

Committee-based Discourse Purpose Assignment: Discourse Structure Annotations of Spontaneous Japanese Monologue

Kazuhiro Takeuchi[†], Katsuya Takanashi[†], Ikuyo Morimoto[†],
Hanae Koiso[‡] and Hitoshi Isahara[†]

[†]Communications Research Laboratory
[‡]The National Institute for Japanese Language

ABSTRACT

Identifying the regularity of middle level discourse is useful for annotating discourse structure in long monologues. A convincing way to find segments is described on the middle level of discourse structure (sub-story), based on a committee-based decision of discourse purposes and a detection of transitive expressions.

1. PROJECT OVERVIEW

Numerous well-established projects have set the standard for future corpus-based research through the use of large corpora. In so doing, they have revolutionized the study of language.

This paper describes an attempt at discourse structure annotation in constructing a spontaneous corpus of Japanese (hereafter CSJ) [3] as a part of project 'Spontaneous Speech: Corpus and Processing Technology', which is supported by a grant from the Science and Technology Agency. The CSJ primarily consists of monologues with a total speech length of 800 hours (roughly 7M words). One-tenth of the utterances (the "Core") will be manually given orthographic and phonetic transcriptions.

The project will be conducted over a 5-year period in pursuit of the following three major goals.

Goal 1: Manual annotation for the core of the CSJ using the morphological information to construct an analysis program [9]. All or some part of the core will also be tagged with para-linguistic information including intonation.

Goal 2: Acoustic and linguistic modeling for spontaneous speech understanding and summarization using linguistic as well as para-linguistic information in speech.

Goal 3: Construction of a prototype for a spontaneous speech summarization system.

The monologues in CSJ are divided into two types: academic presentations and simulated public speeches. The academic presentations are live recorded data of researchers' presentations in various academic meetings. The simulated public speeches are short monologues (mostly 10 to 15 minutes long) that were recorded specifically for our corpus

by paid non-professional speakers. They were instructed to prepare an outline of their talk instead of a completely pre-fixed text.

We discuss in this paper an annotation scheme for discourse structure, which should be a model for spontaneous monologue understanding and summarization.

2. ANNOTATION OF DISCOURSE STRUCTURE

2.1 Instruction for Annotating Discourse

Some works in the projects of corpus construction have discourse structure that is manually annotated. Nakatani and her colleagues [6] propose instruction guides (Instruction for Annotating Discourse (IAD)) to annotate such corpora. Their instructions are based on the discourse structure model that was proposed by Grosz and Sidner [1] (hereafter the GS-model).

The basic annotation process complying with the instruction of the IAD consists of two major identification: segmenting discourse and assigning the purpose of each segment. According to the IAD, the annotation task is described intuitively:

The annotation task is similar to making an outline of a discourse in that you will provide a hierarchical organization of the purposes in the discourse. The purpose of a discourse segment is somewhat like a topic in an outline.

In this paper, we discuss the problems and the extensions of applying their instruction to annotating our corpus.

2.2 Segment and its Purpose

Because a debate that goes on for a day is clearly divided into sub-discourses, to analyze one of them, we must first analyze how the discourse is constructed: is it made up of one discourse, a compound discourse, or a sequence of small discourses?

We refer to such a sub-discourse as a 'segment' in this paper. In the GS-model, discourse structure shows how such segments constitute a discourse. The purpose of a segment (hereafter, we simply refer to it as the 'purpose') is like the title of the segment. More precisely, it is defined as

the reason a speaker conveys propositions in an uttered segment.

Figure 1 shows an example, which is annotated in its discourse structure following the IAD. In the figure, a segment boundary is shown by inserting a line that begins from ‘WHY?’, and the rest of the ‘WHY?’ line shows the purpose of the segment. Each indentation shows which level the sentence belongs to within the hierarchal structure.

For instance, the segment whose purpose is ‘Explain how to make shrimp mousse’ governs two segments, whose purposes are ‘Describe peeling’ and ‘Tell how to find vein by cutting.’ While the segment ‘Describe peeling’ is directly under the segment ‘Explain how to make shrimp mousse’, the segment ‘Describe peeling’ is directly under the segment ‘Tell how to find vein by cutting.’

All purposes in such a discourse structure will capture not only the content of the discourse, but also the speaker’s motivation (intention) about how and why s/he tells the listener the information contained in it.

<p><i>WHY? Explain how to make shrimp mousse</i> And what you want to do is you want to take the shrimp, okay And you want to peel and devein them <i>WHY? Describe peeling</i> Okay, what you do is you peel the outer shell off <i>WHY? Tell how to find vein by cutting</i> Okay, and then you hold the shrimp and You run a knife down the outside, It’s like the back of the shrimp, okay, :</p>
--

Figure 1: Annotated Discourse Structure

2.3 Annotation Procedure

As a preprocessing of an annotation, monologues are manually transcribed and divided into sentences. In transcribing, not only the words that the speaker uttered but also some the inarticulate sounds (pauses, slight slips of the tongue, etc.) are transcribed into text. The sentences that we divide in advance are minimal units of discourse [7].

In the first step, an annotator is required to listen to the whole discourse. Then, the annotator divides the discourse into segments and assigns a purpose to each segment. Finally, s/he has to check over the annotation from the beginning to the end.

We developed an on-line marking tool, which provides two main advantages for annotations. One is that the annotator may examine any part of pair of the discourse transcription and its sound at any time. The other is that s/he can easily know which part s/he is annotating in the overall structure because the tool displays the discourse in a tree structure.

3. PROBLEMS IN ANNOTATION OF CSJ MONOLOGUES

3.1 Annotation on a Trial Basis

We examined trial annotations in which three annotators analyzed the discourse structure of our six monologues using the IAD. We found two major problems in the trial: it takes longer than we expected to complete the annotation, and different annotations are sometimes put into a segment.

The IAD is an annotation scheme based on the GS-model, which mainly concerns the intention exchanges in the dialogues. While the speaker’s intentions are recognized more easily by the references to a response of the listeners in a dialogue, such responses are rarely observed in the monologue by the annotators. Suppose that the overall purpose of a certain monologue is to explain the speaker’s experience. We can scarcely decide how to decompose the explanation into its sub-purposes because the exchanges of sub-purposes between the speaker and listener are not directly displayed in the transcribed information. Thus, the notion of purposes in the monologues is more abstract than that in the dialogues. It must be one of the causes that makes the monologue annotation a difficult task. The aforementioned problems motivated us to try to extend the IAD.

Although we had a number of problems in analyzing the monologues, we found some particular parts of the annotations were relatively stable among the annotators. We characterized such parts into two types: cohesive patterns and sub-stories.

3.2 Cohesive Patterns

The cohesive patterns are referred to as a few sentences that have a particular local relationship within them. Those patterns were featured by the stable annotations of both segment boundaries and their purposes. The small trial annotations showed us only a few stable local patterns such as, exemplifications, listings, etc. We should be able to improve the stability of annotations for such local patterns by listing more and by characterizing such kinds of patterns in terms of sentence types.

As van Dijk (1997) pointed out, there are various structures on which the representation of discourse should focus. He proposed that one of ways to analyze discourse in these structures starts by considering cohesive patterns to provide a structural description of it. In this paper, this idea of analyzing from the local to the global level is referred to as a bottom-up analysis of discourse.

Our approach of listing the cohesive patterns contributes to the bottom-up analysis of discourse. Rhetorical Structure Theory (hereafter RST) provides us with some useful frameworks for this discourse because the sets of such a local relationship are defined in it [4]. Moser and Moore [5] attempted to explore a synthesis of the GS-model and the

RST. We will take into consideration the theoretical issue they raised in creating our list.

3.3 Sub-stories

Another tendency evident is that all annotators assigned very similar purposes to a particular part. These parts, which we call ‘sub-stories’ in this paper, are clearly distinguished from the cohesive patterns because each of them has a broader part that contains some cohesive patterns. If we regard the monologue as a book, such a sub-story might play a role as a chapter. Such sub-stories will provide good information for developing a full-fledged summarization system, one of the goals in our project. However, to give a formal definition of a sub-story, we have to discuss at least the following problems.

According to the definition of the GS-model, a purpose of a segment must be related to the reason the speaker uttered the segment in the larger segments where it is included. This kind of relationship between sub-stories is more abstract than that of adjacent sentences. Let us regard sub-stories of a monologue metaphorically as chapters in a book. For instance, while we can easily define the relationship of a pair of adjacent sentences such as the pair “I can’t graduate.” and “Because I don’t have enough credits”, it is difficult, in contrast, to distinguish the same type of a relationship between a chapter and the next chapter in a book. Therefore, annotators have to identify the intentions of the sub-stories using a different level from those of the cohesive patterns.

A further problem concerns the identification of boundaries and their purposes in the sub-stories. Although we intuitively found ‘chapter-level’ purposes, it is hard to describe their contents stably in natural language. While this problem is mentioned in major discourse theories, none of them provide an adequate explanation for it.

To address this problem experimentally, we will focus on collecting the manually identified sub-stories and on constraining their features. In the next section, we propose a formalized way that will enable us to collect more stable boundaries and purposes of sub-stories.

4. COMMITTEE-BASED PURPOSE ASSIGNMENT

To identify boundaries of sub-stories and to assign their purpose, we call a ‘committee’, in which the members discuss how to segment a monologue before annotators start a deeper analysis of it. In the discussion of the committee, some ‘guidelines’ help the committee in assigning stable sub-stories. In the actual annotation, after the committee assigns purposes of the sub-stories, annotators find cohesive patterns in each of the given sub-stories through the bottom-up analysis described in Section 3.2.

The guidelines adopt a ‘generalized boundary’, which has a certain type of region in an un-stable boundary (we call the region a transitive expression). If the members of the committee have a different identification of a particular sub-story boundary, the committee has to discuss the role of such a region in the un-stable boundary. In the rest of this paper, we will propose that the role of such a region (transitive expression) relates to a transitive phenomenon between sub-stories. The features and examples of transitive expressions are described in the guidelines.

The definition of the guidelines is based on the results of an experiment, in which two researchers in this area divided a monologue into ‘flat’ sub-stories and assigned their purposes with no hierarchical structure. It contrasts with a hierarchical segmentation of the IAD in the GS-model and is an extended point in this work. In the experiment, the number of sub-stories into which the monologue should be divided was assumed to be from 5 to 15 (each researcher divided the monologues into about 10 sub-stories during practice). As a result, we found that they stably assigned more than half of the boundaries of the sub-stories and that the rest of them can be categorized into the two types indicated in Figure 2.

Figure 2 shows the examples of the discourses, in which the boundaries assigned by an annotator are shown by the ‘Purpose.’ Each of the shadowed regions in the discourses X and Y was bounded by the purposes that all annotators stably assigned (In Figure 2, the pairs of <Purpose A1, Purpose B1> and <Purpose A3, Purpose B3> show such stable boundaries). We could formally distinguish between Patterns 1 and 2 because one of the annotators assigned three or more purposes in the shadowed region in the annotation of Discourse Y.

Tbl.1 shows the averaged numbers of each type of boundary points in which the two researchers (denoted by A and B) assigns purposes of sub-story to 10 monologues. From the table, we found that the examples classified as Pattern 1 cover most of the unstable points, where researchers assigned different boundaries into these patterns.

These findings reveal that the flat segmentation is quite stable because most of the boundaries of the sub-stories were categorized as being either stable or as being equal to those in Pattern 1 of Figure 2. Moreover, we confirmed such a small region in Pattern 1 has a transitive function, which connects adjacent sub-stories and sometimes causes mismatches in boundary assignments among annotators at the same time.

This observation prompts us to assume the following. If we completely identify such transitive expressions between adjacent sub-stories, it will help us to exclude the inadequate segmentation of sub-stories. On this assumption, we are collecting the transitive expressions and characterizing them into some categories. These approaches will enable us to identify the un-stable points like Pattern 1 as equivalent to stable points.

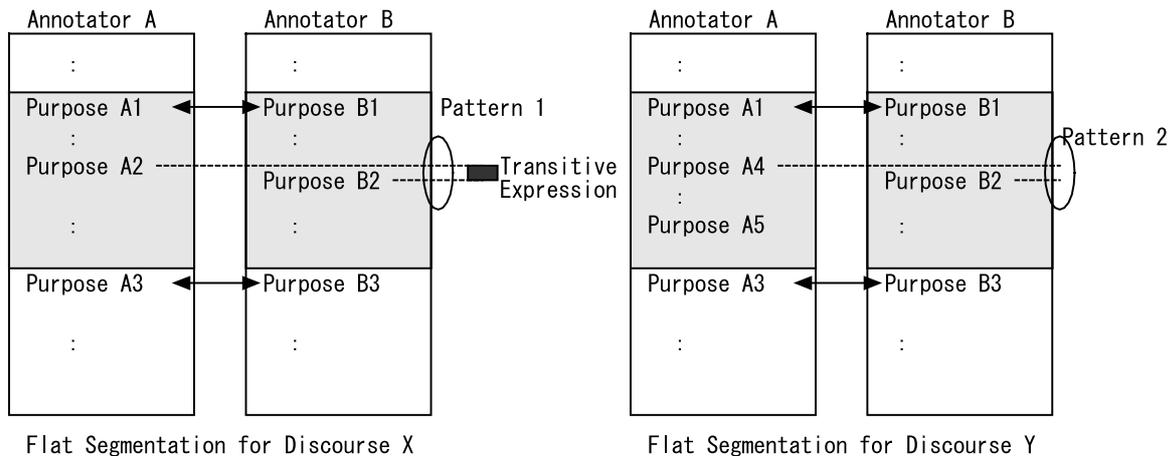


Figure 2. Difference in purpose assignment of sub-stories between annotators

Tbl.1 The averages of each type of boundary points

Each researcher assigned		Each type of boundary point	
A	B	Stable	Pattern 1
8.0	9.0	5.4	2.0

Furthermore, the aforementioned generalization of sub-story boundaries contributes to propose a discussion procedure to assign stable purposes for the complex examples like the discourse Y in Figure 2:

- First, the committee discusses whether a transitive expression exists in an un-stable boundary point classified as pattern 2.
- Then, the committee concentrates on discussing whether the adjacent sub-stories should combine. (eg. adjacent sub-stories bounded by the purpose A4 and A5 in the discourse Y)

In this paper, we have shown a committee-based procedure to identify sub-stories. A further investigation of the collected sub-stories by this procedure will lead us to discover what their appropriate purposes really are.

5. DISCUSSION

The corpora constructed in our approach, should contribute not only to developing a prototype summarization system but also to extending the research area of speech recognition.

Hirschberg and Nakatani [2] examined the co-relations between discourse segments and their prosodic features by use of a corpus annotated by the IAD. Our corpus

contributes to these kinds of studies because the discourse structures we described provide information on the larger constituents of a discourse than the cue phrases that Hirschberg and Nakatani mainly focused on.

In addition, such discourse information is likely to relate to human activities such as spontaneous story telling. We believe the use of our corpus will further this research area.

REFERENCES

- [1] B.J. Grosz and C.L. Sidner, "Attention, Intention, and the Structure of Discourse". *Computational Linguistics*, 12(3), 175-204, 1986.
- [2] J. Hirschberg and C.H. Nakatani, "A Prosodic Analysis of Discourse Segments in Direction Giving Monologues", *In Proceedings of the 34th Annual Meeting of the ACL*, 286-293, 1996.
- [3] K. Maekawa, "Corpus of Spontaneous Japanese: Its Design and Evaluation", This volume.
- [4] W.C. Mann and S.A. Thompson, "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization", *Text*, 8(3), 243-281, 1988.
- [5] M. Moser and J.D. Moore, "Toward a Synthesis of Two Account of Discourse Structure". *Computational Linguistics*, 22(3), 409-419, 1996.
- [6] C.H. Nakatani et al. *Instructions for annotating discourse*, Technical Report 21-95, Center for Research in Computing Technology, 1995.
- [7] K. Takanashi, T. Maruyama, K. Uchimoto, and H. Ishahara, "Identification of 'Sentences' in Spontaneous Japanese. -Detection and Modification of Clause Boundaries-". This volume.
- [8] T.A. van Dijk, "The Study of Discourse", In van Dijk (ed.) *Discourse as Structure and Process*, 1-34, 1997.
- [9] K. Uchimoto et al. "Morphological Analysis of Corpus of Spontaneous Japanese". This volume.