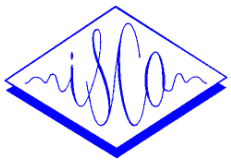

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

NATURAL UTTERANCE SEGMENTATION AND DISCOURSE LABEL ASSIGNMENT

Mutsuko TOMOKIYO

ATR Interpreting Telecommunications Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

Abstract

This paper proposes an approach based on pragmatics of spontaneously-spoken Japanese dialogue. Input to the system presented here appears as an unbroken stream of "utterance" *i.e*, a sequence of clauses or clause fragments separated by particles and other transitional elements, uttered by one speaker. To cope with data of this sort, a three-step procedure is employed. The aim of the present paper is motivate and discribe step1 and 2 of this three-step process. As step 1, the input stream is automatically segmented into discourse units and in step 2, discourse labels are assigned.
Test results of 1742 utterances show that the procedures presented here effected discourse segmentation correctly on average of 95.32% of the time, and they assigned discourse labels correctly on average of 85.12% of the time.

Introduction

This paper presents step 1, 2 of a three-step analysis of Japanese spoken dialogue in the context of telephone inquiries in travel information services. The focus of this discussion is on "single-inter locutor utterances", *i.e.* uninterrupted speech by one particular conversational participant in the context of a give-and-take conversation in which the parties interrupt each other's speech simultaneously, and so forth.
Japanese utterances consist of short clauses which are connected by an auxiliary sequence, particles, conjunctions or adverbs. One of the difficulties of Japanese-English machine translation is that the concept of subordinate and main clause, as defined within standard English grammar, does not apply to Japanese utterance analysis.
The utterances with which we are dealing also contain repetitions, revisions, and insersions of clauses or fragments. So, the concept of "sentence", *e.g.* within Chomskian grammar doesn't apply to utterance analysis.

In order to cope with this situation, a three-step procedure is employed. As step 1, we segment input utterances into sentence-like units which we call *star* by searching the predetermined character cues in the input uttrance. In step 2, each star is labelled with a descriptive label representing its discourse function, because the original discourse relation among the *stars* recovered from a given utterance should be preserved. We can call step1 and 2 of a three-step analysis discourse analysis.

The motivation behind this "repacking" of the input utterance is as follows:
First, the utterance is recast into simple syntactic units, useful for the purpose of, *e.g.*, translation. And the discourse function of each of these units is made apparent.

We use an analyzer, called KK analyzer for utterance segmentaions and distributions of discourse labels through discourse analysis.[17] The KK analyzer contains non-lexical rules, *viz* rules concerning *turn taking* of utterances.The discourse analysis is different from the syntactic analysis(step 3) in necessary knowledge to be used and results to be obtained.

The structure of this article is as follows: In section 1, pecularities of Japanese natural utterance are investigated. In Section 2, the concept of discourse function labels is motivated. An actual example of the discourse analysis under discussion is stepped through in Section 3. Section 4 presents a performance evaluation of the proposed discourse analysis by KK analyzer.

1. Pecularities of Japanese natural utterance

We investigate here the pecularities of spontaneously spoken Japanese and discuss possibility of utterance segmentation and assignment of discourse function labels in purpose of J-E machine translation.
An Utterance, as defined in this paper is an entire speech which is not interrupted by conversational party, so *turn taking* in dialogues produces *an* utterance.

(1) Natural utterances in Japanese consist of multiple sentences which are connected by auxiliary sequence as well as particles, conjunctions or adverb sequence.These strings express speaking behaviour, grammatical style of sentences, modality of sentences, discourse,..etc. It occurs in regular order from left-hand side to right-hand side at the end of a sentence: voice -> aspect -> mood -> negate -> tense

-> sentence final particles.

e.g.

ロイヤルホテルに泊まるんですけれども

(P1=I am going to stay in the Royal Hotel.)

会議場から遠いでしょうか

(P2=Is it far way from the conference hall?)

(I am going to stay in the Royal Hotel. So is it far way from the conference hall?)

The utterance can be divided into two independent sentences without changing its meaning.

んですけれども "ndesukeredomo" is composed with nominalization particle ん (n), formal auxiliary です (desu) and sentence final particle けれども (keredomo), and it introduces the proposition *P2* while explaining the background of the speech act in *P2* by *P1* at the point of view of pragmatics. In fact such string cues as the auxiliary sequence functions as discourse marker between *P1* and *P2* within the utterance.

(2) Natural utterances contain frequently insertion clause which is semantically independent of other clauses, but which is connected by conjunctions, adverbs and so forth. It is possible to put the case that the insertion clause be independent clause syntactically in an utterance.

e.g.

開園 いわゆる 開園時間というのでしょうか
開園時間がありますか

(Do you know if there is any, well can I say opening time?, any opening time or?)

いわゆる (iwayuru=so to speak) is used to evoke general knowledge from conversational party and lets his conversational party know that he is going to talk about something different from previous topics.
We can segment the utterance into 3 independent sentences without changing its meaning.
From the utterances as cited above, we confirm there are string cues which function not only as segmentation trigger of an utterance, but also as discourse marker.

We discovered 165 different string cues from 183 utterances in the investigation of EMMI dialogue corpus at ATR.[15][16] The discovered 165 string cues are used in step 1 to segment an utterance as segmantation triggers.

2. Discourse function label, DL

We finished the previous section by suggesting that we can recast the string cues in discourse function labels.

After having taken account of following points, the discovered string cues were recast into 15 different discourse labels manually.[7]

(1) Lexical items: set of auxiliaries, conjunctions, adverbs, particles or some verbs.

(2) Meanings of verb in the proposition of a sentence.

(3) Grammatical properties of proposition: person of the subject, tense..etc.

(4) *Turn taking*: We don't think *turn taking* is directly relevant to identification of *discourse* function as Edmondson mentioned.[3] However, *turn taking* helps us identify an answer to a question.

e.g. A: How is it? OK? and

A: How is it? B: OK.

(5) Stress and intonation contour.[12]

(6) Fragments expressing speaker's hesitation or vacillation in utterance.

One of problems is that the relation between the string cues and the discourse function labels is *many-to-many* rather than the one-to-one relationship. So, when *star*s of an utterance are labelled by one of 15 DLs in step 2, disambuguation should be made. Some of 15 DLs are presented in Appendix 1.

3. Discourse analysis

We focus on the utterance segmentation and the assignment of the DL for *star*s as the discourse analysis.The input of the discourse analysis is incremental tokens which are transcribed without any punctuations. We use following knowledge for the discourse analysis.

(1) DL Data List and segmentation triggers catalogue(STC): 165 string cues as mentioned in previous section.

(2) Rules concerning *turn taking*: It will occur or occured.

(3) Combination List of DLs (CL): DL as defalt value

(4) History of DLs of *star*s.

The discourse analysis is illustrated in figure 1.

First, an utterance is segmented and transformed into a set of *star*s by searching the character sequence of the input utterance with STC(discovered string cues) by longest string match.
Second, the *star*s are examined one by one as to whether they have only one or multiple DL candidates, by refering the DL Data List. If *star* has only one DL candidate, the DL indicated in it is assigned and the analysis is achieved. Third, if *star* has multiple candidates of DL, the history of labelled *star*s or occurance concerning *turn taking* are checked.

When a *star* isn't labelled through whole process indicated in the DL Data List, DL assignment for the *star* is skipped. The same procedure is continued until the last *star* in the utterance has been handled.
Finally, when all attempts to use the DL Data List for the utterance have completed and there still exists a *star* which isn't labelled, the history of *star*s of 2

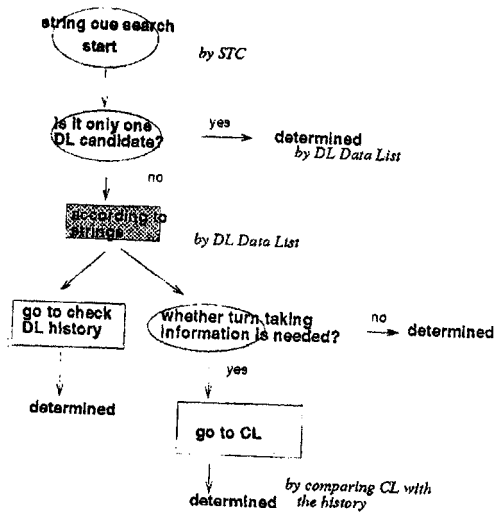

Figure 1: step 1 and step 2

adjacent utterances in focus is compared with CL. When the DL combination pattern of the previous utterance in the history is the same pattern as a DL combination pattern in CL, the unlabelled *star* is labelled by refering DL combination sequence in CL. Thus, all *star*s are labelled by DLs.
In the previous example, the first *star* withんですけれども (ndesukeredomo) has two candidates of DL:*attention request* and *information offer*.
In order to determine the DL for the first *star*, information as to whether turn taking will occur after the *star*, is needed as indicated in DL Data List. *Turn taking* does not occur, so the *star* is labelled as *attention request*. The string cue for the next*star*,

でしょうか (desyouka) has only one DL candidate. Consequently, the utterance is transformed into two independent sentences with the DLs of *attention request* and *information request*, respectively. Results of the discourse analysis is applied as input of the syntactic analysis.

{ndesukeredomo <-->

IFTCandidates(attention request,

information offer),

CheckHistory(no),

CheckTurnTaking((2,occur),(1,not))}

DL Data List for ndesukeredomo

4. Evaluation of the discourse analysis

We have based the discourse analysis upon a set of several telephone task dialogues, averaging 10 minutes each in length: 183 utterances as training data and 1742 utterances as test data for the sake of a performance evaluation of the discourse analysis.
Test results of 1742 utterances show that the procedure presented here effected discourse segmentation correctly on average of 95.32%, and they assigned discourse labels correctly on average of 85.12%.
The segmentation errors are produced by (hai=yes, right, OK), and label assignment errors are produced especially by default DLs.

Conclusion

The discourse analysis of the three-step analysis was presented in order to model Japanese natural dialogue. The discourse analysis is performed by using non-linguistic rules and serves to make *star*s from an utterance and to automatically assign discourse labels to them as the objectivization of natural utterance analysis.

Although the discourse analysis is performed by using mainly surface information, it conduces not only to the prevention of parsing failure on the analysis task, but also to ease of use, because *star*s are more manageable sentence units and DLs enable us to make apparent discourse relations among *star*s.The advantage of the three-step analysis is to enable us to analyse fragmental clauses or complex-type utterances as well as ordinary sentences.

When utterances are segmented into *star*s which are assigned by the discourse label, it is necessary to be described explicitly the relation among *star*s within an utterance, for the sake of translation into English. So, in step 3 we will use a grammar for representing relations among stars using connectives.

References

[1] P.Szatrowski Nihongo no danwano kouzoubunseki Kuroshio Tokyo 1993

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

[3] W.Edmondson Spoken Discourse Longman London and New York 1981

[4] P.Larreya Enoncé performatifs Présupposition Université Nathan information formation Fernand Nathan France 1979

[5] Kokuritukokugokenkyuusyo Nihongokyouiku-eiga Kisohen Sougoubunkenhyou Nihonsiniseru Tokyo 1987

[6] M.Sakuma Bundanninteinoichikijyun University of Tukuba Tukuba 1987

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

[8] M.Tomokiyo and T.Morimoto Communicative Functions of Spoken Japanese and Its Meaning Interpretation on MT System TR-I-0260 ATR technical report ATR Kyoto 1992

[9] M.Kasiwazaki Nihongoniokerukouisiji-katahyougennokinou Kuroshio Tokyo 1993

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

[11] E.Bilange A task independent oral Dialogue Model ACL proceedings Fifth European Conference Berlin 1991

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

[13] M.Nagata and K.Kogure HPSG-based Lattice Parser for Spoken Japanese in a Spoken Language Translation System Proc. of the 9th ECAI, 1990

[14] Jacqueline C.Kowtko, Stephen D.Isard and Gwyneth M. Doherty Conversational Games Within Dialogue HCRC Publications University of Edinburgh 1992

[15] Kyung-ho Loken-Kim, Fumihiro Yato, Tsuyoshi Morimoto 4-16 AV Information processing meeting IPSJ March, 1993

[16] Kyung-ho Loken-Kim, Fumihiro Yato, Laurie Fais, Kazuhiko Kurihara, Yoshihiro Kitagawa Transcription of spontaneous speech collection using a multimodal simulator-EMMI in a dirction finding task (Japanese-Japanese English-English) TR-IT-0029 ATR technical report ATR Kyoto 1993

[17] Kyung-ho Loken-Kim, Mark Seligman Yoshihiro Kitagawa Bidirectional Chart Parser: C language version (II) Manual TR-IT-0015 ATR technical report ATR Kyoto 1993

Appendix 1: Discourse Labels

The discourse function labels which are recast from the discovered string crues are as follows:

(1) discourse indication: such expressions as permit for a hearer to suppose contents of next utterance.

(2) conjunctive relation: conjunctions which are constituted of a pronoun and auxilliary. The pronoun has an anaphora in previous utterance.

(3) deictic indication: temporal or spacial indication in conversation in progress.

(4) information offer: answers to questions: Grammatical tense form must be present.

ます (masu, polite declarative verb ending)

(5) information request: to ask information. In general, sentente final particle か (ka=interrogative verb ending) is used:

(6) agreement request:to take an agreement: (desune=is it ok?) ですね

(7) agreement offer: positive answer to agreement request: ええ はい (I think so.)

(8) suggest:demand, suggest or order from speaker's party: てください (tekudasai=,please)

Appendix 2: Combination of DLs

When a *star* is labelled by a DL, an utterance is represented by the DL sequence, bacause utterances consist of one *star* or more *star*s. A dialogue consists of utterances, consequently it can be represented as DL.

utterance -> *star*1....*star*n

dialogue -> utterance1....utterancen

We can not know how many DL combination varieties for an utterance exist in dialogues, but we have currently 27 of different DL combination varieties, after having investigation of the dialogue corpus. And they are used as defalt DL list, when all atemptes to assign DL are failed. DL combination list is a sequence of adjacent *star* DLs within an utterance.

e.g. attention request -> attention request -> information request

We recognize there are some principles about DL sequence by which a dialogue stream is controlled.

(1) In the context that *information request* occurs after *attention request* in caller-side utterance, speaker explains the background to hearer before his demande of information. This means there is a principle concerning speaking behabiours

(2) *Attention offer* occurs before *information offer* or *agreement* offer in agent-side utterance. This means there is a principle concerning cooperative manner or sincerity which affects dialogue stream.

(3) In combination of *information offer* -> *information offer*, the utterance is dominated by speaker-side felicity principle, in which speaker wants to speak breifly and easily.[8]

(4) In a pair of utterances, first *star* in later utterance is corresponding to the last *star* in previous utterance in semantic view.