# A GRAMMAR AND A PARSER FOR SPONTANEOUS SPEECH

Mikio Nakano, Akira Shimazu, and Kiyoshi Kogure

NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, 243-01 Japan
{nakano, shimazu, kogure}@atom.ntt.jp

## ABSTRACT

This paper classifies distinctive phenomena occurring in Japanese spontaneous speech, and proposes a grammar and processing techniques for handling them. Parsers using a grammar for written sentences cannot deal with spontaneous speech because in spontaneous speech there are phenomena that do not occur in written sentences. A grammar based on analysis of transcripts of dialogues was therefore developed. It has two distinctive features: it uses short units as input units instead of using sentences in grammars for written sentences, and it covers utterances including phrases peculiar to spontaneous speech. Since the grammar is an augmentation of a grammar for written sentences, it can also be used to analyze complex utterances. Incorporating the grammar into the distributed natural language processing model described elsewhere enables the handling of utterances including variety of phenomena peculiar to spontaneous speech.

## 1 INTRODUCTION

Most dialogue understanding studies have focused on the mental states, plans, and intentions of the participants (Cohen et al., 1990). These studies have presumed that utterances can be analyzed syntactically and semantically and that the representation of the speech acts performed by those utterances can be obtained. Spontaneously spoken utterances differ considerably from written sentences, however, so it is not possible to analyze them syntactically and semantically when using a grammar for written sentences.

*Spontaneous speech*, a sequence of spontaneously spoken utterances, can be distinguished from well-planned utterances like radio news and movie dialogues. Much effort has been put into incorporating grammatical information into speech understanding (e.g., Hayes et al. (1986), Young et al. (1989), Okada (1991)), but because this work has focused on well-planned utterances, spontaneously spoken utterances have received little attention. This has partly been due to the lack of a grammar and processing technique that can be applied to spontaneous speech. Consequently, to attain an understanding of dialogues it is necessary to develop a way to analyze spontaneous speech syntactically and semantically.

There are two approaches to developing this kind of analysis method: one is to develop a grammar and analysis method for spontaneous speech that do not depend on syntactic constraints as much as the conventional methods for written sentences do (Den, 1993), and the other is to augment the grammar used for written sentences and modify the conventional analysis method to deal with spontaneous speech. The former method would fail, however, when new information is conveyed in the utterances; that is, when the semantic characteristics of the dialogue topic are not known to the hearer. In such cases, even in a dialogue, the syntactic constraints are used for understanding utterances. Because methods that disregard syntactic constraints would not work well in these kinds of cases, we took the latter approach.

We analyzed more than a hundred dialogue transcripts and classified the distinctive phenomena in spontaneous Japanese speech. To handle those phenomena, we develop a computational model called *Ensemble Model* (Shimazu et al., 1993b), in which syntactic, semantic, and pragmatic processing modules and modules that do combination of some or all of those processing analyze the input in parallel and independently. Even if some of the modules are unable to analyze the input, the other modules still output their results. This model can handle various kinds of irregular expressions, such as case particle omission, inversions, and fragmentary expressions.

We also developed *Grass-J* (*Gra*mmar for *s*pontaneous *s*peech in *J*apanese), which enables the syntactic and semantic processing modules of the Ensemble Model to deal with some of the phenomena peculiar to spontaneous speech. Since *Grass-J* is an augmentation of a grammar used to analyze written sentences (*Grat-J*, *Gra*mmar for *t*exts in *J*apanese), *Grass-J*-based parsers can be used for syntactically complex utterances.

There are two distinctive features of *Grass-J*. One is that its focus is on the short units in spontaneous speech, called *utterance units*. An utterance unit instead of a sentence as in *Grat-J* is used as a grammatical category and is taken as the start symbol. A *Grass-J*-based parser takes an utterance unit as input and outputs the representation of the speech act (illocutionary act) performed by the unit. The other distinctive feature is a focus on expressions peculiar to spontaneous speech, and here we explain how to augment *Grat-J* so that it can handle them. Previous studies of spontaneous speech analysis have focused mainly on repairs and ellipses (Bear et al., 1992; Langer, 1990; Nakatani & Hirschberg, 1993; Otsuka & Okada, 1992), rather than expressions peculiar to spontaneous speech.

This paper first describes *Grat-J*, and then classifies distinctive phenomena in Japanese spontaneous speech. It then describes *Grass-J* and presents several analysis examples.

1. Subcategorization rule
   Rule for NP (with particle) -VP constructions.
   M → C H
   ⟨M head⟩ = ⟨H head⟩
   ⟨H subcat⟩ = ⟨M subcat⟩ ∪ ⟨C⟩
   ⟨M adjacent⟩ = nil
   ⟨H adjacent⟩ = nil
   ⟨M adjunct⟩ = ⟨H adjunct⟩
   ⟨M lexical⟩ = –
   ⟨M sem index⟩ = ⟨H sem index⟩
   ⟨M sem restric⟩
   = ⟨C sem restric⟩ ∪ ⟨H sem restric⟩
   Symbols M, C, and H are not names of categories but variables, or identifiers of root nodes in the graphs representing feature structures. M, C, and H correspond to mother, complement daughter, and head daughter. The head daughter's subcat feature value is a set of feature structures.
2. Adjacency rule
   Rule for VP-AUXV constructions, NP-particle constructions, etc.
   M → A H
   ⟨M head⟩ = ⟨H head⟩
   ⟨M subcat⟩ = ⟨H subcat⟩
   ⟨H adjacent⟩ = ⟨A⟩
   ⟨M adjacent⟩ = nil
   ⟨M adjunct⟩ = ⟨H adjunct⟩
   ⟨M lexical⟩ = –
   ⟨M sem index⟩ = ⟨H sem index⟩
   ⟨M sem restric⟩
   = ⟨A sem restric⟩ ∪ ⟨H sem restric⟩
   M, A, and H correspond to mother, adjacent daughter, and head daughter. The head daughter's adjacent feature value is unified with the adjacent daughter's feature structure.
3. Adjunction rule
   Rule for modifier-modifiee constructions.
   M → A H
   ⟨M head⟩ = ⟨H head⟩
   ⟨M subcat⟩ = ⟨H subcat⟩
   ⟨H adjacent⟩ = nil
   ⟨A adjunct⟩ = ⟨H⟩
   ⟨M lexical⟩ = –
   ⟨M sem index⟩ = ⟨H sem index⟩
   ⟨M sem restric⟩
   = ⟨A sem restric⟩ ∪ ⟨H sem restric⟩
   M, A, and H correspond to mother, adjunct daughter (modifier), and head daughter (modifiee). The adjunct daughter's adjunct feature value is the feature structure for the head daughter.

Fig. 1: Phrase structure rules in *Grat-J* .

## 2 A GRAMMAR FOR WRITTEN SENTENCES

*Grat-J*, a grammar for written sentences, is a unification grammar loosely based on Japanese phrase structure grammar (JPSG) (Gunji, 1986). Of the six phrase structure rules used in *Grat-J*, the three related to the discussion in the following sections are shown in Fig. 1 in a PATR-II like notation (Shieber, 1986).[1] Lexical items are represented by feature structures, and example of which is shown in Fig. 2.

*Grat-J*-based parsers generate semantic representations in logical form in Davidsonian style. The semantic representation in each lexical item consists of a variable called an *index* (feature ⟨sem index⟩) and *restrictions* placed on it (feature ⟨sem restric⟩). Every time a phrase structure rule is applied, these restrictions are aggregated and a logical form is synthesized.

For example, let us again consider 'aisuru' (love). If, in the feature structure for the phrase 'Taro ga' ('Taro-NOM'), the ⟨sem index⟩ value is *p and the ⟨sem restric⟩ value is {(taro *p)}, after the subcategorization rule is applied the ⟨sem restric⟩ value in the resulting feature structure for the phrase 'Taro ga aisuru' (Taro loves) is {(taro *x) (love *e) (agent *e *x) (patient *e *y)}.

*Grat-J* covers such fundamental Japanese phenomena as subcategorization, passivization, interrogation, coordination, and negation, and also covers copulas, relative clauses, and conjunctions. We developed a parser based on *Grat-J* by using bottom-up chart parsing (Kay, 1980). Unification operations are performed by using constraint projection, an efficient method for unifying disjunctive feature descriptions (Nakano, 1991). The parser is implemented in Lucid Common Lisp ver. 4.0.

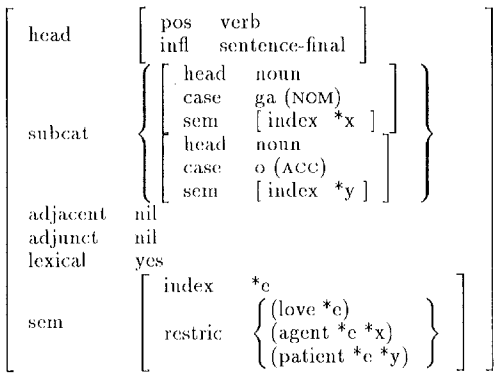



Fig. 2: Feature structure for the word 'aisuru' (love).

## 3 DISTINCTIVE PHENOMENA IN JAPANESE SPONTANEOUS SPEECH

### 3.1 Classification of Phenomena

We analyzed 97 telephone dialogues (about 300,000 bytes) about using LaTeX to prepare documents and 26 dialogues (about 160,000 bytes) obtained from three radio listener call-in programs (Shimazu et al., 1993a). We found that augmenting the grammars and analysis methods requires taking into account at least the following six phenomena in Japanese spontaneous speech.

(p1) expressions peculiar to Japanese spontaneous speech, including fillers (or hesitations).
   (ex.) 'etto aru ndesukedomo ... ' 'kono fairu tte ...' (well, we have them... this file is...)

(p2) particle (case particle) omission
   (ex.) 'sore watashi yarimasu' (I will do it.)

(p3) main verb ellipsis, or fragmentary utterances

[1] Rules for relative clauses and for verb-phrase coordinations are not shown here.

(ex.) 'aa, shinkansen de Kyoto kara.' (uh, from Kyoto by Shinkansen line.)

(p4) repairing phrases
(ex.) 'ano chosya be, chosya no arufabetto jun ni naran da, indekkusu naai?' (well, are there, aren't there indices ordered alphabetically by authors' names?)

(p5) inversion
(ex.) 'kopii shite kudasai, sono ronbun.' (That paper, please copy.)

(p6) semantic mismatch of the theme/subject and the main verb
(ex.) 'rikuesuto no uketsuke jikan wa, 24-jikan jouji uketsuke teori masu.' (The hours we receive your requests, they are received 24 hours a day.)

### 3.2 Treatment of the Phenomena by the Ensemble Model

These kinds of phenomena can be handled by the Ensemble Model. As described in Section 1, the Ensemble Model has syntactic, semantic, and pragmatic processing modules and modules that do combination of some or all of those processings to analyze the input in parallel and independently. Their output is unified, and even if some of the modules are unable to analyze the input, the other modules output their own results. This makes the Ensemble Model robust. Moreover, even if some of the modules are unable to analyze the input in real-time, the others output their results in real-time.

The Ensemble Model has been partially implemented, and Ensemble/Trio-I consists of syntactic, semantic, and syntactic-semantic modules. It can handle (p2) above as described in detail elsewhere (Shimazu et al., 1993b). Phenomena (p3) through (p6) can be partly handled by another implementation of the Ensemble Model: Ensemble/Quartet-I, which has pragmatic processing module as well as the three modules of Ensemble/Trio-I. The pragmatic processing module uses plan and domain knowledge to handle not only well-structured sentences but also ill-structured sentences, such as those including inversion and omission (Kogure et al., 1994).

To make the system more robust by enabling the syntactic and semantic processing modules to handle phenomena (p1) and (p3) through (p6), we incorporated *Grass-J* into those modules. *Grass-J* differs from *Grat-J* in two ways: *Grass-J* has lexical entries for expressions peculiar to spontaneous speech, so that it can handle (p1). And because sentence boundaries are not clear in spontaneous speech, it uses the concept of *utterance unit* (Shimazu et al., 1993a) instead of sentence. This allows it to handle phenomena (p3) through (p6). For example, an inverted sentence can be handled by decomposing it, at the point where the inversion occurs, into two utterance units.

Fig. 3 shows the architecture of Ensemble /Quartet-I. Each processing module is based on the bottom-up chart analysis method (Kay, 1980) and a disjunctive feature description unification method called constraint projection (Nakano, 1991). The syntactic-semantic processing module uses *Grass-J*, the syntactic processing module uses *Grass-J* without semantic constraints such as sortal restriction, the semantic processing module uses *Grass-J* without syntactic constraints such as case information, and the pragmatic processing module uses a plan-based grammar.

A: 1 *anoo kisoken*
well the Basic Research Labs.
*eno ikikata o desune*
to how to go ACC
'well, how to go to the Basic Research Labs.'

B: 2 *hai*
uh-huh
'uh-huh'

A: 3 *chotto shira nai nde*
well know NOT because
'because I don't know well'

4 *oshie teitadaki tai ndesukedo*
tell HAVE-A-FAVOR want
'I'd like you tell me it'

Fig. 4: Dialogue 1.

## 4 A GRAMMAR FOR SPONTANEOUS SPEECH

This section describes *Grass-J*.

### 4.1 Processing Units

'Sentence' is used as the start symbol in grammars for written languages but sentence boundaries are not clear in spontaneous speech. 'Sentence' therefore cannot be used as the start symbol in grammars for spontaneous speech. Many studies, though, have shown that utterances are composed of short units (Levelt, 1989: pp. 23–24), that need not be sentences in written language. *Grass-J* uses such units instead of sentences.

Consider, for example, Dialogue 1 in Fig. 4. Utterances 1 and 3 cannot be regarded as sentences in written language. Let us, however, consider 'hai' in Utterance 2. It expresses participant B's confirmation of the contents of Utterance 1. [2] Each utterance in Dialogue 1 can thus be considered to be a speech act (Shimazu et al., 1993a). These utterances are processing units we call *utterance units*. They are used in *Grass-J* instead of the sentences used in *Grat-J*. One feature of these units is that 'hai' can be interjected by the hearer at the end of the unit.

The boundaries for these units can be determined by using pauses, linguistic clues described in the next section, syntactic form, and so on. In using syntactic form to determine utterance unit boundaries, *Grass-J* first stipulates what an utterance unit actually is. This stipulation is based on an investigation of dialogue transcripts, and in the current version of *Grass-J*, the following syntactic constituents are recognized as utterance units.

- verb phrases (including auxiliary verb phrases and adjective phrases) that may be followed by

[2] The roles of 'hai', an interjectory response corresponding to a back-channel utterance such as *uh-huh* in English but which occurs more frequently in Japanese dialogue, are discussed in Shimazu et al. (1993a) and Katagiri (1993).

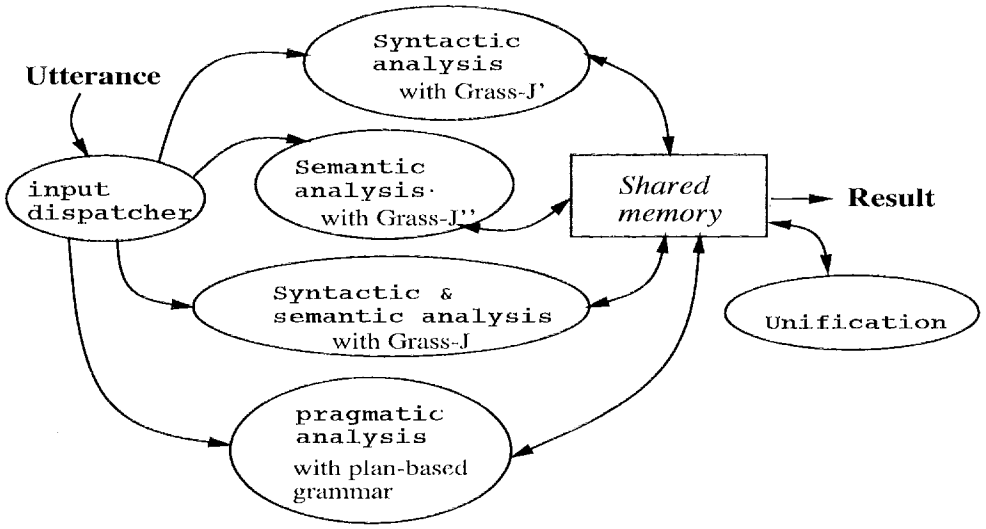



Fig. 3: Architecture of Ensemble/Quartet-I.

conjunctive particles and sentence-final particles

- noun phrases, which may be followed by particles
- interjections
- conjunctions

*Grass-J* includes a bundle of phrase structure rules used to derive speech act representation from the logical form of these constituents. A *Grass-J*-based parser inputs an utterance unit and outputs the representation of the speech act performed by the unit, which is then input to the discourse processing system.

Consider the following simple dialogue.

A: 1 *genkou* o
manuscript ACC
'The manuscript'

B: 2 *hai*
uh-huh
'uh-huh'

A: 3 *okut tekudasai*
send please
'please send me'

The logical form for Utterance 1 is ((manuscript *x)), so that its resulting speech act representation is

(1) ((refer *e) (agent *e *s) (speaker *s) (object *e *x) (manuscript *x))[3]

or, as written in usual notation,

(2) Refer(speaker, ?x:manuscript(?x)).

In the same way, the speech act representation for Utterance 3 is

(3) Request(speaker, hearer, send(hearer, speaker, ?y)).

The discourse processor would find that ?x in (2) is the same as ?y in (3). A detailed explanation of this discourse processing is beyond the scope of this paper.

[3] 'Refer' stands for the *surface* referring in Allen and Perrault (1980).

### 4.2 Treatment of Expressions Peculiar to Spontaneous Speech

**Classification**

The underlined words in Dialogue 1 in Fig. 4 do not normally appear in written sentences. We analyzed the dialogue transcripts to identify expressions that frequently appear in spoken sentences which includes spontaneous speech but that do not appear in written sentences, and we classified them as follows.

1. words phonologically different from those in written sentences (words in parenthesis are corresponding written-sentence words)
   (ex.) 'shinakya' ('shinakereba', if someone does not do), 'shichau' ('shiteshimau', have done)
2. fillers (or hesitations such as *well* in English)
   (ex.) 'etto', 'anoo'
3. particles peculiar to spoken language
   (ex.) 'tte', 'nante', 'toka'
4. interjectory particles (words inserted interjectorily after noun phrases and adverbial/adnominal-form verb phrases)
   (ex.) 'ne', 'desune', 'sa'
5. expressions introducing topics
   (ex.) '(na)ndesukedo', '(na)ndesukedomo', '(na)ndesuga'
6. words appearing after main verb phrases (these words take the sentence-final form of verbs/auxiliary verbs/adjectives)
   (ex.) 'yo', 'ne', 'yone', 'keredo', 'kedo', 'keredomo', 'ga', 'kedomo', 'kara'

Nagata and Kogure (1990) addressed Japanese sentence-final expressions peculiar to spoken Japanese sentences but did not deal with all the spontaneous speech expressions listed above. These expressions may be analyzed morphologically (Takeshita & Fukunaga, 1991). Because some expressions peculiar to spontaneous speech do not affect the propositional

content of the sentences, disregarding those expressions might be a way to process spontaneous speech. Such cascaded processing of morphological analysis and syntactic and semantic analysis disables the incremental processing required for real-time dialogue understanding. Another approach is to treat these kinds of expressions as extra, 'noisy' words. Although this can be done by using a robust parsing technique, such as the one developed by Mellish (1989), it requires the sentence to be processed more than two times, and is therefore not suitable for real-time dialogue understanding. In *Grass-J* these expressions are handled in the same way as expressions appearing in written language, so no special techniques are needed.

**Words phonologically different from corresponding words in written-language**

The words 'teru' and 'ndesu' in 'shit teru ndesu ka' (do you know that?) correspond semantically to 'teiru' and 'nodesu' in written sentences. We investigated such words in the dialogue data (Fig. 5). One way to handle these words is to translate them into their corresponding written-language words, but because this requires several steps it is not suitable for incremental dialogue processing. We therefore regard these words as independent of their corresponding words in written-language, even though their lexical entries have the same content.

**Fillers**

Fillers such as 'anoo' and 'etto', which roughly correspond to *well* in English, appear frequently in spontaneous speech (Arita et al., 1993) and do not affect the propositional content of sentences in which they appear[4]. One way to handle them is to disregard them after morphological analysis is completed. As noted above, however, such an approach is not suitable for dialogue processing. We therefore treat them directly in parsing.

In *Grass-J*, fillers modify the following words, whatever their grammatical categories are. The feature structure for fillers is as follows.

$$\begin{bmatrix} \text{head} & [\text{pos interjection}] \\ \text{subcat} & \{\} \\ \text{adjunct} & [\text{lexical } +] \\ \text{adjacent} & \text{nil} \\ \text{lexical} & + \\ \text{sem} & [\text{restric } \{\}] \end{bmatrix}$$

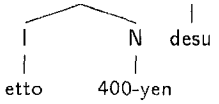
The value of the feature *lexical* is either + or −: it is + in lexical items and − in feature structures for phrases composed, by phrase structure rules, of subphrases. Because these words do not affect propositional contents, the value of the feature ⟨sem restric⟩ is empty.

For example, let us look at the parse tree for 'etto 400-yen desu' (well, it's 400 yen). Symbols I (Interjection), NP, and VP are abbreviations for the complex feature structures.

[4]Although Sadanobu and Takubo (1993) investigated the discourse management function of fillers, we do not discuss it here.

1. expressions related to aspects
   teku (teiku in written-language), teru (teiru), chau (tesimau), etc.
2. expressions related to topic marker 'wa'
   cha (tewa), chaa (tewa), ccha (tewa), ja (dewa), etc.
3. expressions related to conjunctive particle 'ba'
   nakerya (nakereba), nakya (nakereba), etc.
4. expressions related to formal nouns
   n (no), mon (mono), toko (tokoro), etc.
5. demonstratives
   kocchi (kochira), korya (korewa), so (sou), soshitara (soushitara), sokka (souka), socchi (sochira), son (sono), soreja (soredewa), sorejaa (soredewa), etc.
6. expressions related to interrogative pronoun *nani*
   nanka (nanika), nante (nanito), etc.
7. other
   mokkai (mouikkai), etc.

Fig. 5: Words that in spoken language differ from corresponding words in written language.




The filler 'etto' modifies the following word '400-yen' and the logical form of the sentence is the same as that of '400-yen desu'.

**Particles peculiar to spoken language**

Words such as 'tte' in 'Kyoto tte Osaka no tsugi no eki desu yone' (Kyoto is the station next to Osaka, isn't it?) work in the same way as case-marking/topic-marking particles. Because they have no corresponding words in written language, lexical entries for them are required. These words do not correspond to any specific surface case, such as 'ga' and 'o'. Like the topic marker 'wa', the semantic relationships they express depend on the meaning of the phrases they connect.

**Interjectory particles**

Interjectory particles, such as 'ne' and 'desune', follow noun phrases and adverbial/adnominal-form verb phrases, and they do not affect the meaning of the utterances. The interjectory particle 'ne' differs from the sentence-final particle 'ne' in the sense that the latter follows sentence-final form verb phrases. These kinds of words can be treated by regarding them as particles following noun phrases and verbs phrases. The following is the feature structure for these words.

$$\begin{bmatrix} \text{head} & *1 \\ \text{subcat} & \{\} \\ \text{adjunct} & \text{nil} \\ \text{adjacent} & \begin{bmatrix} \text{head} & *1 \\ \text{sem} & [\text{index } *2] \end{bmatrix} \\ \text{lexical} & + \\ \text{sem} & \begin{bmatrix} \text{index} & *2 \\ \text{restric} & \{\} \end{bmatrix} \end{bmatrix}$$

The interjectory particles indicate the end of utterance units; they do not appear in the middle of utterance units. They function as, so to speak, utterance-unit-final particles. Therefore, a noun phrase followed by an interjectory particle forms a (surface) *referring* speech act in the same way as noun phrase utterances. Interjectory particles add nothing to logical forms. For example, the speech act representation of 'genkou o desune' is the same as (2) in Section 4.1.

**Expressions introducing topics**

As in Utterance 4 of Dialogue 1, an expression such as *(na)ndesukedo(mo)* frequently appears in dialogues, especially in the beginning. This expression introduces a new topic. One way to handle an expression such as this is to break it down into $na + ndesu + kedo + mo$. This process, however, prevents the system from detecting its role in topic introduction. We therefore consider each of these expressions to be one word. The reason these expression are used is to make a topic explicit, by introducing a discourse referent (Thomason, 1990). Consequently, an 'introduce-topic' speech act is formed. These expressions indicate the end of an utterance unit as an interjectory particle.

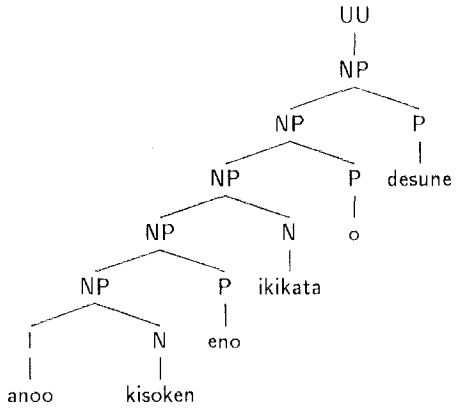
**Words appearing after main verb phrase**

It has already been pointed out that sentence-final particles, such as 'yo' and 'ne', frequently appear in spoken Japanese sentences (Kawamori, 1991). Conjunctive particles, such as 'kedo' and 'kara', are also used as sentence-final particles (Hosaka et al., 1991) and they are treated as such in *Grass-J*. They perform the function of anticipating the hearer's reaction, as a trial expression does (Clark & Wilkes-Gibbs, 1990). They also indicate the end of utterance units.

## 5 ANALYSIS EXAMPLES

Below we show results obtained by using a *Grass-J*-based parser to analyze some of the utterances in Dialogue 1. UU means the utterance unit category.

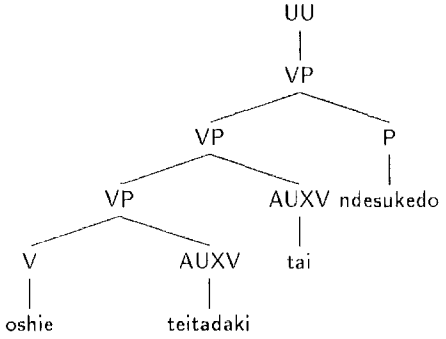
- Utterance 1: 'anoo kisoken eno ikikata o desune' (well, how to go to the Basic Research Labs.)
  parse tree:




speech act representation:

```
index = *X29
restriction =
 ((REFER *X29) (OBJECT *X29 *X30)
  (AGENT *X29 *X31) (SPEAKER *X31)
  (BASIC-RESEARCH-LABS *X32)
  (DESTINATION *X30 *X32)
  (HOW-TO-GO *X30))
```

- Utterance 4: 'oshie teitadaki tai ndesukedo' (I'd like you to tell me it)
  parse tree:




speech act representation:

```
index = *X777
restriction =
 ((INTRODUCE-TOPIC *X777)
  (OBJECT *X777 *X778)
  (AGENT *X777 *X779)
  (SPEAKER *X779)
  (TELL *X780)
  (AGENT *X780 *X808)
  (OBJECT *X780 *X809)
  (PATIENT *X780 *X810)
  (HAVE-A-FAVOR *X784)
  (OBJECT *X784 *X780)
  (AGENT *X784 *X811)
  (SOURCE *X784 *X808)
  (WANT *X778)
  (OBJECT *X778 *X784)
  (AGENT *X778 *X811))
```

## 6 CONCLUSION

We have developed a grammar, called *Grass-J*, for handling distinctive phenomena in spontaneous speech. The grammatical analysis of spontaneous speech is useful in combining the fruits of dialogue understanding research and those of speech processing research. As described earlier, *Grass-J* is used as the grammar for the experimental systems Ensemble/Trio-1 and Ensemble/Quartet-1, which are based on the Ensemble Model. It enables the processing of several kinds of spontaneous speech, such as that lacking particles.

We focused on processing transcripts because a grammar and an analysis method for spontaneous speech can be combined with speech processing systems more accurately than can those for written languages.

Finally, although we focused only on Japanese spontaneous speech, most of the techniques described in this paper can also be used to analyze spontaneous speech in other languages.

## ACKNOWLEDGEMETNS

We thank Chung Pai Ling, Yuiko Otsuka, Miyoko Sou, Kaeko Matsuzawa, and Sanae Nagata, for helping us analyze dialogue data.

## REFERENCES


Allen, J. F., & Perrault, C. R. (1980). Analyzing Intention in Utterances. *Artificial Intelligence, 15*, 143–178.

Arita, H., Kogure, K., Nogaito, I., Maeda, H., & Iida, H. (1993). Media-Dependent Conversation Manners. In *SIG-NL-61, Information Processing Society of Japan.* (in Japanese).

Bear, J., Dowding, J., & Shriberg, E. (1992). Integrating Multiple Knowledge Sources for the Detection and Correction of Repairs in Human-Computer Dialog. In *ACL-92*, pp. 56–63.

Clark, H. H., & Wilkes-Gibbs, D. (1990). Referring as a Collaborative Process. In Cohen, P. R., Morgan, J., & Pollack, M. E. (Eds.), *Intentions in Communication*, pp. 463–493. MIT Press.

Cohen, P. R., Morgan, J., & Pollack, M. E. (Eds.). (1990). *Intentions in Communication.* MIT Press.

Den, Y. (1993). A Study on Spoken Dialogue Grammar. *SIG-SLUD-9302-5, Japanese Society of AI*, 33–40. (in Japanese).

Gunji, T. (1986). *Japanese Phrase Structure Grammar.* Reidel, Dordrecht.

Hayes, P. J., Hauptmann, A. G., Carbonell, J. G., & Tomita, M. (1986). Parsing Spoken Language: A Semantic Caseframe Approach. In *COLING-86*, pp. 587–592.

Hosaka, J., Takezawa, T., & Ehara, T. (1991). Constructing Syntactic Constraints for Speech Recognition using Empirical Data. In *SIG-NL-83, Information Processing Society of Japan*, pp. 97–104. (in Japanese).

Katagiri, Y. (1993). Dialogue Coordination Functions of Japanese Sentence-Final Particles. In *Proceedings of International Symposium on Spoken Dialogue*, pp. 145–148.

Kawamori, M. (1991). Japanese Sentence Final Particles and Epistemic Modality. *SIG-NL-84, Information Processing Society of Japan*, 41–48. (in Japanese).

Kay, M. (1980). Algorithm Schemata and Data Structures in Syntactic Processing. Tech. rep. CSL-80-12, Xerox PARC.

Kogure, K., Shimazu, A., & Nakano, M. (1994). Plan-Based Utterance Understanding. In *Proceedings of the 48th Conference of Information Processing Society of Japan*, Vol. 3, pp. 189–190. (in Japanese).

Langer, H. (1990). Syntactic Normalization of Spontaneous Speech. In *COLING-90*, pp. 180–183.

Levelt, W. J. M. (1989). *Speaking.* MIT Press.

Mellish, C. (1989). Some Chart-Based Techniques for Parsing Ill-Formed Input. In *ACL-89*, pp. 102–109.

Nagata, M., & Kogure, K. (1990). HPSG-Based Lattice Parser for Spoken Japanese in a Spoken Language Translation System. In *ECAI-90*, pp. 461–466.

Nakano, M. (1991). Constraint Projection: An Efficient Treatment of Disjunctive Feature Descriptions. In *ACL-91*, pp. 307–314.

Nakatani, C., & Hirschberg, J. (1993). A Speech-First Model for Repair Detection and Correction. In *ACL-93*, pp. 46–53.

Okada, M. (1991). A Unification-Grammar-Directed One-Pass Search Algorithm for Parsing Spoken Language. In *Proceedings of ICASSP-91.*

Otsuka, H., & Okada, M. (1992). Incremental Elaboration in Generating Spontaneous Speech. *SIG-NLC92-41, Institute of Electronics, Information and Communication Engineers.* (in Japanese).

Sadanobu, T., & Takubo, Y. (1993). The Discourse Management Function of Fillers–a case of "eeto" and "ano(o)"–. In *Proceedings of International Symposium on Spoken Dialog*, pp. 271–274.

Shieber, S. M. (1986). *An Introduction to Unification-Based Approaches to Grammar.* CSLI Lecture Notes Series No. 4. Stanford: CSLI.

Shimazu, A., Kawamori, M., & Kogure, K. (1993a). Analysis of Interjectory Responses in Dialogue. *SIG-NLC-93-9, Institute of Electronics, Information and Comunication Engineers.* (in Japanese).

Shimazu, A., Kogure, K., & Nakano, M. (1993b). An Experimental Distributed Natural Language Processing System and its Application to Robust Processing. In *Proceedings of the Symposium on Implementation of Natural Language Processing.* Institute of Electronics, Information and Communication Engineers/Japan Society for Software Science and Technology. (in Japanese).

Takeshita, A., & Fukunaga, H. (1991). Morphological Analysis for Spoken Language. In *Proceedings of the 42nd Conference of Information Processing Society of Japan*, Vol. 3, pp. 5–6. (in Japanese).

Thomason, R. H. (1990). Accommodation, Meaning, and Implicature: Interdisciplinary Foundations for Pragmatics. In Cohen, P. R., Morgan, J., & Pollack, M. E. (Eds.), *Intentions in Communication*, pp. 325–364. MIT Press.

Young, S. R., Hauptmann, A. G., Ward, W. H., Smith, E. T., & Werner, P. (1989). High Level Knowledge Sources in Usable Speech Recognition Systems. *Communication of the ACM, 32*(2), 183–194.