

# Controlling Interaction in Multilingual Conversation

Christina Alexandris

National University of Athens, Greece  
calexandris@gs.uoa.gr

**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 is proposed to act as a mediator to control the dialog flow, within the modeled framework of the sublanguage-specific and pragmatically related design.

**Keywords:** Templates, Simple Interlinguas, Non Task-related Speech Acts, Skype, subtitles.

## 1 Introduction

The domain of the proposed framework concerns routine business meetings via Skype with an agenda, namely standard, controlled conversation. This domain includes services requiring a less task-oriented form of interaction as well as statement of sentiment or opinion. The proposed framework is not recommended for business meetings concerning negotiations or business deals. Conditions involve multilingual conversation with speakers of less spoken languages and an average or less than average knowledge of English. The interaction involves Skype meetings with or without translated subtitles. Subtitles (text messages) are combined with spoken input. Users may speak at the same time while the corresponding, similar or even different text message is generated. The communicating parties have access to prosodic and paralinguistic information, such as tone of voice, as well as facial expressions and other paralinguistic elements.

The flow of the conversation can be checked by the System or the User, as well as the topics covered. This is achieved by intervening messages by the System, appearing in the screen of the interface. Furthermore, additional free input from the User's utterances may be processed after the interaction. Spoken interaction is processed with the use of Interlinguas (we propose the use of "Simple Interlinguas" – SILTs) [8] generated simultaneously with translated text messages chosen by the User.

The proposed service may be regarded as an alternative to email in routine business meetings, allowing face-to-face interaction and feed-back from paralinguistic elements, such as gestures, facial expression and tone of voice. Furthermore, in multilingual applications involving communication with a standard agenda, the proposed approach may be adaptable and reusable in respect to several languages.

Specifically, the present design concerns a Speech Act based template and agenda. This agenda may be visible or invisible to Users. The interaction takes place in a Directed Dialog [15][16] like form and related to the respective Speech Act. The template contains a set of sublanguage-specific questions, statements and answers, including opinions, some pre-existing, while others are left to be prepared by the User prior to the meeting. The template-agenda contains the topics covered during the interaction and reminds Users, if a topic is not covered. In other words, interaction is controlled by the templates of the System, which acts as a mediator.

## **2 Design Principles and Previous Approaches**

Three factors may be taken into account for the present design. First, Service- Oriented Dialog Systems targeted towards the broad public involve a higher percentage of non-sublanguage specific vocabulary and a lower percentage of terminology and professional jargon. In particular, applications related to business meetings often involve expressions related to emotion and the statement of opinion (1). Second, unlike highly specialized Task-related Dialog Systems, in Service- Oriented Dialog Systems, the Human-Computer interaction taking place is directed towards two equally significant goals, namely the successful performance of the activated or requested task as well as User satisfaction and User-friendliness (2). These goals are related to requirements on the Satisfaction Level in respect to a System's evaluation criteria, namely perceived task success, comparability of human partner and trustworthiness [10]. It should be noted that the more goals to be achieved, the more parameters in the System Design and System Requirements, and, subsequently, Dialog Design are to be considered [14]. The diversity of a multilingual User Group (3) constitutes an additional factor in the present design, ranging from Users belonging to communities where any real experience with computers and electronic devices is restricted to only a minority of Users, to User Groups with an absolute familiarity of society with computers and electronic devices, including cases where a tendency towards attachment [9] or animism of these objects is observed.

### **2.1 Directed Dialogs**

For achieving User-friendliness in multilingual Dialog Systems with a diversity of Users, strategies such as the use of Directed Dialogs [15] [16] using keywords offer a predefined pattern of User interaction with the System in order to prevent an uncontrolled number of possible forms and variations [8] in (a) the expression in each language and in (b) User behavior due to cultural and social factors.

The use of Directed Dialogs and Yes-No questions aims to the highest possible recognition rate of a very broad and varied user group and the use of free spoken input processes the detailed information involved in a complex application. Keyword recognition largely occurs within a Yes-No question sequence of a Directed Dialog.

Within the Directed-Dialog framework of Dialog Systems in the service sector, such as in the present application, additional types of Speech Acts are detected, other than strictly Task-related Speech Acts. These Non-Task-related Speech Acts, [2]

whose determination was based on data from European Union Projects [19][20][21][22] are used for tasks such as “Offer”, “Reminder” or “Manage-Waiting-Time” [2], mostly in messages generated by the System.

## 2.2 Template Generation

The present approach is based on previous practices concerning the use of templates interacting within a Directed Dialog framework. A typical Dialog System involving the use of templates is the CitizenShield Dialog System for consumer complaints [3][11] handling routine tasks involving food and manufactured products (namely complaints involving quality, product labels, prices etc.). The spoken input is automatically entered into templates containing a number of fields related to the categories and types of information concerning the product involved. Free spoken input is recorded within a defined period of time, following a question requiring detailed information and/or detailed descriptions. All spoken input, whether constituting an answer to a Yes-No question or constituting an answer to a question triggering a Free-Input answer, is automatically directed to the respective templates of a complaint form, which are filled in by the spoken utterances, recognized by the System’s Automatic Speech Recognition (ASR) component. The automatic filling-in of complaint forms with spoken input via the consumer organization’s call center is especially helpful to mobile Users and Users that have no internet access. The information contained in each field of the complaint form is automatically or manually processable, according to the type of task to be executed [3][11]. The generation of the template-based complaint forms is also aimed towards the construction of continually updated databases from which statistical and other types of information is retrievable for the use of companies, organizations or authorities or other interested parties [3][11].

In other words, in previous approaches, the templates are both registering and controlling User input [11]. In the present design, the System may also use the templates to check if all issues to be addressed are covered. In other words, the template also behaves as an agenda during the interaction. In addition, according to the indications of the template-agenda, the System may automatically intervene with asking additional questions, until the issue in question is addressed. Most of these questions are of the Non-task related Speech Act type, such as “Offer”, “Reminder” or “Manage-Waiting-Time” [2].

Furthermore, the present approach involves the activation of prepared answers contained in the template activated by the User in the appropriate step in the dialog.

## 2.3 Directed Dialogs and Interlinguas

For Multilingual Dialog Systems using Directed Dialogs, Simple Interlinguas (S-ILTs) are proposed (Table 1) [8], constituting lexical-based alternatives and a simplified form of Interlinguas. S-ILTs may be characterized by a very simple structure and with a weakened “frame” function [8] which in traditional Interlinguas summarizes the semantic content of a spoken utterance [7][13]. In S-ILTs, the semantic content of a spoken utterance is signaled by the respective topic of step in the dialog structure. Specifically, the

use of Directed Dialogs, and the Speech Acts performed in the dialog structure, the proposed lexical-based alternatives are linked to Speech Act types in respect to the steps in dialog context. Thus, the role of the “frame” is weakened and the role of dialog structure is reinforced. With the use of Directed Dialogs in multilingual applications, the proposed “Simple Interlinguas” (S-ILTs) [8] allow the recognition and isolated processing of keywords at the lexical level, a feature which facilitates the compatibility with Systems that support multiple languages and cases of polysemy and multiple grammatical functions of word types. Furthermore, Simple Interlinguas operating on the lexical level may also be directly linked to more language-independent entities such as the Universal Words (UWs) of the UNDL System of the United Nations [5][23] which are linked to each other by semantic relations such as hyperonymy and synonymy.

**Table 1.** Example of a Simple Interlingua (S-ILT) connected to the step in dialog structure and respective Speech Act for applications concerning the expression of opinion. Optional entries at Keyword Content level are functions “Who”, and “When” and respective lexical entries.

SPEECH ACT TOPIC OF STEP IN DIALOG STRUCTURE (SYSTEM):	S-ILT (RESPONSE FROM USER CONTAINING LANGUAGE-SPECIFIC EXPRESSIONS)
TOPIC-OF-STEP {REFUSE} =>	=> [ S-FRAME TOPIC-OF-STEP {REFUSE} YES/NO (no-answer/no-expression) <i>Optional</i> : WHO(person), WHEN(time) ]

However, it should be stressed that the proposed S-ILTs are sublanguage-specific and are modeled according to the Speech Acts in the steps of the dialog structure and keywords of the application domain. In other words, a rigid and controlled nature of a System’s dialog structure allows the successful function of the proposed S-Interlinguas, specifically, Directed Dialogs [15][16], involving Yes-No Questions or questions directed towards Keyword answers with related signaled topic (TOPIC-OF-STEP). The combination of the use of Directed Dialogs and S-Interlinguas allows the control and successful handling of a varied or ambiguous input, in accordance to the criteria of the Utterance Level (Question-Answer-Level) and the Functional Level, especially informativeness and intelligibility (Utterance Level) and the ease of use, interaction control and processing speed/smoothness (Functional Level) [10].

It should additionally be noted that the proposed S-ILTs allow minimum interference of language-specific factors, since they are designed to function within a very restricted sublanguage related to a specific task. They can, therefore, be adapted for a very diverse range of languages and language families, for example, Hindi and Chinese. This S-ILT framework is proposed for the present approach, however, any other Interlingua type, if appropriate for the languages concerned, may be used.

### 3 Interaction

In the present approach, interaction occurs in three levels in respect to interaction type (A) and also in three levels in respect to the chronological process of interaction (B).

The type of interaction (A) concerns the multimodal interaction with spoken utterances with Speech-to-Speech Translation with Interlinguas or S-ILTs (1) which occurs simultaneously with interaction with prepared texts appearing on the screen as subtitles during interaction with Skype (2) or any other form of visual interaction. These texts can be translated by a Text-to-Text translation System. The third type of interaction involves the production of filled-in templates (3) with the issues covered during the multimodal interaction process or pending issues, as well as additional information and comments produced by speakers in the form of free input. The proposed approach is strictly sublanguage and domain-specific, however, it is designed to be compatible with online (written text) Machine Translation systems, such as Google Translate, processing many languages. However, a language-specific choice is necessary for the spoken Machine Translation component in regard to Interlinguas (ILTs) or the proposed Simple Interlinguas (S-ILTs). The design of the proposed approach in respect to interaction type may be depicted in the following table (Table 2).

**Table 2.** Interaction Type: Input and Output

<b>Input Type (During Interaction):</b>	
<b>SCREEN</b> (e.g. Skype)	<b>Template-Agenda ☒</b> [open/close] <b>(3)</b> <ul style="list-style-type: none"> <li>•Registers Free Input</li> <li>•Checks topics covered</li> <li>•Generates messages</li> </ul>
<i>Paralinguistic Data (Prosody, Gestures etc.)</i>	
<b>(1) INTERLINGUAS (ILT OR S-ILT)</b>	
Speech-to-Speech Translation	
<b>(2) SUBTITLES</b>	
Text-to-Text Translation ( <i>e.g. Google Translate</i> )	
<b>(3) SYSTEM MESSAGES TO USER</b>	

Regarding the chronological process of interaction (B), the involvement of the Users of both ends concerns the time span prior to the actual interaction (I), the time span of the actual interaction (II) and the time span after the interaction (III).

### 3.1 Template-Agenda and Preparation of Interaction (I)

In the time span prior to the actual interaction (I), the Users prepare the set of questions, statements and possible set of answers within the framework of the template containing the types of information handled during interaction.

The prepared restricted set of questions, statements and set of possible answers are activated by the User during the time of the actual online interaction. The preparation of topics, agenda and general outline of dialog structure is determined by the type, content and style of routine communication defined according to the tasks and policy of the company or organization. In other words, the sublanguage-specific set of questions, statements and set of possible answers prepared by the Users is also determined by the tasks and policy of the company or organization. This template-agenda contains a set of sublanguage-specific questions, statements and answers, some pre-existing in the sublanguage-specific design, while others are left to be prepared by

the User prior to the meeting. However, the messages to be prepared by the User receive specific content tags related to the sublanguage of the template and the related Interlinguas and must contain words related to them and there are restrictions in respect to the length of the utterances used. In addition, the User may choose from a set of symbolic markers to indicate attitude or emotion, if desired or if necessary, for example “assertive” or “polite”. The prepared answers and other messages contained in the template are activated by the User in the appropriate step in the dialog. Each activated answer in the source language is automatically translated by a conventional Text-to-Text Machine Translation System and appears on the screen of the Receiver in the target language, while at the same time the User in the source language may utter the same or a similar sentence to the utterance activated. As an optional element, any symbolic markers indicating attitude or emotion may also be displayed on the screen of the Receiver. The template-agenda contains the topics covered during the interaction and reminds Users, if a topic is not covered. A similar process is activated if opinions are requested. The general outline and content of the template-agenda may be depicted in the following table (Table 3).

**Table 3.** Outline and content of the template-agenda

Prepared Utterances by User:	System:	Template-Agenda
ISSUE -1 [proposal]: Utterance 1 Utterance 2	ISSUE -1: CHECKED	∅
ISSUE -2 [check]: Utterance 1 Utterance 2	ISSUE -2: CHECKED	∅
ANSWER: Rejection	ANSWERED:	
Utterance 1 [neutral] Utterance 2 [assertive]	YES	∅ NO ∅
Utterance 3 [polite]		

### 3.2 Multimodal Interaction (II)

Apart from the automatically translated prepared utterances activated by the Users on both ends (2), the System also uses a Speech-to-Speech translation System with the use of the previously presented simplified form of Interlinguas, the S-ILTs (2), operating within a strict Directed Dialog framework. During the actual interaction, Users may speak, while at the same time they can activate the corresponding, text message (Table 4). Thus, the Interlingua processes spoken input whose content is identical, similar or even different than the generated text message, allowing both a controlled and a spontaneous type of information to be directed and evaluated by the Receiver. Additionally, it may be noted that written and spoken input may be compared to paralinguistic elements appearing on the screen.

The flow of the dialogs is controlled by the agenda in the template, indicating addressed and pending issues. The template-agenda contains the topics covered during the interaction and reminds Users, if a topic is not covered. A similar process is activated if opinions are requested. The template may be visible to the User, if requested (“open”, “close”). If an issue or a question is not addressed, the agenda on the template informs the User with a respective message, for example “Pending Issue: Terms and Conditions” or “No answer”. To repeat question press “R”. These messages

appear below the screen. In the case of an unaddressed issue, the System may remind the User for a predefined number of times after the response of the Receiver. Issues that are left unaddressed are saved in the agenda of the template. In the case of an unanswered question, especially, if opinions are asked, after a second attempt, the question is marked as unanswered in the agenda of the template.

**Table 4.** Chronological process of interaction

<b>Before Interaction:</b> Preparation/Editing Utterances
<b>During Interaction:</b> Prepared Utterances ->Text-to-Text Translation- Spontaneous Utterances ->Speech-to-Speech Translation
<b>After Interaction:</b> Template with points covered / Issues unaddressed / no answers Free input

### 3.3 Post-Interaction (III)

At the end of the interaction, closing remarks and any additional comments are processed as free input and saved in a wav.file to be processed after the interaction by a speech recognition (ASR) system and are subject to translation. After the interaction, the template also indicates unaddressed issues and unanswered questions, including unexpressed opinions.

## 4 Expression of Opinion

In multilingual HCI applications not limited to Task-Related Speech Acts and allowing the expression of views, opinions and spoken suggestions, as in the case of the present application, the relation of the semantic with the pragmatic level may create problems in some language pairs, since the “opinion” is expressed either directly or indirectly with significant differences between languages. Specifically, it may be noted that opinions in relation to a query or a specific issue may not always be expressed with a direct answer at a “Yes-No” question. Depending on the languages concerned, opinion may be contained within specific expressions or it may be avoided but indirectly expressed at the pragmatic level, for example, by avoiding to specifically address an issue.

In the proposed approach, the use of templates managed by the System allows the intervention of the System with inserted Non Task-Related Speech Acts (a) to control complex types of interaction concerning the expression of opinion in multiple languages and (b) to avoid the processing of additional Speech Acts containing culture-specific politeness features such as compliments and polite refusal in Chinese [12], [17] or highly-context-based politeness in Hindi [6]. Specifically, the template-agenda of the System contains fields to be filled in with specific types of opinion related to specific types of issues or queries. If these fields fail to be filled in, the System makes two

attempts to direct the Speaker to produce an answer. In the case in which opinions may be expressed with a direct answer at a “Yes-No” question or a “Wh-Question”, the rigid nature of the Directed Dialog structure allows the formulation of a Task-Related Speech Act (“Yes-No” question or “Request” Speech Acts). It may be additionally noted that the feasibility of the processing of this type of input by S-ILTs is directly related to the content of the Directed Dialogs, constraining User-input. For example, the System may ask: “What do you think? Please say “Yes” if you agree or “No” if you disagree” or “Please express your answer choosing one of the selected words from the template”. The “Yes” or “No” answer may be replaced by equivalent language-specific expressions in languages concerned.

Non Task-Related Speech Acts may be used by the System for the extraction of opinions, especially in the case in which a Task-Related Speech Act fails, for example, the “Reminder” or even the “Offer” Non Task-Related Speech Acts.

#### **4.1 Indirect Expression of Opinion and Avoiding Expressing Opinion**

In cases of languages processed in which opinions are frequently indirectly stated with the use of specific phrases or expressions, predefined expressions and phrases expressing opinions or views may be signaled and processed as additional keywords in the template-agenda. These expressions may include language-specific word-types related to connotative features such as adjectives and adverbs, verbs containing semantic features related to descriptive features, mode, malignant/benign action or emotional/ethical gravity, as well as some modal verbs [1]. These elements may be tagged as “non-neutral” opinion markers [4] in written text processed by the Text-to-Text Machine Translation System or in the proposed S-ILTs in the Speech-to-Speech Translation System. In respect to the proposed S-ILTs, expressions and phrases expressing opinions or views may be defined at a lexical level and processed by the Simple Interlinguas (S-ILTs). For example, in a Dialog System involving the expression of views, opinions and spoken suggestions [18], User actions are categorized as “propose” (a User proposes an idea with respect to a topic), “comment” (a User comments on a proposal, or answers a question), “acknowledgement” (a User confirms someone else’s comment or explanation, e.g., “yeah,” “uh huh,” and “OK”), “requestInfo” (a User requests unknown information about a topic), “askOpinion” (a User asks someone else’s opinion about a proposal), “posOpinion” (a User expresses a positive opinion, i.e., supports a proposal) and “negOpinion” (a User expresses a negative opinion, i.e., disagrees with a proposal)[18]. These elements may be registered in the System’s template-agenda as a positive, negative or other type of opinion.

In the multilingual interaction, there are cases in which the expression of opinion is completely avoided, for instance, the Users continue in not responding to a question, respond in an unexpected way or introduce a different topic. In this case, as described above, the template indicates unaddressed issues, unanswered questions and unexpressed opinions after the interaction for subsequent evaluation.



## 5 Conclusions and Further Research

For multilingual applications concerning business meetings and other complex types of interaction beyond purely Task-Related dialogs, the System is proposed to act as a mediator to control the dialog flow. The control of the interaction via the template-agenda is shared between the System, controlling interaction, and the Users, checking the issues addressed and activating prepared utterances.

The proposed framework allows the handling of opinion in routine tasks in business meetings to a relatively high extent, excluding the case of negotiations and business deals. However, it cannot cover all cases in which opinions or emotions are expressed, nor can it replace the benefits of an in person meeting with the parties concerned, with or without the presence of an interpreter. On the other hand, this framework also allows the evaluation of the communication immediately after the interaction. It is strictly sublanguage and domain-specific and requires some type of preparation from the Users, which may re-used for the same type of meetings in various languages with minor alterations or no alterations at all. The proposed approach allows reusability for standardized types of communication within a cross-linguistic and cross-cultural framework.

Further research is required in respect to the actual implementation of the proposed design in various languages as well as in relation to possible adaptation to other types of Service- Oriented Dialog Systems in different domains.

## References

1. Alexandris, C.: English, German and the International “Semi-professional” Translator: A Morphological Approach to Implied Connotative Features. *Journal of Language and Translation* 11(2), 7–46 (2010)
2. Alexandris, C.: Speech Acts and Prosodic Modeling in Service-Oriented Dialog Systems. In: *Computer Science Research and Technology*. Nova Science Publishers, Hauppauge (2010)
3. Alexandris, C.: "Show and Tell": Using Semantically Processable Prosodic Markers for Spatial Expressions in an HCI System for Consumer Complaints. In: Jacko, J.A. (ed.) *HCI 2007*. LNCS, vol. 4552, pp. 13–22. Springer, Heidelberg (2007)
4. Alexandris, C.: User Interface Design for the Interactive Use of Online Spoken German Journalistic Texts for the International Public. In: Stephanidis, C. (ed.) *Posters, Part I, HCII 2011*. CCIS, vol. 173, pp. 551–555. Springer, Heidelberg (2011)
5. D’Souza, R., Shivakumar, G., Swathi, D., Bhattacharyya, P.: Natural Language Generation from Semantic Net like Structures with application to Hindi. In: *Proceedings of STRANS- 2001: Symposium on Translation Support Systems*, Kanpur, India (2001)
6. Kumar, R.: A Politeness Recognition Tool for Hindi, with Special Emphasis on Online Texts. In: *Proceedings of the WWW PhD Symposium*, Hyderabad, India, March 28–April 1 (2011)
7. 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*, Seattle, WA (April 2000)

8. Malagardi, I., Alexandris, C.: Verb Processing in Spoken Commands for Household Security and Appliances. In: Stephanidis, C. (ed.) UAHCI 2009, Part II. LNCS, vol. 5615, pp. 92–99. Springer, Heidelberg (2009)
9. Matsumoto, N., Ueda, H., Yamazaki, T., Murai, H.: Life with a Robot Companion: Video Analysis of 16-Days of Interaction with a Home Robot in a “Ubiquitous Home” Environment. In: Jacko, J.A. (ed.) HCI International 2009, Part II. LNCS, vol. 5611, pp. 341–350. Springer, Heidelberg (2009)
10. Moeller, S.: Quality of Telephone-Based Spoken Dialog Systems. Springer, New York (2005)
11. Nottas, M., Alexandris, C., Tsopanoglou, A., Bakamidis, S.: A Hybrid Approach to Dialog Input in the CitizenShield Dialog System for Consumer Complaints. In: Proceedings of HCI 2007, Beijing, Peoples Republic of China (2007)
12. Pan, Y.: Politeness in Chinese Face-to-Face Interaction. Advances in Discourse Processes series V. 67. Ablex Publishing Corporation, Stamford (2000)
13. 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)
14. Wieggers, K.E.: Software Requirements. Microsoft Press, Redmond (2005)
15. Williams, J.D., Witt, S.M.: A Comparison of Dialog Strategies for Call Routing. International Journal of Speech Technology 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, Lisbon (September 2005)
17. 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)
18. Zhu, H., Wei, L., Yuan, Q.: The sequential organization of gift offering and acceptance in Chinese. Journal of Pragmatics 32, 81–103 (2000)
19. The Agent-DYSL Project, <http://www.agent-dysl.eu/>
20. The HEARCOM Project, <http://hearcom.eu/main.html>
21. The ERMIS Project, <http://www.image.ntua.gr/ermis/>
22. The SOPRANO Project, <http://www.soprano-ip.org/>
23. The UNDL Foundation, The United Nations, <http://www.undlfoundation.org/undlfoundation/>