Segmentation and Aggregation of Utterances
by Using Speech Act Labels
toward a discourse analysis

Mutsuko Tomokiyo
ATR Interpreting Telecommunications Laboratories

The article presents segmentation and aggregation of utterances by using *speech act* labels for a discourse structure analysis for Japanese naturally-spoken dialogues. The utterance segmentation consists in segmenting utterances into some units by using cue patterns which are discovered through empirical studies of Japanese dialogues corpus at ATR. The segmented units are automatically assigned *speech act* labels.

The aggregation consists in aggregating segmented units in groups by using a grammar which is designed to use the *speech act* labels as terminal categories of the grammar. Each of the aggregated groups indicates an attentional state of the participants as the discourse unfolds. The discourse structure of dialogues is represented as transitions of the attentional state of the participants.

The article focuses on *speech act* labels, grammar rules and representation of discourse structure of spoken Japanese dialogue.

発話意図ラベルによる自然発話の分割と再構成
- 日本語談話解析にむけて

友清睦子
ATR 音声翻訳通信研究所

本稿は、自然に話された日本語の談話構造の解析を目的として、発話の自動分割とラベルの付与および分割された単位の再構成について述べるものである。発話分割は、点や丸のない書き起こし文を入力として、対話文コーパスの分析によってあらかじめ決められた文字列を使うことによってなされる。発話は一つのターンよりもより短い単位に分割され、その一つ一つに発話行為ラベルが自動的に割り当てられる。こうして分割された発話全体は、発話行為ラベルを終端記号とする書き変え規則によって、最終的にいくつかのグループにまとめられる。談話構造は、その談話の時間的流れにそって、話者双方が展開する話題の推移の仕方として表現される。

本稿は、主として発話行為ラベルとそれによる文法規則、および談話構造の意味表現について述べる。

Introduction

The article aims to present a discourse structure analysis and its representation for Japanese spoken dialogue in the context of telephone inquiries about travel information, while proposing an utterance segmentation and a label assignment to segmented units.

Japanese utterances consist of short clauses which are connected by auxiliary sequence, conjunctions or adverbs. The utterances with which we are dealing also contain repetitions, revision, and insertions of fragments or clauses, or consist of mere fragments. Difficulties of Japanese dialogue analysis consist in these situations as the concepts of subordinate and main clause, as defined within standard English grammar, do not apply to it.

It's the reason why we proposed an automatic segmentation of utterances and discourse label assignment to the segmented units, called *stars*.[14] The segmentation of utterances enables us to make manageable units for the discourse analysis. In addition, the discourse label assignment make discourse relation between *stars* apparent.[11]

After having made the segmentation and label assignment, a discourse structure analysis is undertaken by using rewriting rules which consist of *speech act* and *connectives* labels. The rewriting rules is used to aggregate the *stars* into some groups.

Representation of discourse structure consists in an abstraction of "the state of focus of the participants attention as the discourse unfolds".[5] The state of focus of the participants attention transfers from *Outset*, *On-and-on*, *Go-ahead*, *Repetition*, *Volt-face* and *Flashback*, *New* to *Upshot* in a dialogue.

The article focuses on the merits of using CAs and *connectives* labels to analyze the discourse structure of spoken Japanese dialogue and on presenting a discourse structure representation.
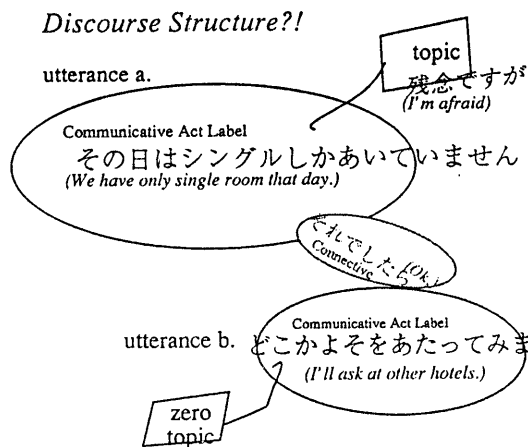
1 Discourse structure analysis

1.1 Aims of discourse structure analysis

Aims of the discourse analysis are two told in the context of the task of a machine translation or an interactive man-machine system:

1. to determiner the reference scope for pronouns or demonstratives, and the search range for supplying grammatical ellipsis

2. to represent topic transitions

1.2 Basic idea for the discourse structure analysis

Discourse Structure?!

utterance a.

topic
残念ですが
(I'm afraid)

Communicative Act Label
その日はシングルしかあいていません
(We have only single room that day.)

それでし
Connective た。OK

utterance b.
Communicative Act Label
どこかよそをあたってみま
(I'll ask at other hotels.)

zero topic

DS ->   ((topic CA) (Connective) (topic CA))

(Leech 1983. Grosz 1986. Szatrowski 1993)

The discourse analysis consists in aggregating *stars* in such a way as tracks participants attention in a dialogue, so-called topic transition. To aggregate the *stars* or to make groups of *stars*, we use rewriting rules we developed.

Basic thought is that the discourse structure depends on rethorical factors, intentional factors of speakers in utterances and turn takings in a dialogue. The rethorical one relates to how to express what a speaker wishes to convey. It's expressed by distinctive cue patterns which belong to various grammatical class; conjunctions, adverbs, interjections, and parenthetical clauses. We extracted them from EMMI corpus[8] and recast them into labels, called *connectives* in the article.

The intentional one relates to what a speaker wishes to convey or what kind of reactions from his conversational party he expects. It's expressed also by distinctive cue patterns which are located as predicate phrases at the end of clauses in Japanese. We also extracted them from EMMI corpus[ibid.] and recast them into labels, called *Communicative Act* (CA) in the article.

We hypothesize the discourse structure be represented as the sort of montage of the labels of *connectives* and CAs.

1.2.1 Communicative acts labels

A *communicative act* is a communicative goal or aim which can be expressed in a language L by a distinctive and descriptive set of conventional cue patterns in specified discourse contexts. *Communicative acts* are thus similar to speech acts, and similar to the pragmatic categories often called *illocutionary force* type. 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. Here are some examples for CAs.[13][7][12]

e.g.

A: お名前をお願いできます
か (Could I have your name?)
→*action-request*
C: 鈴木和子と言います (Suzuki, Kazuko Suzuki) →*inform*

A sentence final cue pattern "できます か"(Could I...?) expresses a interrogative sentence to which the conversational party replies "yes" or "no" as a response at the grammatical point of view on one hand, but on the other hand what the speaker wants to get isn't a response of 'yes' or 'no'. His expect in his utterance is to get the name of his conversational party. Hence, できますか is labeled in a different way from such an interrogative sentence as 東京は日本の首都 ですか (Is Tokyo the capital of Japan?).

> C:
> 鈴木和子です (Suzuki, Kazuko Suzuki) →*inform*
> A: 鈴木和子様で すね (OK, Ms.Kazuko Suzuki.) →*confirmation-question*

"ね" is a sentential final particle that indicates the speaker's request for confirmation or agreement from the conversational party about some shared knowledge. Hence, "ですね" is recast into different label from "です" at the point of the communicative goal.

CAs frequently depend on the tense or aspect of utterances.

> (a) こちらからホテルに部屋の 予約電話をしておき ます (I'll phone a hotel for accommodations for you.) →*promise*

> (b) こちらからホテルに部屋の 予約電話をしておき ました (I phoned a hotel for accommodations for you.) →*inform*

The utterance (a) in the present tense expresses a promise by the speaker, but the utterance (b) in the past tense expresses a kind of report by the speaker. Hence, the two utterances are differently assigned CAs at the point of view of the communicative goal, although there is no difference between the two except their tenses at the grammatical point of view.

CAs frequently depend on the context uttered in dialogues.

> *e.g.* (a)
> A: そこを左に曲がっていただき まして (Turn to the left there and..)
> C: はい(Hum) →acknowledge

> *e.g.* (b)
> A: 京都駅はご存知ですか (Do you know Kyoto Station?)
> C: はい(Yes, I do.) →yes

"はい" in (a) is as an acknowledge and one in the (b) is a response to yes-no question at the point of view of the communicative goal, although the two are the same cue patterns.

CA labels are recast from such surface cue patterns as above mentioned. Thus, we discovered 276 surface cue patterns in the EMMI corpus at ATR [8] and recast them into 27 CAs.

1.2.2 Connectives

Connective is defined cue patterns such as "つまり (after all)" and "えーと (well)", which serve as connectors between utterances. The *connectives* also

belong to various grammatical classes; conjunctions, adverbs, interjections, and parenthetical clauses[1]. Especially the parenthetical clauses contains expressions which previously explain the content of the utterance which will be uttered in next utterance.

e.g.

これから電話番号を
申し上げます
(I'll give you my phone number from now on).

These *connectives* are classified into 12 categories according to their functions. Consequently, we have 27 CAs and 12 *connectives* labels [2].

2   Rewriting rules for aggregating *stars*

A rule expresses two aspects at the same time: the immediate dominance between mother and daughter nodes and linear precedence between sister nodes like ordinary context free grammars. There are 39 categories, which consist of the *connectives* and CA labels as terminal symbols in the rewriting rules. The pro-

---

[1]Some of the expressions taken as *connectives* have both an ordinary use and a discourse use. We don't try to make differences between these two types of usages clear.

[2]There are ambiguities for the automatic assignment of a CA label to a *star*. The performance of CA assignments is 85.75% on the average.

jection along the bar-level from preterminals to the starting symbol is done through three strata at maximum. The top level represents a transferring discourse state in the context of a dialogue. The second level branches in *stimulus* and *reaction*. The *stimulus* or *reaction* is a *star* or a sequence of *stars*. The third level contains *reaction*s to the *reaction*. There are eight different kinds of *stimulus* and four different kinds of *reactions*. For example, utterances labeled with *question* or *action-request*, in principle belong to the stimulus. Utterances labeled with *inform* or *acknowledge* belong to, in principle the *reaction*.

There are rewriting rules which are applied only to *stars* within a turn, and rewriting rules which are applied only beyond turns. The rewriting rules containing the *connectives* labels are useful to determine a relation between *stars* beyond a turn: either to connect to previous *star* or to separate from it.

e.g.

A: いまどちらにいらっしゃいますか (wh-question)
    (Where are you located?)
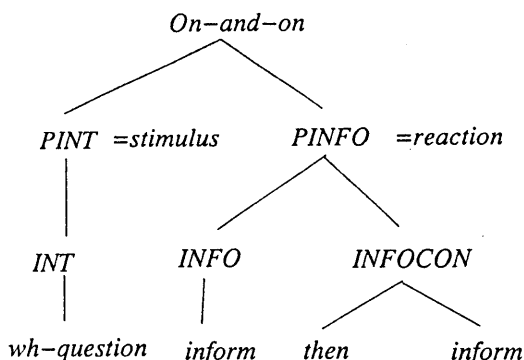C: 京都駅です (inform)
    (I'm at Kyoto Station.)
A: それでは (then)
    (So)
    地図のこのあたりです (inform)
    (You are here in the map.)

```
(<On-and-on> <--> (<PINT><PINFO>))
(<PINT> <--> (<INT>))
(<INT> <--> (wh-question))
(<PINFO> <--> (<INFO><INFOCON>))
(<INT> <--> (wh-question))
(<INFOCON> <--> (then inform))
(<INFO> <--> (inform))
```

```
                    On-and-on
                   /         \
                  /           \
         PINT =stimulus    PINFO =reaction
              |               /      \
              |              /        \
            INT          INFO        INFOCON
             |             |          /    \
             |             |         /      \
        wh-question     inform    then      inform
```

2.1 Representation of the discourse structure

The discourse structure is represented as results of an automatic description of a sequence of transferring discourse states in the context of a dialogue. The transferring discourse state is defined as a state of focus of the participants attention as the discourse unfolds.

The state of a focus of the participants attention is represented as one of 10 states including open and close of the dialogue: *Outset, On-and-on, Go-ahead, Repetition, Volt-face* and *Flashback, New* to *Upshot* and open/close-of-conversation. The analysis result for a short dialogue is shown in Appendix

Conclusion

One of the discourse analysis purposes is mentioned to determiner the reference scope for pronouns or demonstratives and the search range for grammatical ellipsis. The reference of demonstratives and pronouns falls 89% on average within an aggregated group, for 10 dialogues and 199 turns in total. So, CA and *connectives* labels are considered to be useful from segmenting utterances and assigning labels through the discourse structure analysis. However, it is an important question how to evaluate the discourse structure analysis. If the analysis is considered to be correctly performed, when topic transitions are correctly traced, in the sense, we need much more experiments to evaluate it.

References

[1] J.R.Searle Expression and meaning : Studies in the theory of Speech acts Cambridge University Press 1979 London

[2] Willis Edmondson Spoken Discourse Longman London and New York 1981

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

[4] Stephen C. Levinson Pragmatics Cambridge University Press Cambridge 1983

[5] Barbara J.Grosz Attention, Intention, and the Structure of Discourse Computational Linguistics Volume 12 Number 3 July-September 1986 175-205

[6] G.N.Leech 池上嘉彦他訳 語用論 紀伊国屋書店 1987 Tokyo

[7] Mutsuko Tomokiyo and Tsuyoshi Morimoto Communicative Functions of Spoken Japanese and Its Meaning Interpretation on MT System TR-I-0260 ATR technical report ATR 1992 Kyoto

[8] ローケン・キム他 マルチモーダル・シュミレータ EMMI を用いた道案内データベースのテキスト TR-IT-0029 ATR Technical Report 1993.12

[9] 柏崎雅世 日本語における行為指示型表現の機能 くろしお出版 1993 Tokyo

[10] Polly Szatrowski 日本語の談話の構造分析 くろしお出版 Tokyo 1993

[11] 友清睦子 語用論的分析に基づく自然発話の長文分割 TR-IT-0019 ATR Technical Report Sept.1993 Kyoto

[12] 友清睦子 対話行為ラベルとその自動付与 TR-IT-0069 ATR Technical Report Sept.1994 Kyoto

[13] Mark Seligman, Laurel Fais and Mutsuko Tomokiyo A Bilingual Set of Communicative Act Labels for Spontaneous Dialogues TR-IT-0081 ATR technical report ATR 1994 Kyoto

[14] Mutsuko Tomokiyo Natural Utterance Segmentation and Discourse Label Assignment Proceedings of ICSLP94 1994 Yokohama

Appendix

Analysis result

```
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-OF-CONVERSATION>]
  [A [<OPEN-OF-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)
  [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 <On-and-on>]
　 [C [<INT> then- じゃあ <AR>]]
　 [C [<AR> action-request]]
　 [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> 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-OF-CONVERSATION>]
　 [C [<THANK> thank]]
　 [A [<THANKR> thanks-response]]]]

N.B. 30 lines in the middle are omitted from output result.