

Automatic Slide Generation Based on Discourse Structure Analysis

Tomohide Shibata and Sadao Kurohashi

Graduate School of Information Science and Technology, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{shibata, kuro}@kc.t.u-tokyo.ac.jp

Abstract. In this paper, we describe a method of automatically generating summary slides from a text. The slides are generated by itemizing topic/non-topic parts that are extracted from the text based on syntactic/case analysis. The indentations of the items are controlled according to the discourse structure, which is detected by cue phrases, identification of word chain and similarity between two sentences. Our experiments demonstrates generated slides are far easier to read in comparison with original texts.

1 Introduction

A presentation with slides is so effective to pass information to people in many situations, such as an academic conference or business. Although some softwares, such as PowerPoint and Keynote, help us with making presentation slides, it is still cumbersome to make them from scratch.

Some researchers have developed a method of (semi-)automatically making presentation slides from a technical paper or a news article [1, 2]. However, input texts of their systems were supposed to be documents whose structure is annotated: in [1], \TeX source and in [2], semantically annotated documents by GDA tag.

In this paper, we propose a method of automatically generating summary slides from a raw text. An example of a text is shown in Figure 1 and an example slide that is generated from the text is shown in Figure 2 (the translated slide is shown in Figure 3). In a slide, topic/non-topic parts that are extracted from the original text are itemized and each item is indented based on the discourse structure of the text. In particular, a big contrast/list structure in the text is an important clue for producing an easy-to-read slide.

The outline of our procedure is as follows:

1. Input sentences are processed by Japanese morphological analyzer, JUMAN [3], and are parsed by Japanese syntactic analyzer, KNP [4].
2. Each sentence is segmented into clauses and the discourse structure of the text is analyzed.
3. Topic/non-topic parts are extracted from the text.
4. Summary slides are generated by displaying the topic/non-topic parts based on the discourse structure.

大阪と神戸を結ぶ J R 神戸線、阪急電鉄神戸線、阪神電鉄本線の 3 線の不通により、一日 45 万人、ラッシュ時最大 1 時間 12 万人の足が奪われた。J R 西日本東海道・福知山・山陽線、阪急宝塚・今津・伊丹線、神戸電鉄有馬線の不通区間については、震災直後から代替バスによる輸送が行われた。国道 2 号線が開通した 1 月 23 日から、同国道と山手幹線を使って、大阪～神戸間の代替バス輸送が実施された。1 月 28 日からは、国道 2 号、43 号線に代替バス優先レーンが設置され、効率的・円滑な運行が確保された。(Due to the interruption of the three train services, JR Kobe-line, Hankyu Express Kobe-line and Hanshin Electric Railway, which connected between Osaka and Kobe, 450,000 people per day, 120,000 people per hour at the peak of rush, had no transportation. At the interruption sections in West Japan Railway Tokaidou Line, Sanyou Line, Hankyu Takarazuka, Imazu and Itami Line and Kobe-Electric Arima-line, transportation by alternate-bus was provided just after the earthquake occurred. From January 23th, when National Route 2 was opened, transportation by alternate-bus between Osaka and Kobe was provided. From January 28th, the alternate-bus priority lane was set up and smooth transportation was maintained.)

Fig. 1. An example of a text

鉄道の復旧 (1)

- 大阪と神戸を結ぶ J R 神戸線、阪急電鉄神戸線、阪神電鉄本線の 3 線の不通
 - 一日 45 万人、ラッシュ時最大 1 時間 12 万人の足が奪われた
- J R 西日本東海道・福知山・山陽線、阪急宝塚・今津・伊丹線、神戸電鉄有馬線の不通区間
 - 震災直後から
 - * 代替バスによる輸送
 - 国道 2 号線が開通した 1 月 23 日から
 - * 同国道と山手幹線を使って、大阪～神戸間の代替バス輸送が実施
 - 1 月 28 日から
 - * 国道 2 号、43 号線に代替バス優先レーンが設置され、円滑な運行が確保

Fig. 2. An example of a slide

Our method not only helps us with making presentation slides but also creates a full-automatic presentation. That is to say, input texts are spoken via text-to-speech engine while presenting automatically generated summary slides. We call this system “text-to-presentation”, as illustrated in Figure 4. For the input of text-to-speech engine, written texts are not appropriate, because unnatural speech might be produced due to difficult words or long compound nouns, which are unsuitable for speech synthesis. Therefore, written texts are automatically converted into spoken texts based on paraphrasing technique [5, 6] and then are inputted into speech synthesis.

The rest of this paper is organized as follows. Section 2 introduces how to analyze discourse structure. Section 3 explains how to extract topic/non-topic parts and Section 4 introduces how to generate a slide. Next, in Section 5, we describe our implemented system and report the evaluation. Finally, in Section 6, we discuss the related work, and in Section 7, we conclude this paper.

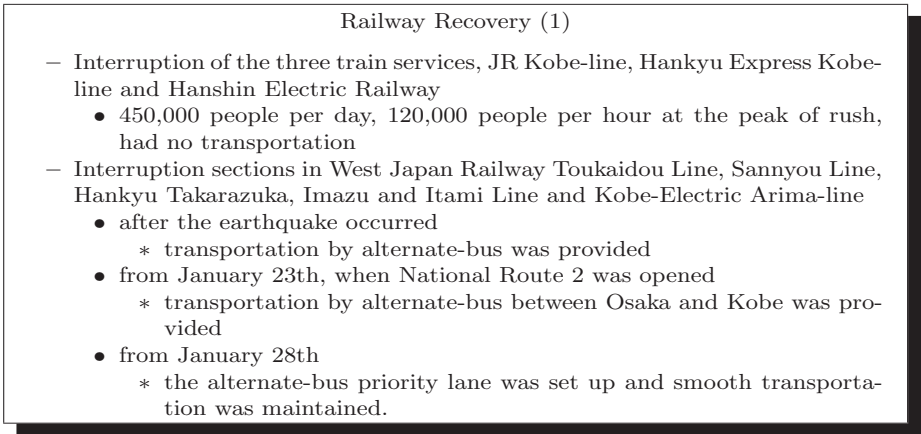


Fig. 3. An example of a slide (in English)

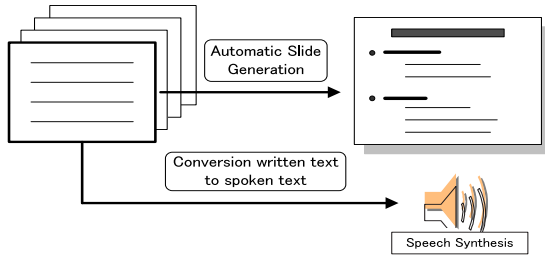


Fig. 4. Overview of text-to-presentation system

2 Discourse Structure Analysis

2.1 Model of Discourse Structure

We consider a clause and a sentence as a discourse unit and take into account the following two types of coherence relations:

1. coherence relations between two clauses in a sentence (four types)
 - list, contrast, additive, adversative
2. coherence relations between two sentences (eight types)
 - list, contrast, topic-chaining, topic-dominant chaining, elaboration, reason, cause, example

An example of discourse structure is shown in Figure 5. The arrows mean connection of clauses/sentences and the labels on the arrows mean coherence relation.

In our model, as an initial state, the discourse structure has a starting node. A sentence connecting to the starting node has the coherence relation “start”, which means this sentence is the beginning of a new topic.

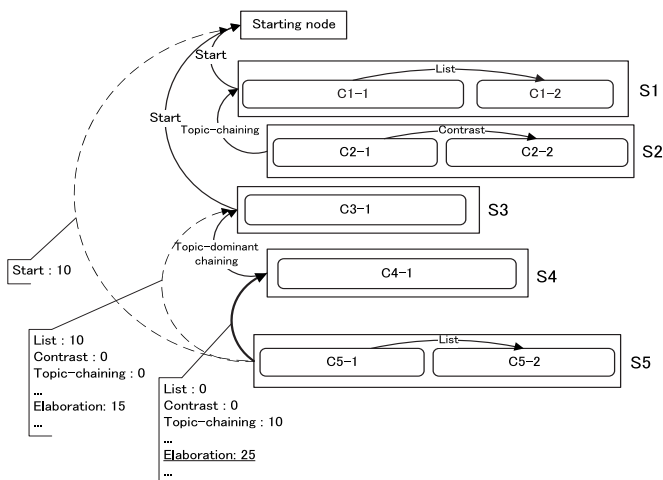


Fig. 5. The model of discourse structure

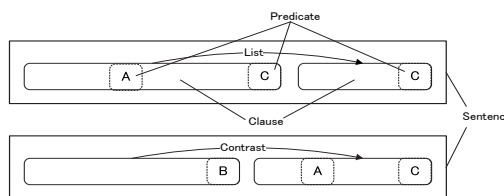


Fig. 6. Segmenting a sentence into discourse units

In the subsequent subsections, we describe how to analyze the discourse structure. This method is based on [7]. We start with the starting node and build a discourse tree in an incremental fashion. An input sentence is first parsed and segmented into clauses, and relations between clauses in a sentence are analyzed. Then, the sentence is connected to the most related preceding sentence and the coherence relation between the sentences is determined.

2.2 Segmenting a Sentence into Discourse Units

First, an input sentence is parsed by KNP and segmented into discourse units, which are the basic units of not only discourse structure but also extraction of topic/non-topic parts described in Section 3.

Japanese is a head-final language and a predicate is placed at the end of the clause. Therefore, each predicate in a sentence can be a discourse-unit boundary. We determine whether the predicate is regarded as a discourse-unit boundary according to the strength of the clause, which is classified into three levels depending on the subsuming relation of the scope¹. The strength of a clause can be detected by clause-end patterns.

¹ The level A has the narrowest scope and the level C has the broadest scope.

Level C (e.g., ...*ga*(although)) always divided

Level B (e.g., ...*te*(and)) divided in case that the clause and its parent clause are similar² or the length of both the clause and its parent clause exceeds a threshold³.

Level A (e.g., ...*nagara*(while)) not divided

2.3 Detection of Relation Between Clauses in a Sentence

After a sentence is segmented into discourse units, we analyze relations between clauses. First, the parent of each clause is simply determined based on syntactic structure of a sentence. Next, the coherence relation of two clauses is determined as follows:

- Two clauses are similar
 - contrast/list
- Two clauses are not similar
 - additive (...*te*(and),renyou-form)
 - adversative (...*ga,keredomo*(although))

Additive or adversative is recognized by the clause-end patterns. The method of calculating the similarity between two clauses and how to recognize contrast/list relation are described in detail below.

Calculation of similarity between two clauses. The parser calculates similarity between two arbitrary word-strings to detect coordinate structures in a sentence. Similarity between clauses is calculated as follows. First, the similarity value between two words is calculated according to exact matching, matching of their parts of speech, and their closeness in a thesaurus dictionary [8]. Then, the similarity value between two word-strings is calculated by the dynamic programming method: combining the similarity values between words in the two word-strings [4].

If two clauses have a certain similarity, the coherence relation between clauses can be two cases: list or contrast. In case of list relation, a thing/person has two different properties, as shown in (1-a), and, in case of contrast relation, two similar but different things have similar properties, as shown in (1-b).

- (1) a. John had a beer and eat martini. (list)
 b. John went to Paris and Mary went to London. (contrast)

The judgment whether the coherence relation is contrast or list is performed according to similarity between topics of two clauses: if the similarity exceeds a threshold, the coherence relation is determined to contrast relation, otherwise list relation. In Japanese, topics are often marked with *wa*. In the example (2),

² The parser calculates similarity between two clauses. The method of calculating similarity is described in the next subsection.

³ This threshold depends on a width of a slide and a font size.

Table 1. Examples of rules for discourse structure analysis

Coherence relation	Score	Applicable range	Patterns for a connected sentence	Patterns for a new sentence
start	10	*	*	<i>sate(then)</i> ...
list	5	1	*	<i>soshite(and)</i> ...
list	30	1	<i>daiichini(first)</i> ...	<i>dainini(second)</i> ...
contrast	30	1	*	<i>mushiro(rather than)</i> ...
elaboration	15	1	*	<i>tokuni(especially)</i> ...
reason	30	1	*	... <i>karada(because)</i>

because *tousho* (at first) and *3-gatsusue-made* (until the end of March) have a certain similarity, the coherence relation is determined to contrast.

- (2) *Daitaibus riyousya-wa* *tousho-wa 1-nichi atari*
 Alternate-bus user number-TOP first-TOP per day
3 kara 5 mannin-deattaga *3-gatsusue made-wa 1-nichi*
 from 30 to 50 thousands people the end of March-until per day
yaku 20-mannin-ga *riyoushita*
 about 20 thousands people used

(The number of alternate-bus users per day was from 30 to 50 thousands at first and was about 20 thousands until the end of March.)

2.4 Detection of Relation Between Two Sentences

As a new sentence comes in, by checking surface information, we find a connected sentence and the coherence relation between them by calculating reliable scores for all relations by the following three points: (1) cue phrases, (2) word/phrase chain, and (3) similarity between two sentences. As a final result, we choose the connected sentence and the relation that have the maximum score. Note that we make the assumption that a new sentence can be connected to the sentences on the right most edge in the discourse tree (in Figure 5, S5 is not allowed to connect to S1 and S2).

(1) Cue phrases. Examples of rules for matching cue phrases are shown in Table 1. Each rule specifies a condition for a pair of a new sentence and a possible connected sentence: the range of possible connected sentences (how far from the new sentence) and patterns for the two sentences. If a pair meets these condition, the relation and score in the rule are given to it.

(2) Word/phrase chain. A sentence consists of a topic part and a non-topic part. Words in a phrase whose head word is marked with “*wa*” are regarded as a topic part, and words except a topic part as a non-topic part.

In two sentences, if word-chaining is identified between a topic of a connected sentence and a topic of a new sentence, some score is given to topic chaining relation. Similarly, if word-chaining is identified between a non-topic of a connected

sentence and a topic of a new sentence, some score is given to topic-dominant relation.

(3) Similarity between two sentences. Similarity between two sentences is calculated by the similar method to one utilized for calculating similarity between two clauses in a sentence described in subsection 2.3. If the topics of two sentences have a certain similarity, the normalized similarity score between two sentences is given to contrast relation: otherwise, list relation.

In the following example, because *1-gatsu-23-nichi-kara*(from January 23th) in the sentence (3-a) and *1-gatsu-28-nichi-kara*(from January 28th) in the sentence (3-b) have a certain similarity, some score is given to contrast relation.

- (3) a. *1-gatsu-23-nichi kara* *Oosaka Kobe kan-no*
 January 23th-kara Osaka Kobe between
daitai bus yusou-ga *jisshisareta*
 alternate bus transportation-NOM provided
 (From January 23th, transportation by alternate-bus between Osaka and Kobe was provided.)
- b. *1-gatsu-28-nichi-kara-TOC* *enkatuna unten-ga* *kakuho-sareta*
 January 28th-kara smooth transportation maintained
 (From January 28th, the alternate-bus priority lane was set up and smooth transportation was maintained.)

3 Extraction of Topic Part / Reduction of Non-Topic Part

In this section, we describe how to extract what are displayed in a slide from an original text. As mentioned in Subsection 2.4, a sentence consists of a topic part and a non-topic part. Considering a clause as a basic unit, we extract a topic part and a non-topic part. Since, in general, non-topic part is relatively long, we reduce a non-topic part to make the produced slide easy-to-read. These procedures are illustrated in Figure 7 (T denotes the topic part and N denotes the non-topic part).

3.1 Extraction of Topic Part

A phrase whose head word is marked with a topic marker *wa* is extracted as a topic. A topic is extracted also by several cue phrases. When there are multiple

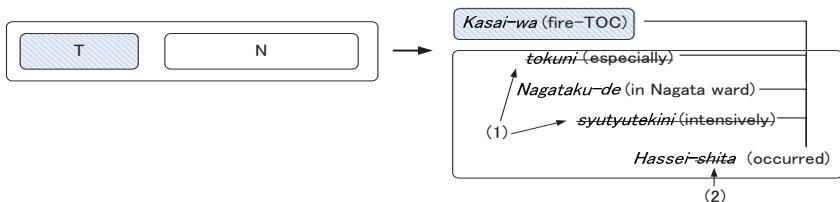


Fig. 7. Extraction of a topic part and reduction of a non-topic part

Table 2. Predicates and their case for key points

Case	Predicate
<i>ga</i>	<i>zyuuyouda</i> (important), <i>kagi-da</i> (key factor), <i>taisetsuda</i> (important), <i>yuuekida</i> (beneficial)
<i>wo</i>	<i>jyuushi-suru</i> (attach importance to), <i>akirakani-suru</i> (clarify)
<i>ni</i>	<i>tyakumoku-suru</i> (focus attention on), <i>jyuutenwo-oku</i> (stress)

cue phrases in a clause, a topic is extracted only by the first cue phrase. Some of them are shown below:

- *syukkagenin-ga-hanmei-shita-kasai ni-oite* ... (In the fire whose cause was revealed, ...)
- *3-sen-no-futsuu ni-yori* ... (Due to the interruption of the three train services, ...)

Note that the following cases are not regarded as a topic:

- phrases such as “wareware-wa” (we are) and “honronbun-dewa” (in this paper)
- a phrase does not depend on the end of the clause

3.2 Reduction of Non-Topic Part

In order to make a slide easy-to-read, it is important to reduce a text as long as the original meaning is preserved. We reduce a non-topic part by (1) pruning extraneous words/phrases based on syntactic/case structure and (2) deleting extraneous parts of the main predicate by some rules.

1. Pruning extraneous words/phrases

Extraneous words/phrases are pruned based on syntactic/case structure.

The following materials are pruned:

- conjunction
- adverb
- A-level clause
- adverb phrase
e.g., *computer-teishi-no-tame data-hikitsugi-mo-konnandatta*. (Because of the computer down, it is difficult to turn over the data.)
- appositive phrases
e.g., *nourinsuisansyou, kokudotyoo-nado kuni-no-kakukikan*. (Agencies, such as Agriculture, Forestry and Fisheries Ministry and National Land Agency)

2. Removing extraneous parts of a main predicate

Extraneous parts of a main predicate are removed by the following rules:

- deverbative noun-*suru/sareta* -> *suru/sareta* is deleted
ex.) *jisshi-sareta* -> *jisshi* (carried out)
- deverbative noun-*ga-okonawareta* -> *ga-okonawareta* is deleted
ex.) *yusou-ga-okonawareta* -> *yusou* (transport)
- noun-copula -> copula is deleted
ex.) *genin-da* -> *genin* (cause)

Note that extraneous parts of the main predicate are not removed if the predicate has a negation expression.

3.3 Extraction of Key Points

When the main predicate of the clause is the one listed in Table 2 and has the specified case component, we regard the non-topic part as key points. Key points are emphasized when they are placed in the slides as described in the next section.

4 Slide Generation

As illustrated in Figure 8, topic/non-topic parts that are extracted in Section 3 are placed in a slide based on the discourse structure that is detected in Section 2. Heuristic rules for generating slides are the following:

- If the text has a title, it is adopted as the title of the slide; otherwise, let the first topic in the text be the title of the slide.
- If there is a topic in a clause, the topic is displayed and in the next line the non-topic parts are displayed in the subsequent indented level. If there is not a topic in a clause, only non-topic parts are displayed. If non-topic parts are regarded as key points, they are emphasized.
- The level of the clause in the same sentence is set to be equal and the indent of each sentence is determined according to the coherence relation to its parent sentence:
 - **start:** the level is set to 0 because a new topic starts.
 - **contrast/list:** the same level.
 - **topic chaining:** if a topic is equal to that of the parent sentence, the level is set to the same level and the topic is not displayed; otherwise, the level is decreased by one and the topic and non-topic parts are displayed.
 - **otherwise:** the level is increased by one.

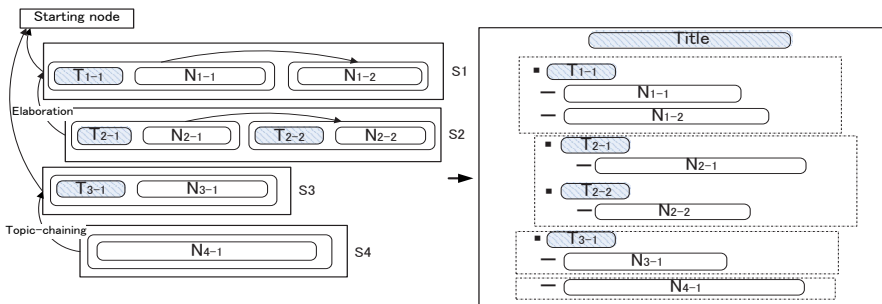


Fig. 8. Slide generation based on discourse structure

Table 3. Evaluation of discourse structure analysis

	accuracy
relation between clauses	30 / 39 (76.9%)
relation between sentences	60 / 89 (67.4%)

If the quantity of displayed texts exceeds a threshold, multiple slides are generated by splitting so that the number of lines in each slide is less than 12.

As a number of researchers have pointed out [9, 10], the discourse structure can be a clue to summarization: units found closer to the root of a discourse structure tree are considered to be more important than those found at lower levels in the tree. Along with such an idea, it can be considered that topic/non-topic parts in a lower unit (deep in a discourse structure) are not placed in a slide. However, in case that automatically generated slides are presented along with speech synthesis, the above treatment is not applied because speech without any corresponding description in the slide is not natural.

5 Implementation and Evaluation

5.1 Implementation of Text-to-Presentation System

We implemented a text-to-presentation system, in which a user can ask a question and enjoy the presentation about the query. We adopted *Hanshin-Awaji Daishinsai Kyoukunn Siryousyuu* (The Collection of Data Regarding the Lessons from the Great Hanshin-Awaji Earthquake Disaster)⁴ as a collection of text. The number of the unit of text is 400, which contains an average of 3.7 sentences with an average length of 50 characters.

First, the system retrieves a text that is the most similar to the user's query [11]. Then, the system converts the (written) text into spoken languages and feeds them to a speech synthesis engine, while presenting summary slides generated by the method described in this paper. The system runs on a Web browser as shown in Figure 9. When multiple slides are generated from a text, the system switches the slide to the next slide in synchronization with speech synthesis.

5.2 Evaluation

We generated slides from 30 queries and performed two evaluations: one is evaluation of discourse structure detection, the other is evaluation of automatically generated slides. As for evaluation of paraphrasing and retrieving texts, see [6], [11]. The average reduction rate when comparing the original texts to the generated slides was 0.797.

The accuracy of detecting discourse structure is shown in Table 3. Performance was evaluated by labeled attachment of clauses/sentences. Major errors of discourse structure detection are caused by word-chain mis-identification due to

⁴ <http://www.hanshin-awaji.or.jp/kyoukun/>

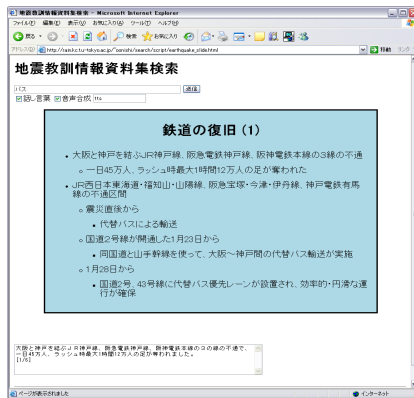


Fig. 9. A screenshot of our system

expression gaps. To deal with this problem, we are planning to apply a method of recognizing semantic equivalence of these gaps. Recognition errors of contrast relations between two clauses/sentences are attributed to a large variety of sentence structure and calculation failure of similarities between topics of two clauses/sentences by thesaurus.

The work of Marcu et al. is well known in the field of discourse parsing [12, 13]. They developed a discourse-annotated corpus and learned a discourse-parsing algorithm from it. In contrast, our discourse analyzer was based on generic heuristic rules. Actually, its application to the earthquake domain texts worked well for producing a summary slide. We do not have any plan for constructing discourse-annotated corpus in this domain because annotating discourse structure of texts takes much costs.

As for automatically generated slides, out of 30 texts, 15 was good, 12 was partially good and 3 was bad (judged by authors). The criterion for evaluation is whether a slide prevents a user to understand a presentation or not. Errors of generating slides are caused by the failure of detecting discourse structure and deleting non-topic part because of syntactic/case analysis errors. The heuristic rules of generating slides described in Section 4 do not cause an error.

Converting original texts into multimodal presentation (summary slides and speech synthesis) is significantly better in comparison with presenting original texts, even if there are some errors in generated slides. In particular, a slide in which a contrast/list structure is detected is far easier to read than an original text, as shown in Figure 9.

6 Related Work

Utiyama and Hasida presented a method of generating slide shows from documents which are semantically annotated by GDA [2]. The GDA is an XML tagset which allows machines to automatically infer the semantic structure underlying the raw documents. The system picks up important topics in the input

document on the basis of the semantic dependencies and coreferences identified from the tags and generates a slide by extracting relevant sentences and paraphrasing them to an itemized summary. Although it is possible to generate slide shows from semantically annotated documents even if they are relatively long, the manual annotation costs too much.

In the field of automatic summarization, for improving the quality of summarization, the sentence reduction system has been proposed [14]. Their system was constructed from pairs of articles and human-written summaries. This idea can be applied into our system, utilizing an alignment technique between technical papers and presentation sheets [15].

7 Conclusion

In this paper, we have proposed the method of automatically generating a summary slide from a text. Our method of generating slides consists of discourse structure analysis, extraction of topics/non-topic parts and displaying them based on discourse structure.

While we are improving the accuracy of detecting discourse structure and reducing the non-topic parts, we are planing to integrate text-to-presentation system with embodied conversational agents to enhance the presentation contents.

References

1. Yoshiaki, Y., Masashi, T., Katsumi, N.: A support system for making presentation slides (in japanese). *Transactions of the Japanese Society for Artificial Intelligence* **18** (2003) 212–220
2. Utiyama, M., Hasida, K.: Automatic slide presentation from semantically annotated documents. In: *1999 ACL Workshop on Coreference and Its Applications*. (1999)
3. Kurohashi, S., Nakamura, T., Matsumoto, Y., Nagao, M.: Improvements of Japanese morphological analyzer JUMAN. In: *Proceedings of the International Workshop on Sharable Natural Language*. (1994) 22–28
4. Kurohashi, S., Nagao, M.: A syntactic analysis method of long japanese sentences based on the detection of conjunctive structures. *Computational Linguistics* **20** (1994)
5. Kaji, N., Kawahara, D., Kurohashi, S., Sato, S.: Verb paraphrase based on case frame alignment. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. (2002) 215–222
6. Kaji, N., Okamoto, M., Kurohashi, S.: Paraphrasing predicates from written language to spoken language using the web. In: *Proceedings of the Human Language Technology Conference*. (2004) 241–248
7. Kurohashi, S., Nagao, M.: Automatic detection of discourse structure by checking surface information in sentences. In: *Proceedings of 15th COLING*. Volume 2. (1994) 1123–1127
8. Ikehara, S., Miyazaki, M., Shirai, S., Yokoo, A., Nakaiwa, H., Ogura, K., Oyama, Y., Hayashi, Y., eds.: *Japanese Lexicon*. Iwanami Publishing (1997)

9. Ono, K., Sumita, K., Miike, S.: Abstract generation based on rhetorical structure extraction. In: Proceedings of the 15th COLING. (1994) 344–348
10. Marcu, D.: Discourse trees are good indicators of importance in text. In I.Mani, M.Maybury, eds.: Advances in Automatic Text Summarization. The MIT Press (1999) 123–136
11. Kiyota, Y., Kurohashi, S., Kido, F.: Dialog navigator: A question answering system based on large text knowledge base. In: Proceedings of 19th COLING. (2002) 460–466
12. Marcu, D.: The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics* **26** (2000) 395–448
13. Carlson, L., Marcu, D., Okurowski, M.E.: Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue. (2001)
14. Jing, H.: Sentence reduction for automatic text summarization. In: Proceedings of the sixth conference on Applied natural language processing. (2000) 310–315
15. Hayama, T., Nanba, H., Kunifuji, S.: Alignment between a technical paper and presentation sheets using hidden markov model. In: Proceedings of the 2005 International Conference on Active Media Technology. (2005)