# Controlling Interaction in Multilingual Conversation Revisited: A Perspective for Services and Interviews in Mandarin Chinese

Jiali Du[1,3(✉)], Christina Alexandris[1,2], Dimitrios Mourouzidis[2],
Vasilios Floros[2], and Antonios Iliakis[2]

[1] Guangdong University of Foreign Studies, Guangzhou, China
`dujiali68@126.com`
[2] National and Kapodistrian University of Athens, Athens, Greece
`calexandris@gs.uoa.gr, {dmour,florbas,antiliak}@di.uoa.gr`
[3] Nanjing University, Nanjing, China

Abstract. The present approach targets to provide a framework for facilitating multilingual interaction in online business meetings with an agenda as well as in similar applications in the service sector where there is a less task-oriented form of interaction. A basic problem to be addressed is the control of the topics covered during the interaction and the expression of opinion. In the proposed template-based approach, the System acts as a mediator to control the dialog flow, within the modeled framework of the sublanguage-specific design.

Keywords: Templates · Speech Acts · Skype · Subtitles · Editors

1 Introduction: Assisting Skype Communication

The System involves the strategies employed for the integration of Spoken Mandarin Chinese in a Human-Computer Interaction framework for multilingual applications in routine business meetings as well as in short interviews via Skype. Based on previous approaches, the application concerns Skype communication with subtitles in the foreign language and the possibility of Machine Translation of subtitles and spoken text. The System designed in previous approaches concerns a Speech Act based template and agenda where all interaction is registered and controlled by the System, acting as a mediator between the communicating parties via Skype.
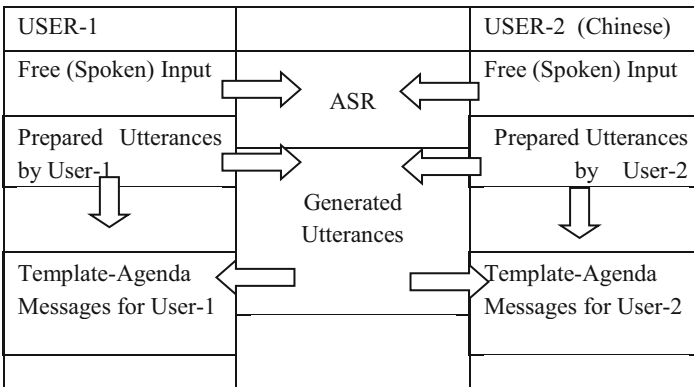
The application processes conversations of a standard and controlled nature, namely routine business meetings via Skype with an agenda as well as short interviews with fixed topics and agenda. Both types of communication are of a less task-oriented type which may include statement of sentiment or opinion. It should be stressed that business meetings or interviews whose main purpose is to persuade or to exercise pressure to obtain information are not handled by the present System.

Face-to-face interaction via Skype allows access to prosodic and paralinguistic information and feed-back from elements such as gestures, facial expression and tone of voice [1].

As presented in previous research [1], the application concerned, intended to be adaptable and reusable for various languages, involves a Speech Act based template and agenda, with interaction occurring within a Directed Dialog [2, 15, 16] connected to a respective Speech Act [1], however, with a Mixed Initiative [2, 4]. The Users communicate with each other with the assistance of the System-mediator. Interaction and turn-taking may be considered "push-to-talk conversations" [13].

Interaction is controlled by the mediating Speech Act based template and agenda of the System containing the topics covered during the interaction and checking the flow of the conversation by intervening messages appearing in the screen of the interface (Table 1). In this Mixed Initiative type of interaction, if a topic is not covered, the Users are alerted by the System [1]. The Speech Act based template and agenda contains a predefined set of sublanguage-specific questions/answers and statements and answers incorporated in the template-agenda by the User before the interaction and/or Skype meeting [1]. These (written) utterances may already be subjected to Machine Translation by online and/or commercial Machine Translation tools. Additional free input from the User's utterances is processed by Speech Recognition (ASR) and subsequently by a Machine Translation System after the interaction [1].

Table 1. Overview of system framework.

USER-1		USER-2 (Chinese)
Free (Spoken) Input	⇒ ASR ⇐	Free (Spoken) Input
Prepared Utterances by User-1	⇒ ⇐	Prepared Utterances by User-2
⇓	Generated Utterances	⇓
Template-Agenda Messages for User-1	⇐ ⇒	Template-Agenda Messages for User-2

The System is adapted to handle typical problems encountered in Spoken Mandarin Chinese in a Human-Computer Interaction framework with an English-speaking International Public.

Typical problems concern the tendency for native Mandarin Chinese speakers to express themselves implicitly and economically (1, Implicit statements), also with limited syntactic information, including the omission of syntactic elements in Chinese language (2, Omission) as well as the management of lexical ambiguity (3, Ambiguity). Silence, as an effective way of showing modest behavior (4, Silence) as well as a tendency to seldom take the role of the dialog pace-setter and to play a more passive role (5, Passive role in Dialog), constitutes additional problems in international communication with native speakers of Mandarin Chinese.

2   Mandarin Chinese - Design Parameters and Challenges

English and Chinese share some linguistic similarities. For example, both belong to the analytic language branch, which is contrary to inflected and agglutinative one; both linguistic systems work properly on a basis of word order and functional words, etc. However, the difference of syntactic form and semantic expression between English and Chinese makes international communication difficult to remain efficient and effective.

2.1   Implicit Statements

English is considered a low-context language in which nearly all the information has to be shown clearly and openly, especially for the syntactic elements. English has distinctive morphological changes and verbal conjugations as tense, aspect or voice markers. English is more like a subject-verb language for the fully syntactic structure can help readers grasp the grammatical framework easily.

Chinese is regarded as a high-context language in which some of syntactic elements can be omitted without influencing the semantic expression in oral or written practice. It is common for Chinese people to express themselves implicitly and economically. Chinese people tend to mean more than what they say. Principle of Least Effort can be well applied to the Chinese performance in which less words can mean more sense. Sometimes, even the Chinese functional words are omitted without altering the meaning. Chinese has few inflectional phenomenon, and even a Chinese character can yield its full meaning. Characters can convey more information than English words do, and limited Chinese characters can be applied to encapsulate the same meaning conveyed by more English words. In short, Chinese is more like a topic-focused language, and you have to guess the potential meaning by the limited syntactic information.

2.2   Omission

Anaphora is an important means of Chinese discourse cohesion. Zero anaphora in Chinese is sometimes considered to be Chinese empty category (ec). The omission of syntactic elements in Chinese language may greatly influence the quality of bilingual MT. English systems have to be familiar with this kind of Chinese omission and try to make the slots filled with appropriate elements. The cultural context, context of situation, and linguistic context are helpful for regaining the omitted forms [3].

Chinese people are used to omitting the subject, especially the animate subject; Chinese antecedent is usually considered to be the subject of a sentence, and the omitted anaphora phenomenon often appears here and there. Therefore, international English speakers who have access to SKYPE, if possible, have to regain the omitted forms in real-time situation to make the communication clear.

2.3 Ambiguity

If good communication is expected, performers have to pay attention to the ambiguity which may lead to misunderstanding or confusion. Both the lexical and syntactic ambiguity can be resolved by real-time paraphrase.

The application of paraphrase is helpful for ambiguity-free. For example, in the domain of lexical ambiguity, many Chinese speakers can be confused by the expression of "biweekly", which means appearing or happening both "every two weeks" and "twice a week", in the sentence of "the engines in a safe and good working condition should be taken out of service for maintenance work biweekly at least." To avoid the confusion, speakers have to paraphrase "biweekly" clearly, namely, to paraphrase it into every two weeks or twice a week. In the domain of syntactic ambiguity, "the horse raced past the barn fell" is a local ambiguous sentence. In order to alleviate the cognitive suffering, you have to paraphrase the sentence. Both "the horse that was raced past the barn fell" and "the horse drawn past the barn fell" are syntactic ambiguity-free sentences. With the development of translation technology, successful disambiguation is feasible and practical even though the application of SKYPE is now facing the challenge in China: "Meet the challenge and make the change".

2.4 Silence and Passive Role in Dialog

Silence is an effective way to show the modest behavior. Based on Confucianism, the modesty can be shown by keeping silence, and speaking aloud can obviously betray your weakness. Chinese speakers even believe that sometimes the outstanding usually bear the brunt of attack, and that common assumed knowledge is seldom to blame. This is one of the reasons why sometimes Chinese people seldom take the role of dialogue pace-setter in the international communication via SKYPE. According to the statistical data, Chinese scholars find that most Chinese speakers are shame in communicating with foreigners, and seldom lead the conversation in SKYPE-based international communication [19]. Besides focusing on cultural influence on language and behavior, many linguists highlight the ontology of Chinese (e.g. zero anaphora and omission) and try to pave the way for MT across languages.

2.5 Design Parameters

The foreign speakers who have access to SKYPE system have to respect both the cultural difference and linguistic difference between English and Chinese. Meanwhile, they can improve the performance by the suggestions shown as follows:

Providing options is an effective method. During the international communication, misunderstanding and embarrassment may occur here and there. The foreigners who have limited Chinese cultural and linguistic knowledge are suggested to write down the questions and alternative answers in advance before talking. Thus, the alternatives make Chinese speakers have to give a definite reply without considering the cultural and linguistic influence. However, the questions and answers prepared in advance should be logical and unambiguous. Besides that, Chinese people are familiar with the hierarchical

and numeral outline. If possible, international speakers are suggested to show the addressees the list of key points of conversation and the questions they expect to be answered.

Bilingual translation systems are helpful for improving the communication skill levels. In China, both SYSTRAN (http://www.systran-software.cn) and GOOGLE (http://translate.google.cn) translation systems are available. Many Chinese speakers try to use the systems to help them deal with some linguistic problems in real-time communication. By analyzing the translation results for 90 English business-related sentences, Chinese scholar [8] finds the fact that SMT (statistical MT) is better than RBMT (rule-based MT) when lexical ambiguity occurs (such as homograph, polysemy, as well as lexical transfer problems), while the effectiveness of two systems is the same when structural ambiguity happens.

3 Interaction and Design for Chinese Users for Services

Preparation before interaction and control of input and output to the Users-Participants are key features of previously designed approaches which are presently adapted to the requirements of Mandarin Chinese native speakers. Specifically, Users determine the types of information contained in the interaction, in the form of a list of prepared questions and possible types of responses and statements [1]. This preparation occurs before the actual interaction and is assisted by pre-determined topic and sublanguage-specific questions, statements and answers. The topics and overall sublanguage-specific framework are determined by the Users, according to the content and agenda of the meeting.

During interaction, the translated message appears on the screen of the User-Receiver in the target language. The topics covered during interaction are registered by the template-agenda. If a topic is not addressed, the template-agenda generates a reminder-message [1].

The construction of the present System with the above-described interaction is based on implemented applications for Spoken Dialog Systems in the Service Sector (Call Centers for mobile telephones) [2], with Directed Dialogs and registration of the path of the interaction with the respective Speech Acts, keywords and free input [2]. The implemented modules of the Dialog System applications [2] are adapted to the needs of Business Meetings (I) and Interviews (II).

3.1 Avoiding Implicit Statements, Omission and Ambiguity

The preparation of utterances to be activated is assisted by an editor controlling the (1) length of the utterances (sentences with a maximum of 30 words to facilitate Machine Translation), confirming the User's choice of topic (TOPIC) and selected Speech Act type (SPEECH-ACT) (2), as well as 1–3 words selected to be highlighted as "keywords" (3) and related to recognizable utterances and acceptable answers.

The defined topics, selected Speech Act types (2) and keywords to be highlighted (3) are the types of pre-determined sublanguage-specific information handled during interaction and are contained in templates filled in by the Users before interaction.

The series of topics and respective templates constitute the template-agenda, activated by the Users during the time of the actual online interaction. Templates related to Speech-Acts constituting questions are designed to contain the phrases such as "Please answer with Yes or No" or "Please reply with X, Y or Z" (X, Y and Z constituting keywords).

Each response and/or message is activated by the User in the appropriate step in the dialog. The activated response or message may be previously translated by an online conventional Text-to-Text Machine Translation System [1].

In particular, the Speaker's input is limited to the topic of the activated prepared messages and a positive or negative answer or a keyword-specific answer, restricting the possibilities of Ambiguity, Implicit Statements and Omission. Additionally, in the case of complications during interaction, a form of "stepping stone" [7] strategy is employed where activated Speech Acts requesting clarification may remind the User to answer in the form of keywords, resembling a typical type of interaction within a strict Directed Dialog and/or traditional Interlingua (ILT) framework [6, 12], targeting to bypass the phenomena of Ambiguity, Implicit Statements and Omission. In other words, the functions of traditional ILTs are performed during the preparation process of the templates with the assistance of the editor. Also, unlike traditional ILTs, the sublanguage of the interaction is not limited to one specialized field but the sublanguage-related subject area and topics can be determined each time by the Users of the System prior to interaction.

The present model of User interaction with the System and Directed Dialog framework [15, 16] on which the previously designed approaches are based [1], aims to prevent an uncontrolled number of possible forms and variations [8] in (a) the expression in the language concerned and in (b) User behavior due to cultural and social factors. For spoken Mandarin Chinese, the above-described restrictions and specifications also contribute to the limitation of phonologically similar keywords but with a different tone.

3.2   Handling Silence and Passive Role in Dialog

Targeting to user-friendliness [14], explanatory repetitions as an alternative version of the so-called "stepping stones" [7] assist the User during interaction [2, 7] as additional Task-related and Non-task related Speech Acts integrated in dialog structure. These Non-Task-related Speech Acts [1], whose determination was based on data from European Union Projects [9], are used for tasks such as "Offer", "Reminder" or "Manage-Waiting-Time" [9], mostly in messages generated by the System.

In the present Mixed-Initiative type of interaction, the activated Non-Task-related Speech Acts (NTRs) may extend the length of the dialog; however, they encourage the participation of the Speakers and contribute to the management of deliberate silence of the Speakers, as in the case of spoken Mandarin Chinese and the native speakers of the language. In the application concerned, NTRs are activated by the User; however, if there is no response or if there is another type of complication during communication, the System's template-agenda reminds the Users to activate an NTR- Speech Act selected from the set of prepared utterances (Table 2).

Table 2. Prepared utterances and Topics (from 1 to X), activated non-task related Speech Acts and template-agenda for interviews.

Prepared Utterances by USER:	SYSTEM: Alert-Messages
TOPIC -1 [A] SPEECH ACT [inform]: ☐ Utterance 1 ☐ Utterance 2	ALERT: SYSTEM MESSAGE: NO ANSWER OR UNPROCESSABLE INPUT
IF: NO-INPUT: NTR- SPEECH ACT Offer-Explain	
TOPIC -1 [A] [question]: ☐ Utterance 1 ☐ Utterance 2	ALERT: SYSTEM MESSAGE: NO ANSWER OR UNPROCESSABLE INPUT
IF: NO-INPUT: NTR- SPEECH ACT Reminder-Explain	
[….] TOPIC -X [W] SPEECH ACT [[proposal]: ☐ Utterance 1 ☐ Utterance 2 TOPIC -X [W] [question]: ☐ Utterance 1 ☐ Utterance 2	
[END –OF- INTERACTION] SYSTEM: Template-Agenda	
TOPIC -1: CHECKED ∅ ANSWERED: YES ∅ NO ∅ Confirmed-Silence ∅ [….] TOPIC -X: CHECKED ∅ ANSWERED: YES ∅ NO ∅ Confirmed-Silence ∅	

For spoken Mandarin Chinese, the Non-Task-related Speech Acts (NTRs) are in pre-existing templates prepared by the User and are designed to already contain phrases with explanatory content, in addition to any modifications the User wishes to make. For example, the activated NTR-Speech Act "Offer-Explain" in System output is related to phrases such a "Would you like me to proceed with the next thing you want to tell me?" or the activated NTR-Speech Act "Reminder-Explain" in output is related to phrases such as "I have no information from you about {X}. You must tell me if you want {X}". The "Offer-Explain" NTR-Speech Act is activated in the case of an unanswered question and/or unaddressed TOPIC, after a second attempt of receiving an answer, giving the

impression that interaction is "moving forward", according practices in spoken Dialog Systems [7]. The "Reminder-Explain" NTR-Speech Act is activated at the end of a part of the interaction or at the end of the entire interaction, as an additional attempt to receive an answer and/or to address a TOPIC. Both the "Offer-Explain" and the "Reminder-Explain" Non-Task-related Speech Act can be activated before Speech Acts managing silences.

The use of Non-Task-related Speech Acts also allows the insertion and integration of Chinese [10, 17, 18, 20] pragmatically-related politeness forms, without affecting the basic form and content of the original dialog structure. This possibility may also be adapted to languages such as Arabic [11] and Hindi [5] for handling pragmatically-related politeness in the inserted Non-Task-related Speech Acts not containing stand-ardized and cross-linguistic dialog content.

Additional Task-related Speech Acts are introduced such as "Manage-Silence" with System output such as "Please confirm that you have not answered this question by pressing OK", where the System by-passes a situation causing the native speaker of Mandarin Chinese to feel uncomfortable, while at the same time allows the signalization of a deliberate silence as a response from the native speaker of Mandarin Chinese. In contrast to previous approaches [1], unaddressed TOPICs saved in the agenda of the template can either be marked as "Confirmed-Silence" or the default option, "Unad-dressed". TOPICs marked with "Confirmed-Silence" are evaluated by the participants.

4   Interaction and Design for Interviews

A similar Mixed-Initiative type of interaction is employed on the case of interviews (Interviews, II) with native speakers of Mandarin Chinese, provided that the interviews are short and contain a specific agenda. In this case, there is the additional possibility of the ad hoc generation of not previously prepared short messages (Table 3). These messages are subjected to online Machine Translation and appear at the bottom of the screen, along with the prepared questions and answers from both parties.

Table 3. Overview of system framework for interviews.
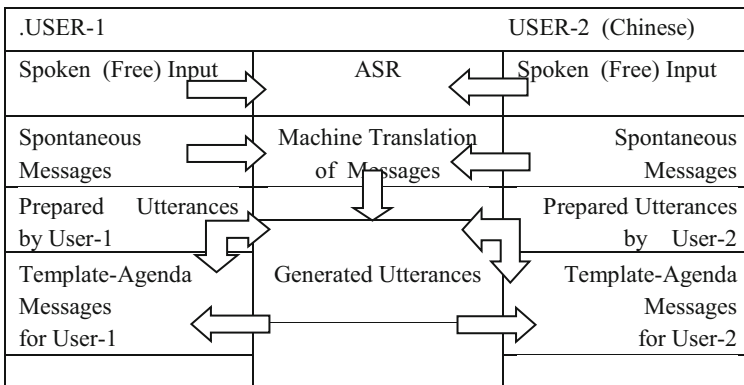
For short interviews with a specific agenda, the above-presented "stepping stone" option [7], activating Speech Acts requests the User to answer in the form of the determined keywords or with an answer corresponding to a "Yes" or a "No". In other words, the User is directed to a list of possible or acceptable answers, targeting to by-pass the possibilities of Ambiguity, Implicit Statements and Omission. Task-related and Non-Task-related Speech Acts are activated if the content of the produced utterances is unclear or ambiguous, if there is a deliberate silence (Silence) from the Speaker or if the passive role of the Speaker creates complications in the interaction (Passive Role).

Furthermore, in Interviews (II) the content of the Non-Task-related Speech Acts may differ from that of Business Meetings (I). For example, the activated NTR-Speech Act "Offer-Explain" in System output is related to phrases such a "Would you like me to proceed with the next TOPIC?" or the activated NTR-Speech Act "Reminder-Explain" in output is related to phrases such as "TOPIC {X} has not been discussed yet. We must talk about this TOPIC {X}".

As applied in Business Meetings (I), in Interviews (II) the "Offer-Explain" Non-Task-related Speech Act is activated in the case of an unanswered question and/or unaddressed TOPIC, after a second or even a third attempt of receiving an answer, and the "Reminder-Explain" Non-Task-related Speech Act is activated at the end of a part of the interaction or at the end of the entire interaction, as an additional attempt of receiving an answer or resolving an issue concerning a defined TOPIC.

Also, as applied in Business Meetings (I), these Non-Task-related Speech Acts can be activated before the "Manage-Silence" Speech Act. In particular, in the case of Interviews (II), the "Manage-Silence" Speech Act may activate the generation of messages such as "Please confirm that you wish this question to remain unanswered by pressing No-answer", or, if opinions are requested "No opinion". The template-agenda marks the respective choices corresponding to the Speaker's response.

5   Conclusions and Further Research

The presented System concerns Skype communication with Machine Translation of subtitles and generated spoken text with parameters involving spoken Mandarin Chinese. Based on previous approaches, the System contains a Speech Act based template and agenda registering and controlling all interaction and acting as a mediator between the communicating parties. Features of spoken Mandarin Chinese may create complications in communications with the international public, especially speakers of the Western culture, even if conversations of a standard and controlled nature are processed. The applications in the presented framework concern routine business meetings and short interviews with an agenda. Based on standard and Directed Dialog based practices of Spoken Dialog Systems, interaction of the Users with the System is targeted to bypass typical problems in communication related to linguistic and cultural issues. At the same time, the System allows the Users to determine the type of topics and agenda prior to the meeting or the interview, without excluding the processing of free-input.

The limitations of the framework presented is the necessity to prepare utterances with the aid of the System prior to the interaction and to process and evaluate free input

after the interaction. In addition, all or most of the topics to be addressed should be defined by both communicating parties. This process may be time-consuming, but on the other hand, it allows the possibility of determining different topics and sublanguage by the Users of the System each time the System is used. Especially in the case of interviews with an agenda, Users are also allowed to determine the content and appropriate style of the activated Speech Acts and messages, including the Non-Task-related Speech Acts assisting interaction. Furthermore, use of already-existing tools in Machine Translation and the adaptation of modules from implemented applications for Spoken Dialog Systems in the Service Sector (Call Centers for mobile telephones) reduce the cost and time involved in building the application.

The designed applications are to be evaluated by a larger user-group to be determined. The envisioned further development includes the design and implementation of an improved user-friendly interface and possible adaptions to other languages.

References

1. Alexandris, C.: Controlling interaction in multilingual conversation. In: Kurosu, M. (ed.) HCI 2013. LNCS, vol. 8007, pp. 3–12. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39330-3_1

2. Floros, V., Mourouzidis, D.: Multiple Task Management in a Dialog System for Call Centers. Masters thesis, Department of Informatics and Telecommunications, National University of Athens, Greece (2016)

3. Hou, M., Sun, J.J.: Zero anaphora in Chinese and how to process it in Chinese-English MT. J. Chin. Inf. Process. **1**, 14–20 (2005)

4. Jurafsky, D., Martin, J.: Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall series in Artificial Intelligence, 2nd edn. Pearson Education, Upper Saddle River (2008)

5. Kumar, R.: A politeness recognition tool for Hindi, with special emphasis on online texts. In: Proceedings of the WWW Ph.D. Symposium, March 28–April 1, Hyderabad, India (2011)

6. Levin, L., Gates, D., Lavie, A., Pianesi, F., Wallace, D., Watanabe, T., Woszczyna, M.: Evaluation of a practical interlingua for task-oriented dialog. In: Proceedings of ANLP/NAACL-2000 Workshop on Applied Interlinguas, April 2000, Seattle, WA USA (2000)

7. Lewis, J.R.: Introduction to Practical Speech User Interface Design for Interactive Voice Response Applications. IBM Software Group, USA, Tutorial T 2009 presented at HCI 2009 San Diego, CA, USA (2009)

8. Ma, J.: A comparative analysis of the ambiguity resolution of two English-Chinese MT approaches: RBMT and SMT. Dalian Univ. Technol. J. **31**(3), 114–119 (2010)

9. Nottas, M., Alexandris, C., Tsopanoglou, A., Bakamidis, S.: A hybrid approach to dialog input in the citzenshield dialog system for consumer complaints. In: Proceedings of HCI 2007, Beijing, Peoples Republic of China (2007)

10. Pan, Y.: Politeness in Chinese Face-to-Face Interaction. Advances in Discourse Processes series, vol. 67. Ablex Publishing Corporation, Stamford (2000)

11. Shammas, N.A.: Lingua-pragmatic politeness and translatability. Damascus Univ. J. **21**(3+4), 23–56 (2005)

12. Schultz, T., Alexander, D., Black, A., Peterson, K., Suebvisai, S., Waibel, A.: A Thai speech translation system for medical dialogs. In: Proceedings of the conference on Human Language Technologies (HLT-NAACL), Boston, MA, USA (2004)

13. Taboada, M.: Spontaneous and non-spontaneous turn-taking. Pragmatics **16**(2–3), 329–360 (2006)
14. Wiegers, K.E.: Software Requirements. Microsoft Press, Redmond (2005)
15. Williams, J.D., Witt, S.M.: A comparison of dialog strategies for call routing. Int. J. Speech Technol. **7**(1), 9–24 (2004)
16. Williams, J.D., Poupart, P., Young. S.: Partially observable markov decision processes with continuous observations for dialog management. In: Proceedings of the 6th SigDial Workshop on Discourse and Dialog, September, Lisbon (2005)
17. Yin, L.: Cultural differences of politeness in English and Chinese. Asian Soc. Sci. **5**(6), 154–156 (2009)
18. Yu, Z., Yu, Z., Aoyama, H., Ozeki, M., Nakamura, Y.: Capture, recognition, and visualization of human semantic interactions in meetings. In: Proceedings of PerCom, Mannheim, Germany (2010)
19. Zhao, C.R., Liu, R.Q.: An exploration on distance English oral test- an empirical study on English oral test through web phone Skype. Mod. Educ. Technol. **22**(2), 95–98 (2012)
20. Zhu, H., Wei, L., Yuan, Q.: The sequential organization of gift offering and acceptance in Chinese. J. Pragmatics **32**, 81–103 (2000)