TR-IT-0150

Japanese Discourse Representation using Communicative Act Labels

Mutsuko TOMOKIYO

March 1996

Abstract

We previously investigated Japanese and English utterances focusing on their Communicative Acts, and presented a paraphrase-based methodology for discovering and revising sets of cue-based CAs for labeling spontaneous dialogues. This report extends the investigation of automatic CA analysis in two directions. First, we note a significant relation between segmentation based upon cue patterns and segmentation based upon natural pauses: 77% of the pause-bounded segments in our corpus coincide with segments defined by CAs. Second, we describe attempts to parse CA groupings, and thus to analyze discourse structures, using context-free rules. Experiments on ten spontaneous Japanese dialogues show that the obtained groupings are useful for resolution of referring expressions of ellipses and that pro-forms and their referents fall within certain groupings about 78.5% of the time.

Keywords

communicative goal, pause, CA units, pause units, discourse structure representation, speech act

ATR Interpreting Telecommunications Research Laboratories

Acknowledgments

I express my appreciation to Dr. Y.Yamazaki, president of ATR-ITL and Mr. T.Morimoto, department head, for their support in this research. Thanks also should go to Dr. M.Seligman, Mr. T.Iwadera and Mr. Y.Kitagawa for their contributions to this research.

Contents

List of Tables

List of Figures

1 Introduction

This report aims to describe discourse structure analysis using Communicative Act labels (CAs), which we have established for use in processing naturally-spoken dialogue. In general, dialogue processing involves speech recognition, syntactico-semantic analysis, discourse analysis, etc. Discourse analysis is especially important in machine translation or in a man-machine interactive system. However, for discourse analysis, one needs new information from a different paradigm like pragmatics in addition to syntactic or semantic information. So, speech acts of utterances are often used as a way of analyzing discourse structures. We began by investigating collected dialogues from the point of view of their speech acts for Japanese and English, and made a tentative set of CA labels.

We previously presented a paraphrase-based methodology for discovering and revising sets of cue-based CAs for labeling spontaneous dialogues. The discovery procedure was applied to 16 English and 16 Japanese spontaneous dialogues concerning direction finding and hotel accommodations. Labels were then compared cross-linguistically by observing translation relationships among cue phrases. It was found that 27 CAs were needed to cover both corpora at the desired level of specificity: 25 bilingual and 2 monolingual. Following up on this study carried out by hand, [15] [18] [20] and [31] described attempts to use cue patterns for automatic utterance segmentation, and CA assignment to segmented units.

This report extends the investigation of automatic CA analysis in two directions. First, we note a significant relation between segmentation based upon cue patterns and segmentation based upon natural pauses [23]; 77% of the pause-bounded segments in our corpus coincide with segments defined by CAs [29]. Second, we describe attempts to parse CA groupings, and thus to analyze discourse structures, using context-free rules. Experiments on ten spontaneous Japanese dialogues show that the obtained groupings are useful for resolution of referring expressions or ellipses and that pro-forms and their referents fall within certain groupings about 78.5% of the time [28].

2 Overview of CA Label Set

A Communicative Act is a communicative goal which can be expressed in a given language by a distinctive set of conventional cue patterns in specified discourse contexts. Communicative Acts are similar to speech acts or illocutionary force types [1]. However, we restrict our attention to communicative goals which can be explicitly expressed via conventional surface cue patterns, thus excluding goals which can only be defined in terms of relations between utterances [3][10].
A tentative set of CAs is verified by being assigned to Japanese dialogues and being compared with pause units from the waveform in speech recognition for the same dialogues.

2.1 CA definition and its discovering methods

We defined the Communicative Act as follows[1]:

> A Communicative Act is a communicative goal which can be expressed in a given language by a distinctive set of conventional cue patterns in the specified discourse context [12][19][32].

Inform, Request and Yn-question are typical Communicative Acts. The communicative goals which they respectively represent are roughly "speaker wishes to convey new information to hearer"; "speaker wishes to convey to hearer that speaker wishes hearer to perform some action"; and "speaker wishes to convey to hearer that speaker wishes hearer to convey to speaker whether a specified proposition is true or not". According to the shared conventions of Japanese and English, Inform can be expressed by a declarative clause syntax like "[noun +] となっています (tonatteimasu)". Request can be expressed by using expressions like "[bare infinitive +] たいのですが (tainodesuga)" in Japanese and "would/will/could you [+ verb phrase or bare infinitive]" in English. Yn-question can be expressed by using expressions like "[verb +] ますか (masuka)" or "[noun +] ですか (desuka)" in Japanese and an inverted interrogative clause syntax and/or special prosody in English.

Communicative Acts are similar to speech acts or illocutionary force types [1][3][4]. However, we use this new terminology to stress several differences in principle.
The most important difference arises because we wish to explore the limits of discourse structure analysis based on surface cues. We restrict our attention to only those communicative goals which can be expressed using conventional linguistic cue patterns, that is, fixed cue patterns which can be memorized and used repeatedly as part of the speakers' shared knowledge of a given language. We reserve the term Communicative Act for only such conventionally expressible goals. Communicative goals that cannot be described as Communicative Acts include utterance goals which are expressed non-conventionally (using one-time-only combinations); goals which are expressed only implicitly; or goals which can only be defined in terms of relations between utterances [5].
The methods to discover a CA set for corpus C in a given language, are mentioned in detail in [19] and [32].

CA must be configurations of cues which are used *repeatedly* in a corpus to express a given communicative goal, rather than expressions composed one time only according to the productive capacity of the language.
We assume that the automatic mapping between a cue pattern and a communicative goal is listed for a program, as an element of linguistic competence.

[1]This section is overlapped with [19].

2.2 Tentative CAs set

We have determined 27 CA labels for Japanese and English. Japanese cue patterns currently number 246; English cue patterns indeed are comparatively less. Confirmation-question and Action-request from the set of CA labels are displayed and explained below [12][15]. → *Vide* Appendix 2, also.

- explanation: A gloss giving the approximate sense of the label: S means Speaker , and H means Hearer.
- illocutionary verbs: Verbs or verb phrases that can be used to talk about the Communicative Act.
- pattern: Cue patterns for the CA.
- parameter: Dimensions along which the CA can vary while still preserving its essential function.
- example: Taken from the EMMI dialogue corpus [16][17].

explanation	S informs H that S wants H to provide information about the correctness or incorrectness of a quoted proposition.
illocutionary verbs	確認する (kakuninsuru), request confirmation
patterns	ですね (desune), [clause], is it right?, tag question
parameters	politeness
example	十六日にご出発ですね (16 nichini gosyuppatsu desune): You will leave on the 16th, right?)

Table 1: Confirmation-question

explanation	S informs H that S wants H to perform some action, but without authority.
illocutionary verbs	要求する (youkyuusuru), request
patterns	てください (tekudasai), たいのですが (tainodesuga), 願います (negaimasu), will/would/can/could you VP ?
parameters	politeness, formality
example	そのプリントアウトがいただきたいのですが (sono purintoautoga itadakitainodesuga): Please give me a printout of this map.

Table 2: Action-request

3 Communicative Act Units

3.1 Utterance segmentation and label assignment

We intend to use the proposed label set for various purposes. One is to get more manageable utterance units by segmenting utterances and labeling the segmented units for Japanese spoken dialogues with a view toward representing their discourse structure.
This idea arose from the following Japanese utterance peculiarities:

- Japanese utterances consist of multiple sentences which are connected by auxiliary sequences as well as conjunctions or adverbs. From this one difficulty in spoken Japanese analysis emerges [7][11][14].
- Distinctive cue patterns which express communicative goals are, in principle, located at the end of sentences or clauses in Japanese from the point of view of word order.

Cue patterns yield units similar to *sentences* or *clauses* in Japanese standard grammar, when used to segment utterances. So, the segmentation of utterances by CA cue patterns enables us to get convenient, labeled units, thus making any dialogue analysis easier.
Segmentation of utterances and CA label assignment to CA units are automatically performed on an analyzer called the KK analyzer [20]. This analyzer uses rules which disambiguate the many-to-many mapping relationships existing between cue patterns and CA labels.
The performance of automatic utterance segmentation has an average success rate of 98.5 %. The performance of automatic CA label assignment has an average success rate of 86.5 % [15][18]. The 1.5% of segmentation errors involve ambiguities related to the short pattern "はい (hai)", which can express a positive response, an acknowledgment, etc.

> *e.g.*
> そちらへ は い ろんな行き方があります (sochiraheha ironnaikikatagaarimasu.)
> (You can travel a number of different ways.)

In the example, は of は い is a particle which makes the noun phrase そちらへは, and い of は い is the head phoneme of adjective いろんな. We have no grammar, nor lexicon for the segmentation, so we can not check ろんな, which is not word. These segmentation experiments have been carried out using transcripts prepared off line. However, for spoken language systems, the ultimate goal is to process utterances as they are spoken. For such on-line processing, it would be desirable to exploit some aspects of prosody to aid in the analysis. Thus it would be useful to seek correlations between prosodic and pragmatic aspects of utterances.

3.2 Comparison of CA units with pause units

CA units are defined as follows from the point of view of standard Japanese grammar:

if $\exists$ simple sentence in a turn,
then: cue pattern $\subseteq$ CA unit $\equiv$ simple sentence,
if not, then: CA unit $\equiv$ a cue pattern or : CA unit $\equiv$ a turn

The pause as defined here is a silent period in a speech utterance of more than 100 milliseconds in a waveform of speech recognition [21][23]. So, a pause unit is a speech period from a pause to the next pause or to the end of turn[2] in an utterance.

3.2.1 Used data

Table 8 displays the data used to compile correlations between CA units and pause units [33][34]. Tagged Japanese phonemes and their onset and ending times appear in the first column. Reading down, we see the word "m o sh i m o sh i" ("hello ?"), which forms a cue pattern for the Greet CA, followed by a pause. The second column gives hiragana transcriptions[3]. The appropriate CA label appears in the third column.

→ *Vide* Appendix 1, also.

speech recognition results	transcription	CA labels
260.0 a 370.0	あ	Expressive
370.0 pau 505.0	*pause*	
505.0 m 525.0	も	
525.0 o 585.0		
585.0 sh 665.0	し	
665.0 i 695.0		
695.0 m 735.0	も	
735.0 o 800.0		
800.0 sh 935.0	し	Greet
935.0 i 1080.0		
1080.0 pau 1265.0	*pause*	

Table 3: Data used for comparing CA units & pause units

[2]The end of a turn is considered to be a kind of pause.

[3]The used data called TAC22011 comes from the ATR dialogue corpus.

3.2.2 Collation of CA units with pause units

In the subcorpus studied, containing 151 CAs and 176 segments bounded by pauses of at least 100 milliseconds or turn boundaries, 77% of the pause units coincided with CA units. That is, among the 176 pause units 135 pause units ended with CA final character sequences, and 41 units ended in other ways.
Table 4 shows the percentages of each CA unit which is bounded by pauses.

given CA label	paused/total	%
Expressive	32/34	94%
Greet	2/2	100%
Temporizer	4/6	67%
Acknowledge	10/16	62%
Inform	9/38	24%
YES	12/14	86%
Thank	1/1	100%
Alert	1/1	100%
Apology	1/1	100%

Table 4: Percentage of each CA coinciding with pause units

undefined category	(un)paused/total	%
n-spell,n-num	13/18	72.0%
mod-n	4/28	14.0%
pp-particle	22/59	37.0%
p-kakari-wa	1/7	14.0%
p-rentai-no	1/29	3.45%
adv	2/13	15.0%
np	3/126	2.38%
(p-conj-syusi)	1/20	5.00%
(cl)	4/62	6.45%
???	1/1	100%

Table 5: Percentage of each non-CA constituent coinciding with pause-units

The 23% of the pause-bounded segments in our sample which did not contain CAs contained instead a variety of syntactic constituents. Occurrence counts appear in Table 5 (as the leftmost numbers of the middle column). The most frequent pause-bounded constituents were postpositional particles (pp-particle), with 22 out of 176 pause units. Next, with 13 occurrences, were nouns related to spelling or telephone numbers. Table 5 also shows for each constituent type the proportion of occurrences bounded by pauses. Spelling or phone numbers were pause-bounded in most instances (13 out of 18, or 68%), while most of the other constituents were pause-bounded quite rarely. For example, noun phrases were pause-bounded in only 3 of 126 cases. Overall, our impression was that

those pause units not associated with CAs could be divided among (1) constituents purposely paused to allow listener processing, such as telephone numbers and (2) constituents paused only irregularly, perhaps for reasons related to the speaker's own processing. Hence, if we can could write a grammar sensitive to pauses, we consequently could apply this grammar to the syntactico-semantic analysis of CA units. Thus, we would be able to use the same grammar for speech recognition and syntactico-semantic parsing for Japanese dialogues, and this would enable us to make a linkage of the two processings.

4 Using a Pause-based Grammar

We have a speech recognition grammar at ATR, called *subtree grammar* [25], which is an ordinary CFG, but sensitive to pauses.

This paused-based grammar is different from ordinary syntactical analysis grammars, *e.g.* HPSG style grammar and Bunsetsu[4]-based grammar, *e.g.* speech recognition CFG style grammar on ASURA at ATR, in the following respect.

- The end of a sentence[5] is not a comma, but a pause. Consequently, noun fragments or noun phrases are *sentences*, when they are bounded by a pause or it is the end of the turn. Adjectives, adjectival phrases, adverbs and adverbial phrases also are *sentences*, when they are bounded by a pauses or it is the end of the turn.

- The following nouns are subcategorized: spelling, family name, first name, telephone number, ward name, prefecture name, country name, etc., because these nouns are bounded often by pauses.

This pause-based grammar aims neither to make syntactico-semantic parsing nor discourse structure analysis, but to filter ambiguities in speech recognition. However, it is possible to use this grammar for syntactico-semantic parsing. Therefore, our basic thought was to connect a speech recognition phase with a syntactico-semantic parsing phase and discourse analysis, by using a grammar. As such, we applied this grammar to the syntactico-semantic analysis of CA units[6], because CA units are often bounded by pause units as mentioned above.

desuThe following displays a syntactico-semantic parsing process for a CA unit by using the pause-based grammar. "pau" in the rules means pauses in an incoming utterance.

[4]The Bunsetsu is a constituent unit of Japanese phrases, and it consists of *jiritsugos* (verbs, nouns, adjectives, adverbs and interjections) or *jiritsugos* followed by functional words.

[5]The *sentence* is defined here as a processing unit of utterances in the syntactico-semantic analysis.

[6]For this purpose, we must grammarize the difference between pause units and CA units:

- to specify noun fragments such as telephone numbers, addresses, administrative districts, etc., because noun fragments are not CAs[7].

- to specify an adjectival phrase containing cardinal or ordinal numbers, because adjectival phrases are not CA units.

- Utterance: このホテルですね (kono hoteru desune)(This hotel, right ?)
- Used rules:

```
(<start> <--> (<_start>))
(<_start> <--> (q1 <sent> q2))
(<sent> <--> (<cl> pau))
(<cl> <--> (<vp-sfp-dir-obj>))
(<vp-sfp-dir-obj> <--> (<vaux-sfp-dir-obj>))
(<vaux-sfp-dir-obj> <--> (<vaux-dir-obj-syusi> <aux-sfp>))
(<vaux-dir-obj-syusi> <--> (<vaux-np-syusi>))
(<vaux-np-syusi> <--> (<np> <aux-cop-desu-syusi>))
(<aux-cop-desu-syusi> <--> (<auxstem-desu> <vinfl-spe-su>))
(<np> <--> (<n-hutu>))
(<n-hutu> <--> (<mod-n> <n-hutu>))
(<mod-n> <--> (<rentai>))
(<rentai> <--> (k o n o) ("この" "この" "連体詞"))
(<n-hutu> <--> (h o t e r u) ("ホテル" "ホテル" "普通名詞"))
(<auxstem-desu> <--> (d e) ("で" "です" "助動詞" "語幹"))
(<vinfl-spe-su> <--> (s u) ("す" "です" "助動詞" "語尾"))
(<aux-sfp> <--> (n e) ("ね" "ね" "終助詞"))
```

- Input: q1 k o n o h o t e r u d e s u n e pau q2
- Output:

```
parse(q1) [0]
 [shift]
parse(k) [0 1]
 [shift]
parse(o) [0 1 3]
 [shift]
parse(n) [0 1 3 15]
 [shift]
parse(o) [0 1 3 15 25]
 [shift]
parse(h) [0 1 3 15 25 31]
 [reduce] (12) <rentai> --> k o n o
 [reduce] (11) <mod-n> --> <rentai>
 [shift]
parse(o) [0 1 5 6]
 [shift]
parse(t) [0 1 5 6 17]
 [shift]
parse(e) [0 1 5 6 17 26]
 [shift]
```

```
parse(r) [0 1 5 6 17 26 32]
 [shift]
parse(u) [0 1 5 6 17 26 32 34]
 [shift]
parse(d) [0 1 5 6 17 26 32 34 35]
 [reduce] (13) <n-hutu> --> h o t e r u
 [reduce] (10) <n-hutu> --> <mod-n> <n-hutu>
 [reduce] (9) <np> --> <n-hutu>
 [shift]
parse(e) [0 1 8 18]
 [shift]
parse(s) [0 1 8 18 27]
 [reduce] (14) <auxstem-> --> d e
 [shift]
parse(u) [0 1 8 19 28]
 [shift]
parse(n) [0 1 8 19 28 33]
 [reduce] (15) <vinfl-spe-su> --> s u
 [reduce] (8) <aux-cop-desu-syusi> --> <auxstem-desu> <vinfl-spe-su>
 [reduce] (7) <vaux-np-syusi> --> <np> <aux-cop-desu-syusi>
 [reduce] (6) <vaux-dir-obj-syusi> --> <vaux-np-syusi>
 [shift]
parse(e) [0 1 10 21]
 [shift]
parse(pau) [0 1 10 21 30]
 [reduce] (16) <aux-sfp> --> n e
 [reduce] (5) <vaux-sfp-dir-obj> --> <vaux-dir-obj-syusi> <aux-sfp>
 [reduce] (4) <vp-sfp-dir-obj> --> <vaux-sfp-dir-obj>
 [reduce] (3) <cl> --> <vp-sfp-dir-obj>
 [shift]
parse(q2) [0 1 13 23]
[reduce] (2) <sent> --> <cl> pau
 [shift]
parse($) [0 1 14 24]
 [reduce] (1) <_start> --> q1 <sent> q2
*[accept]
Success [1]
[1]
  (12) <rentai> --> k o n o
  (11) <mod-n> --> <rentai>
  (13) <n-hutu> --> h o t e r u
  (10) <n-hutu> --> <mod-n> <n-hutu>
  (9) <np> --> <n-hutu>
  (14) <auxstem-desu> --> d e
  (15) <vinfl-spe-su> --> s u
```

```
(8) <aux-cop-desu-syusi> --> <auxstem-desu> <vinfl-spe-su>
(7) <vaux-np-syusi> --> <np> <aux-cop-desu-syusi>
(6) <vaux-dir-obj-syusi> --> <vaux-np-syusi>
(16) <aux-sfp> --> n e
(5) <vaux-sfp-dir-obj> --> <vaux-dir-obj-syusi> <aux-sfp>
(4) <vp-sfp-dir-obj> --> <vaux-sfp-dir-obj>
(3) <cl> --> <vp-sfp-dir-obj>
(2) <sent> --> <cl> pau
(1) <_start> --> q1 <sent> q2
```

4.1 CA unit representation

When represented in the form of f-structures of attributes and values, the above analysis result is as follows [27]:

```
[[SEM [[RELN *ne*]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJ [[RELN *desu*]
             [OBJE [[RELN *hoteru*]
                    [RESTR [[RELN *kono*]]]]]]]]]]
```

The f-structures are rewritten into the following f-structures integrated with a CA label on the RWS system [9][12]. In these f-structures, the value of RELN (relation name), *i.e.* Confirmation-question, is the communicative goal of the unit. [AGEN *SPEAKER*] and [RECP *HEARER*] are set up as the environment in which the utterance is uttered. The syntactico-semantic f-structures of the unit are embedded as the value of [OBJ .....]. In this way, a pause unit is rewritten into the representation of a CA unit [22].

```
[[SEM [[RELN Confirmation-question]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJ ?OBJE]]]]
```

N.B.

"ですね"(desune) is ambiguous between Confirmation-question and Inform when its communicative act is assigned, if there was no prosody information. Therefore, f-structures of a Confirmation-question unit containing "ですね" should actually be as follows:

In Fig. 1, Confirmation-question is assigned to this CA unit, and its prosodic cue is "ね" which is uttered with rising intonation. The "hump's-height" is

```
[[SEM [[RELN Confirmation-question]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJ ?OBJE]
       [RESTR [[RELN F0 contour]
               [prosodic-cues [[cue ね]]
               [hump's-height Raising]
               [speech-rate !Z]]]]]]
```

Figure 1: F-structures of Confirmation-question unit containing "ですね"

a waveform presenting speech frequencies for the x-axis and time for the y-axis. It is interpreted as rising intonation with a high numerical value and as a falling intonation with a low numerical value. The "speech-rate" is the number of morae for a fixed range in the utterance. It is interpreted as a high waveform according to the increasing mora count.

Incidentally, the Action-request CA also is as follows:

e.g.
ホテルにはこのように行って下さい (hoteruniha konoyouni ittekudasai)
(Go to the hotel like this.)

```
[[SEM [[RELN *下さい (kudasai)*]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJ [[RELN *行 (iku)*]
             [AGEN *HEARER*]
             [GOAL *ホテル (hoteru)*]
             [RESTR [[MANN *このように (konoyouni)*]]]]]]]]
```

These syntactico-semantic f-structures are rewritten into the following structures, by adding the CA label to the top level of f-structures.

```
[[SEM [[RELN  Action-request]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJ ?OBJ]]]]
```

When a CA label is integrated with the syntactico-semantic representation of a CA unit, it follows that the communicative goal of the CA unit is also represented. This enables us to make a link from syntactico-semantic analysis to discourse analysis.

5 Discourse Structure Representation

We finished the previous section by introducing a syntactico-semantic representation including communicative goals for CA units.
Let's sum up the process of our discourse structure analysis.
First, utterances are segmented into units by conventional cue patterns. Second, each unit is automatically assigned a CA label. Finally, the units are aggregated into groups in a way that the discourse structure of the dialogue is made apparent.
What is a discourse structure (DS) representation? We consider that a discourse structure can be represented by building a sequence of *states of focus of the participants' attention as the discourse unfolds* [8].
The state of focus of the participants' attention involves diverse constituent elements of dialogues: syntactico-semantics of CA units, the contextual environment, and turn-taking in the dialogue as well as the communicative goals of the units.
The discourse structure should hence be represented while containing these elements:

DS $\longrightarrow$ (CA, syntactico-semantics, turn-taking)

5.1 Previous studies for designing discourse grammar

5.1.1 CA units relationship within a turn

We found special features of CA unit links in a dialogue, when we investigated a subcorpus from the ATR dialogue corpus[8]. There are 826 CA units and 181 turns in total in our corpus.

1. A turn consists of 2 or 3 CA units. There are sequences of two adjacent CA units which occur repeatedly in dialogues. Table 6 shows the quantative number of CA sequences in a turn. In our corpus, 83% of the turns in a dialogue consist of 1 to 3 CA units.

   The two adjacent CA sequences which occur frequently in dialogues are shown in Table 7.

[8]Each file in the subcorpus is called: TAC22011, TAC22012, TAC22013, TAC22014, TAC23031, TAC23032, TAC23033, TAC23034, TAS12001, TAS12002, TAS12005, TAS22002 and TBS12001.

CA units	occurence / total turns	%
1 CA	45/296	15%
2 CAs	124/296	42%
3 CAs	77/296	26%
4 CAs	27/296	9.12%
5 CAs	10/296	3.38%
6 CAs	7/296	2.36%
7 CAs	2/296	0.68%
10 CAs	1/296	0.34%
11 CAs	1/296	0.34%
12 CAs	1/296	0.34%
16 CAs	1/296	0.34%

Table 6: CA units sequences in a turn

YES, NO, Greet and Temporizer are very often located at the beginning of turns and are able to co-occur with Inform. In addition, the communicative goals of these CA units are completed by the Inform unit which follows them. That is, these CA units can be unified with the Inform unit.

(⟨INFO⟩ → YES Inform)
(⟨INFO⟩ → NO Inform)
(⟨INFO⟩ → Temporizer Inform)
e.g.

A: ほかに変更はございませんか。(Yn-question)
(Will there be other changes~?)
C: いいえ、(NO)(no) ほかにはありません。(Inform)(I haven't)

There are also CA units that achieve their communicative goals with adjacent CA units. For example, a CA unit like Acknowledge is often used to maintain communications and to add politeness without strong indication of understanding or agreement. These CA units can be unified with succeeding CA units for the same reason as CA units like YES, NO, etc.

(⟨INFO⟩ → Acknowledge Inform)
e.g.

A: カードの期限はいつまでになっているかお分かりですか。(Wh-question)
(Do you know when your card will expire?)
C: はい、(Acknowledge)(YES) 1995年の4月までです。(Inform)
(It will expire on April of 1995.)

2. In natural dialogues, there are turns which momentarily break the conversation. These CA units should be specified in the discourse structure representation.

adjacent CA units	occurrence / total CA units	%	examples
Inform-Inform	76/826	9.2%	ビザカードです。(Inform) ナンバーは.... です。(Inform)
Acknowledge-Inform	40/826	4.8%	はい、(Acknowledge) 電話番号のほうが.... です。(Inform)
Confirmation-question-Confirmation-question	24/826	2.9%	シングルルームがお一つで、(Confirmation-question) 八月十日にご到着になられて (Confirmation-question) 十六日にご出発ですね。(Confirmation-question)
Inform-Action-request	18/826	2.18%	お調べいたしますので、(Inform) 少々そのままでお待ちくださいませ。(Action-request)
Acknowledge-Acknowledge	18/826	2.18%	はい (Ack.) 分かりました (Ack.)
YES-YES	16/826	1.93%	はい (YES) そうです (YES)
Greet-Inform	12/826	1.45%	もしもし、(Greet) ニューヨークシティホテルでございます。(Inform)
YES-Inform	11/826	1.33%	はい、(YES) 確かにそうでございますが、(Inform)
Apology-Inform	10/826	1.21%	お待たせいたしました。(Apology) その日はシングルルームがご用意できますが、(Inform)
topic-Inform	8/826	0.9%	あいにくですが、(topic) 八日、九日はツインがもう満室となっております。(Inform)

Table 7: Adjacent CA units sequences

(⟨DEICTIC⟩ → (Inform Action-request- お待ち下さい))
e.g.

A: 空き室状況をお調べしますので、(Inform)
少々お待ちください。(Action-request- お待ちください)
(Just a moment please, I'll check the reservation list.)

3. There are CA units which are used to explain the succeeding CA unit content for helping the hearer with understanding. These CA units should be specified in the discourse structure representation.

(⟨CONFIRM⟩ → (topic- 確認させていただきます Confirmation-question))
e.g.

A: では、(Then)
確認させていただきます。(topic- 確認させていただきます)
(Let me confirm.)
シングルルームがお一つで、(Confirmation-question)
(One single room and)
八月十日にご到着になられて、(Confirmation-question)
十六日にご出発ですね。(Confirmation-question)
(You will arrive on the 10th of August
and leave on the 16th of August?)

4. There is a sequence of identical CA units which are connected with connectives. These CA units can be unified.

(⟨INFO⟩ (→ Inform/ ので -reason Inform))
e.g.

C: わたくし一人です ので、(Inform/ ので -reason)
(Watashi hitoridesunode) (I'm alone)
シングルでよろしいんですけど。(Inform)
(Singurude yoroshiinodesuga)(A single is OK.)

5.1.2 CA units relationship over turns

We found special features of CA unit links over turns in a dialogue, also.

1. CAs can be roughly classified into initiate CAs or response CAs. Initiate CAs are CAs which evoke a new discourse state like Wh-question, Yn-question or Action-request. Response CAs are CAs which respond to initiate CAs like Acknowledge, YES and NO. The distribution of these two sorts of CAs is shown in Tables 8 and 9, respectively.

CA	occurence / total units	%
Action-request	68/826	8.23%
Confirmation-question	49/826	5.93%
Wh-question	41/826	4.96%
Yn-question	32/826	3.87%

Table 8: Initiate CA units distribution

CA	occurence / total units	%
Inform	182/826[9]	22%
Acknowledge	97/826	12%
YES	44/826	5.33%
NO	1/826	0.12%

Table 9: Response CA units distribution

2. Cohesion of the initiate CA and the response CA over turns is shown in Tables 10 and 11: 86% of WH-question CA units are followed by Inform CA units. Also, 87% of Confirmation-questionCA units are followed by YES CA units. These high percentages should be taken account into for discourse grammar design.

   In Tables 10 and 11, others (1) include Temporizer, Action-request, Yn-question, etc., others (2) Inform, Thanks, etc., others (3) Thanks, Inform, etc. and others (4) unspecified CAs.

adjacent turn	occurance / total turn	%	examples
Confirmation-question and YES	13/15	87%	A: 大人が一名様、子供が二名様でよろしいですね。(Confirmation-question) C: はい、(YES) ‖
Confirmation-question and others (4)	2/15	13%	A: 八月十日にご到着になられて、十六日にご出発ですね。(Confirmation-question) C: すいません、(Alert) 先ほど、[えー] シングルルームのほうは十三、十四、十五はないので、ツインということでなかったですか。(Yn-question)
Wh-question and Inform	42/49	86%	A: 場所と、それからどれぐらいの価格のホテルをお望みでしょうか。(Wh-question) C: 一泊百ドルぐらいのがいいですね。(Inform) ‖
Wh-question and others (1)	7/49	14%	A: どのような情報をお望みですか。(Wh-question) C: ホテルの紹介を幾つかしてください。(Action-request)

Table 10: Cohesion over turns of CA units (1)

adjacent turn	occurance / total turn	%	examples
Yn-question and YES/NO	18/28	64%	C: ホテルの中に土産物か何か売ってますか。(Yn-question) A: はい、(YES) 当ホテルにはギフトショップはございます。(Inform) ‖
Yn-question and others (2)	10/28	36%	A: 地下鉄の近くがよろしいでしょうか。(Yn-question) C: そうですね (temporizer) 市街の中心地がいいですね。(Inform)
Action-request and Acknowledge	44/64	69%	C: できましたら二階か三階までの部屋にしていただきたいんですけど。(Action-request) A: はい、(Acknowledge) かしこまりました。‖
Action-request and others (3)	25/64	31%	C: コンチネンタル式のほうをお願いします。(Action-request) A: コンチネンタル式の朝食を一名分ですか、それともお子様の分もご用意しましょうか。(Yn-question)

Table 11: Cohesion over turns of CA units (2)

5.2 Grammar design

We propose rewriting rules using CA labels as terminal symbols for the purpose of grouping CA units. CA units are unified and classified into groups in a way specifying the above-mentioned characteristics of CA sequences and cohesion. That is, one CA unit may be a prompt and another CA unit may be its response. So, cohesion between CA units is made in a way that prompt units cohere with their response units [28]. In this way, some unit groups are obtained from the dialogue. We consider each group to express the attentional focus state of participants.

The attentional focus state is denoted as one of ten states including Open and Close-conversations: Open-conversation, Outset, On-and-on, Go-ahead, New, Volt-face, Flashback, Close-conversation, Repetition and Upshot.
Roughly speaking, Outset denotes CA sequences containing a Greet; Go-ahead are topic sequences derivated from a previous group; Volt-face denotes CA sequences containing deictic units; Flashback are CA sequences after Volt-face; Repetition denotes CA sequences containing Confirmation-questions; New denotes a transition from an attention state; On-and-on denotes embedded structures of Question and Inform units; and Upshot denotes CA sequences containing Thanks or Good-wishes units.

The rewriting rules consist of 27 terminal symbols[10], 55 symbols as non-terminal symbols

[10]As a matter of fact, there were 27 CA labels, and some subcategories including connectives and

of move[11] level, and 25 symbols as non-terminal symbols of exchange[12] level, thus having three strata of the grammar.

5.2.1 CA units grouping process

The process of CA unit grouping is mapped out below:

e.g.
A: で (So,) (THEN)
　何時頃チェックインの御予定でしょうか (Wh-question)
　(What time will you check in?)
C: 6時頃になると思います (Inform)(About 6 p.m. maybe.)
A: 分かりました (RMK)(Very good, sir.)

THEN[13], Wh-question, Inform and RMK[14] are CA units, and the semantic representation for each unit is as follows:

```
[[SEM [[RELN *THEN*]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJE [[RELN *で(de)*]]]]]]

[[SEM [[RELN *Wh-question*]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJE [[RELN *でしょうか(desyouka)*]
              [OBJE [[WH WHAT-TIME]]]
              [IDEN [[RELN *チェックイン(check in)*]]]]]]

[[SEM [[RELN *Inform*]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJE [[RELN *と思います(toomoimasu)*]
              [OBJE [[RELN *なる(naru)*]
                     [OBJE [[RELN *6時頃(rokujigoro)*]]]
                     [IDEN *UNSPECIFIED*]]]]]]]
```

topics were used in our experiments on discourse analysis.

[11]The *move* is defined here as the smallest significant element by means of which a conversation is developed.

[12]The *exchange* is defined here as sequences of *moves*. Exchanges of different types exhibit different sorts of linkages, thus combining to form a state of conversation.

[13]THEN is derived from Topic.

[14]RMK is derived from Acknowledge.

```
[[SEM [[RELN *RMK*]
       [AGEN *SPEAKER*]
       [RECP *HEARER*]
       [OBJE [[RELN *分かりました(wakarimasita)*]]]]]]
```

The following rewriting rules are applied to group these four units.

```
(<NEW> <--> (<STEPINT> <PINFO>))
(<STEPINT> <--> (THEN Wh-question))
(<PINFO> <--> (Inform RMK))
```

In Fig. 2, an attentional state is represented at the top level of the discourse feature-structures as ⟨NEW⟩. ⟨NEW⟩ denotes that a discourse state had shifted into another state as the discourse had unfolded. ⟨STEPINT⟩ and ⟨PINFO⟩ are non-terminal symbols of the exchange level. They roughly mean a prompt and its response, respectively, while grouping CA units. THEN, Wh-question, Inform and RMK are CA units labeled with THEN, Wh-question, Inform and RMK, respectively. A and C indicate turn taking of the CAs' sequence.

6 Experiment Evaluation of Discourse Structure Analysis

Our discourse structure analysis aims to limit the search range for taking referents of anaphoric expressions and for complementing elliptical expressions in utterances in the context of machine translation. With these possibilities in mind, we conducted automatic discourse analysis experiments on ten Japanese-Japanese dialogues from the ATR dialogue corpus [24].

6.1 Referring expressions and their Referents

Eighteen non-zero referring expressions have been observed as Japanese pro-forms apart from personal pronouns. They are also used as deictic indicators. However, we confined our current investigation to their anaphoric usage. In Table 12, the first column shows Japanese pro-forms, the second each pro-form having a referent within the same exchange level and the total occurence, and the third its percentages.
According to judgments by native speakers, referents were found within the same exchange-level grouping to be relevant pro-forms 78.5% of the time.

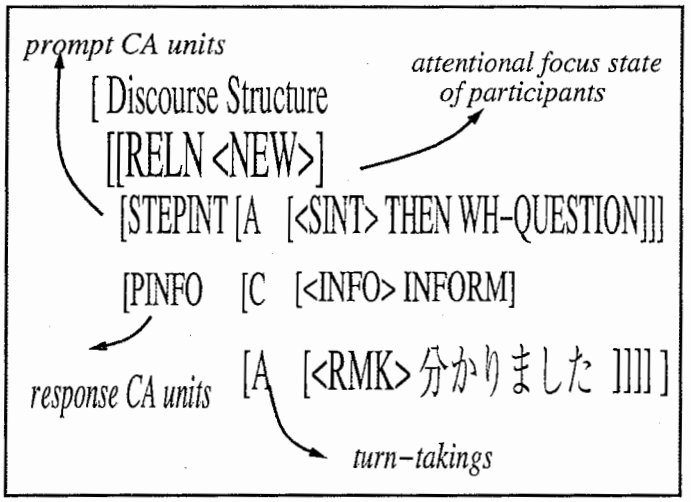

Figure 2: Discourse structure representation

6.2 Ellipses

Several types of ellipses are common in Japanese dialogues. Observed have been ellipses including antecedents within the dialogue and ellipses not-having antecedents within the dialogue. In examples 1 and 2, ellipses of known information from the point of view of discourse grammar are observed, and they have antecedents. In example 3, an ellipsis particular to the Japanese spoken language, called the *zero pronoun* can be observed, and it has no antecedent. We confined our current investigation to ellipses having antecedents.

e.g. 1. ellipsis of known information

A: お支払方法ですが、どのようになさいますか。
(oshiharaihouhoudesuga donoyouni nasaimasuka)(How will you pay?)
C: カードでお願いいたします (ka-dode onegaiitasimasu) (By credit card, please)
←instead of 支払はカードで....

e.g. 2. ellipsis of known information

C: カードは［あの］ビザカードですけど。(ka-doha visa ka-dodesukedo)
(My card is a Visa card.)
A: 期限はいつまでになっているかお分かりですか
(kigenha itsimadeninatteiruka owakaridesuka)
(Do you know the expiration date of your card?)
←instead of そのカードの期限は...

Japanese pronouns	referents/total	%
あれ (are)(that one)	0/0	
これ (kore)(this)	2/2	100%
それ (sore) (this one)	17/23	74%
あっち (acchi)(that one)	0/0	
こっち (kocchi)(this way)	0/0	
そっち (socchi)(that way)	0/0	
あの (ano)(the)	0/0	
この (kono)(the)	2/2	100%
その (sono) (the)	5/7	71%
あそこ (asoko)(there)	0/0	
ここ (koko)(here)	0/0	
そこ (soko)(there)	0/0	
あちら (achira)(there)	0/0	
こちら (kochira)(here)	0/1	0%
そちら (sochira)(there)	2/2	100%
ああ (aa)(like that)	0/0	
こう (kou)(like this)	1/1	100%
そう (sou)(like that)	29/35	83%

Table 12: Percentages of each pro-form having referent within same exchange

e.g. 3. ellipsis of pronoun

C: そちらのホテルの予約をしたいのですが (sochirano hoteruno yoyakuwo sitainodesuga)(I'd like to make a reservation to stay at your hotel.)
←instead of わたしはそちらの....

In our sample, such antecedents were found within the same exchange level as with their ellipses 61.5% of the time. This figure is considerably lower than the 78.5% figure for pro-forms, but may still indicate a useful constraint.
Also of interest were the distance in CAs between ellipses and antecedents, and whether the ellipses and antecedents were in the same turn. Table 13 presents this information. For each distance-turn combination, we also list the percentage of ellipses having antecedents in the same exchange. The table shows a marked tendency for ellipses and antecedents to be found within two or three CAs of each other: among all 94 of our ellipses, 79, or 84%, are within 3 CAs. In that short a range, ellipses and antecedents are quite likely to occur within the same exchange.

6.3 Comparing with dialogues analyzed by hand

Our current automatic analyses of CAs and CA groupings do appear to provide useful constraints for pro-form and ellipsis resolution. But of course the means of evaluating the analyses is incomplete: it argues that the limits of exchanges are being usefully recognized, but does not indicate whether their labeling (as Open-conversation, Outset, Go-ahead,

distance-turn	ant. within exchange/total	%
1-same turn	21/22	95%
1-different turn	14/14	100%
2-same turn	8/8	100%
2-different turn	20/24	83%
3-same turn	4/4	100%
3-different turn	4/7	57%
4-same turn	1/1	100%
5-same turn	2/4	50%
5-different turn	1/2	50%
6-same turn	2/2	100%
6-different turn	1/3	33%
7-same turn	0/1	0%
9-different turn	0/1	0%
11-different turn	0/1	0%

Table 13: Ellipses-to-antecedent distances in CAs

Volt-face, etc.) is equally useful. To address this issue, we made a comparison between automatically-analyzed dialogues and dialogues analyzed by hand.

Here is an automatically-analyzed dialogue[15]:

```
A:(1th)
  はい (Greet) ;1
  ニューワシントンホテルでございます。(Inform) ;2
  担当のメアリ・フィリップスです。(Inform) ;3

        ;(1 - 3)
        [Discourse Representation
         [[RELN <OPEN-CONVERSATION>]
          [A [<OPEN-CONVERSATION> Greet <INFO>]]
          [A [<INFO> <INFO> Inform]]
          [A [<INFO> Inform]]]]

C:(2th)
  もしもし (Greet) ;4
  [あの] 部屋の予約をお願いしたいんですけれども。(Action-request1) ;5

A:(3th)
  はい (Acknowledge) ;6
  いつがご希望でしょうか。(Wh-question) ;7
C:(4th)
```

[15]The example TAS22001 comes from the ATR dialogue database.

［えー］八月の十日から十二日で、シングルルームでお願いします。（Inform）;8

```
;(4 - 8)
[Discourse Representation
 [[RELN <Outset>]
  [C [<PAR1> Greet Action-request1]]
  [A [<FINT> Acknowledge Wh-question]]
  [C [<INFO> Inform]]]]
```

A:(5th)
少々お待ちくださいませ。（Action-request- お待ちくださいませ）;9

```
;( 9 )
[Discourse Representation
 [[RELN <Volt-face>]
  [A [<DEIX> Action-request- お待ちくださいませ]]]]
```

普通のシングルルームは満室となっております。（Inform）;10
シングルのシャワー付きのお部屋が一泊八十ドルで、（Inform）;11
ツインのバス付きのお部屋が一泊百四十ドルでございますが。（Inform）;12
C:(6th)
［あ］そうですか。（Acknowledge- そうですか）;13

```
;(10 - 13)
[Discourse Representation
 [[RELN <Flashback>]
  [C [<Flashback> <PINFO> Acknowledge- そうですか]]
  [A [<INFO> Inform]]
  [A [<INFO> Inform <INFO>]]
  [A [<INFO> Inform]]]]
```

じゃあ (then- じゃあ) ;14
シングルのシャワー付きの部屋をお願いします。 (Action-request) ;15
A:(7th)
分かりました。 (Explain- 分かりました) ;16

;(14 - 16)

[Discourse Representation
[[RELN ⟨CLOSE-CONVERSATION⟩]
[A [⟨CLOSE-CONVERSATION⟩ then- じゃあ ⟨PAR⟩]]
[C [⟨AR⟩ Action-request]]
[A [⟨RMK⟩ Explain- 分かりました]]]
;(14 - 16)
[Discourse Representation
[[RELN ⟨On-and-on⟩]
[C [⟨INT⟩ then- じゃあ ⟨AR⟩]]
[C [⟨AR⟩ Action-request]]
[A [⟨RMK⟩ Explain- 分かりました]]]

そちらのお名前と電話番号をお願いいたします。(Action-request) ;17
C:(8th)
はい (Acknowledge) ;18
[え] 鈴木和子と言います。(Inform) ;19
今、ニューヨークシティホテルに滞在しています。(Inform) ;20
[えー] ホテルの電話番号ですが、(topic) ;21
[え] 二零三の四四三の一七零零です。(Inform) ;22
A:(9th)
分かりました。(Explain- 分かりました) ;23

```
;(17 - 23)
[Discourse Representation
 [[RELN <On-and-on>]
  [A [<AR> Action-request]]
  [C [<INFO> <INFO> Inform]]
  [C [<RMK> Acknowledge]]
  [C [<INFO> Inform]]
  [C [<INFO> topic Inform]]
  [A [<RMK> Explain- 分かりました]]]]
```

だいたい何時ごろチェックインのご予定でしょうか。(Wh-question) ;24
C:(10th)
たぶん、六時ぐらいになると思います。(Inform) ;25
A:(11th)
分かりました。(Explain- 分かりました) ;26
;(24 - 26)

[Discourse Representation
[[RELN ⟨On-and-on⟩]
 [A [⟨INT⟩ Wh-question]]
 [C [⟨INFO⟩ Inform]]
 [A [⟨RMK⟩ Explain- 分かりました]]]

予約を確認させていただきます。(topic- 確認させていただきます) ;27
鈴木和子様 (Confirmation-question) ;28
八月の十日から十二日まで、シングルルームシャワー付き二泊ですね。
(Confirmation-question) ;29
現在、ニューヨークシティホテルにお泊まりですね。
(Confirmation-question) ;30

;(27 - 30)
[Discourse Representation
 [[RELN <Repetition>]
  [A [<INTCONFIRM> topic- 確認させていただきます]]
  [A [<CONFIRM> <CONFIRM> Confirmation-question]]
  [A [<CONFIRM> <CONFIRM> Confirmation-question]]
  [A [<CONFIRM> Confirmation-question]]]]

電話番号は二零三、四四三、一七零零でよろしいでしょうか。(Yn-question) ;31
C:(12th)
はい (YES) ;32
そうです。(YES) ;33

;(31 - 33)
[Discourse Representation
 [[RELN <On-and-on>]
  [A [<INT> Yn-question]]
  [C [<INFO> YES YES]]]]

[あ] それと (add- それと) ;34
トラベラーズチェックは使えますか。(Yn-question) ;35
A:(13th)
もちろんです。(YES) ;36
パスポートの提示をお願いすることになりますけれども。(Permission-request) ;37
C:(14th)

分かりました。(Explain-分かりました) ;38

```
;(34 - 38)
[Discourse Representation
 [[RELN <New>]
  [C [<NINT> add-それと Yn-question]]
  [A [<INFO> YES]]
  [A [<PR> Permission-request]]
  [C [<RMK> Explain-分かりました]]]]
```

どうもありがとう。(Thank) ;39

A:(15th)
ニューワシントンホテルをご利用いただきましてありがとうございます。
(Thanks-response) ;40

```
;(39 - 40)
[Discourse Representation
 [[RELN <CLOSE-CONVERSATION>]
  [C [<THANK> Thank]]
  [A [<THANKR> Thanks-response]] ]]
```

As a matter of fact, there were a lot of segment ambiguities for individual states, if all analysis outputs were displayed. We decided to adopt the longest discourse segment for each state of discourse. That is, when there was the possibility of several discourse segments, we took the longest segment for a discourse state from the opening CA unit of the dialogue in the given order. So, if there were three posibilities as follows, discourse segment 3 was taken.

discourse segment 1: [1,2][3][4,5,6][7,8,9,10]
discourse segment 2: [1][2,3,4][5,6,][7,8,9][10]
discourse segment 3: [1,2][3,4,5,6][7,8,9,10]

Among 268 discourse segments containing the same outputs, 43 sorts of discourse segments were output for a dialogue which consisted of 40 CA units and 15 turns. Only the longest sequences for each state of the discourse were chosen from all outputs, and the discourse structure was represented with these segments and their state labels.

According to native judgments, sequences enclosed by a rectangle in the automatically-analyzed dialogue were not correctly analyzed (They are marked with the symbols $\star$ in Table 14.). The first sequence in the rectangle had two different outputs: On-and-on and Close-conversation, though they are not the longest sequence, so the sequence are not taken. For the second sequence in the rectangle, Go-ahead was better than On-and-on as the state name, because Go-ahead is used to express a deviating sequence from previous one. In fact the second sequence gave us an impression that the participants' attention was going to shift to another state here.

discourse segments	outputs	state	discourse segments	outputs	state
(1 - 2)	2		(17 - 22)	17	
(1 - 3)	8	OPEN	(17 - 22)	17	
(4 - 5)	1		(17 - 22)	17	
(4 - 6)	5		(19 - 23)	6	
(4 - 8)	2	Outset	(20 - 23)	2	
(7 - 8)	2		(24 - 25)	2	
(9)	2	Volt-face	(24 - 26)	3	⋆On-and-on
(10 - 13)	16	Flashback	(27 - 28)	2	
(11 - 13)	8		(27 - 29)	4	
(12 - 13)	4		(27 - 30)	8	Repetition
(14 - 16)	3	⋆	(31 - 32)	2	
(14 - 19)	2		(31 - 33)	6	On-and-on
(14 - 20)	6		(34 - 36)	2	
(14 - 23)	6	On-and-on	(34 - 37)	6	
(15 - 16)	2		(34 - 38)	4	New
(15 - 19)	2		(35 - 36)	2	
(15 - 20)	6		(35 - 37)	2	
(15 - 23)	15		(35 - 38)	4	
(16 - 23)	6		(36 - 37)	2	
(17 - 18)	2		(34 - 36)	2	
(17 - 19)	6		(39 - 40)	1	CLOSE
(17 - 20)	21				

Table 14: discourse segments occured for a state

6.4 Related works

We have a related work at ATR to our discourse structure analysis[16]. [31] has independently attempted to automatically group cue-based symbols similar to CAs, and like us has used a discourse rule set. Both the rules and procedures of [31] differ significantly from ours, however.

Our grammar uses the 27 current CAs directly as terminals; non-terminals function on two levels, with 55 symbols at the move level and 25 symbols at the exchange level. That is, the grammar subcategorizes turns and exchanges rather specifically. In contrast, [31] uses 15 CAs, and the rules aim mainly to identify initiate-response relationships among turns. While essentially the same terminal objects as ours are used, [31] introduces a level of symbols called acts at an immediately dominating level whose purpose is to categorize terminals into only two types: *initiate* and *response*[17]. *Initiate* (open exchange) and *response* (maintain or close exchange) acts are grouped into turn-level symbols called moves. Again, there are only two move types: *initiate* moves and *response* moves. Groupings of

[16]This chapter was added after having consulted with Mr. Iwadera about his paper [31].

[17]In reality, it categorizes terminals into three types: initiate, response and neutral. However, the third is used only in preprocessing.

moves are called exchanges, and once again there are only two types: minimal exchanges (having one initiate move and one response move) and non-minimal exchanges (having more than one response). The differences in grammar design between the two are listed in Table 15.
The grouping rules on the act, move and exchange levels in [31] should be applied in a fixed order; in contrast, our grouping rules can be applied in no special order.

As for procedures, the two research groups use very similar pattern matching techniques to segment and identify candidate terminal symbols.
The procedure in [31] is :

- to segment into unlabeled act units using predefined surface form patterns
- to use a mapping table to assign act labels to act units
- to build a structure:
  to find the act class using the table → to build moves →to build exchanges

From there, however, the respective techniques are distinct. As described above, we disambiguate terminals using the techniques of [20] and then apply our CFG using standard parsing techniques on the Kitagawa chart parser. [31] in contrast, describes no techniques for disambiguating terminals, but describes special-purpose procedures for using contextual symbols to categorize act- and move-level symbols as *initiate* or *response* for grouping terminals.

Otherwise, [31] conducts topic transitions in a way of searching for particles は (ha) which gives a long-term topic, and が (ga), を (wo), に (ni) which give short-term topics within the current exchange.

comparison	[31]	Ours
terminal symbols	15 symbols	27 symbols
act level	3 symbols : initiate, response, (neutral)	55 symbols: Open-conversation, Close-conversation, INFO, INFOCON, CONSEQINFO, INT, STEPINT, etc.
move level	3 symbols : initiate, response, (follow-up)	
exchange level	minimal exchanges, non-minimal exchanges	25 symbols: PINFO, PINT, PINFOCON, DEIX, etc.
discourse state	0 symbol	10 symbols: Outset, On-and-on, Volt-face, Repetition, Go-ahead, New, Flashback, Upshot, Open-conversation, Close-conversation

Table 15: Comparison of our discourse grammar with [31]

7 Conclusion

We have extended an earlier work on Communicative Acts, or cue-based speech acts, in two directions. First, we have noted a significant relation between segmentation based upon cue patterns and segmentation based upon natural pauses ("pause-units"): Some 77% of the pause-bounded segments in our corpus coincide with segments defined by CAs. Second, we have described attempts to parse CA groupings, and thus to analyze the discourse structure, using context-free rules. Experiments on ten spontaneous Japanese dialogues show that the resulting groupings can be useful for pronoun disambiguation and ellipsis recovery. Pro-forms and their referents fell within exchange-level groupings 78.5% of the time. Certain ellipses and their referents fell within exchange-level groupings 61.5% of the time. We introduced a set of CA labels and its application to dialogue processing.

The advantages of using the CA labels are as follows with regard to semantic analysis and discourse structure analysis for naturally-spoken Japanese:

- Based on a set of CA labels, the CA unit is obtained. The CA unit is a convenient unit to handle Japanese dialogues.
- CA units have been shown to fall on pause units in segmentation experiments, so we can apply the pause-based grammar to the syntactico-semantic analysis for CA units. Additionally, the CA unit is rewritten into a communicative act representation preserving its syntactico-semantic representation, by merging CA labels to the grammar. It follows that a speech recognition grammar can be applied to syntactic analysis; thus successive processing from speech recognition to discourse analysis can be made.

- There is a close relationship between CA units such as prompts and responses. The relationship involves the states of the participants' attention in the dialogue. So, when CA units are aggregated into groups by using rules expressing the relationship, the discourse structure can be made apparent.

With regard to speech recognition, CA labels are also useful.

- One of the motivations of using the 27 CAs is to constrain predictions of upcoming symbols as tightly as possible: since the symbols are linked to surface structures, predicting specific symbols means predicting specific surface structures. We hope such tight predictions of surface structures will provide useful constraints, particularly for speech recognition.
  For example, if we can predict the relative probability that the current utterance is a Yn-question as opposed to an Inform, we may be able to differentiate utterance-final か (*ka*) (a question particle) and utterance-final が (*ga*) (a conjunction or attenuation particle), which are often very similar phonetically.

- Once spontaneous data can be labeled, speech recognition researchers can try to recognize prosodic cues to aid in CA recognition and disambiguation. For instance, they can try to distinguish Inform and Yn-question according to their F_0 contours - a distinction which would be especially useful for recognizing Yn-questions with no morpho-syntactic markings.

- Similarly, speech synthesis researchers can try to provide more natural prosody by exploiting CA information. Once relations between prosody and CA have been extracted from corpora labeled with CA information, they can attempt to supply natural prosody for synthesized utterances according to the specified CA. For instance, Yn-questions and Confirmation-questions can be made to sound more natural [30].

Appendix 1 Collation of CA units with pause units

speech recognition results	transcription	CA labels
260.0 a 370.0	あ	Expressive
370.0 pau 505.0	*pause*	
505.0 m 525.0	も	
525.0 o 585.0		
585.0 sh 665.0	し	
665.0 i 695.0		
695.0 m 735.0	も	
735.0 o 800.0		
800.0 sh 935.0	し	Greet
935.0 i 1080.0		
1080.0 pau 1265.0	*pause*	
1265.0 w,a 1330.0	わ	
1330.0 t 1385.0	た	
1385.0 a 1455.0		
1455.0 sh 1605.0	し	
1605.0 i 1725.0		
1725.0 t 1805.0	た	
1805.0 a 1850.0		
1850.0 n 1890.0	な	
1890.0 a 1965.0		
1965.0 k 2010.0	か	
2010.0 a 2060.0		
2060.0 h 2110.0	ひ	
2110.0 i 2150.0		
2150.0 r 2165.0	ろ	
2165.0 o 2235.0		
2235.0 k 2275.0	こ	
2275.0 o 2335.0		
2335.0 t 2380.0	と	
2380.0 o 2435.0		
2435.0 i,i 2495.0	いい	
2495.0 m 2545.0	ま	
2545.0 a 2615.0		
2615.0 s 2695.0	す	
2695.0 u 2735.0		
2735.0 g 2765.0	が	Inform
2765.0 a 2870.0		
2870.0 pau 3340.0	*pause*	

Table 16: Data used for comparing CA units & pause units

Appendix 2 CA Label Set

explanation	S gives H information.
illocutionary verbs	言う (iu), 述べる (noberu), assert, state, tell, inform
patterns	Declarative clause syntax: となっています (tonatteimasu), のようです (noyoudesu), と思います (toomoimasu), ... ます (masu), ... です (desu)
parameters	politeness, possessor of information
example	いろいろな行き方があります (iroirona ikikataga arimasu): You can travel a number of different ways.

Table 17: Inform

explanation	S informs H that S wants H to provide information about the truth or falsehood of a proposition.
illocutionary verbs	尋ねる (tazuneru), ask
patterns	... ますか (masuka), ... ですか (desuka), でしょうか (desyouka), Yes-no interrogative clause syntax with wide variety of aspect and mood.
parameters	politeness
example	地図を見ていますか (chizuwo miteimasuka): Are you looking at the map?

Table 18: Yn-question

explanation	S informs H that the response to H's Yn-question, Confirmation-question, or Do-you-understand-question is affirmative.
illocutionary verbs	答える (kotaeru), answer, reply
pattern	はい (hai), ええ (ee), そうです (soudesu), yes, yea
parameters	politeness
example	C: 10日はあいてますか A: はい、あいています (hai aiteimasu): Yes, it's available.

Table 19: YES

explanation	S informs H that information given by H has been heard, received, understood. Acknowledge is often used to maintain communication and be polite, without strong indication of understanding or agreement.
illocutionary verbs	合づちをうつ (aizuchiwoutsu), acknowledge, recognize
patterns	ええ (ee), はい (hai), I see, uhum, That's fine., OK
parameters	politeness
example	A: 三条で乗り換えていただきまして (sanjyode orikae teitadakimasite) C: ええ (ee): uhum

Table 20: Acknowledge

explanation	Fixed expressions used by S to get H's attention.
illocutionary verbs	注意を引く (chuuiwohiku), get one's attention
patterns	ちょっと (cyotto), すいません (suimasen), excuse me, hey
parameters	politeness
example	C: あ、ちょっと (chotto)、 会員でない場合はどうなりますか: Excuse me, ...

Table 21: Alert

REFERENCES

[1] J.Searle, Speech Acts, Cambridge University Press, Cambridge, 1969

[2] G.Leech and J.Svartvik, A Communicative Grammar of English, Longman, Singapore, 1975

[3] P.Larreya, Enoncé performatifs Présupposition, Université Nathan information, Fernand Nathan, France, 1979

[4] W.Edmondson, Spoken Discourse, Longman, London & New York, 1981

[5] G.N.Leech, Principles of Pragmatics, Longman, London, 1983

[6] S.C.Levinson, Pragmatics, Cambridge University Press, Cambridge, 1983

[7] M.Nakau, Structure and functions of the sentence Semantics, Taisyukan, Tokyo, 1983

[8] B.Grosz & C. Sidner, Attention, Intention, and the Structure of Discourse, Computational Linguistics, 1986

[9] T.Hasegawa, A Rule Application Control Method in a Lexicon-driven Transfer Model of a Dialogue Translation System, In Research Activities of the Natrual Language Understanding Department of the Data Processing Department for Nov. 1989 March 1991, ATR Technical Report TR-I-0231, kyoto, 1991

[10] L.Fais & G. Kikui, Determining Surface Form for Indirect Speech Acts in English, ATR Technical Report TR-I-0235, Kyoto, 1991

[11] M.Seligman, Generating Discourses from Networks Using an Inheritance-based Grammar, Dissertation, University of California, Berkeley, 1991

[12] M.Tomokiyo, Transfert de la langue parlée japonais-anglais dans le système de traduction automatique ASURA, TA-TAO, UREF-AUPELF, Montréal, 1993

[13] P.Szatrowski, Discourse Structure Analysis in Japanese Spoken dialogue, Kuroshio, Tokyo, 1993

[14] M.Kasiwazaki, Function of act indication type expression in Japanese, Kuroshio, Tokyo, 1993

[15] M.Tomokiyo, Segmentation of Naturally-Spoken Japanese Dialogues based on Pragmatics, TR-IT-0019 ATR Technical Report, Kyoto, 1993

[16] K-H.Loken-Kim et al., EMMI-ATR Environment for MulitiModal Interaction, TT-IT-0081 ATR Technical Report, Kyoto, 1993

[17] K-H.Loken-Kim et al., Transcription of spontaneous speech collected using a multimodal Simulator-EMMI in a direction finding task, TR-IT-0029 ATR Technical Report, Kyoto, 1993

[18] M.Tomokiyo, Communicative labels and its automatic assignment TR-IT-0069 ATR Technical Report, Kyoto, 1994

[19] M.Seligman et al., A Bilingual Set of Communicative Act Labels for Spontaneous Dialogues, TR-IT-0081 ATR Technical Report, Kyoto, 1994

[20] M.Tomokiyo, Natural Utterance Segmentation and Discourse Label Assignment, Proc. of ICSLP94 Vol.3 Yokohama, 1994

[21] H.Singer et al., A Modular Speech Recognition System Architecture, Proc. of Acoustic Soc. Tokyo, 1994

[22] A.Stenström, An Introduction to Spoken Interaction, Longman, London and New York 1994

[23] J.Hosaka et al., Pause as a Phrase Demarcator for Speech and Language Processing, Proc. of COLING-94, Kyoto, 1994

[24] T.Morimoto et al., A Speech and Language Database for Speech Translation, Proc. of ICSLP94 Vol.4, Yokohama, 1994

[25] T.Takezawa et al., A Japanese Grammar for Spontaneous Speech Recognition Based on Subtrees, TR-IT-0110 ATR Technical Report, Kyoto, 1995

[26] T.Shimizu et al., Time-Synchronous Continuous Speech Recognizer Driven by a Context-Free Grammar Proc. of ICASSP95, Yokohama, 1995

[27] T.Tashiro & T. Morimoto, Parsing Experiments of Spoken-Style Japanese Sentences NLC-95-24 The institute of Electronics Information and Communication Engineers, Tokyo, 1995

[28] M.Tomokiyo, Aggregation of Segmented Utterance Units by using a Speech Act Label Grammar, Seminar Report, Korea, 1995

[29] M.Tomokiyo, Star units and pause units, TR-IT-0125 ATR Technical Report, Kyoto, 1995

[30] A.Black & N.Campbell, Predicting the intonation of discourse segments from examples in dialogue speech, Proc. of ESCA Workshop on Spoken Dialogue Systems, Denmark, 1995

[31] T.Iwadera et al., Recognizing an interactional Structure and Topics of Task-oriented Dialogues, Proc. of ESCA Research Workshop on Spoken Dialogue Systems, Denmark, 1995

[32] M.Seligman et al., Cue-base Communicative Acts for English and Japanese Dialogues, Submitted to Computational Linguistics Special Issue on Empirical Studies in Discourse Interpretation and Generation, 1995

[33] M.Tomokiyo, Investigation for the Star and Pause units ATR technical report ATR Kyoto 1995

[34] M.Tomokiyo, Pause units and Communicative Act units - toward Japanese Discourse Representation, Proc. of NL-111, Tokyo, 1996

Correspondence should be sent to Mutsuko Tomokiyo, ATR Interpreting Telecommunications, Hikaridai 2-2, Seikacho, Sorakugun, Kyoto 619-02, Japan. (Email:tomokiyo@itl.atr.co.jp)
On and after April of 1996, it should be sent to her, GETA-CLIPS-IMAG-campus, BP 53, rue de la Chimie, 38041, Grenoble, France. (Email:tomokiyo@imag.fr)