

## **TITLE: A COLLOCATION ANALYSIS OF TOPIC CHANGE UTTERANCES IN MULTI-PARTY MEETING CONVERSATIONS**

**Presentation details:** Poster

**Author name:** Shunichi Ishihara

**Institutional affiliation:** Faculty of Asian Studies, College of Asian and the Pacific, the Australian National University

**Abstract:** The aim of this paper is to identify what sorts of expressions, i.e. formulaic phrases and multi-word expressions, are significantly associated with topic boundaries in multi-party meeting conversations. Previous work on automatic topic segmentation and topic changes in discourse analysis has found that cue phrases, such as *okay*, *anyway*, and *alright*, provide valuable information in identifying topic boundaries (Grosz and Sidner, 1986; Howe, 1991; Passonneau and Litman, 1997). In this study, both grammatical and lexical single words that have statistical significance in relation to topic boundaries are extracted to investigate their co-occurrence network. The International Computer Science Institute (ICSI) Meeting Recorder Dialogue ACT (MRDA) corpus (Shriberg et al., 2004) is used in this study. The MRDA corpus is a hand-annotated version of the ICSI Meeting corpus (Morgan et al., 2001) which consists of 75 naturally occurring multi-party meetings, each approximately one hour in length. Altogether 53 different speakers appear in the corpus, with an average of approximately six speakers per meeting. The annotation of the MRDA corpus provides three types of information: marking of dialogue act (DA), marking of DA segment boundaries, and marking of correspondences between DAs (= adjacency pairs). In the ICSI Meeting corpus, a stream of dialogue is segmented in terms of *utterances*, each of which constitutes prosodically one unit. One of the DA tags used in the MRDA corpus is *topic change* (*tc*), which indicates the locations where a topic change occurs. There are 679 *tc* locations in the MRDA corpus.

First of all, in this study, those single words (one-word phrases, etc) that are significantly correlated with topic boundaries are selected. Then, secondly, the collocations of these single words which are significantly correlated with topic boundaries are identified. *Yates' chi-square* test is employed for the identification of these single words and the co-occurrences of them. The network of the identified collocated words is graphically presented in this study in order to illustrate what types of expressions are significantly associated with topic changes.

We will demonstrate in this study that the derived network of collocated words supports previous studies on topic changes in the area of discourse analysis in some points, whereas we will also report that some findings made in discourse analysis are not confirmed from the empirical results of our study (perhaps due to the unique nature of multi-party meeting conversations).

As future research, we plan to investigate if there are any expressions which are specifically correlated with the opening or ending of a topic. It will also be interesting to see how collocational information, as opposed to cue words, improve the performance of automatic topic segmentation systems.

### **List of references:**

- Grosz, B. and Sidner, C. (1986) Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3), 175-204.
- Howe, M. (1991) *Topic Changes in Conversation*. PhD. thesis, University of Kansas.
- Morgan, N., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Janin, A., Pfau, T., Shriberg, E. and Stolcke, A. (2001) The meeting project at ICSI. *Proceedings of the 1<sup>st</sup> International Conference on Human Language Technology Research*, 1-7.
- Passonneau, R. and Litman, D. (1997) Discourse segmentation by human and automated means. *Computational linguistics*, 23(1), 103-139.
- Shriberg, R., Dhillon, S., Bhagat, S., Ang, J. and Carvey, H. (2004) The ICSI meeting recorder dialog act (MRDA) corpus. *Proceedings of 5<sup>th</sup> SIGdial Workshop on Discourse and Dialogue*, 97-100.

**Keywords:** collocation, topic change, meeting, multi-party conversation

**Contact details:** Shunichi Ishihara, Faculty of Asian Studies, College of Asian and the Pacific, the Australian National University, Canberra, ACT, 0200; (02) 6125 4656; (02) 6125 0745; shunichi.ishihara@anu.edu.au