AN ANNOTATION SCHEME OF SPOKEN DIALOGUES WITH TOPIC BREAK INDEXES

Yoichi Yamashita *Michiyo Murai*

Dep. of Computer Science, Ritsumeikan University
1-1-1, Noji-Higashi, Kusatsu-shi, Shiga, 525-8577 Japan
{yama,michiyo}@slp.cs.ritsumei.ac.jp

ABSTRACT

This paper proposes a scheme of annotating spoken dialogues with discourse level information in terms of the discourse segment. Dialogues are coded with topic break index (TBI), which indicates the degree of topic break between the discourse segments, instead of marking a beginning and an ending utterances of the segment. TBI is graded by two levels, 1 and 2, and TBI=2 indicates a large change of the topic. Two methods are tried for assigning a TBI value for segment boundaries. In the method-I, the coder directly describes TBI according to the difference of contents between the adjacent segments. In the method-II, the coder classifies relative change of the topic break between the adjacent segments into three categories. Then, the relative changes are automatically converted into TBIs by extraction of local maximum change of the topic break. Two annotation methods are evaluated with the agreement score and the relation to prosodic parameters.

1. INTRODUCTION

Many spoken dialogue systems have been studied in recent years. Dialogue level knowledge is necessary to develop dialogue-oriented applications. Dialogue corpora annotated with several types of discourse information as well as syntactic information are indispensable to understanding and modeling characteristics of the dialogue. The success of many corpus-based approaches for spoken language processing is increasing the importance of developing corpora. Several activities have started in order to standardize annotated information and to construct dialogue corpus[3, 7, 6].

A dialogue is composed of several discourse segments in which dialogue participants talk about a topic[4, 5]. The structure of the discourse segment is important both to understand and to generate dialogues. This paper describes methods for annotating spoken dialogues at the level of discourse structure in terms of the discourse segment(DS). Annotation results are evaluated based both on the agreement score between multiple coders and on relation to prosodic parameters in utterances.

2. ANNOTATION OF DISCOURSE SEGMENT BOUNDARIES

We can often find clear structure of DS with the nesting in monologues. A monologue has well-organized structure because it is generated by one speaker. On the other hand, a dialogue is generated by interaction of two speakers and topic is not controlled by one speaker. Such a process results in complicated structure of the DS in many dialogues. In the previous discourse model[4], a DS

- has a beginning and an ending utterances, and
- may have smaller DSs in it.

In a preliminary experiment of coding DSs for spoken dialogues, there were a lot of disagreements on the granularity or the relation of the DS. It was very difficult to identify an ending utterances of the DS and the nesting structure, especially for utterances having both initiate and response functions. The utterance 11 in Figure 1 is a such example, which responds to a question of the utterance 7 and makes a new requirement.

This paper proposes an alternative scheme of coding the discourse structure in order to avoid difficulties in the previous scheme and describe discourse level information about the DS. A new proposed scheme annotates spoken dialogues with boundary marking of the DS, instead of identifying a beginning and an ending utterance of each DS. A DS boundary is coded with topic break index (TBI) which indicates the degree of topic break between adjacent segments. This scheme does not directly describe nesting structure of the DS, but provides structural information of the DS with relational information of adjacent segments in terms of TBI. Figure 1 is a sample dialogue annotated with TBI. In this figure, a DS tag is denoted as

[TBI : *topic_name* : *segment_relation*].

TBI is graded by two levels, 1 and 2, and TBI=2 indicates a large change of the topic. The *topic_name* is a name of the topic which starts after the boundary. The *segment_relation* is optional, and it describes relational information between two segments preceding and following the boundary. The *segment_relation* can optionaly include one of the three attributes, clarification, interruption, and return, in the current annotation scheme.

[2:greeting:]
1 L: '*Kochira chiri-annai shisutemu desu.*' <I>
(This is the route guidance system.)
[2:destination:]
2 R: '*Osaka hoteru ni ikitaino desu ga...*' <I>
(I want to go to the Osaka hotel...)
3 L: '*Hai.*' <R>
(Yes.)
[1:place at present:]
4 R: '*Ima Osaka kuukou ni irun desu yo.*' <I>
(I'm at the Osaka airport right now.)
5 L: '*Hai.*' <R>
(Yes.)
[1:routing:]
6 R: '*Dou yattara ikerun de shou ka?*' <I>
(How can I go to the Osaka hotel?)
[1:choice of routing:clarification]
7 L: '*Basu wo tsukau houhou to densha wo tsukau houhou ga arimasu keredomo*' <I>
(Which do you want to go by buss or train?)
[1:faster way:clarification]
8 R: '*Hayai houhou wo shiritaino desuga.*' <I>
(I want to know a faster way.)
9 L: '*Hayai houhou desu to densha ni nari masu.*' <R>
(The faster way is by train.)
10 R: '*Hai.*' <F>
(Yes.)
[2:Route by train:return]
11 R: '*Soshitara, densha no ikikata wo oshiete morae masu ka.*' <R&I>
(Then, please tell me a way by train.)
[1:Subway:]
12 L: '*Hommachi made chikatetsu de itte, ...*' <R>
(Please take a subway to Hommachi station, ...)
...

Figure 1. A sample dialogue annotated with TBI.

3. CODING METHODS

3.1. Detection of Segment Boundaries

The DS boundaries are automatically determined by identification of exchange structure in the dialogue. Each utterance unit is classified into four major types of the utterance, initiate, response, follow-up, and response with initiate, based on a decision tree approach[6, 2]. An exchange is a sequence of the initiate, the response with initiate, the response, and the follow-up utterances. The response with initiate and the follow-up are repeatable and optional in an utterance sequence. The proposed scheme uses an exchange as a building block of discourse segments regarding that an initiating utterance always starts a discourse segment. The DS boundaries are always placed before the initiate and the response with initiate utterances, and they are annotated with the DS tag. In Figure 1, the symbols of <I>, <R>, <F>, and <R&I> at the end of the utterance indicate the utterance type of the initiate, the response, the follow-up, and the response with initiate, respectively. The DS tags are placed before the utterances 1, 2, 4, 6, 7, 8, 11, and 12.

3.2. Assignment of TBI

TBI is described for the DS boundaries after detection of the DS boundaries. We tried two methods of assigning a TBI value for the DS boundary.

3.2.1. Method-I

In the first method (Method-I), the coders directly assign a TBI value for a DS boundary according to degree of topic break between the preceding and the following segment of the DS boundary. The threshold for differentiating TBI=1 and =2 relies on the coders' criteria for measuring the degree of topic break.

3.2.2. Method-II

In the second method (Method-II), the coders describe relative change of the topic break without directly assigning a TBI value. The coders judge whether degree of the topic break of a DS boundary increases, decreases, or equals to that of the preceding DS boundary, and they marks the relative change of the topic break with a symbol, `+`, `-`, or `=`, respectively. Then, the relative changes of the topic break are automatically converted into TBIs by extraction of local maximum change of the topic break. After the '`=`' is replaced by the same symbol '`+`' or '`-`' of the preceding boundary, the TBI is set to 2 for the last '`+`' boundary in a consecutive '`+`' sequence, and TBI=1 for the others. In preliminary experiments based on the method-I, the coders tended to match topic structure of the dialogue with his/her own knowledge on the dialogue domain. The difference of understanding of the dialogue domain caused disagreements of the tags. The method-II is designed for the coders to use the local context of the utterance and to judge the change of topic break by paying attention to only two boundaries.

Figure 2 shows examples of relative change markings of the topic break, which are indicated as [[ ]], and converted DS tags with TBI, which are at right-hand of →. The '`=`' before the utterance 12 is replaced by the '`+`' because its preceding symbol of change is '`+`', which marks the boundary before the utterance 9. The last '`+`' in a consecutive '`+`' sequence, that includes symbols before the utterance 9 and 12, is converted into TBI=2, and the other are converted into TBI=1.

4. EVALUATION

4.1. Coding Experiments

To evaluate the coding methods, 9 and 9 coders annotated 3 spoken dialogues based on the method-I and -II, respectively. They coded the DS tags for transcriptions of the dialogue without listening the speech. The size of the dialogue and the number of DS boundary in the dialogue are listed in Table 1. The dialogue tasks of 99kyo01, 99osa02, and 99atr01, are the meeting room

[[-]] → [1::]
3 L: '*Dono kaigishitsu wo shiyou shimashou ka?*' <I>
(What room do you use?)
[[-]] → [1::]
4 R: '*Mazu kaigishitsu no shuuyou ninzuu wo oshiete kudasai.*' <I>
(Could you tell me the capacity of rooms?)
[[-]] → [1::]
5 L: '*Dono kaigishitsu no shuuyou ninzuu de shou ka?*' <I>
(What room?)
6 R: '*Ichiou zenbu onegai shimasu.*' <R>
(All rooms, please.)
7 L: '*Shou-kaigishitsu ga 8 nin, ... koushuu-shitsu ga 25nin to natte imasu.*' <R>
(8 persons for Small Meeting Room, ... 25 persons for Lecture Room.)
8 R: '*Hai.*' <F>
(Yes.)
[[+]] → [1::]
9 R: '*Eto shiyou ryou wa dou natte masu ka?*' <I>
(Well, how about charge?)
10 L: '*Shou-kaigishitsu ga 1 jikan 3,000 en ... to natte ori masu.*' <R>
(3,000 yen for Small Meeting Room, and ...)
11 R: '*Hai.*' <F>
(Yes.)
[[=]] → [2::]
12 R: '*12 gatsu 19 nichi no 14 ji kara kaigi wo shitain desu ga.*' <I>
(I want to have a meeting from 14:00 on December 19.)
[[-]] → [1::]
13 R: '*Dono kaigishitsu ga aite imasu ka?*' <I>
(What rooms may I use?)
14 L: '*14 ji kara aite iru nowa ...*' <R>
(Rooms available from 14:00 are ...)
...

Figure 2. The relative changes of topic break and converted TBIs

Table 1. Dialogue data

ID	number of utterance	number of DS boundary
99kyo01	37	16
99osa02	221	66
99atr01	43	18
total	301	100

arrangement, the travel consultation, and the hotel reservation, respectively. The number and the place of the DS boundary are the same for all coders because the type of each utterance unit was pre-determined.

4.2. Agreement

The first evaluation criterion is the agreement score. The agreement of TBI among multiple coders is evaluated with reliability in terms of the kappa coefficient K[8, 1]. The *topic_name* and the *segment_relation* in the DS tag are ignored in the evaluation. The results are shown in Table 2. The Kappa coefficients for the TBI annotation are 0.4 for both methods. The two methods showed almost the same performance as for the agreement. The kappa coefficients are not so high, and an improved scheme or a new coding method are necessary to obtain the consistent DS annotation.

Table 2. Tag agreement in the kappa coefficient.

dialogue ID	method-I	method-II
99kyo01	0.452	0.420
99osa02	0.403	0.334
99atr01	0.338	0.452
average	0.398	0.402

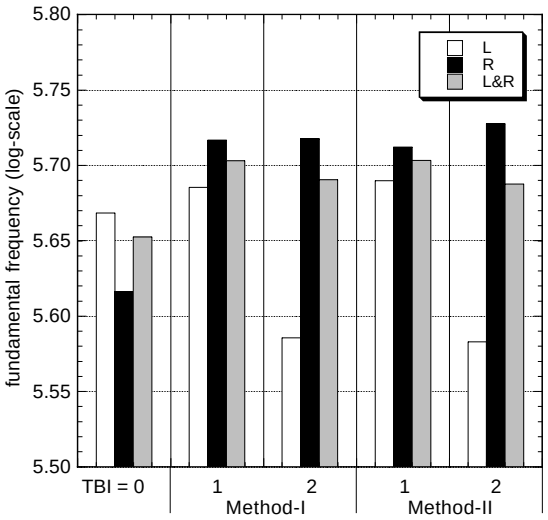

Figure 3. Average F0 for TBIs

4.3. Relation to prosodic parameters

Discourse annotation should be evaluated by viewpoints of applications as well as agreement between multiple coders. Prosodic parameters are important features for developing spoken dialogue systems including speech synthesis and speech recognition. In this paper, relation between TBI and prosodic parameters, the fundamental frequency(F0) and the power, is investigated for two coding methods.

The typical TBI is extracted for each DS boundary from tagging results by 9 coders based on the majority decision. All utterances are classified into three categories, the utterances following the DS boundary with TBI=2, the utterances following the DS boundary with TBI=1, and the other utterances which are the response or the follow-up. The F0 and the power of the utterance are obtained in the log scale by averaging for the begging 0.5[s] of the utterance. The average of the parameter for each utterance category is calculated after the speaker normalization.

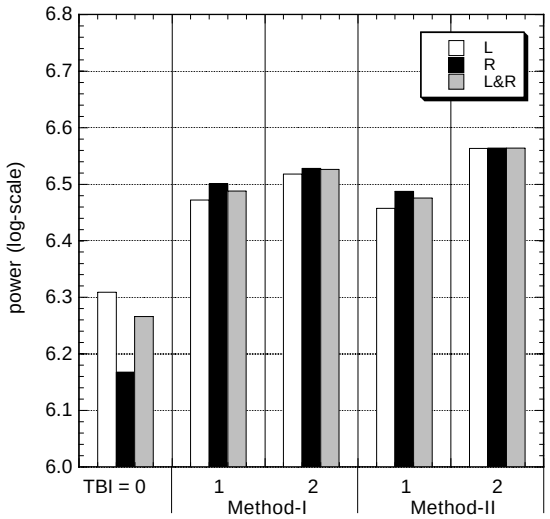

Figure 4. Average power for TBIs

Table 3. The number of the utterance for TBIs

(a) Method-I

TBI	0		1		2	
speaker	L	R	L	R	L	R
99kyo01	11	10	3	7	2	4
99osa02	116	39	18	27	1	20
99atr01	13	12	8	3	4	3
total	140	61	29	37	7	27

(b) Method-II

TBI	0		1		2	
speaker	L	R	L	R	L	R
99kyo01	11	10	3	9	2	2
99osa02	116	39	18	30	1	17
99atr01	13	12	7	4	5	2
total	140	61	28	43	8	21

Figure 3 and 4 shows the average parameters for each utterance category. TBI=0 means the utterance category of no DS boundaries and it is independent of the coding methods. The parameters for the speaker cluster are also shown. The speaker `L` and `R` played a role of a consultant (system) and its user, respectively, in all tasks. The `L&R` indicates the average of both utterances. The number of the utterance for each utterance category is shown in Table 3. Both prosodic parameters increase in the utterances with the DS boundary (TBI$\neq$0), the power also increases more in TBI=2 than TBI=1. In the comparison of the coding methods, there are clear differences between TBI=1 and =2 for the method-II. TBIs based on the method-II showed high correlation to the prosodic parameters, especially to the power.

The F0 frequency of the speaker `L` is very low for TBI=2, shown in 3. The utterances of the speaker `L` include back-channels, such as *'hai'*(Yes), and they are coded with TBI=0. There are a few `L`'s utterances with TBI=2, many of which begin with a discourse marker, such as *'dewa'*(then). These characteristics decrease F0 for the `L`'s utterances with TBI=2. For the `R`'s utterances, F0 is higher for TBI=2 and the difference between TBI=2 and =1 is larger in the method-II.

5. CONCLUSIONS

Difficulties in annotating the discourse segments with TBI are namely divided into two kinds of factors. The first one is a criterion for coders to measuring the topic break, that is a threshold between TBI=1 and =2. The method-II intended to resolve this difficulty by introducing relative judgment for topic change. The agreement scores by two methods are almost same, but the method-II gave higher relationship of TBI to the prosodic parameters, especially the power.

The second factor is information which a coder pays attention to for measuring the topic break, such as cue words, overlap of the words, and so on. Although it is not considered in this paper, it is important to give the coders a common guideline for judging the difference of topic in order to expect different coders to annotate the consistent tags for describing discourse segments.

REFERENCES

[1] Carletta, J. : Assessing Agreement on Clasification Tasks: The Kappa Statistic, *Computational Linguistics*, Vol.22, No.2, pp.249–254 (1996).

[2] Carletta, J., Isard, A., Isard, S., Kowkto, J.C., Doherty-Sneddon, G., and Anderson, A.H. : The Reliability of A Dialogue Structure Coding Scheme, *Computational Linguistics*, Vol.23, No.6, pp.13-31 (1997).

[3] http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html

[4] Grosz, B. J. and Sidner, C. L. : Attention, Intentions, and the Structure of Discourse, *Computational Linguistics*, Vol.12, No.3, pp.175–204 (1986).

[5] Hirschberg, J. and Grosz, B. J. : Intonational Features of Local and Global Discourse Structure, *Proc. of the DARPA Workshop on Speech and Natural Language*, pp.441–446, Morgan Kaufmann (1992).

[6] Ichikawa, A., et al. : Evaluation of Annotation Schemes for Japanese Discourse, Proc. of ACL '99 Workshop on Towards Standards and Tools for Discourse Tagging, pp.26–34 (1999).

[7] Nakatani, C. H., Grosz, B. J., Ahn, D. D., and Hirschberg, J. : Instructions for Annotating Discourse, Technical Report TR–21–95, Center for Research in Computing Thechnology, Harvard University (1995).

[8] Siegel, S. and Castellan, N. J., Jr. : Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, second edition (1988).