Dialogue Structure Annotation of In-car Speech Corpus based on Speech-Act Tag

Shingo Kato
Dept. of Information Engineering,
Graduate School of Information Science,
Nagoya University
Furo-cho, Chikusa-ku, Nagoya
464-8601, Japan
gotyan@el.itc.nagoya-u.ac.jp

Shigeki Matsubara
Yukiko Yamaguchi
Nobuo Kawaguchi
Information Technology Center,
Nagoya University
Furo-cho, Chikusa-ku, Nagoya
464-8601, Japan

Abstract

This paper describes the dialogue annotation of an in-car speech corpus. According to the observations of CIAIR restaurant guide task, we introduced a new category and expressed the dialogue structure as a binary tree. 789 dialogues consisting of 8150 utterances are annotated.

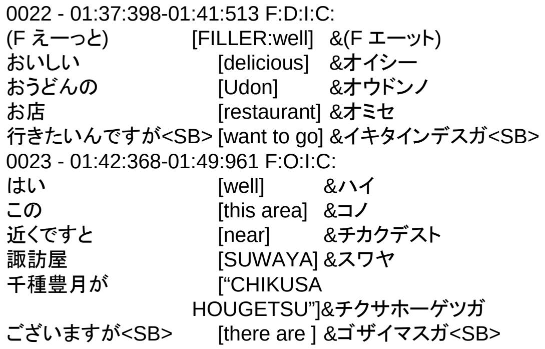
1 Introduction

With the improvement of speech processing technologies, some researches about spoken dialogue systems have been studied.

Spoken dialogue systems are required to understand the intentions of a user's utterances, the purpose of the dialogue, and its achievement state to execute a dialogue appropriately and cooperatively (Litman, 1990). We suppose that the system can figure out these things through the incremental building of the dialogue structure in real time. By using the structural rules and an existing technique for natural language processing, the dialogue structure can be built. One of the ways to acquire the rule is statistically dealing with the structurally annotated corpus.

In this paper, we describe the structural annotation of a spoken dialogue corpus. We use the restaurant guide dialogues in the CIAIR in-car spoken dialogue corpus (Irie, 2003; Kawaguchi, 2004; Kawaguchi, 2005). The speech-act tags which indicates the speaker's intention was provided for the transcription of the corpus. We describe the dialogue structure as a binary tree based on the tags. We semi-automatically annotated 789 dialogues consisting of 8150 utterances.

In section 2, we explain the CIAIR in-car spoken dialogue corpus and the speaker's intention tags. In sections 3 and 4, we discuss the design policy of a structurally annotated spoken dialogue corpus and the construction of the corpus.



```
0022 - 01:37:398-01:41:513 F:D:I:C:
(F えーっと)              [FILLER:well]  &(F エーット)
おいしい                    [delicious]   &オイシー
おうどんの                  [Udon]        &オウドンノ
お店                        [restaurant]  &オミセ
行きたいんですが<SB>        [want to go]  &イキタインデスガ<SB>
0023 - 01:42:368-01:49:961 F:O:I:C:
はい                        [well]       &ハイ
この                        [this area]  &コノ
近くですと                  [near]       &チカクデスト
諏訪屋                      [SUWAYA]     &スワヤ
千種豊月が                  [“CHIKUSA
                         HOUGETSU”]&チクサホーゲツガ
ございますが<SB>            [there are ] &ゴザイマスガ<SB>
```

Figure 1: Transcription of in-car speech dialogue

2 Spoken Dialogue Corpus and Layered Intention Tags

The Center for Integrated Acoustic Information Research (CIAIR), Nagoya University, has compiled a database of in-car speech and dialogue since 1999, in order to achieve robust spoken dialogue systems in actual usage environments (Kawaguchi, 2004; Kawaguchi, 2005). All dialogue data were transcribed according to transcription standards in compliance with CSJ (Corpus of Spontaneous Japanese) (Maekawa, 2000) and were assigned discourse tags such as fillers, hesitations, and slips. An example of a transcript is shown in Figure 1. Utterances were divided into utterance units by a pause of 200 ms or more.

These dialogues are annotated by speech act

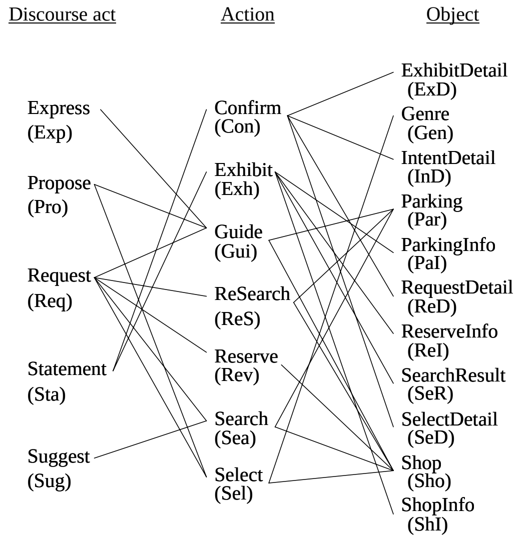


Figure 2: A part of the LIT

Table 1: Type and substance of POD's

POD	Substance
GENRE	choosing style of cuisine.
GUIDE	guidance to restaurant or parking.
P_INFO	extracting parking information such as vacant space, neighborhood.
P_SRCH	searching for a parking space.
S_INFO	extracting shop information such as price, reservation, menu, area, fixed holiday.
SLCT	selecting a restaurant or parking space.
SRCH	searching for a restaurant.
SRCH_RQST	requesting a search.
RSRV	making a reservation.
RSRV_DTL	extracting reservation information such as time, number of people, etc.
RSRV_RQST	requesting a reservation.

tags called Layered Intention Tags (LIT) (Irie, 2004(a); Irie, 2004(b)), which indicate the intentions of the speaker's utterances. LIT consists of four layers: "Discourse act", "Action", "Object", and "Argument". Figure 2 shows a part of the organization of LIT. As Figure 2 shows, the lower layered intention tag depends on the upper layered one. In principle, one LIT is given to one utterance unit. In this research, we use parts of the restaurant guide dialogues between a driver and a human operator. An example of the dialogue corpus with LIT is shown in Table 2. In the *Speaker* column, "D" means a driver's utterance and "O" means an operator's one. Because the "Argument" layer is too detailed to express the dialogue structure, we omitted it. So, we used the Discourse act, Action, and Object layers and extended them with speaker symbols such as *"D+Request+Search+Shop"*. There are 41 types of extended LIT.

3 Dialogue Structure Description

3.1 Dialogue structure

In this research, we assume that the fundamental unit of a dialogue is an utterance to which one LIT is given. We defined a category called POD (Part-Of-Dialogue), according to the observations of the restaurant guide task, that was especially focused on what subject was dealt with. As a result, 11 types of POD were built (Table 1). We express the dialogue structure as a binary tree because of the following two points. One is that these dialogues were had by two participants, a driver and a human operator. Another is to make the structural analyzing process of the dialogue more easy. Each node of a structural tree is labeled with a POD or LIT. The dialogue structural tree of Table 2 is shown in Figure 3.

3.2 Design of dialogue structure description

Before the annotation was started, repairs and corrections should be eliminated. Because we considered a dialogue as a LIT sequence, and LIT couldn't be provided for them.

The annotation of the dialogue structure was done in the following way.

Merging utterances: When two adjoining utterances such as request and answer, they seem to be able to pair up and merge with an appropriate POD. In Table 2, for example, the utterance "Should I make a reservation?" (#286) is a request and the answer to #286 is "No, a reservation is not necessary"(#287). In this way, utterances are combined with the POD "S_INFO".

When the LIT's of two adjacent utterances are corresponding, these utterances are supposed to be paired and merged with the same LIT. Utterance "Fresh and roe" (#280) and "I want to have Hotpot" (#281) are related to choosing the style of cuisine, so they were provided with the same LIT.

Table 2: Example of the dialogue corpus with LIT

Utterance Number	Speaker	Transcription	LIT		
			First layer (Discourse Act)	Second layer (Action)	Third layer (Object)
277	D	kono hen de tai ga tabera reru tokoro nai kana. (I’d like to eat some sea bream.)	Request	Search	Shop
278	O	hai. (Let me see.)	Statement	Exhibit	IntentDetail
279	O	o ryori wa donna o ryouri ga yorosi katta desuka. (Which kind do you like?)	Request	Select	Genre
280	D	nama kei ga ii kana. (Fresh and roe.)	Statement	Select	Genre
281	D	Nabe ga tabe tai desu. (I want to have a Hotpot.)	Statement	Select	Genre
282	O	hai kono tikaku desu to tyankonabe to oden kaiseki ato syabusyabu nado ga gozai masu ga. (Well, there are restaurants near here that serve sumo wrestler’s stew, Japanese hotpot, and sliced beef boiled with vegetables.)	Statement	Exhibit	SearchResult
283	D	oden kaiseki ga ii. (I love Japanese Hotpot.)	Statement	Select	Genre
284	O	hai sou simasu to “MARU” to iu omise ni nari masu ga. (“MARU” restaurant is suitable.)	Statement	Exhibit	SearchResult
285	O	yorosi katta de syou ka. (How about this?)	Request	Exhibit	IntentDetail
286	D	yoyaku wa hituyou ari masu ka. (Should I make a reservation?)	Request	Exhibit	ShopInfo
287	O	a yoyaku no hou wa yoyoku sare naku temo o mise ni wa hairu koto ga deki masu ga. (No, a reservation is not necessary.)	Statement	Exhibit	ShopInfo
288	D	a zya soko made annai onegai si masu. (I see. Please guide me there.)	Request	Guide	Shop
289	O	kasikomari masi ta. (Sure.)	Statement	Exhibit	IntentDetail
290	O	sore dewa “MARU” made go annnai itasi masu. (Now, I’m navigating to “MARU”)	Express	Guide	Shop
291	D	hai. (Thanks.)	Statement	Exhibit	IntentDetail

Therefore they are combined with the LIT “*D+Statement+Select+Genre*”.

Merging partial dialogues: When two adjoining partial dialogues (i.e. a partial tree) are composing another partial dialogue, they are merged with a proper POD. In Table 2, for example, a search dialogue (from #277 to #285, SRCH) and a shop information dialogue helping search (from #286 to #287, S_INFO) are combined and labeled as the POD “SLCT”.

When the POD’s of two adjacent partial dialogues are corresponding, these dialogues are merged with the same POD. Two search dialogues (one is from #277 to #282, other is from #283 to #285) are combined with the same POD “SRCH”.

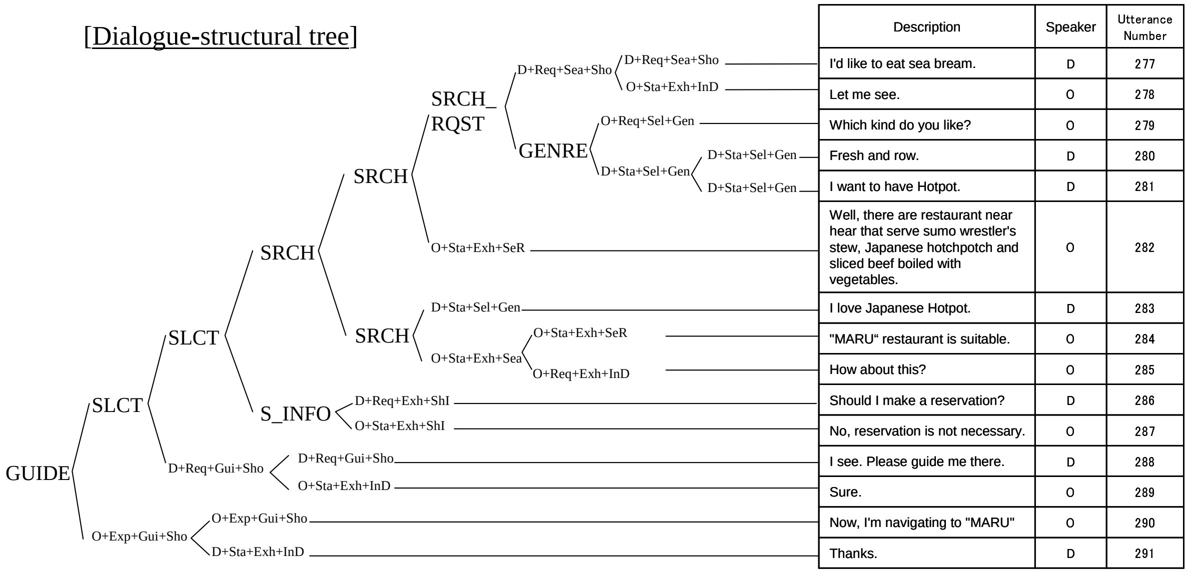
The root of the tree: The POD of the root of the tree is “GUIDE”, because the domain of the corpus is restaurant guide task.

4 Dialogue Structure Annotation

4.1 Work environment and procedures

We made a dialogue parser as a supportive environment for annotating dialogue structures.

Applying the dialogue-structural rules, which


[Dialogue-structural tree]

Description	Speaker	Utterance Number
I'd like to eat sea bream.	D	277
Let me see.	O	278
Which kind do you like?	O	279
Fresh and row.	D	280
I want to have Hotpot.	D	281
Well, there are restaurant near hear that serve sumo wrestler's stew, Japanese hotchpotch and sliced beef boiled with vegetables.	O	282
I love Japanese Hotpot.	D	283
"MARU" restaurant is suitable.	O	284
How about this?	O	285
Should I make a reservation?	D	286
No, reservation is not necessary.	O	287
I see. Please guide me there.	D	288
Sure.	O	289
Now, I'm navigating to "MARU"	O	290
Thanks.	D	291

[Dialogue-structural rules]

GUIDE→SLCT O+Exp+Gui+Sho
SLCT→SLCT D+Req+Gui+Sho
SLCT→SRCH S_INFO
SRCH→SRCH SRCH
SRCH→SRCH_RQST O+Sta+Exh+SeR
SRCH_RQST→D+Req+Sea+Sho GENRE
D+Sta+Exh+SeR→O+Sta+Exh+SeR O+Re+Exh+InD
GENRE→O+Req+Sel+Gen D+Sta+Sel+Gen
S_INFO→D+Req+Exh+ShI O+Sta+Exh+ShI
D+Sta+Sel+Gen→D+Sta+Sel+Gen D+Sta+Sel+Gen
D+Req+Gui+Sho→D+Req+Gui+Sho O+Sta+Exh+InD
D+Req+Sea+Sho→D+Req+Sea+Sho O+Sta+Exh+InD
O+Exp+Gui+Sho→O+Exp+Gui+Sho D+Sta+Exh+InD

Figure 3: Dialogue-structural tree and rules for Table 2

are obtained from annotated structural trees (like Figure 3.), the parser analyzes the inputs of the LIT sequences and outputs all available dialogue-structural trees. An annotator then chooses the correct tree from the outputs. When the outputs don't include the correct tree, the annotator should rectify the wrong tree rewriting the list form of the tree. In this way, we make the annotation more efficient.

The dialogue parser was implemented using the bottom-up chart parsing (Kay, 1980). The structural rules were extracted from all annotated dialogues. In the environment outlined above, we have worked at bootstrap building. That is, we

1. outputed the dialogue structures through the parser.

2. chose and rectified the dialogue structure using an annotator.

3. extracted some structural rules from some dialogue-structural trees.

Table 3: Corpus statistics

number of dialogues	789
number of utterances	8150
number of structural rules	297
utterances per one dialogue	11.61
number of dialogue-structural tree types	659
number of LIT sequence types	657

We repeated these procedures and increased the structural rules incrementally, so that the dialogue parser improved it's operational performance.

4.2 Structurally annotated dialogue corpus

We built a structurally annotated dialogue corpus in the environment described in Section 4.1, using the restaurant guide dialogues in the CIAIR corpus. The corpus includes 789 dialogues consisting of 8150 utterances. One dialogue is composed of 11.61 utterances. Table 3 shows them in detail.

5 Conclusion

In this paper, we described the dialogue annotation of in-car speech corpus based on speech-act tag. From observating the restaurant guide dialogues, we designed the policy of the dialogue structure and annotated 789 dialogues consisting of 8150 utterances.

6 Acknowledgments

The authors would like to thank Ms. Yuki Irie for her valuable comments about the design of the dialogue structure. This research was partially supported by the Grant-in-Aid for Scientific Research (No. 15300045) of JSPS.

References

D. J. Litman and J. F. Allen : Discourse Processing and Commonsense Plans. Phillip R. Cohen, Jerry Morgan, Martha E. Pollack, editors. Intentions in Communication. pp.365-388, MIT Press, Cambridge, MA, 1990.

K. Maekawa, H. Koiso, S. Furui, and H. Isahara: Spontaneous speech corpus of Japanese, LREC-2000, pp.947-952, 2000.

M. Kay: Algorithm Schemata and Data Structures in Syntactic Processing, TR CSL-80-12, Xerox PARC, 1980.

N. Kawaguchi, K. Takeda, and F. Itakura: Multimedia corpus of in-car speech communication. J. VLSI Signal Processing, vol.36, no.2, pp.153-159, 2004.

N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura: CIAIR In-Car Speech Corpus -Influence of Driving States-. IEICE Trans. on Information and System, E88-D(3), pp.578-582, 2005.

Y. Irie, N. Kawaguchi, S. Matsubara, I. Kishida, Y. Yamaguchi, K. Takeda, F. Itakura and Y. Inagaki: An Advanced Japanese Speech Corpus for In-car Spoken Dialogue Research, Proceedings of Oriental COCOSDA-2003, pp.217-224, 2003

Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki: Design and Evaluation of Layered Intention Tag for In-Car Speech Corpus, Proceedings of Oriental COCOSDA-2004, pp.82-86, 2004

Y. Irie, S. Matsubara, N. Kawaguchi, Y. Yamaguchi, and Y. Inagaki: Speech Intention Understanding based on Decision Tree Learning, Proceedings of 8th International Conference on Spoken Language Processing, Cheju, Korea, 2004.