2 Natural Language

2-1 Construction of the *Corpus of Spontaneous Japanese* and Annotation Techniques

UCHIMOTO Kiyotaka, ISAHARA Hitoshi, TAKANASHI Katsuya, TAKEUCHI Kazuhiro, NOBATA Chikashi, MORIMOTO Ikuyo, and YAMADA Atsushi

This paper describes annotations for the *Corpus of Spontaneous Japanese*. The information we annotated to the corpus includes morphemes, clause units, dependency structures, summaries, and discourse structures. They are integrated in the form of XML. Morphological information was semi-automatically annotated to the transcribed text by reducing the human labor cost within the framework of morphological annotation that we proposed. Next, clause unites were detected based on the morphological information as basic units for our annotation. Then, dependency structures, summaries, discourse structures were annotated based on the clause units.

Keywords
Spontaneous speech corpus, Morphological analysis, Clause unit, Dependency structure, Summary, Discourse structure, XML

1 Introduction

This paper introduces the *Corpus of Spontaneous Japanese* (CSJ) and the technologies used to create this corpus. As part of a project to establish a science of spontaneous speech engineering based on elucidation of the linguistic and paralinguistic structures of spontaneous speech (FY 1999–FY 2003)[1] — an open, joint research project funded by the Japanese government's Special Coordination Funds for Promoting Science and Technology — this corpus has been constructed jointly with the National Institute for Japanese Language. The CSJ is a large-scale corpus for spontaneous Japanese, primarily covering monologues such as lectures. This corpus includes not just audio data but also transcribed text. Moreover, the transcribed text has been annotated with a wide range of verbal information. Figure 1 shows an outline of the verbal annotations used with the CSJ.

Data collection and transcription, and annotation with morphemes and prosodic information, was conducted chiefly at the National Institute for Japanese Language. The National Institute of Information and Communications Technology (NICT; formerly the Communications Research Laboratory) annotated the transcribed text with a wide range of verbal information, including morphemes, clause units, dependency structures, summaries, and discourse structures. Morphological annotation of small-scale transcribed text was carefully con-

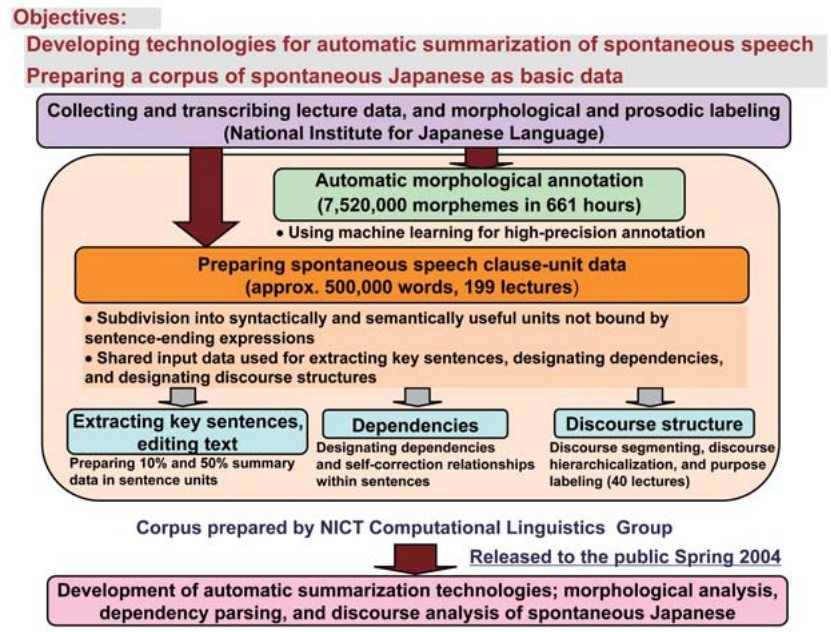


Fig.1 *An outline of the verbal annotations used with the Corpus of Spontaneous Japanese*

ducted by hand at the National Institute for Japanese Language. NICT then "trained" a morphological analysis system using all of the foregoing results as training data, and this system was used for morphological annotation of the remaining transcribed text[2]. This morphological annotation is described in greater detail in section **2**. Next, we identified clause units as the basic units for annotation, based on the morphological annotations[3]. This identification of clause units is described in section **3**. We then proceeded with annotation of dependency structures[4], preparation of summary data[5], and annotation of discourse structures[6] for these clauses. Section **4** describes dependency structural annotation, section **5** summary data, and section **6** discourse-structure analysis. In addition, section **7** provides an outline of a structure enabling the use of XML to integrate, describe, and store this data[7].

2 Morphological annotation

The transcribed text is annotated with two types of morphemes, as defined by the National Institute for Japanese Language: those in short units and those in long units. Those in short units are referred to as "**short words**", defined similarly to the headwords in an ordinary dictionary. Morphemes in long units are referred to as "**long words**", defined to include a broad range of compound words. These two units differ in length and as parts of speech, with long words defined to include short words. The publicly released corpus includes a total of approximately 7,520,000 short words. Since a long word is made up of one or more short words, the number of long words is approximately 20% smaller than this figure. About 1/8 of long and short words have been annotated with morpheme (morphological) information such as parts of speech, conjugation types, and conjugation forms, all by hand. The accuracy of the morphological information for this approximately 1/8 amount has been estimated at approximately 99.9%, using random sampling. Morphological annotation for the remaining approximately 7/8 was conducted semi-automatically.

In this paper, we refer to the entire process of morphological analysis and maintenance of the corpus as "**morphological annotation**". Figure 2 shows a graphical representation of

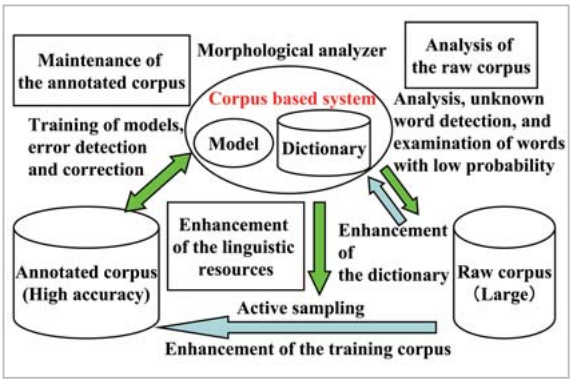


Fig.2 *Morphological annotation framework*

this framework. The goal of this framework is to improve the precision of morphological information for the entire corpus while minimizing labor costs for the case in which an annotated corpus, a raw corpus subject to analysis, and a morphological analyzer are available. During preparation of the corpus, we adopted a morphological analyzer using methods based on a corpus that is robust with respect to changes in corpus definitions, since structures and definitions change frequently. The biggest stumbling block in ordinary morphological analysis is the presence of unknown words — that is, morphemes not included either in dictionaries or in the training corpus. Two primary means of responding to this problem have been employed to date. One method involves automatically collecting unknown words and registering them in a dictionary; the other method entails the preparation of a model enabling analysis even of unknown words. We have proposed a method of morphological analysis that exploits the benefits of both of these responses, based on a maximum entropy model[8]. Since a model using this method could in principle estimate as a probability value the possibility that any given text string is a morpheme, it is highly likely that such a model could solve the problem of unknown words. For this reason, we employed this model in morphological analysis of the CSJ as well. Moreover, in this project we treated phenomena characteristic of spontaneous speech as described below.

• Presence of fillers and disfluencies

Fillers and disfluencies, which are phenomena characteristic of spontaneous speech, are difficult to identify because they could appear in any position. Since in the CSJ fillers and disfluencies are tagged by hand, we removed them before morphological analysis and then reinserted them later.

• Pronunciation forms

As one type of morpheme-related data, information on the pronunciation forms actually uttered is vital to preparing a linguistic model for speech recognition. However, it would not be practical to supplement information on these pronunciation forms using dictionary data. For this reason, we annotated actual pronunciation forms corresponding to the pronunciation-form field in the CSJ transcribed text and the results of morphological analysis.

As described below, the morphological annotation framework consists of preparing an annotated corpus, preparing a raw corpus subject to analysis, and expanding the annotated corpus. With the CSJ, within this framework we checked by hand approximately 2% of the detection and registration of unknown words and approximately 1% of active sampling. As a result, for the approximately 7/8 of the corpus not morphologically annotated by hand, the final precision level of automatic analysis for short words and long words, expressed in F-score, has been estimated at roughly 98.2% and 96.5%, respectively.

• Enhancement of the linguistic resources

When using a corpus-based analysis system, if the corpus includes a large number of errors there is a general tendency for accuracy to decline due to overfitting to the errors. To avoid this situation, errors in the training corpus need to be detected and corrected. With the CSJ, we corrected by hand errors detected using the following method. First, we conducted morphological annotation of the training corpus by hand, trained a model for morphological analysis, and then calculated using the model the probability of each difference arising between the results of analysis and the training corpus; we then replaced the relevant

portions of the training corpus with the corresponding portions of the results of analysis.

• Analysis of the raw corpus

If the raw corpus subject to analysis includes unknown words — that is, words that do not appear in the dictionary or in the training corpus — the likelihood that errors arising at left and right context of the unknown words. In many cases, the number of errors increases by more than the number of unknown words. In such a case, the precision of the corpus overall is increased by detecting and registering in the dictionary the unknown words in the corpus subject to analysis and, furthermore, by checking for low-probability words by hand[9]. If the corpus contains both types of morphemes — long and short units — then (to the extent long units are formed of short units) extracting unknown words from the shortest units and checking low-probability words will increase the precision of the long units[9].

• Enhancement of the linguistic resources

In general, a corpus-based model of a morphological analyzer will often require a large training corpus. However, in many cases simply increasing the size of the training corpus leads only to slight improvements in precision relative to the increase in size. This is because models for morphological analysis usually learn the connections between adjacent words, so if the added data is already connected in a way that is easy for the model to estimate, increasing the corpus in this manner will have practically no effect. For this reason, the training corpus needs to be expanded by extracting from a large-scale corpus subject to analysis beneficial data, including numerous sequences of words that are difficult for the model to analyze. Ultimately the aim is to ensure that precision will be improved substantially with the smallest possible additions. Accordingly, in this project we reduced labor costs by expanding the training corpus through active sampling[10].

3 Clause boundary detection

In past studies of written language, sentences were used as units subject to annotation. However, when studying spontaneous speech, sentences are not necessarily clearly defined units. For the CSJ, using sentences as units presents a number of problems, including the following:

- While with written language the writer him or herself uses periods to decide where to separate sentences, such information is not available in speech.
- A monologue is characterized by an individual speaking continuously; in terms of grammar, the speaker will not necessarily use sentence-ending forms frequently. Extremely long sentences may result.
- In spontaneous speech, sometimes it is difficult to determine the scope of a sentence due to factors such as self-correction, rephrasing, and stops; moreover, speech may sometimes consist only of fragments of words and sentences.

In light of the foregoing, we must be able to employ some method to detect syntactic and semantic units corresponding to sentences in written language that will address these problems. Our response to this problem was to employ the clause as a unit, instead of the sentence.

In the Japanese language, it is possible to detect a wide range of clause boundaries automatically, based solely on limited morphological information such as predicate conjugation and conjunctions. By revising the CBAP tool for automatic detection of clause boundaries[11], we prepared a set of rules called CBAP-csj for automatic detection of CSJ clause boundaries. CBAP-csj reads from one to three words before and after a morpheme, detects any clause-boundary types included in these words, and inserts labels corresponding to these types. The morphological information annotated in CSJ is expressed in the following four descriptions: “appearance”, “part of speech”, “conjugation type”, and “conjugation form”. These rules are written as regular expressions using pattern matching: when a morpheme string corresponding to a registered clause pattern is discovered, a label is then

inserted immediately following the string. This clause-boundary detection uses clause-boundary labels grouped into three categories: absolute boundaries ([ ]), strong boundaries (/ /), and weak boundaries (< >), depending on the size of the break immediately following the clause. An absolute boundary corresponds to an expression at the end of a sentence that is clearly defined formally. Although a strong boundary is not a sentence ending, it is nevertheless a clause boundary that can be considered a major break in speech. Although a weak boundary is a clause boundary, it is one that is not an ordinary break in speech. Furthermore, default units consisting of one or more clauses were each detected automatically using absolute boundaries and strong boundaries only as default boundaries of speech. These two types of clause boundaries serve as major breaks in speech. Since syntactic and semantic groupings have been prepared, these boundaries of units prove beneficial in a wide range of analysis and processing techniques. In theory, these categories group subordinate clauses into multiple classes based on differences in clause-boundary forms. The categories have been corrected through experiential knowledge, based on Minami[12] categories, which are associated with degrees of syntactic and semantic independence. Using these categories, it is possible to estimate (to some degree) grammatical behaviors that differ by clause type (such as sharing of subjects and cases and differences in scope of modalities). As result, it is possible to avoid the generation of default units that are not independent syntactically and semantically.

Since CBAP-csj detects boundaries in reference only to limited morpheme strings, it cannot detect special clause boundaries such as noun clauses. Nor can it handle characteristic phenomena of spontaneous speech (such as slips of the tongue and utterance stops), nor problems arising in relation to discourse structure. In order to detect units that are appropriate both syntactically and semantically, default units must be corrected by hand, with reference to actual speech data. To this end, we defined the following three types of tasks and used these tasks to effect these corrections. Some 40 types have been defined as standard for tasks conducted by hand.

- Connecting two or more default units with "+"
- Splitting default units using "–"
- Enclosing elements in "( )", "{ }", and "<< >>", indicating insertion, citation, and inversion, respectively

4 Dependency structure annotation

In this project, we annotated the CSJ as well with syntactic-construction information, in order to respond to a wide range of research and development needs. Since the corpus covers the Japanese language, we employed the dependency structure between clauses as a syntactic construction. In Japanese, word order is relatively flexible, making it difficult in many cases to identify dependency relations between clauses. However, in order to understand the meaning of a sentence it is very important to identify dependency relations. For this reason, in processing Japanese the dependency structure between clauses is often employed as a syntactic construction, with particular attention paid to information for which identification is difficult but nevertheless important. A typical tagged text corpus available to us, the Kyoto University Corpus[13], which is used for a wide range of processing such as machine translation, information extraction, summarization, and questions and answers, employs such a structure.

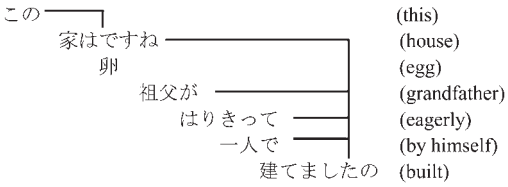
In principle, the dependency structure between clauses in the CSJ complies with the standards of the Kyoto University Corpus. However, phenomena often differ between text and speech, so that these standards alone cannot cover all cases. For this reason, we established the following new standards for phenomena characteristic to speech.

- **Utterance stops**

Although in principle these stops are eliminated through the process of identifying clause

boundaries, in certain cases — with dependencies that cross over utterance stops, for example — these are not eliminated. In such cases, utterance stops are not considered to have attached phrases.

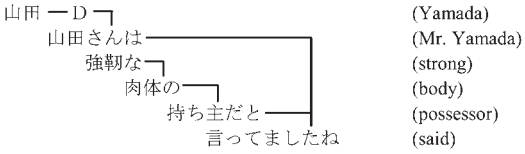
ex) "卵(egg)" is an utterance stop, and it has no modifiee.



• Self-correction, rephrasing

We established a new standard for handling self-correction and rephrasing within a clause unit. Although there are many different conceivable types of self-correction and rephrasing as well, for the CSJ dependency structure we focused on identifying the general scopes of self-correction and rephrasing without categorizing types in detail. All of them were labeled with D.

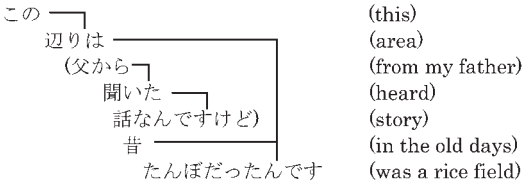
ex) "山田(Yamada)" is corrected as "山田さん(Mr.Yamada)" by the speaker.



• Inserted clauses

Dependencies are enclosed within inserted clauses. Inserted clauses are identified in the process of detecting clause boundaries.

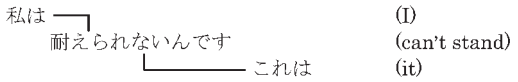
ex) "父から聞いた話なんですけど (which is a story that I heard from my father)" is an inserted clause.



• Inversion

This refers to a dependency that flows from right to left.

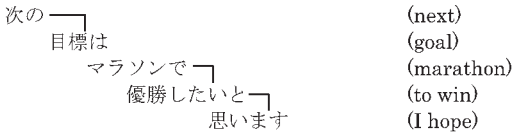
ex) "これは (it)" is an inversion.



• Distortion

Changes in speech plans frequently lead to unnatural syntactic constructions. In principle, such cases are handled by assuming no attached phrase exists. Also, in some cases — such as when a major break is located immediately following the expression of a subject — a clause boundary may be detected, resulting in the identification of a separate unit.

ex) The sentence is distorted after "目標は (goal)".



The actual annotation process was conducted by hand, using this tool. We used the following structure: two persons annotated each lecture, and one checker inspected the annotation. The subject was comprised of 199 lectures with clause units identified, with conversations and rereading not covered.

5 Preparation of various types of summary data

Traditionally, automatic summarization using computers has been based on extraction of key sentences or key passages. In other words, a summary often is seen as a collection of extracted key passages. Against this background, in this project as well we prepared key-sentence selection data summarizing conversations in the CSJ. However, with a view to contributing to the future development of natural-language processing technology, we also prepared summary data using methods other than selection of key sentences. Specifically, we prepared the following three types of summary data. This data is provided for the 199 lectures with clause units identified.

• key-sentence selection data

We prepared key-sentence selection data for each lecture by selecting key sentences with summarization rates of 50% and 10%. As

used here, a designated summarization rate of, for example, 10% refers to selection by the operator of units corresponding to only 10% of the character count of the transcribed text assigned to him or her. In selecting key sentences, the units used were clauses using the clause-boundary information described in the preceding section. First of all, key sentences with summarization rates of 50% were selected and then, from these, key sentences with summarization rates of 10% relative to the original lecture were selected.

• Free-summary data

In addition to key-sentence selection data, we also prepared, from the transcribed text, free-summary data summarizing the lectures in the form of directly written text. For free-summary data as well, we prepared two types of data for each conversation: one with a summarization rate of 50% and one with a summarization rate of 10%. In preparing data with a summarization rate of 50%, we restricted the editing task, preparing summaries only through extraction of key sections and changing expressions within each section. Although in principle data with a summarization rate of 10% was prepared chiefly through tasks similar to those used for the data with a summarization rate of 50%, when we were unable to include through this means sufficient content as needed we permitted rewriting using free expression and replacement of sections.

• Text-editing data

Text-editing data differs from free-summary data in that it is data in which brief summaries have been prepared by having operators rewrite key sentences in key-sentence selection data using specific operations only. In other words, this can be seen as an intermediate topic that fills the gap with existing key-sentence selection methods after considering issues related to automatic generation of free summaries by computers.

In principle, operations that can be applied to data in which operators have selected key sentences include the deletion of words and clauses. Insertion of new words and expressions not present in the transcribed text is limited to cases in which the text would be grammatically incorrect without such insertion. Rewriting key-sentence selection data through operations such as this is intended primarily to prepare more concise summaries by eliminating redundancy, while also maintaining ease of reading as a summary of key-sentence selection data.

6 Discourse-structure analysis

This section introduces discourse-structure tags annotated to the CSJ and related annotation methods. Discourse-structure tag annotation used in this project is based on Grosz and Sidner's theory (GS)[14]. GS states that a speaker's intentions and purposes are reflected in the surface language structure of discourse. According to GS, a speaker's intentions and purposes concern the following:

- Why is the operation of discourse (rather than some other behavior) being attempted?
- Why is the content of this discourse (rather than some other content) being communicated?

Furthermore, we believe that an entire conversation has more than one purpose. Each of the multiple conversation segments making up a conversation has its own purpose ("conversation purpose" hereinafter) that could serve as part of the purpose of the entire conversation.

A number of previous studies use GS as a conversation model. From such studies, we focused on the manual (IAD) of Nakatani et al. [15], which assigned discourse-structure tags to actual data. Next, we decided to assign discourse structures to CSJ conversations by expanding the IAD through sorting the issues that arise in applying the IAD to CSJ conversations. The IAD categorizes discourse-structure tagging into three tasks: (1) identifying segment boundaries, (2) identifying the ranks between segments, and (3) describing conversation segment purposes. However, since it was not clearly stated as a result of preliminary conversation tagging of the IAD in which order upper-level tasks should be conducted in

the IAD, it was clear that it would be easy for disparities in tagging results to arise between operators, due to mixing of tasks. For example, operators tried to identify segment boundaries and segment ranks simultaneously.

We addressed this issue by separating the task of identifying conversation segments into the following two tasks:

- Task 1) Separating a single conversation into conversation segments with no hierarchy, like a chapter in a novel. This task was conducted while listening to speech, with a single conversation analyzed by multiple operators. Conversation segments identified consistently by multiple operators are referred to as sections.
- Task 2) Compiling together clauses identified through clause identification, to discover patterns of connection between clauses with consistent content. Conversation segments identified through this task are referred to as episodes.

In this way we separated, at a task level, identification of conversation segment layers that subdivide the entire conversation in general terms and identification of those that achieve semantic combinations at the clause level. As a result, we identified only a very small number of episodes crossing section boundaries. Also, to maintain consistency between the results of these two tasks, we described the conversation purposes in Task 2 and defined section conversation purposes based on these. As a result, in the open data a single conversation consists of multiple sections, with each section including one or more episodes.

7 Using XML for integration of annotation

The wide range of information annotated to the CSJ was integrated, described, and stored using XML. For this purpose, we first created XML data for the information on clauses, dependencies, clause boundaries, key sentences, and discourse structures with which the transcribed text was annotated and described mutual relationships between these elements. Next, we integrated this information with phonological annotation information used separately for annotation by the National Institute for Japanese Language. For linguistic information, we expressed the lectures in a hierarchical structure consisting of clause units and clauses and annotated each clause unit with information on the conversation and key sentences and annotated each clause with dependency information. Combining this approach with a structure consisting of elements of basic transcription units separated by silent breaks of a certain length or longer, it is possible that the basic transcription units may intersect with clause units or clauses. For this reason, as a means of combining this data without losing information, we applied a method of annotating the leading short word of each relevant structural element with the clause unit and clause information, instead of expressing the clause units and clauses within their hierarchy. In restoring the hierarchical structure of clause units and clauses from a structure consisting of basic transcription units as structural elements, for all short words crossing the boundaries of basic transcription units we decided to collect short words having no attributes concerning clause units or concerning clauses that succeeded those with such attributes. Using XML made it possible to express in an efficient manner data logically belonging to different layers and mutually dependent data.

The XML data thus prepared was used in two forms: to study data within a single lecture and for cross-sectional study of multiple lectures. Data within a single lecture is easy to use and study by extracting information and constructions corresponding to the purpose of the original XML instance and converting it to a different format. Using XML-related techniques such as XSLT makes this conversion easy. A database system is required when studying multiple lectures. Under current conditions, three methods are available: using a native XML database to store the data as XML data, storing the data converted to an appropriate data structure for use in a relation-

al database, and using XML data on the front end and a relational database on the back end.

8 Conclusions

In this paper, we summarized the annotation of the *Corpus of Spontaneous Japanese* (CSJ) by the National Institute of Information and Communications Technology (NICT; formerly the Communications Research Laboratory). Following completion of this project, we have continued research into speech annotation and analysis technologies such as enhancing the tools developed in connection with preparation of the CSJ, as well as improving the precision of natural-language processing such as detection of sentence boundaries, morphological analysis, and dependency analysis[16][17], using the information annotated to the CSJ in related studies.

References

1 K. Maekawa, H. Koiso, S. Furui, and H. Isahara, "Spontaneous speech corpus of Japanese", Proceedings of the Second International Conference of Language Resources and Evaluation, 2, pp.947-952, 2000.

2 K. Uchimoto, K. Takaoka, C. Nobata, A. Yamada, S. Sekine, and H. Isahara, "Morphological Analysis of the Corpus of Spontaneous Japanese", IEEE Transactions on Speech and Audio Processing, Vol.12, No.4, pp.382-390, 2004.

3 K. Takanashi, T. Maruyama, K. Uchimoto, and H. Isahara, "Identification of "sentences" in spontaneous Japanese - Detection and modification of clause boundaries", in SSPR-2003.

4 K. Uchimoto, R. Hamabe, T. Maruyama, K. Takanashi, T. Kawahara, and H. Isahara, "Dependency-structure Annotation to Corpus of Spontaneous Japanese", In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), pp.635-638, 2006.

5 C. Nobata, K. Uchimoto, K. Takanashi, and H. Isahara, "Development of Summary Data in the Corpus of Spontaneous Japanese", In Proceedings of the Third Spontaneous Speech Science and Technology Workshop, pp.99-104, 2004.

6 K. Takeuchi, K. Takanashi, I. Morimoto, H. Koiso, and H. Isahara, "Committee-based discourse purpose assignment: Discourse structure annotations of spontaneous Japanese monologue", in SSPR-2003.

7 A. Yamada, K. Takanashi, K. Uchimoto, K. Takeuchi, C. Nobata, I. Morimoto, and H. Isahara, "Annotation Integration for the Corpus of Spontaneous Japanese", In Proceedings of the Third Spontaneous Speech Science and Technology Workshop, pp.33-38, 2004.

8 K. Uchimoto, S. Sekine, and H. Isahara. "The Unknown Word Problem: a Morphological Analysis of Japanese Using Maximum Entropy Aided by a Dictionary", In Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing, pp.91-99, 2001.

9 K. Uchimoto, C. Nobata, A. Yamada, S. Sekine, and H. Isahara, "Morphological Analysis of a Large Spontaneous Speech Corpus in Japanese", ACL, pp.479-488, 2003.

10 K. Uchimoto and H. Isahara, "Morphological Annotation of a Large Spontaneous Speech Corpus in Japanese", IJCAI, pp.1731-1737, 2007.

11 T. Maruyama, H. Kashioka, T. Kumano, and H. Tanaka, "Development and evaluation of Japanese Clause Boundaries Annotation Program", Journal of Natural Language Processing, 11(3):39.68, 2004. (in Japanese)

12 F. Minami, "Gendai Nihongo no Kohzoh", Tokyo : Taishukan Shoten, 1974. (in Japanese)

13 S. Kurohashi and M. Nagao, "Building a Japanese Parsed Corpus while Improving the Parsing System", In Proceedings of the NLPRS, pp.451-456, 1997.

14 B. J. Grosz and C. L. Sidner, “Attention, intention, and the structure of discourse”, Computational Linguistics, Vol.12, No.3, pp.175-204, 1986.
15 C. H. Nakatani et al, “Instructions for annotating discourse”, Technical Report, 21-95, Center for Research in Computing Technology, Harvard University Press, 1995.
16 K. Shitaoka, K. Uchimoto, T. Kawahara, and H. Isahara, “Dependency Structure Analysis and Sentence Boundary Detection in Spontaneous Japanese”, In Proceedings of the 20th International Conference on Computational Linguistics (COLING2004), pp.1107-1113, 2004.
17 R. Hamabe, K. Uchimoto, T. Kawahara, and H. Isahara, “Detection of Quotations and Inserted Clauses and Its Application to Dependency Structure Analysis in Spontaneous Japanese”, COLING-ACL, pp.324-330, 2006.

***UCHIMOTO Kiyotaka**, Ph.D.*

Senior Researcher, Computational Linguistic Group, Knowledge Creating Communication Research Center (Former: Senior Researcher, Computational Linguistics Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Department)

Natural Language Processing

***ISAHARA Hitoshi**, Ph.D.*

Group Leader, Computational Linguistic Group, Knowledge Creating Communication Research Center (Former: Group Leader, Computational Linguistic Group, Keihanna Human Info-Communication Research Center, Information and Communications Department)

Natural Language Processing

TAKANASHI Katsuya

Asistant Professer, Academic Center for Computing and Media Studies, Kyoto University (Former: Expert Researcher, Computational Linguistics Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Department)

Communication Science

***TAKEUCHI Kazuhiro**, Ph.D.*

Lecturer, Dept of Engineering Informatics, Osaka Electro-Communication University (Former: Expert Researcher, Computational Linguistics Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Department)

Natural Language Processing

***NOBATA Chikashi**, Ph.D.*

Research Associate, School of Computer Science, The University of Manchester (Former: Expert Researcher, Computational Linguistics Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Depertment)

Natural Language Processing

***MORIMOTO Ikuyo**, Ph.D.*

Associate Professer, School of Low and Politics, KWANSEI GAKUIN University (Former: Expert Researcher, Computational Linguistics Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Department)

Conversation Analysis

***YAMADA Atsushi**, Dr. Eng.*

Chief Researcher, Advanced Software Technology & Mechatronics Research Institute of Kyoto (Former: Expert Researcher, Computational Linguistics Group, Keihanna Human Info-Communication Research Center, Information and Network Systems Department)

Natural Language Processing