

Annotating Dialogue Acts to Construct Dialogue Systems for Consulting

Kiyonori Ohtake Teruhisa Misu Chiori Hori Hideki Kashioka Satoshi Nakamura

MASTAR Project, National Institute of Information and Communications Technology

Hikaridai, Keihanna Science City, JAPAN

kiyonori.ohtake (at) nict.go.jp

Abstract

This paper introduces a new corpus of consulting dialogues, which is designed for training a dialogue manager that can handle consulting dialogues through spontaneous interactions from the tagged dialogue corpus. We have collected 130 h of consulting dialogues in the tourist guidance domain. This paper outlines our taxonomy of dialogue act annotation that can describe two aspects of an utterance: the communicative function (speech act), and the semantic content of the utterance. We provide an overview of the Kyoto tour guide dialogue corpus and a preliminary analysis using the dialogue act tags.

1 Introduction

This paper introduces a new dialogue corpus for consulting in the tourist guidance domain. The corpus consists of speech, transcripts, speech act tags, morphological analysis results, dependency analysis results, and semantic content tags. In this paper, we describe the current status of a dialogue corpus that is being developed by our research group, focusing on two types of tags: speech act tags and semantic content tags. These speech act and semantic content tags were designed to express the dialogue act of each utterance.

Many studies have focused on developing spoken dialogue systems. Their typical task domains included the retrieval of information from databases or making reservations, such as airline information e.g., DARPA Communicator (Walker et al., 2001) and train information e.g., ARISE (Bouwman et al., 1999) and MASK (Lamel et al., 2002). Most studies assumed a definite and consistent user objective, and the dialogue strategy was usually designed to minimize the cost of information access. Other target tasks include tutoring and trouble-shooting dialogues (Boye, 2007).

In such tasks, dialogue scenarios or agendas are usually described using a (dynamic) tree structure, and the objective is to satisfy all requirements.

In this paper, we introduce our corpus, which is being developed as part of a project to construct consulting dialogue systems, that helps the user in making a decision. So far, several projects have been organized to construct speech corpora such as CSJ (Maekawa et al., 2000) for Japanese. The size of CSJ is very large, and a great part of the corpus consists of monologues. Although, CSJ includes some dialogues, the size of dialogues is not enough to construct a dialogue system via recent statistical techniques. In addition, relatively to consulting dialogues, the existing large dialogue corpora covered very clear tasks in limited domains.

However, consulting is a frequently used and very natural form of human interaction. We often consult with a sales clerk while shopping or with staff at a concierge desk in a hotel. Such dialogues usually form part of a series of information retrieval dialogues that have been investigated in many previous studies. They also contain various exchanges, such as clarifications and explanations. The user may explain his/her preferences vaguely by listing examples. The server would then sense the user's preferences from his/her utterances, provide some information, and then request a decision.

It is almost impossible to handcraft a scenario that can handle such spontaneous consulting dialogues; thus, the dialogue strategy should be bootstrapped from a dialogue corpus. If an extensive dialogue corpus is available, we can model the dialogue using machine learning techniques such as partially observable Markov decision processes (POMDPs) (Thomson et al., 2008). Hori et al. (2008) have also proposed an efficient approach to organize a dialogue system using weighted finite-state transducers (WFSTs); the system obtains the

Table 2: Overview of Kyoto tour guide dialogue corpus

dialogue type	F2F	WOZ	TEL
# of dialogues	114	80	62
# of guides	3	2	2
avg. # of utterance / dialogue (guide)	365.4	165.2	324.5
avg. # of utterance / dialogue (tourist)	301.7	112.9	373.5

structure of the transducers and the weight for each state transitions from an annotated corpus. Thus, the corpus must be sufficiently rich in information to describe the consulting dialogue to construct the statistical dialogue manager via such techniques.

In addition, a detailed description would be preferable when developing modules that focus on spoken language understanding and generation modules. In this study, we adopt dialogue acts (DAs) (Bunt, 2000; Shriberg et al., 2004; Bangalore et al., 2006; Rodriguez et al., 2007; Levin et al., 2002) for this information and annotate DAs in the corpus.

In this paper, we describe the design of the Kyoto tour guide dialogue corpus in Section 2. Our design of the DA annotation is described in Section 3. Sections 4 and 5 respectively describe two types of the tag sets, namely, the speech act tag and the semantic content tag.

2 Kyoto Tour Guide Dialogue Corpus

We are currently developing a dialogue corpus based on tourist guidance for Kyoto City as the target domain. Thus far, we have collected itinerary planning dialogues in Japanese, in which users plan a one-day visit to Kyoto City. There are three types of dialogues in the corpus: face-to-face (F2F), Wizard of OZ (WOZ), and telephonic (TEL) dialogues. The corpus consists of 114 face-to-face dialogues, 80 dialogues using the WOZ system, and 62 dialogues obtained from telephone conversations with the interface of the WOZ system.

The overview of these three types of dialogues is shown in Table 2. Each dialogue lasts for almost 30 min. Most of all the dialogues have been manually transcribed. Table 2 also shows the average number of utterances per a dialogue.

Each face-to-face dialogue involved a profes-

sional tour guide and a tourist. Three guides, one male and two females, were employed to collect the dialogues. All three guides were involved in almost the same number of dialogues. The guides used maps, guidebooks, and a PC connected to the internet.

In the WOZ dialogues, two female guides were employed. Each of them was participated in 40 dialogues. The WOZ system consists of two internet browsers, speech synthesis program, and an integration program for the collaborative work. Collaboration was required because in addition to the guide, operators were employed to operate the WOZ system and support the guide. Each of the guide and operators used own computer connected each other, and they collaboratively operate the WOZ system to serve a user (tourist).

In the telephone dialogues, two female guides who are the same for the WOZ dialogues were employed. In these dialogues, we used the WOZ system, but we did not need the speech synthesis program. The guide and a tourist shared the same interface in different rooms, and they could talk to each other through the hands-free headset.

Dialogues to plan a one-day visit consist of several conversations for choosing places to visit. The conversations usually included sequences of requests from the users and provision of information by the guides as well as consultation in the form of explanation and evaluation. It should be noted that in this study, enabling the user to access information is not an objective in itself, unlike information kiosk systems such as those developed in (Lamel et al., 2002) or (Thomson et al., 2008). The objective is similar to the problem-solving dialogue of the study by Ferguson and Allen (1998), in other words, accessing information is just an aspect of consulting dialogues.

An example of dialogue via face-to-face communication is shown in Table 1. This dialogue is a part of a consultation to decide on a sightseeing spot to visit. The user asks about the location of a spot, and the guide answers it. Then, the user provides a follow-up by evaluating the answer. The task is challenging because there are many utterances that affect the flow of the dialogue during a consultation. The utterances are listed in the order of their start times with the utterance ids (UID). From the column “Time” in the table, it is easy to see that there are many overlaps.

Table 1: Example dialogue from the Kyoto tour guide dialogue corpus

UID	Time (ms)	Speaker	Transcript	Speech act tag**	Semantic content tag
56	76669–78819	User	<i>Ato</i> (And.) <i>Ohara ga</i> (Ohara is) <i> dono henni</i> (where) <i> narimasuka</i> (I'd like to know)	WH-Question.Where	null (activity),location (activity),(demonstrative),interr (activity),predicate
57	80788–81358	Guide	<i>kono</i> (here) <i> hendesume</i> (is around)	State_Answer→56	(demonstrative),kosoa (demonstrative),noun
58	81358–81841	Guide	<i>Ohara ha</i> (Ohara) <i> Chotto</i> (a bit)	State_Inversion	location
59	81386–82736	User	<i> hanaresugitemasume</i> (is too far)	State_Evaluation→57	(transp),(cost),(distance),adverb-phrase (transp),(cost),(distance),predicate
60	83116–83316	Guide	<i>A</i> (Yeah.) <i> Kore demo</i> (it)	Pause_Grabber	null null
61	83136–85023	User	<i>ichinichi dewa</i> (in a day) <i> doudeshou</i> (Do you think I can do)	Y/N-Question	(activity),(planning),duration (activity),(planning),(demonstrative),interr
62	83386–84396	Guide	<i>Soudesune</i> (right.) <i> Ichinichi</i> (One day)	State_Acknowledgment→59	null (activity),(planning),(enlity),day-window
63	85206–87076	Guide	<i>areba</i> (is) <i> jubuN</i> (enough) <i> ikemasu</i> (to enjoy it.)	State_AffirmativeAnswer→61	(activity),(planning),predicate (consulting),(activity),adverb-phrase (consulting),(activity),action
64	88392–90072	Guide	<i>Oharano</i> (Ohara is) <i> sugoku</i> (very) <i> kireidesuyo</i> (a beautiful spot)	State_Opinion	(activity),location (recommendation),(activity),adverb-phrase (recommendation),(activity),predicate
65	89889–90759	User	<i> tidesume</i> (that would be nice.)	State_Acknowledgment→64 _Evaluation→64	(consulting),(activity),predicate

* Tags are concatenated using a delimiter '.' and omitting null values.
The number following the '→' symbol denotes the target utterance of the function.

3 Annotation of Communicative Function and Semantic Content in DA

We annotate DAs in the corpus in order to describe a user's intention and a system's (or the tour guide's) action. Recently, several studies have addressed multilevel annotation of dialogues (Levin et al., 2002; Bangalore et al., 2006; Rodriguez et al., 2007); in our study, we focus on the two aspects of a DA indicated by Bunt (2000). One is the communicative function that corresponds to how the content should be used in order to update the context, and the other is a semantic content that corresponds to what the act is about. We consider both of them important information to handle the consulting dialogue. We designed two different tag sets to annotate DAs in the corpus. The speech act tag is used to capture the communicative functions of an utterance using domain-independent multiple function layers. The semantic content tag is used to describe the semantic contents of an utterance using domain-specific hierarchical semantic classes.

4 Speech Act Tags

In this section, we introduce the speech act (SA) tag set that describes communicative functions of utterances. As the base units for tag annotation, we adopt clauses that are detected by applying the clause boundary annotation program (Kashioka and Maruyama, 2004) to the transcript of the dialogue. Thus, in the following discussions, "utterance" denotes a clause.

4.1 Tag Specifications

There are two major policies in SA annotation. One is to select exactly one label from the tag set (e.g., the AMI corpus¹). The other is to annotate with as many labels as required. MRDA (Shriberg et al., 2004) and DIT++ (Bunt, 2000) are defined on the basis of the second policy. We believe that utterances are generally multifunctional and this multifunctionality is an important aspect for managing consulting dialogues through spontaneous interactions. Therefore, we have adopted the latter policy.

By extending the MRDA tag set and DIT++, we defined our speech act tag set that consists of six layers to describe six groups of function: *General*, *Response*, *Check*, *Constrain*, *ActionDiscussion*, and *Others*. A list of the tag sets (excluding the *Others* layer is shown in Table 3. The *General* layer has two sublayers under the labels, *Pause* and *WH-Question*, respectively. The two sublayers are used to elaborate on the two labels, respectively. A tag of the *General* layer must be labeled to an utterance, but the other layer's tags are optional, in other words, layers other than the *General* layer can take null values when there is no tag which is appropriate to the utterance. In the practical annotation, the most appropriate tag is selected from each layer, without taking into account any of the other layers.

The descriptions of the layers are as follows:

General: It is used to represent the basic form

¹<http://corpus.amiproject.org>

Table 3: List of speech act tags and their occurrence in the experiment

Tag	Percentage(%)		Tag	Percentage(%)		Tag	Percentage(%)		Tag	Percentage(%)	
	User	Guide		User	Guide		User	Guide		User	Guide
(General)			(Response)			(ActionDiscussion)			(Constrain)		
Statement	45.25	44.53	Acknowledgment	19.13	5.45	Opinion	0.52	2.12	Reason	0.64	2.52
Pause	12.99	15.05	Accept	4.68	6.25	Wish	1.23	0.05	Condition	0.61	3.09
Backchannel	26.05	9.09	PartialAccept	0.02	0.10	Request	0.22	0.19	Elaboration	0.28	4.00
Y/N-Question	3.61	2.19	AffirmativeAnswer	0.08	0.20	Suggestion	0.16	1.12	Evaluation	1.35	2.01
WH-Question	1.13	0.40	Reject	0.25	0.11	Commitment	1.15	0.29	(Check)		
Open-Question	0.32	0.32	PartialReject	0.04	0.03				RepetitionRequest	0.07	0.03
OR-after-Y/N	0.05	0.02	NegativeAnswer	0.10	0.10				UnderstandingCheck	0.19	0.20
OR-Question	0.05	0.03	Answer	1.16	2.57				DoubleCheck	0.36	0.15
Statement==	9.91	27.79							ApprovalRequest	2.01	1.07

of the unit. Most of the tags in this layer are used to describe forward-looking functions. The tags are classified into three large groups: “Question,” “Fragment,” and “Statement.” “Statement==” denotes the continuation of the utterance.

Response: It is used to label responses directed to a specific previous utterance made by the addressee.

Check: It is used to label confirmations that are along a certain expected response.

Constrain: It is used to label utterances that restrict or complement the target of the utterance.

ActionDiscussion: It is used to label utterances that pertain to a future action.

Others: It is used to describe various functions of the utterance, e.g., Greeting, SelfTalk, Welcome, Apology, etc.

In the *General* layer, there are two sublayers:— (1) the *Pause* sublayer that consists of Hold, Grabber, Holder, and Releaser and (2) the *WH* sublayer that labels the WH-Question type.

It should be noted that this taxonomy is intended to be used for training spoken dialogue systems. Consequently, it contains detailed descriptions to elaborate on the decision-making process. For example, checks are classified into four categories because they should be treated in various ways in a dialogue system. *UnderstandingCheck* is often used to describe clarifications; thus, it should be taken into account when creating a dialogue scenario. In contrast, *RepetitionRequest*, which is used to request that the missed portions of the previous utterance be repeated, is not concerned with the overall dialogue flow.

An example of an annotation is shown in Table 1. Since the *Response* and *Constrain* layers are not

necessarily directed to the immediately preceding utterance, the target utterance ID is specified.

4.2 Evaluation

We performed a preliminary annotation of the speech act tags in the corpus. Thirty dialogues (900 min, 23,169 utterances) were annotated by three labellers. When annotating the dialogues, we took into account textual information, audio information, and contextual information. The result was cross-checked by another labeller.

4.2.1 Distributional Statistics

The frequencies of the tags, expressed as a percentages, are shown in Table 3. In the *General* layer, nearly half of the utterances were *Statement*. This bias is acceptable because 66% of the utterances had tag(s) of other layers.

The percentages of tags in the *Constrain* layer are relatively higher than those of tags in the other layers. They are also higher than the percentages of the corresponding tags of MRDA (Shriberg et al., 2004) and SWBD-DAMSL (Jurafsky et al., 1997).

These statistics characterize the consulting dialogue of sightseeing planning, where explanations and evaluations play an important role during the decision process.

4.2.2 Reliability

We investigated the reliability of the annotation. Another two dialogues (2,087 utterances) were annotated by three labelers and the agreement among them was examined. These results are listed in Table 4. The agreement ratio is the average of all the combinations of the three individual agreements. In the same way, we also computed the average Kappa statistic, which is often used to measure the agreement by considering the chance rate.

A high concordance rate was obtained for the *General* layer. When the specific layers and sublayers are taken into account, Kappa statistic was

Table 4: Agreement among labellers

	General layer	All layers
Agreement ratio	86.7%	74.2%
Kappa statistic	0.74	0.68

0.68, which is considered a good result for this type of task. (cf. (Shriberg et al., 2004) etc.)

4.2.3 Analysis of Occurrence Tendency during Progress of Episode

We then investigated the tendencies of tag occurrence through a dialogue to clarify how consulting is conducted in the corpus. We annotated the boundaries of episodes that determined the spots to visit in order to carefully investigate the structure of the decision-making processes. In our corpus, users were asked to write down their itinerary for a practical one day tour. Thus, the beginning and ending of an episode can be determined on the basis of this itinerary.

As a result, we found 192 episodes. We selected 122 episodes that had more than 50 utterances, and analyzed the tendency of tag occurrence. The episodes were divided into five segments so that each segment had an equal number of utterances. The tendency of tag occurrence is shown in Figure 1. The relative occurrence rate denotes the number of times the tags appeared in each segment divided by the total number of occurrences throughout the dialogues. We found three patterns in the tendency of occurrence. The tags corresponding to the first pattern frequently appear in the early part of an episode; this typically applies to Open-Question, WH-Question, and Wish. The tags of the second pattern frequently appear in the later part, this typically applies to Evaluation, Commitment, and Opinion. The tags of the third pattern appear uniformly over an episode, e.g., Y/N-Question, Accept, and Elaboration. These statistics characterize the dialogue flow of sightseeing planning, where the guide and the user first clarify the latter’s interests (Open, WH-Questions), list and evaluate candidates (Evaluation), and then the user makes a decision (Commitment).

This progression indicates that a session (or dialogue phase) management is required within an episode to manage the consulting dialogue, although the test-set perplexity², which was calcu-

²The perplexity was calculated by 10-fold cross validation of the 30 dialogues.

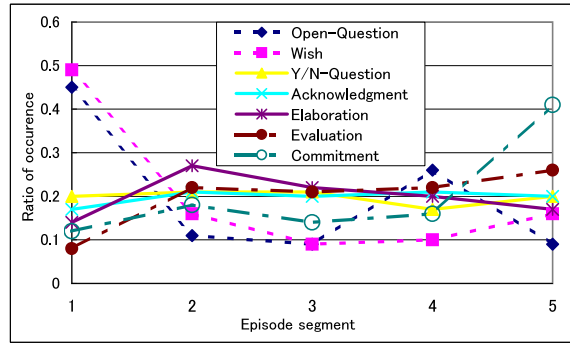


Figure 1: Progress of episodes vs. occurrence of speech act tags

lated by a 3-gram language model trained with the SA tags, was not high (4.25 using the general layer and 14.75 using all layers).

5 Semantic Content Tags

The semantic content tag set was designed to capture the contents of an utterance. Some might consider semantic representations by HPSG (Pollard and Sag, 1994) or LFG (Dalrymple et al., 1994) for an utterance. Such frameworks require knowledge of grammar and experiences to describe the meaning of an utterance. In addition, the utterances in a dialogue are often fragmentary, which makes the description more difficult.

We focused on the predicate-argument structure that is based on dependency relations. Annotating dependency relations is more intuitive and is easier than annotating the syntax structure; moreover, a dependency parser is more robust for fragmentary expressions than syntax parsers.

We introduced semantic classes to represent the semantic contents of an utterance. Semantic class labels are applied to each unit of the predicate-argument structure. The task that identifies the semantic classes is very similar to named entity recognition, because the classes of the named entities can be equated to the semantic classes that are used to express semantic content. However, both nouns and predicates are very important for capturing the semantic contents of an utterance. For example, “10 a.m.” might denote the current time in the context of planning, or it might signify the opening time of a sightseeing spot. Thus, we represent the semantic contents on the basis of the predicate-argument structure. Each predicate and argument is assigned a semantic category.

For example, the sentence “I would like to see

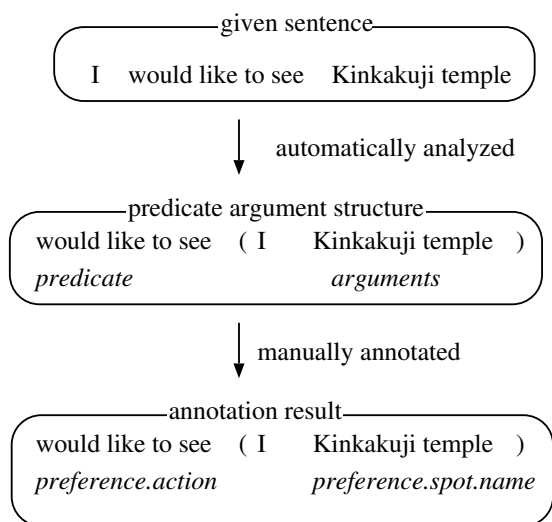


Figure 2: Example of annotation with semantic content tags

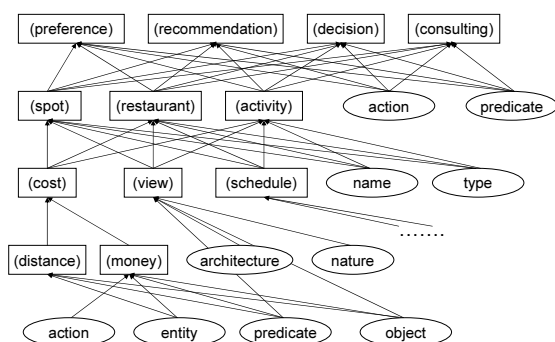


Figure 3: A part of the semantic category hierarchy

Kinkakuji temple.” is annotated as shown in Figure 2. In this figure, the semantic content tag *preference.action* indicates that the predicate portion expresses the speaker’s *preference* for the speaker’s action, while the semantic content tag *preference.spot.name* indicates the *name* of the *spot* as the object of the speaker’s *preference*.

Although we do not define semantic the role (e.g., object (*Kinkakuji temple*) and subject (*I*)) of each argument item in this case, we can use conventional semantic role labeling techniques (Gildea and Jurafsky, 2002) to estimate them. Therefore, we do not annotate such semantic role labels in the corpus.

5.1 Tag Specifications

We defined hierarchical semantic classes to annotate the semantic content tags. There are 33 labels (classes) at the top hierarchical level. The la-

bels are, for example, **activity**, **event**, **meal**, **spot**, **transportation**, **cost**, **consulting**, and **location**, as shown in Figure 3. There two kinds of labels, nodes and leaves. A node must have at least one child, a node or a leaf. A leaf has no children. The number of kinds for nodes is 47, and the number of kinds for leaves is 47. The labels of leaves are very similar to the labels for named entity recognition. For example, there are “year, date, time, organizer, name, and so on.” in the labels of the leaves.

One of the characteristics of the semantic structure is that the lower level structures are shared by many upper nodes. Thus, the lower level structure can be used in any other domains or target tasks.

5.2 Annotation of semantic contents tags

The annotation of semantic contents tags is performed by the following four steps. First, an utterance is analyzed by a morphological analyzer, ChaSen³. Second, the morphemes are chunked into dependency unit (*bunsetsu*). Third, dependency analysis is performed using a Japanese dependency parser, CaboCha⁴. Finally, we annotate the semantic content tags for each *bunsetsu* unit by using our annotation tool. An example of an annotation is shown in Table 1. Each row in column “Transcript” denotes the divided *bunsetsu* units.

The annotation tool interface is shown in Figure 4. In the left side of this figure, the dialogue files and each utterance of the dialogue information are displayed. The dependency structure of an utterance is displayed in the upper part of the figure. The morphological analysis results and chunk information are displayed in the lower part of the figure.

At present, the annotations of semantic content tags are being carried out for 10 dialogues. Approximately 22,000 paths, including paths that will not be used, exist if the layered structure is fully expanded. In the 10 dialogues, 1,380 tags (or paths) are used.

In addition, not only to annotate semantic content tags, but to correct the morphological analyze results and dependency analyzed results are being carried out. If we complete the annotation, we will also obtain these correctly tagged data of Kyoto tour guide corpus. These corpora can be used to develop analyzers such as morphological analyz-

³<http://sourceforge.jp/projects/chasen-legacy/>

⁴<http://chasen.org/~taku/software/cabocha/>

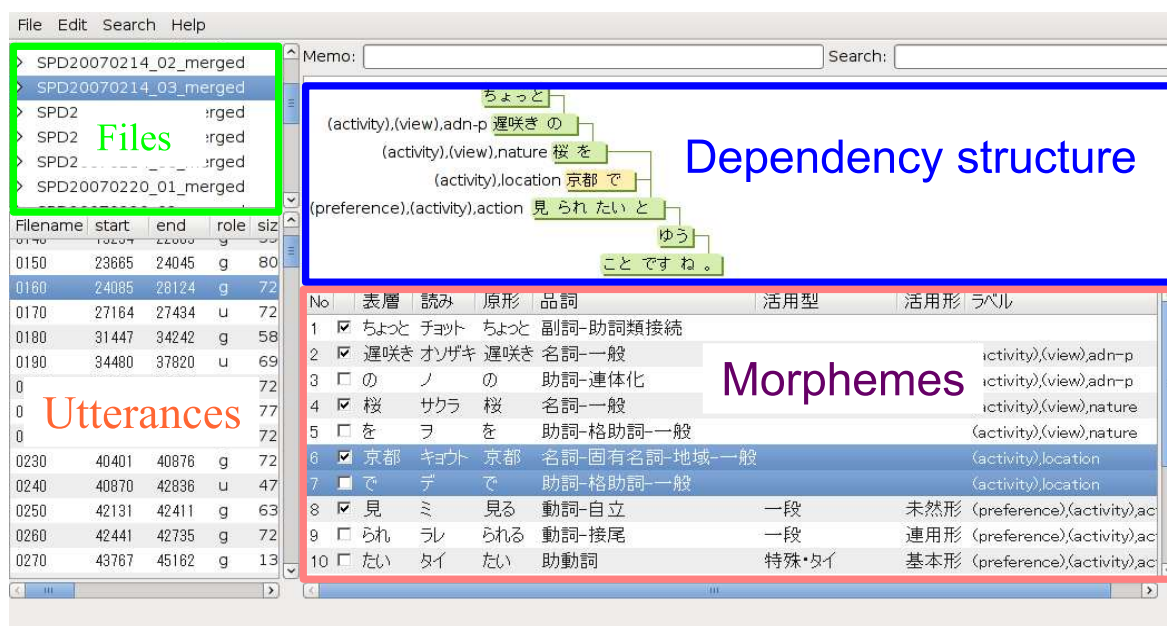


Figure 4: Annotation tool interface for annotating semantic content tags

ers and dependency analyzers via machine learning techniques or to adapt analyzers for this domain.

6 Conclusion

In this paper, we have introduced our spoken dialogue corpus for developing consulting dialogue systems. We designed a dialogue act annotation scheme that describes two aspects of a DA: speech act and semantic content. The speech act tag set was designed by extending the MRDA tag set. The design of the semantic content tag set is almost complete. If we complete the annotation, we will obtain speech act tags and semantic content tags, as well as manual transcripts, morphological analysis results, dependency analysis results, and dialogue episodes. As a preliminary analysis, we have evaluated the SA tag set in terms of the agreement between labellers and investigated the patterns of tag occurrences.

In the next step, we will construct automatic taggers for speech act and semantic content tags by using the annotated corpora and machine learning techniques. Our future work also includes a condensation or selection of dialogue acts that directly affect the dialogue flow in order to construct a consulting dialogue system using the DA tags as an input.

References

- Srinivas Bangalore, Giuseppe Di Fabbrizio, and Amanda Stent. 2006. Learning the structure of task-driven human-human dialogs. In *Proceedings of COLING/ACL*, pages 201–208.
- Gies Bouwman, Janienke Sturm, and Louis Boves. 1999. Incorporating Confidence Measures in the Dutch Train Timetable Information System Developed in the ARISE Project. In *Proc. ICASSP*.
- Johan Boye. 2007. Dialogue Management for Automatic Troubleshooting and Other Problem-solving Applications. In *Proc. of 8th SIGdial Workshop on Discourse and Dialogue*, pages 247–255.
- Harry Bunt. 2000. Dialogue pragmatics and context specification. In Harry Bunt and William Black, editors, *Abduction, Belief and Context in Dialogue*, pages 81–150. John Benjamins.
- Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III, and Anni e Zaenen, editors. 1994. *Formal Issues in Lexical-Functional Grammar*. CSLI Publications.
- George Ferguson and James F. Allen. 1998. TRIPS: An intelligent integrated problem-solving assistant. In *Proc. Fifteenth National Conference on Artificial Intelligence*, pages 567–573.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- Chiori Hori, Kiyonori Ohtake, Teruhisa Misu, Hideki Kashioka, and Satoshi Nakamura. 2008. Dialog Management using Weighted Finite-state Transducers. In *Proc. Interspeech*, pages 211–214.

- Daniel Jurafsky, Elizabeth Shriberg, and Debra Bisca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13. Technical report, University of Colorado at Boulder & SRI International.
- Hideki Kashioka and Takehiko Maruyama. 2004. Segmentation of Semantic Unit in Japanese Monologue. In *Proc. ICSLT-O-COCOSDA*.
- Lori F. Lamel, Samir Bennacef, Jean-Luc Gauvain, H. Dartigues, and J. N. Temem. 2002. User evaluation of the MASK kiosk. *Speech Communication*, 38(1):131–139.
- Lori Levin, Donna Gates, Dorcas Wallace, Kay Peterson, Along Lavie, Fabio Pianesi, Emanuele Pianta, Roldano Cattoni, and Nadia Mana. 2002. Balancing expressiveness and simplicity in an interlingua for task based dialogue. In *Proceedings of ACL 2002 workshop on Speech-to-speech Translation: Algorithms and Systems*.
- Kikuo Maekawa, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara. 2000. Spontaneous speech corpus of Japanese. In *Proceedings of the Second International Conference of Language Resources and Evaluation (LREC2000)*, pages 947–952.
- Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press.
- Kepa Joseba Rodriguez, Stefanie Dipper, Michael Götze, Massimo Poesio, Giuseppe Riccardi, Christian Raymond, and Joanna Rabięga-Wisniewska. 2007. Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus. In *Proc. Linguistic Annotation Workshop*, pages 148–155.
- Elizabeth Shriberg, Raj Dhillon, Sonali Bhagat, Jeremy Ang, and Hannah Carvey. 2004. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In *Proc. 5th SIGdial Workshop on Discourse and Dialogue*, pages 97–100.
- Blaise Thomson, Jost Schatzmann, and Steve Young. 2008. Bayesian update of dialogue state for robust dialogue systems. In *Proceedings of ICASSP '08*.
- Marilyn A. Walker, Rebecca Passonneau, and Julie E. Boland. 2001. Quantitative and Qualitative Evaluation of DARPA Communicator Spoken Dialogue Systems. In *Proc. of 39th Annual Meeting of the ACL*, pages 515–522.