Pause Units and Communicative Act Units
- toward Japanese Discourse Representation

Mutsuko Tomokiyo

ATR Interpreting Telecommunications Research Laboratories

We previously investigated Japanese and English utterances focusing on their communicative acts in order to create a set of Communicative Act labels for the purpose of discourse analysis for English-Japanese machine translation or a man-machine interactive system. This paper extends investigation of CA analysis in two directions. First, a comparison was made between units yielded by CA label assignment to dialogues and pause units in the waveform output from speech recognition of the same dialogues, and 77% of the two boundary were coincided. Second, we conducted discourse analysis experiments using rewriting rules based on CA labels, including utterance segmentation by CA cue patterns and CA label assignment. We demonstrate the advantages of using CA labels for naturally-spoken dialogue processing.

ポーズの単位とコミユニケーティブアクト単位の比較
- 日本語談話構造の意味表現にむけて

友清睦子

ATR 音声翻訳通信研究所

我々は、先に日本語と英語の自然会話文を対象に発話意図に関する調査を行ない、発話意図が成就される単位を見い出し、発話意図ラベルを開発した。本稿は、まず発話意図が成就される単位と、会話文の音声認識から得られるポーズで区切られる単位を比較検討する。両者は 77% の割合いで一致し、この事実は自然会話文の解析にポーズに敏感な文法を適用することが好ましいのではないかと推測させる。この推測に基づき、次に発話意図ラベルを意味表現のトップレベルにもつポーズ文法による談話解析を提案する。最後にその実験結果を示し、発話意図ラベルの適用が日本語談話の解析に有効であることを報告する。

Introduction

This paper aims to describe the advantages of 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 introduced as a way of analyzing discourse structure.

We previously investigated Japanese and English utterances focusing on their communicative acts in order to create a set of Communicative Act labels for the purpose of discourse analysis for English-Japanese machine translation or a man-machine interactive system. This paper extends investigation of CA analysis in two directions. First, a comparison was made between units yielded by CA label assignment to dialogues and pause units in the waveform output from speech recognition of the same dialogues, and 77% of the two boundary were coincided. Second, we conducted discourse analysis experiments using rewriting rules based on CA labels, including utterance segmentation by CA cue patterns and CA label assignment.

The structure of the paper is ; we describe Communicative Acts (CAs) and application of the proposed CA set in Section 1: Section 2 we focus on comparison of CA units with pause units. In Section 3, discourse structure analysis using

rewriting rules based on the set of CA labels is introduced. In conclusion, we emphasize advantages of using CA labels for naturally-spoken dialogue processing, especially discourse analysis.

1 CA Label Set

1.1 CA Definition

We defined Communicative Act as follows:

> A Communicative Act is a communicative goal or aim which can be expressed in language L by a distinctive set of conventional cue patterns in specified discourse context.[11]

INFORM, ACTION-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 the 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 declarative clause syntax, sometimes with various sorts of ellipsis. ACTION-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 「verb + ですか (desuka)」 in Japanese and inverted interrogative clause syntax and/or special prosody in English.

Communicative Acts are similar to speech acts or to illocutionary force types.[1][2][3] 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 speakers' shared knowledge of a given language. We reserve the term Communicative Act for only such conventionally expressible goals. Communicative goals which cannot be described as Communicative Acts include utterance goals which are expressed non-conventionally (using one-time-only combinations); or goals which are expressed only implicitly; or goals which can only be defined in terms of relations between utterances.[2][3]

1.2 Utterance Segmentation and CA 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 to 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. One difficulty in spoken Japanese analysis arises out of this. [7][10]

- 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.[8]

Cue patterns yield units, when used to segment utterances. So, the segmentation of utterances by CA cue pattern enables us to get convenient, labeled units, thus making dialogue analysis easier.

> N.B. Segmentation of utterances and CA label assignment to CA units are automatically performed on an analyzer called the KK analyzer.[12] It 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 %.[9][12].

2 Integration of Pause Units and CA Units

2.1 Comparison of CA units with pause units

We discovered that in Japanese, CA units are often coterminous with units which are segmented by pauses. The pause as defined here is a silent period in a speech utterance more than 100 milliseconds in the waveform of speech recognition.[18]. That is, we know that cue patterns which are located just before silent periods in the waveform of speech recognition are the same as CA cue patterns. This means a communicative goal is achieved within a pause unit for Japanese utterances.[13][14]

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

> if ∃ simple sentence in a turn
> then: cue pattern ⊆ CA unit ≡ simple sentence,
> if not
> then: CA unit ≡ a cue pattern or : CA unit ≡ a turn

The Table A shows a comparison between CA units and pause units for a dialogue. The first column shows speech recognition results. The numbers are time in milliseconds for each phoneme. The second column shows transcription in Japanese, and the third column CA labels which are automatically assigned to the dialogue. The last column shows categories in the subtree grammar for the pause units.[18]

The dialogues taken from the ATR dialogue corpus[20] for the comparison contain 151 CA units and 176 pause units. 77% of CA units fall on pause units. 23% of the inconsistency includes 9% function words which are segmented by pauses, thus 14% of the inconsistency contains special expressions of such noun phrases as telephone numbers and expressions concerning money or time. These are uttered with pauses and without any particles. Table B shows the consistency between CA units and pause units. The Table C shows the inconsistency of segmentation between CA units and the pause units. There are found pause units which are not labeled with a CA label.

phoneme & duration	transcription	CA label	cat. in grammar
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	interj
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		

Table A: Comparison of CA units with pause units

Pause units are convenient processing units for speech recognition of Japanese dialogues. In fact, we are using a phrase structure grammar, called a *subtree grammar* which is sensitive to pauses for disambiguation of speech recognition.[14] When pause units fall on CA units, we consequently can apply this subtree grammar to the syntactico-semantic analysis of CA units. This means we can use the same grammar for speech recognition and syntactico-semantic analysis based on CA units for Japanese dialogues, and it enables us to make a connected processing for both.

CA label	occur./total	%
expressive	32/34	94%
greet	2/2	100%
temporizer	4/6	67%
acknowledge	10/16	62%
topic	1/2	50%
inform	9/38	24%
conjunctive	3/7	43%
yes	12/14	86%
thank	1/1	100%
alert	1/1	100%
apology	1/1	100%

Table B: Consistency between CA and pause units

undefined cat.	occur./total	%
n-spel,n-num	13/18	68%
mod-n	4/28	14%
pp-particle	22/59	37%
p-kakari-wa	1/7	14%
p-rentai-no	1/29	0.35%
p-conj-syusi	1/20	0.5%
adv	2/13	15%
np	3/126	0.24%
cl	4/62	0.65
???	1/1	100%

Table C: Inconsistency between CA and pause units

2.2 CA Unit Representation

When CA labels are assigned to the syntactico-semantic representation of an utterance, it follows that the communicative goal of the utterance is also represented. This enables us to make a link from syntactico-semantic to discourse analysis. In actual fact, when we apply the subtree grammar to CA units as a first step of the analysis, the following feature structures are obtained as an interim result:[15][16]

```
e.g. ホテルにはこのように行って下さい
    (hoteruniha konoyouni ittekudasai)

[[SEM [[RELN * 下さる ( kudasaru)*]
      [AGEN *SPEAKER*]
      [RECP *HEARER*]
      [RESTR [[RELN * 行く ( 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 CA labels to the top level.

```
[[SEM [[RELN  ACTION-REQUEST]
      [AGEN *SPEAKER*]
      [RECP *HEARER*]
      [OBJ [[RELN * て下さい ( tekudasai)*]
           [RESTR [[RELN * 行く ( iku)*]
                  [AGEN *HEARER*]
                  [GOAL * ホテル
                        ( hoteru)*]
                  [RESTR
```

```
[[WAY * このように
( konoyouni)*]
           ]]]]]]]]]
```

In these f-structures, the value of RELN (relation name), *i.e.* ACTION-REQUEST, 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.[17]

3 Discourse Structure Analysis

We finished the previous section by introducing a syntactico-semantic representation containing a communicative goal 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 which the discourse structure of the dialogue makes apparent.

3.1 Discourse Structure Representation

We hypothesize that discourse structure representation consists in making apparent a *state of focus of participants attention as the discourse unfolds.*[4]
The state of focus of participants' attention (AS) involves diverse constituent elements of dialogues: syntax and semantics of CA units, the contextual environment, and turn-taking in the dialogue as well as the communicative goals of CA units. Hence, discourse structure should be represented containing these elements.

$$(AS \longrightarrow (CAs, \text{syntactico-semantics, turn-taking}))$$

In Figure 1, there is an attention state name at the top level of the discourse f-structures. THEN, WH-QUESTION, INFORM, or RMK are CA units labeled with THEN, WH-QUESTION, INFORM and RMK, respectively. The syntactico-semantic f-structures for each of four CA units are substituted for corresponding CA units; thus their syntactico-semantics are represented. A and C indicate speak-
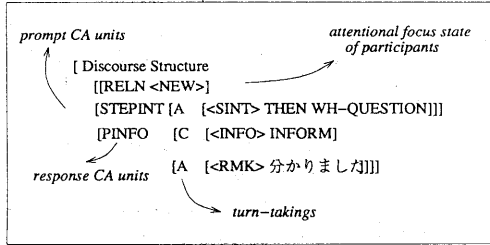
ers of the CA units; thus turn-taking is represented.

Figure 1: Discourse structure representation

3.2 Rewriting Rules for Units Aggregation

We wrote rewriting rules based on CA labels for CA unit aggregation, where CA units are unified and classified into groups. The rewriting rules consist of 39 terminal symbols[1], and three strata, from the pre-terminal symbol level to the top level in discourse tree structure. They make up a context dependent grammar. The rewriting rules are aimed only at aggregating CA units.

The basic motivation for the aggregation rules is: that CA units are often closely related. For instance, one CA unit may be a prompt and another CA unit may be its response. So, cohesion between CA units is made in the way prompt units cohere with their response units.[7][17] In this way, some unit groups are obtained from the dialogue. We consider each group to express an attentional focus state of dialogue participants as the discourse unfolds.

The process of CA unit aggregation 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 (Very good.Sir.) (RMK)

---

[1]There are 27 CA labels, but some subclassification including connectives was used in the experiments on discourse analysis.

THEN[2], WH-QUESTION, INFORM and RMK[3] 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*]
                    ]]]]]]]]

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

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

```
(<NEW> <--> (<STEPINT> <PINFO>))
(<STEPINT> <--> (THEN WH-QUESTION))
(<PINFO> <--> (INFORM RMK))
```

⟨NEW⟩ is an attentional state of dialogue participants, and it means that the attention of participants is shifted into a new state. ⟨STEPINT⟩ and ⟨PINFO⟩ are non-terminal categories of the second stratum of the grammar, and roughly mean a prompt and its response, respectively, in the aggregation. THEN, WH-QUESTION, INFORM and RMK are CA units labeled with THEN, WH-QUESTION, INFORM and RMK, respectively.

Discourse analysis experiments were conducted for 10 files of 5-10 minute Japanese dialogues taken from

---

[2]THEN is derivated from TOPIC.
[3]RMK is derivated from ACKNOWLEDGE.

the ATR dialogue corpus.[20] In order to evaluate the analysis results, we investigated whether references of anaphora were properly handled or not, *i.e.*, whether anaphoric pronouns and their reference words were within a focus state of attention. Comparing analysis results with human judgment for three pronouns, *i.e.*, そう (sou), それ (sore), その (sono), the analysis had an average success rate of 85 %.[17]

Conclusion

We introduced a set of CA labels and its applications 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 dialogue.

- CA units have been shown to fall on pause units in segmentation experiments, so we can apply the subtree 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 states of participants' attention in the dialogue. So, when CA units are aggregated into groups by using rules expressing the relationship, discourse structure can be made apparent.

References

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

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

[3] S.C.Levinson Pragmatics Cambridge University Press Cambridge 1983

[4] B.Grosz and Candy Sidner Attention, Intention, and the Structure of Discourse. *Computational Linguistics* 1986

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

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

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

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

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

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

[11] M. Seligman, L. Fais, M. Tomokiyo. A Bilingual Set of Communicative Act Labels for Spontaneous Dialogues TR-IT-0081 ATR Technical Report Kyoto 1994

[12] M. Tomokiyo. Natural Utterance Segmentation and Discourse Label Assignment Proc. of IC-SLP94 Vol.3 Yokohama 1994

[13] Singer H, Bepppu T, Nakamura A, Sagisaka Y A Modular Speech Recognition System Architecture Proc. of . Soc. Tokyo 1994

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

[15] T. Monzen S, Matsunaga S. Singer H. Time-Synchronous Continuous Speech Recognizer Driven by a Context-Free Grammar Proc. of ICASSP95 Japan 1995

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

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

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

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

[20] T.Morimoto, & al. A Speech and Language Database for Speech Translation Research Proc. of ICSLP94 Vol.4 Japan 1994