need for a self-affirm tag arising from tourism indicates that communicative functions represented in a DA scheme can be specific to a certain domain, project, or application.

5) It can be noticed that the mapping between the four DA schemes is not always one-to-one. For example, the DA type "statement" in CASIA-CASSIL corresponds to two DA types in SWBD-DAMSL (i.e. "statement-opinion" and "statement-non-opinion"). A possible reason is that CASIA-CASSIL is not interested in the fine-grained separation of statements or simply that it may not be easy to distinguish between a statement with opinion and the one without. Another case is that "accept" in CASIA-CASSIL corresponds to "accept" and "yes-answer" in SWBD-DAMSL. Similarly, "reject" in CASIA-CASSIL corresponds to "reject" and "no-answer" in SWBD-DAMSL.

4.2 MTCC Coding Scheme

The Mandarin Topic-oriented Conversation Corpus (MTCC) has also been annotated with dialogue act information. According to Tseng (2005b), its DA scheme is designed with reference to the Verbmobil-2 annotation scheme (Alexandersson *et al*. 1998), which focuses on a global dialogue structure rather than the functions. There are 34 DA tags in MTCC[3], categorized into five phases of the dialogue structure. Table 6 presents an initial mapping between the DA labels in the MTCC and those in the Verbmobil-2.

Table 6 MTCC vs. Verbmobil-2

MTCC			*Verbmobil-2*
Dialogue Structure		DA type	DA type
Opening		opening	INIT
Topic-negotiation		suggest_topic	INIT
		accept_topic	ACCEPT; FEEDBACK_POSITIVE
		comment_topic	DEFER
Topic-introduction		introduce_topic	INIT
Main discussion	Sub-topics management	begin_statement	INFORM
		connect_statement	/

[3] Tseng (2005b) noted that 37 DA tags were designed for MTCC but the latest version of guideline consists of 34 DA tags (http://mmc.sinica.edu.tw/).

		explain	INFORM (CLARIFY; GIVE_REASON)
		give_example	
	Opinion expression	agree	FEEDBACK_POSITIVE
		agree_part	FEEDBACK_POSITIVE
		oppose	FEEDBACK_NEGATIVE
		oppose_part	FEEDBACK_NEGATIVE
		comment_by_self	INFORM
		comment_by_other	FEEDBACK
	Sentential supplementation	confirm	CONFIRM
		correct	FEEDBACK
		rephrase	INFORM; FEEDBACK
		repeat	
	Feedback	feedback_understanding	FEEDBCK_POSITIVE
		feedback_non_understanding	FEEDBCK_NEGATIVE
		backchannel	BACKCHANNEL
	Action/info requests	request	REQUEST
		question	REQUEST
		answer	FEEDBACK
		question_request_answer	REQUEST
		rhetorical_question	INFORM
		rhetorical_question_answer	INFORM
	Sentential completion	completion_by_self	INFORM
		completion_by_other	FEEDBACK
	Exclamation	exclamation	/
	Hesitation	hesitation	/
Closing		closing	BYE; CLOSE
Sentential fragments		not_classified	NOT_CLASSIFIABLE

The differences in the DA coding scheme between the two corpora can be summarized as follows:

1) According to Tseng (2005b), the DA coding structure in the MTCC draws on the experience of Verbmobil-2. Nevertheless, the five phases in the two coding schemes are not the same.

MTCC: Opening, Topic-negotiation, Topic-introduction, Main discussion, Closing

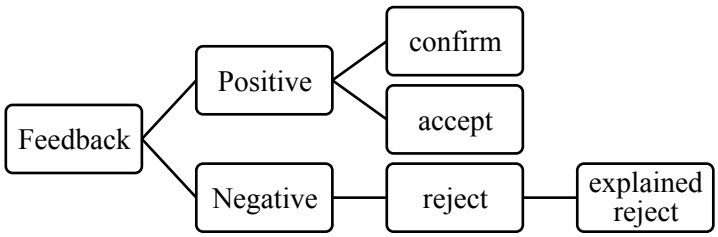
Verbmobil-2: Hello, Opening, Negotiation, Closing, Good_Bye

Such a difference may result from the differences in speaker relationships and conversation topics. First, according to the coding manual, two participants in the MTCC are friends or spouses, while participants in the Verbmobil-2 are not acquaintances. Therefore, the phases such as Hello and Good_Bye do not occur in the MTCC. Secondly, the participants in the MTCC are asked to choose the events of 2001 as the topic in the face-to-face interaction, and therefore it is natural for the speakers to decide the topic

before moving on to the actual discussion. As a result, MTCC has two extra phases, namely, Topic-negotiation and Topic-introduction, that are not represented in Verbmobil-2, thus again highlighting the fact that DA tags are often *ad hoc* and specific to practical tasks and applications.

2) In terms of the individual DA types, it can be noted that quite a few MTCC DA tags can be mapped to one Verbmobil-2 DA type. For example, begin_statement, connect_statement, explain and give_example in MTCC all correspond to "INFROM" in Verbmobil-2. There are two possible reasons. One is that discussions over news events may involve more types of statements than travel planning or appointment scheduling. A second reason is that since Verbmobil-2 is designed for application to multiple languages such as German, English and Japanese, its DA types are more generic.

3) MTCC is less hierarchical than Verbmobil-2. As shown in Table 6, the phase of Main discussion in MTCC has been further divided into sub-categories and yet the coding is nonetheless in a linear fashion. Meanwhile, according to the annotation manual of Verbmobil-2, a few DA categories have internal structures of multiple levels. For example, the DA of INFORM carries a mother-daughter-granddaughter branch (i.e. INFORM – CLARIFY – CLARIFY_ANSWER). The DA of feedback has a four-level branch:



An obvious advantage of hierarchical multi-level annotation is that it permits more flexibility in coding the actual utterance in that the annotation of a higher-level tag can be employed where there is no suitable lower-level tag.

4) Finally, exclamation and hesitation are unique to MTCC. They tend to be related to speakers' attitudes in the expression of their opinions during event discussions focused on in the project. They are not represented in projects like Verbmobil-2, which are more focused on information exchange. This

difference again suggests that communicative functions in interactive speech are closely associated with specific domains and interactive tasks. Event discussions entail a substantial degree of assessments and evaluations and thus differ considerably from information exchanges in terms of the way of expression, which leads to variations in the two taxonomies of communicative functions.

4.3 Discussions

Earlier in this section, we have taken a close look at the two DA coding schemes employed in two Chinese interactive speech corpora (i.e. CASIA-CASSIL and MTCC) respectively. The preliminary examination has also revealed notable similarities and differences between these two DA schemes.

The similarities are found in the following three aspects.

1) Both DA coding schemes are applied to spontaneous telephone conversations between two parties, and more importantly, both are domain specific rather than domain independent.
2) Although in different domains, both schemes cover the fundamental communicative functions such as statements, questions, answers, and feedback, which are also observed in all the other spoken corpora regardless of domain and language.
3) Both schemes are relatively flat in structure instead of adopting a hierarchical multi-level coding style.

Apart from the similarities in the two DA coding schemes, differences can also be observed in the following three aspects:

1) The two DA coding schemes are applied to different domains. To be more specific, CASIA-CASSIL is a collection of hotel reservations data in the field of tourism, and MTCC is restricted to discussions over social events. The domain difference between the two also results in a preference for different types of DAs. For instance, three types of request (i.e. *Request Affirmation, Request Detail, Request Suggestion*) are annotated in the CASIA-CASSIL, suggesting requests are common in the process of information seeking and providing. In contrast, these dialogue acts do not generally occur in the event discussions collected in MTCC. Thus, the design of a DA scheme and the implementation of its component tags are subject to the parameter of the task domain.
2) The participants in the two corpora have different interpersonal relations. In CASIA-CASSIL, the two participants are strangers while in MTCC they are either friends or spouses. Such a difference in

participant relations leads to a different preference for communicative functions and strategies. For example, politeness mechanisms such as "say hello" and "bye" are prominently featured in CASIA-CASSIL while they are not observed at all in the conversations sampled in MTCC.

3) The two DA schemes are designed with a different focus. While the DA tagset for CASIA-CASSIL is organized on the basis of the communicative functions, the DA tagset in the MTCC is "to sketch a global dialogue structure from a top-down perspective" (Tseng 2005b:205).

The above-mentioned comparison also reveals the current situation of DA coding schemes in the Chinese language.

- First, the available corpora of interactive spoken Chinese are not in a small number given the large magnitude of efforts to construct a spoken corpus and yet there is a limited number of corpora with DA annotation. With the development of technology and joint efforts around the world, further processing of the existing corpora with DA annotation could be beneficial to both linguistic research and speech engineering.
- Secondly, the Chinese spoken corpora with DA annotations are by far confined to specific domains, which has meant that the DA schemes are mostly domain specific while requirements arising from research and application have called for the design of a domain-independent DA scheme as a generic analytical framework for spoken corpora of a more varied range of domains and tasks in order to capture a broader understanding about spoken Chinese in an interactive setting.
- Thirdly, the existing DA coding schemes in Chinese are still in a linear fashion. They do meet the specific needs of the particular corpora but the value of being a reference to future DA annotation in Chinese may be limited. A more hierarchical structure of DA types should perhaps be adopted with a reference to the ISO DA coding scheme (24617-2 2012), which represents an international standard for DA analysis.

5. Conclusion

This chapter is a description of the semantic/pragmatic annotation of dialogue acts in corpora of interactive speech. In particular, it reported the current situation in the Chinese context and surveyed twelve Chinese

interactive speech corpora, four compiled in Mainland China (SCCSD, CADCC, CASIA-CASSIL and Chinese Language Acoustic Database Resources), two in Taiwan (TMC and NCCU), two in Hong Kong (HKCAC and HKUST/MTS), and four internationally in other countries (CALLHOME, CALLFRIEND, LLSCC, and WSC). While most of these corpora serve the primary purpose of linguistic research, four corpora (i.e. CADCC_2, CASIA-CASSIL, HKUST/MTS, and CALLHOME) are constructed for practical applications to develop dialogue systems and to improve speech recognition. A common feature of the Chinese spoken corpora is that they have been orthographically and/or phonetically transcribed. In addition, they have been annotated with linguistic and paralinguistic information such as dysfluencies, repairs, and overlapping and non-verbal behaviours such as laughing and coughing.

A close examination was conducted of the existing DA coding schemes in the Chinese context. This chapter performed a detailed analysis of the DA coding schemes for CASIA-CASSIL and MTCC. The analysis shows that the two schemes are domain-specific, suggesting that further DA annotation could be carried out on more generic data of Chinese interactive speech. The analysis also illustrates that the coding of the DA types is linear rather than hierarchical in both schemes. The downside of a linear DA coding scheme is that the exclusiveness lacks flexibility when applied to more complex spoken data. A possible enhancement is for the future design of DA coding schemes to refer to ISO 24617-2, a new international standard for DA annotation released in 2012, which adopts a generic hierarchical DA taxonomy. At the same time, the arrival of an international standard for DA annotation has also strongly suggested a requirement arising from research and application for the design of a domain-independent DA scheme as a generic analytical framework for spoken corpora in order to capture a broader understanding about spoken Chinese in an interactive setting. Indeed, such a generic hierarchical framework for Chinese will help to highlight the fact that dialogue acts are radial in nature, with a set of core functions and dimensions that can be extended to other, more peripheral ones. This fact can be held true not only to different dialogue scenarios but also to different languages such as English and Chinese, which, while sharing a set of core communicative functions, nonetheless may expectedly exhibit a range of extended dialogue strategies specific to linguistic and cultural settings.

As research on DA around the world has shown, DA annotation is becoming increasingly vital to the in-depth investigation into speech as the primary mode of human communication and to a wider range of applications especially in speech engineering in general and man-machine dialogue systems in particular. It is hoped that a comprehensive survey of the resources of spoken Chinese and the existing DA coding schemes will serve as a fresh starting point to facilitate refined studies of interactive Chinese with a view to identifying

the fundamental mechanisms that underpin interactive communications in the Chinese context.

Acknowledgement

Research described in this paper was supported in part by grants received from the General Research Fund of the University Grant Council of the Hong Kong Special Administrative Region, China (Project No. CityU 142711) and City University of Hong Kong (Project Nos. 7004755, 7004491, 9618007, 7004333, and 7004091).

References

Adolphs, S., and Carter, R. 2013. *Spoken corpus linguistics: from monomodal to multimodal*. Routledge.

Alexandersson, J., B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz and M. Siegel. 1998. Dialogue acts in VERBMOBIL-2 (second edition). Verbmobil Report 226. DFKI, Saarbrücken.

Allen, J., Chambers, N., Ferguson, G., Galescu, L., Jung, H., Swift, M., and Taysom, W. 2007. Plow: A collaborative task learning agent. In *Proceedings of the National Conference on Artificial Intelligence*, 1514-1519.

Allen J. and Core, M. 1997. DAMSL: *Dialogue Act Markup in Several Layers (Draft 2.1)*. Technical Report. University of Rochester, Rochester, NY.

Anderson, A. H., Bader, M., Bard, E. G., et al. 1991. The HCRC map task corpus. *Language and Speech*, 34(4), 351-366.

Armstrong, S., Clark, A., Coray, G., Georgescul, M., Pallotta, V., Popescu-Belis, A., Portabella, D., Rajman, M., and Starlander, M. 2003. Natural language queries on natural language data: A database of meeting dialogues. In *NLDB* 2003 *(8th International Conference on Applications of Natural Language to Information Systems)*, 14-27.

Augmented Multiparty Interaction Consortium (AMI). 2005. Guidelines for dialogue act and addressee annotation version 1.0, Unpublished manuscript.

Bunt, H. 2009. The DIT++ taxonomy for functional dialogue markup. In *Proceedings of EDAML/AAMAS Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts"*, Budapest, pp. 13--24.

Bunt, H., 2010. *DIT++ Taxonomy of Dialogue Acts*. Available at http://dit.uvt.nl/. Accessed on 26 December 2016.

Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Fang, A.C., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., and Traum, D. 2010. Towards an ISO Standard for Dialogue

Act Annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, MALTA, 17-23 May 2010.

Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A.C., Hasida, K., Petukhova, V., Popescu-Belis, A. and Traum, D. 2012. ISO 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of LREC* 2012, 23-25 May 2012, Istanbul.

Bunt, H., A.C. Fang, X. Liu, J. Cao and V. Petukhova. 2013. Making Switchboard dialogue annotations ISO-compliant. In *Proceedings of the 9th Workshop on Interoperable Semantic Annotation*, Potsdam, 19-20 March 2013.

Bunt, H., V. Petukhova, Malchanau, A., Fang, A.C. and Wijnhoven, K. 2016. The DialogBank. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, May 2016.

Canavan, A., and Zipperlen, G. 1996. CALLHOME mandarin Chinese speech. LDC96S34. DVD. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC96S34. Accessed 26 December 2016.

Canavan, A., and Zipperlen, G. 1996a. CALLFRIEND Mandarin Chinese-Mainland dialect LDC96S55. Web Download. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC96S55. Accessed 26 December 2016.

Canavan, A., and Zipperlen, G. 1996b. CALLFRIEND Mandarin Chinese-Taiwan dialect LDC96S56. Web Download. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC96S56. Accessed 26 December 2016.

Carletta, J., Isard, A., Isard, S., Kowtko, J., and Doherty-Sneddon, G. 1996. HCRC dialogue structure coding manual. Technical Report HCRC/TR-82.

Chui, K., and Lai, H. L. 2008. The NCCU corpus of spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, 6(2), 119-144.

Core, M.G. and Allen, J. 1997. Coding dialogs with the DAMSL annotation scheme. In *Proceedings of AAAI Fall Symposium on Communicative Action in Humans and Machines*.

Crowdy, S. 1993. Spoken corpus design and transcription. *Literary and Linguistic Computing* 8(4): 259-65

Dhillon, R., Bhagat, S., Carvey, H., and Shriberg, E. 2004. Meeting recorder project: dialogue labelling guide. ICSI Technical Report TR-04-002.

Fang, A.C., Bunt, H., Cao, J. and Liu, X. 2012a. Collaborative Annotation of Dialogue Acts: Application of a New ISO Standard to the Switchboard Corpus. In *Proceedings of EACL 2012 Workshop on Innovative Hybrid Approaches to the Processing of Textual Data*, 23 April 2012, Avignon, France.

Fang, A.C., Cao, J., Bunt, H. and Liu, X. 2012b. The Annotation of the Switchboard Corpus with the New ISO Standard for Dialogue Analysis. In *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, 3-5 October 2012, Pisa.

Fang, A.C., Cao, J., Liu, X. and Bunt, H. 2013. Lexical Characteristics of Dialogue Acts in the Switchboard Corpus of Telephone Conversations. *Journal of Foreign Language Teaching and Research, No 3, Vol 44*.

pp 28-40.

Fang, Zheng, Pengju Yan, Hui Sun, Mingxing Xu, and Wenuhu Wu. 2002. Collection of a Chinese spontaneous telephone speech corpus and proposal of robust rules for robust natural language parsing. *Joint International Conference of SNLP-O-COCOSDA*, 60-67.

Fung, P., Huang, S., and Graff, D. 2005. HKUST Mandarin telephone speech, Part 1 LDC2005S15. Web Download. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC2005S15. Accessed on 26 December 2016.

Gu, Y. 2002. Sampling situated discourse for spoken Chinese corpus. *Manuscript available at:* http://www.guyueguo.com/index.php?option=com_remository&Itemid=25&func=startdown&id=8 Accessed on 28 December 2016.

Gu, Y. 2006. Multimodal text analysis: A corpus linguistic approach to situated discourse. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies*, 26(2), 127-167.

Gu, Y. 2007. Compiling Spoken Chinese Corpora of Situated Discourse. Poster Session at the 10th International Pragmatics Association Conference at Gothenburg, Sweden.

Gu, Y. 2009. From real-life situated discourse to video-stream data-mining. *International Journal of Corpus Linguistics,* 14(4), 433-466 (34).

Huang, S., Bian, X., Wu, G., and McLemore, C. 1997 CALLHOME Mandarin Chinese lexicon LDC96L15. Web Download. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC96L15. Accessed 26 December 2016.

ISO 24617-2. 2012. Language resource management–semantic annotation framework (SemAF), Part 2: Dialogue acts. ISO, Geneva

Jurafsky, D., Shriberg, E. and Biasca, D. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, Draft 13. University of Colorado, Boulder Institute of Cognitive Science Technical Report 97-02.

Leung, M. T., and Law, S. P. 2001. HKCAC: The Hong Kong Cantonese adult language corpus. *International Journal of Corpus Linguistics*, 6(2), 305-325.

Li Yu-ming 李宇明. 2010. On the establishment of the audio database of Chinese language resources 论中国语言资源有声数据库的建设. *Chinese Language 中国语文* 4: 356-363.

Li, Aijun, Zhigang Yin, Maolin Wang, Bo Xu, and Chengqing Zong. 2002. Spontaneous conversation corpus CADCC. Phonetic report. Available at http://paslab.phonetics.org.cn/wp-content/files/research_report/2002/2002_13.pdf. Accessed on 26 December 2016.

Liu, Y., Fung, P., Yang, Y., Cieri, C., Huang, S., and Graff, D. 2006. HKUST/MTS: A very large scale mandarin telephone speech corpus. In *Chinese spoken language processing*, ed. Ma and Li, 724-735. Springer Berlin Heidelberg.

McEnery, Tony, and Richard Xiao. 2008. CALLHOME Mandarin Chinese transcripts - XML version LDC2008T17. Web Download. Philadelphia: Linguistic Data Consortium.

https://catalog.ldc.upenn.edu/LDC2008T17. Accessed on 26 December 2016.

Murray, G., Renals, S., Carletta, J., and Moore, J. 2006. Incorporating speaker and discourse features into speech summarization. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, 367-374.

Newman, J., Lin, J., Butler, T., and Zhang, E. 2007. The Wenzhou spoken corpus. *Corpora, 2*(1), 97-109.

Ranganath, R., Jurafsky, D., & Mcfarland, D. 2009. It's not you, it's me: Detecting flirting and its misperception in speed-dates. In *Proceedings of EMNLP*, 334–342.

Ritter, A., Cherry, C., and Dolan, B. 2010. Unsupervised modeling of twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, 172–180, Los Angeles, California, June 2010.

Sridhar, V. K. R., S. Narayanan, and S. Bangalore. 2008a. Enriching spoken language translation with dialog acts. In *Proceedings of ACL 2008*, Short Papers (Companion Volume). 225-228.

Sridhar, V. K. R., S. Narayanan, and S. Bangalore. 2008b. Incorporating discourse context in spoken language translation through dialog acts. *IEEE Spoken Language Technology*. 2008.

Stolcke, A. Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* 26(3): 339-373.

Traum, D. R. 1996. Coding schemes for spoken dialogue structure. Unpublished manuscript.

Traum, D. R. and Hinkelman, E. A. 1992. Conversation acts in task-oriented spoken dialogue. TRAINS technical report 425. Appear in *Computational Intelligence* 8(3):575-599.

Tseng, S. C. 2004. Processing spoken Mandarin corpora. *Traitement automatique des langues. Special Issue: Spoken Corpus Processing* 45(2): 89-108.

Tseng, S. C. 2005a. Syllable contractions in a Mandarin conversational dialogue corpus. *International Journal of Corpus Linguistics*, 10(1), 63-83.

Tseng, S. C. 2005b. Mandarin topic-oriented conversations. *Computational Linguistics and Chinese Language Processing,* 10(2), 201-218.

Tseng, S. C. 2008. Spoken corpora and analysis of natural speech. *Taiwan Journal of Linguistics,* 6(2), 1-26.

Tseng, S. C. 2013. Lexical coverage in Taiwan mandarin conversation. *Computational Linguistics and Chinese Language Processing,* 18(1), 1-18.

Wheatley, B. 1996. CALLHOME Mandarin Chinese transcripts LDC96T16. Web Download. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC96T16. Accessed on 26 December 2016.

Xiao, Zhonghua and Hongyin Tao. 2006. *The Lancaster Los Angeles Spoken Chinese Corpus*. Lancaster: UCREL, Lancaster University.

Xie, Guodong, Chengqing Zong and Bo Xu. 解国栋, 宗成庆, 徐波. 2002. Chinese spoken language analyzing facing the middle semantic representation 面向中间语义表示格式的汉语口语解析方法, *Journal of Chinese Information Processing 中文信息学报* 17(1):1-6.

Zhou, Keyan 周可艳. 2010. Research on Dialog-Act Understanding and Spoken Language Translation 对话行为理解与口语翻译方法研究. PhD thesis. Institute of Automation Chinese Academy of Sciences 中科院自动化研究所.

Zhou, Keyan and Chengqing Zong 周可艳, 宗成庆. 2010. Apply dialogue act information in spoken language translation 对话行为信息在口语翻译中的应用，*Journal of Chinese Information Processing 中文信息学报* 24(6):57-63.

Zhou, Keyan, Aijun Li, Zhigang Yin, and Chengqing Zong. 2010a. CASIA-CASSIL: A Chinese telephone conversation corpus in real scenarios with multi-leveled annotation. *International Conference on Language Resources and Evaluation*, LREC 2010.

Zhou, Keyan, Aijun Li, and Chengqing Zong. 2010b. Dialogue-act analysis with a conversational telephone speech corpus recorded in real scenarios. In *Proceedings of the Oriental COCOSDA Workshop (COCOSDA: International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques) (O-COCOSDA)*, Kathmandu, Nepal, November 24-25, 2010.