Communicative Effectiveness in Multimodal and Multilingual Dialogues

Erica Costantini
ITC-irst
Trento, Italy
costante@itc.it

Susanne Burger
Carnegie Mellon University
Pittsburgh, PA, U.S.A
sburger@cs.cmu.edu

Fabio Pianesi
ITC-irst
Trento, Italy
pianesi@itc.it

1 Introduction

Multilingual communication enabled by a multimodal speech-to-speech translation system may differ from ‘ordinary’ monolingual conversation in the conversational structure and in the way gestures are integrated in speech. We describe the second of two user studies conducted within the NESPOLE![1] project investigating these issues. NESPOLE! exploited a client-server architecture to allow an English, French or German-speaking user, while browsing through the web pages of a service provider on the Internet, to connect to an Italian-speaking human agent. Speech-to-speech translation (STST) is provided so that both speakers can use their own native languages.

2 NESPOLE! User Study2: Method

The second NESPOLE! user study (Burger et al., 2003) was designed to deeper investigate certain results of the first study (Costantini et al., 2002). Multilingual dialogues (English/ Italian, using the STST system as translation) were compared with monolingual (Italian/Italian) dialogues, using the system with and without push-to-talk mode (PTT). We devised three experimental conditions:

- **STST condition**: multilingual, PTT mode;
- **PTT condition**: monolingual, PPT mode
- **Non-PTT condition**: monolingual, free talk

We expected the multilingual condition to be different from the monolingual conditions with respect to dialogue length, spoken input, dialogue structure and speech-gesture integration patterns. The PTT mode would also play a role, resulting in differences between the two monolingual conditions.

The scenario featured a customer connecting with a human agent to find information about winter holidays. She had to choose a destination and a tourist package in compliance with a given specification, while the agent had to provide the explicitly requested information. We recorded 7 dialogues for the STST condition and 16 monolingual dialogues, half in PTT condition and half in Non-PTT condition. The interface allowed speakers to see each other, to share images and to point at portions of the image by pen-based gestures. The recorded dialogues were transcribed according to the VERBMOBIL conventions[2], and included annotations for gestures. Special annotations were added following an extended version of the Dialogue Structure Coding Scheme (DSCS) from the HCRC research group[3].

DSCS was developed for the Map Task Corpus (Carletta et al. 1997). It classifies single utterances according to their discourse goals and captures the higher-level structure in terms of *games*. Conversational *games* are associated with mutually understood conversational goals, e.g. obtaining information. *Games* consist of conversational *moves* which are different kinds of initiations and responses classified according to their purposes.

Table 1 displays the modified annotation schema; a star marks the newly added moves. The *proposal*, *disposition*, *action* and *information* moves are subclasses of the DSCD’s *information* move.

[1] Project web-site at http://nespole.itc.it

[2] http://www.is.cs.cmu.edu/trl_conventions/

[3] http://www.hcrc.ed.ac.uk/Site/

Initiation Moves	
Align	checks transfer successfulness
Check	checks confirmation
Query-yn *Query-w*	yes/no questions (yn) open questions (w)
Request	requests (former *instruct* move)
Proposal	proposal or offer
Disposition	needs or interests
Action	description of actions
Information	spontaneous information, not elicited
Response Moves	
Acknowledge	confirming
Reply-y *Reply-n* *Reply-w* *Reply-amp*	yes/no answers, answers to open questions (w), answers adding not requested information (-amp: former *clarify* move)
**Problem*	negative feedback (notification of non-successful communication)
**Other*	speaker misunderstood the question, talked about different things
Other Moves	
Preparation	expressing readiness to start
**Comment*	out of domain comments
**Noise*	turns without linguistic content

Table 1. Move Annotation Schema

3 Results

The results for all three conditions reported in Table 2 show that the dialogues in STST condition lasted longer, but had an even lower percentage of actual dialogue contributions. 87% of the time was taken by the STST system's delays, transfer, translation and PTT mode (the PTT condition still shows 30% of non-speech part compared with the Non-PTT condition). The STST condition is also characterized by more repetition turns. Analyzing the involved moves ascribes these repetitions to meta-communicative concepts supposed to resolve misunderstandings. The system failed to translate these. Furthermore, the STST dialogues show: shorter dialogue games, fewer nested games; more questions, more replies, less spontaneously provided, non-elicited information and fewer *acknowledgment* moves. In STST dialogues the speakers focused on 'essential' information, reduced the dialogue complexity and tried to adhere to a question/answer pattern. The number of gestures was similar in all conditions, but in the STST condition, gestures were performed before and more frequently after talking. This suggests that the speech-gesture integration can be lost as soon as the interaction becomes more complex, when more tasks such as PTT, translation and drawing must be handled in parallel. The results for the PPT condition were usually intermediate between those of the Non-PTT condition and those of the STST condition, proving that PTT has an additional effect on STST condition.

Measures	**STST**	**PTT**	***NonPTT***
Dialogue length (min)	23	9.85	8.87
% non-speech partition	87%	49%	19%
% repetition turns	24%	6%	1.3%
Moves per game	4.6	4.6	5.6
% of nested games	10%	26%	23%
% of questions	35%	23%	14%
% of replies	24%	21%	16%
% of *information*	8%	12%	15%
% of *acknowledge*	11%	17%	33%
Gestures during speech	14%	61%	96%

Table 2. Results for all three conditions

4 Conclusions

The results show the existence of adaptive communication strategies to the different contexts of communication. Using Dialogue Structure Analysis seems to be a sufficient method of discovering, understanding and clarifying the phenomena. The revealed communicational structures should be of great interest to the STST research community, both, for evaluation of dialogue effectiveness, but also for the design of appropriate scenarios and choice of training materials covering the linguistic phenomena which are expected to be found during the interaction with an actual translation system.

References

Erica Costantini, Fabio Pianesi & Susanne Burger. 2002. The Added Value of Multimodality in the NESPOLE! Speech-to-Speech Translation System: an Experimental Study. In *Proceedings of ICMI'02*, Pittsburgh, PA.

Jean Carletta et al. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23 (1) 13-21.

Susanne Burger, Erica Costantini & Fabio Pianesi. 2003. Communicative Strategies and Patterns of Multimodal Integration in a Speech to Speech Translation System. To appear in *Proceedings of MT-summit*, New Orleans, LA.