Gesture, Prosody and Lexicon in Task-oriented Dialogues: Progress Report

Ewa Jarmołowicz, Maciej Karpiński, Zofia Malisz, Michał Szczyszek (Center for Speech and Language Processing AMU)

An increasing number of attempts is made to encompass more extra- and paralinguistic components in the analysis of interpersonal communication [Massaro 1987, Antas 2001, Gibbon *et al*. 2000]. In many situations, the linguistic form and content of utterances fail to fully account for the meaning and results of conversation [Mahrebian 1972].

The aim of the *DiaGest* project is to track down, describe and analyze the relations among the gestural, lexical and prosodic components of dialogue. Our pilot studies have proven that in dialogues subjects tend to use more numerous and complex gestures than in monologue tasks. Dialogue tasks are based on the prototypical situation of speech and gesture use, and deliver rich, spontaneous, but also relatively structured, verbal and non-verbal material.

The project involves three main stages: dialogue task design and recording, data labeling and data analysis. In the first stage, a number of dialogue tasks were designed and tested. Three of them were finally selected for further recordings: "spatial arrangement", "origami" and "cartoon". In the first two tasks, strong collaboration between the participants was necessary. One of them played the role of instructor, while the other realized a manual task (re-constructing a spatial arrangement of (a) a set of objects and (b) an origami-like figure). In the third task, partially based on [McNeill 1992], our subjects were asked to answer questions about a cartoon they had been presented. In the currently realized stage of data labeling, the recordings are being transcribed orthographically and tagged on a number of levels. The lexical content, syntactic forms, intonational realizations, rhythmic structures as well as the gestural component are labeled in selected portions of the material. Dialogue contributions are tagged as realizations of dialogue acts. The categorization of dialogue acts is partially based on the Pol'n'Asia [Juszczyk *et al*. 2006] system and extended to include the gestural component. In the process of multimedia data labeling, ELAN [Hellwig, Uytvanck 2004] video data tagging software is used. A configuration allowing for simultaneous viewing and labeling of two video files has been prepared for this purpose. It includes a multi-layered structure for all the annotation levels. The system for gesture labeling is based on a couple of existing systems, including CoGest [Gut 2003] and Mumin [Allwood 2005]. As next steps, the analysis of morphological, syntactic and prosodic phenomena will be conducted. Praat [Boersma, Wenink 2006] will be used as a software tool for intonation and rhythm analysis.

The results of the project will be of twofold value. The knowledge on the interaction among various components of dialogue flow will be substantially extended. Simultaneously, we expect that to find applications in speech technology, including the design and localization of avatars within the Polish and Polish-English Literacy Tutor project [Dziubalska-Kołaczyk, *et al*. 2006].

References

Allwood, J., Cerrato, L., Dybkjaer, L., Jokinen, K., Navaretta, C., Paggio, P. 2005. The MUMIN Multimodal Coding Scheme. NorFA Yearbook.

Antas, J. 2001. Co mówią ręce. Wprowadzenie do komunikacji niewerbalnej. [In:] Przybylska, R., Przyczyna, W. (Eds.) *Retoryka dziś. Teoria i praktyka*. Kraków.

Boersma, P., Wenink, D. 2006. Praat. Doing Phonetics by Computer (a computer program; version 4.4 and later).

Dziubalska-Kołaczyk, K., Bogacka, A., Wypych, M., Pietrala, D., Krynicki, G. 2006. "PELT: An English language tutorial system for Polish speakers" [In:] *Proceedings of the MultiLing2006*.

Gibbon, D., Mertins, I., Moore, R. K. (Eds.) 2000. *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. Kluwer Academic Publishers, Dordrecht.

Gut, U., Looks, K., Thies, A., Gibbon, D.: CoGesT 2003. Conversational Gesture Transcription System. Version 1.0. Technical report. Bielefeld University.

Hellwig, B., Uytvanck, D. 2004. EUDICO Linguistic Annotator: ELAN, Version 3.0 Manual software manual.

Juszczyk, K., Karpiński, M., Kleśta, J., Szalkowska, E. (abstract submitted to PLM 2007) Categorization and Labelling of Dialogue Acts in Pol'n'Asia Project.

Massaro, D.W. 1987. *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Lawrence Earlbaum Assoc.

McNeill, D. 1992. *Hand and mind*. The University of Chicago Press.

Mehrabian, A. 1972. *Nonverbal communication*. Aldine-Atherton, Chicago, Illinois.