

Categorization and Labelling of Dialogue Acts in Pol'n'Asia Project

Konrad Juszczyk, Maciej Karpiński, Janusz Kleśta, Emilia Szalkowska, Marcin Włodarczak (Adam Mickiewicz University, Poznań)

There exists an impressive number of approaches to defining and categorizing dialogue units. Many of them are anchored in the tradition of Speech Act Theory and related to the intentional level of analysis [Austin 1962; Searle 1969, 1976; Boella *et al.*]. The number and variety of available systems has resulted in attempts to propose commonly acceptable and applicable standards [e.g., Core, Allen 1999; Klein 1999; Bunt, Girard 2005]. However, to meet the demands of specific corpora and research aims, new, specialized systems are still developed. The aim of our presentation is to show the process of design and application of a dialogue act labelling system in *Pol'n'Asia* corpus.

Pol'n'Asia project is focused on comparative studies of dialogue intonation. Two intonational (Korean and Polish) and two tonal languages (Thai and Vietnamese) are investigated. In order to compare the prosodic realization of various categories of utterances, the material has been annotated according to a new system of dialogue acts, based on the multidimensional approach [Bunt 1996, 2006; Bunt, Girard 2005].

The following factors have influenced the design of the *Pol'n'Asia* dialogue act system: a) multilingual material coming from regions of vividly varying cultures; b) structural and culture-related divergency of the languages; c) the type of the dialogue task (a relatively complex map-task) resulting in the use of specific dialogue strategies, vocabulary and grammatical structures; d) twofold application of the corpus: research (comparative studies on intonation and structure of task-oriented dialogues) and education (teaching/learning intonation and tone in foreign languages).

In the *Pol'n'Asia* system, four "aspects" or "dimensions" of dialogue acts are considered: *External Action Control (EAC)*; related to extralinguistic, mostly task-related actions), *Information Transfer (IT)*; related to the transfer of information between the interlocutors), *Dialogue Flow Control (DFC)*; related to the control of the flow of conversation) and *Attitudinal Content (AC)*; related to the emotional and attitudinal aspect of utterances). An additional *Modality Index* is also introduced in order to describe the surface-grammatical modality of each utterance. This is mostly meant for the learners who will look for cues about possible intonational realizations of certain typical utterances. In each dimension, each act takes one value from a predefined list. As a result, each dialogue act can be represented as a sequence of five values (*EAC, DFC, IT, AC, MI*).

At the stage of practical application, dialogue act systems often show their weak points even more vividly [Traum, Hinkelmann 1996; Traum 2000]. In order to solve these problems, we provide detailed definitions and procedures for determining the categories of dialogue acts and the ranges of their realizations. Nevertheless, we stress that one should expect and accept an amount of subjectivity and a degree of uncertainty in the categorization of the intentional level units on the basis of linguistic utterances' analysis.

References

- Austin, J. 1962. *How to Do Things with Words*, Oxford.
- Boella, G., Damiano, R., Lesmo, L., Ardissono, L. 1999. Conversational Cooperation: The Leading Role of Intentions. *Proceedings of Amstelogue '99*, Amsterdam.
- Bunt, H. 1996. Dynamic interpretation and dialogue theory. [In:] M. M. Taylor, F. Neel, D. G. Bouwhuis [Eds.] *The Structure of Multimodal Dialogue*, Volume 2, John Benjamins, pp. 139-166.
- Bunt, H. 2006. Dimensions in dialogue act annotation. *Proceedings Fifth International Conference on Language Resources and Evaluation (LREC 2006)*.
- Bunt, H., Girard, Y. 2005. Designing an Open, Multidimensional Dialogue Act Taxonomy. *The Proceedings of DIALOR'2005 Workshop*, Nancy.

- Core, M. G., Allen, J. F. 1997. Coding Dialogues with the DAMSL Annotation Scheme. [In:] D. Traum (red.) *Working Notes of the Fall Symposium in Communicative Action in Humans and Machines*. Menlo Park, CA: AAAI, pp. 28 – 35.
- Klein, M. 1999. Standardisation Efforts on the Level of Dialogue Act in the MATE Project [In:] *Towards Standards and Tools for Discourse Tagging: Proceedings of the Workshop*.
- Searle, J. R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: CUP.
- Searle, J. R. 1976. The classification of illocutionary acts. *Language in Society*, 5, pp. 1 – 24.
- Traum, D. R. 2000. 20 Questions for Dialogue Act Taxonomies. *Journal of Semantics* 17(1), pp. 7 – 30.
- Traum, D. R., Heeman, P. A. 1996. Utterance Units in Spoken Dialogue. *ECAI Workshop on Dialogue Processing in Spoken Language Systems*, Budapest, pp. 84 – 91.