

# Coordination and Feedback: Roles of interjectory utterances in spoken Japanese discourse

Masahito KAWAMORI  
NTT Basic Research Laboratories  
3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 243-01 Japan  
kawamori@atom.ntt.jp

## Abstract

We report an empirical study on interjectory responses, utterances such as *hai*, *ee*, and *un*, corresponding roughly to ‘yes’ or ‘uh-huh’ in English, with special emphasis on the way they help coordinate, and enhance mutual understanding in, dialogue. A corpus of data has been collected from telephone conversations, as well as from radio and television programs. Our analysis demonstrates that these expressions should be interpreted in a highly context dependent manner and that their use involves a complex coordination mechanism that encompasses different hierarchical levels of coordination. We also discuss the way in which appropriate analysis of these interjectory responses is relevant for natural language understanding and speech act recognition.

## 1 Introduction

Naturally occurring dialogues have features that are distinctively their own [5]. The realization of a spontaneous dialogue understanding system presupposes a correct characterization and a way to treat successfully these features of natural dialogues. This paper describes an analysis of Japanese interjectory responses in everyday dialogue.

Interjectory responses (IRs) in Japanese are utterances such as *hai*, *ee*, and *un*, corresponding roughly to ‘yes’ or ‘uh-huh’ in English. They are typically used in response to yes-no questions and requests or are interjected, as feed-backs, between the interlocutor’s utterances to keep the dialogue flowing smoothly. Although IRs are among the most frequently used expressions in spoken dialogue, a proper analysis of their semantic/pragmatic contents as well as their exact discourse functions is non-trivial and involves several different dimensions.

Thus, although *hai*, *ee*, and *un* are conventionally classified as interjections, we call them ‘interjectory responses’ because the conventional classification does not appropriately reflect their functions.

To assess their exact natures and provide empirical grounding, we have conducted a study on IRs. Our analysis is mainly of telephone dialogues about the preparation of test data with L<sup>A</sup>T<sub>E</sub>X [4], a document preparation system. As a supplement, we also analyze dialogues from a telephone hotline broadcast on radio.

In an attempt to crystallize the diversity of the contexts that induce IRs, we then analyzed expressions that elicit IRs. We found that while interjectory responses are elicited by various expressions and grammatical forms, they are most often elicited by sentence final particles following a variety of expressions.

The fact that such a wide range of expressions can elicit interjectory responses demonstrates the highly context dependent manner in which IRs are to be interpreted. Thus, despite their

innocuous appearances, IRs demand some novel rethinking about the way dialogue is treated, especially by machines.

In particular, the results of our analysis have shown that: (1) IRs do not simply function as a way of demonstrating understanding or acknowledgment, but rather as part of a complex coordination mechanism that encompasses different hierarchical levels of coordination; (2) the effective use of IRs entails that a discourse cannot be construed simply as a string of sentences and that the interpretation of speech acts should be associated with more finely-grained discourse units rather than with conventional sentences; (3) since they are closely related to the contents and functions of discourse particles, especially of sentence final particles such as *ne* and *yo*, the uses of IRs involve intricate control strategies governed by inferences about the interlocutor's beliefs and intentions.

## 2 Motives for and Perspectives of the Analysis

Although much research has been carried out on such topics as discourse structure, plan recognition, language generation, semantic interpretation, and syntactic analysis, the lack of theoretical framework makes conventional approaches to natural language understanding unfit for analyzing naturally occurring dialogues.

We need methods for analyzing natural dialogues that can explain how they are processed, and how intentions are generated and understood.

Our study focuses on IRs so that we can find clues to some of the questions about natural dialogue listed below:

1. How does dialogue advance? What are the principles and constraints governing dialogue? How do the speaker and the addressee exchange utterances? How does this relate to the generation and understanding of intention and to language processing?
2. How can we understand the intention of an expression, which is generated with various intentions and motivations.
3. Utterances are not always complete sentences; they have numerous false starts, abrupt ending, and other disfluencies. How should language processing, such as syntax processing which handles various dialogue phenomena, be done?

### 2.1 Analysis Data

The main data of analysis in this paper are taken from dialogues about the preparation of documents with the  $\text{\LaTeX}$  system. The person asked to prepare the documents, which include tables, has little experience with  $\text{\LaTeX}$ , but the person who asked to have the tasks done is fairly knowledgeable. We have also included a series of dialogues in which the person asked to prepare the documents confers with a third party. In the first series of dialogues, whenever a problem arises, the person receiving the requests is motivated to ask questions to facilitate the successful completion of the tasks. Both series of dialogues take place on the telephone so that participants are compelled to focus on language communication and cannot resort to pointing at things or relying on shared visual information. These dialogues are the main objects of analysis, but as a supplement, we also examine telephone hotline dialogues broadcast on radio.

## 3 Analysis

We analyzed 97 dialogues on the preparation of documents with  $\text{\LaTeX}$ . These dialogues contained about 6,200 turns and about 300,000 bytes. We also analyzed three radio hotline programs, which

Roles of Interjectory Utterances in Spoken Japanese Discourse

Rk	EEL	frq	%	Rk	EEL	frq	%	Rk	EEL	frq	%
1	ne (SFP)	338	14.78	24	mo (ADJL)	17	0.74		saa (IJ)	3	0.13
	AUX-F	338	14.78	25	to (CSP)	15	0.66		dake (ADVL)	3	0.09
3	Nominal	141	6.16	26	yo (SFP)	14	0.62		maa (IJ)	3	0.09
4	hai (IJs)	140	6.12	27	tara *4	13	0.57		v-part	3	0.13
5	V-F	126	5.51	28	CONJ	12	0.52		√adj-v	3	0.13
6	te (CJP)	122	5.33		Mid word	12	0.48	50	kara (CSP)	2	0.09
7	desho *1	107	4.68	30	no (SFP)	10	0.44		sono (PRN)	2	0.09
8	ka (SFP)	100	4.37		ee (IJ)	10	0.44		na (SFP)	2	0.09
9	yone (SFP) *2	91	3.98	32	v-i	9	0.39		etto (IJ)	2	0.09
10	kedo *3	88	3.85		ga (CSP)	9	0.39		shi (CJP)	2	0.09
11	moshi moshi (IJ)	58	2.53		de (CJP)	9	0.39		kurai (ADVL)	2	0.09
12	de (CSP)	52	2.32		Adj-S	9	0.39	58	kore (PRN)	1	0.04
13	ni (CSP)	50	2.19	36	soo (IJ)	8	0.35		nomi (ADVL)	1	0.04
14	kara (CJP)	42	1.84		ka (ADVL)	8	0.35		are (PRN)	1	0.04
15	adverb	39	1.70		√ sa-vb	8	0.35		iie (IJ)	1	0.04
16	no (CSP)	36	1.57	39	a (IJ)	7	0.31		sore (PRN)	1	0.04
17	ha (ADJL)	31	1.35		ano (IJ)	7	0.31		haa (IJ)	1	0.04
18	ga (CJP)	29	1.27	41	ja *5	6	0.26		kana (SFP)	1	0.04
19	to (CJP)	28	1.22	42	takke *7	5	0.22		toka (ADVL)	1	0.04
20	node (CJP)	27	1.14		Aux-part	5	0.22		made (CSP)	1	0.04
21	Greet words	22	0.96	44	ba (CJP)	4	0.17		AG	1	0.04
22	un (IJ)	21	0.92		te (CSP) *7	4	0.17		unnto (IJ)	1	0.04
23	wo (CSP)	20	0.87	46	nagara (CJP)	3	0.13		Sum	2287	

AUX-F = auxiliary verb in the end form;  
V-F = verb in the end form  
ADJ-F = adjective in the end form  
AUX-part = auxiliary verb in the participial adjective form  
V-part = verb in the participial form  
V-i = verb in the continuative form  
√ sa-vb = the root of a sa-verb  
√ adj-v = the root of an adjectival verb

CONJ = conjunction  
CSP = case particle  
CJP = conjunctive particle  
SFP = sentence final particle  
IJ = interjection  
AG = attention getter  
PRN = pronoun  
ADJL = adjectival  
ADVL = adverbial

Table 1: Expressions (IEE) that precede *hai*

contained 26 dialogues, about 632 turns, and about 160,000 bytes. We did not count inserted utterances, such as IRs, as turns. As many telephone calls were handled in one radio program, each call handled was counted as one dialogue.

We looked at the kinds of expressions used as IRs, and the kinds of expressions that elicited IRs. *Hai, un*, and *ee* are the primary interjectory responses used. In the  $\text{\LaTeX}$  dialogues, *hai* appears 2,287 times, *un* appears 365 times, and *ee* appears 292 times. In the radio hotline dialogues, *hai* appears 1,065 times, *un* appears 247 times, and *ee* appears 226 times.

We call that expression that comes immediately before an interjectory response an *interjectory eliciting expression* (abbreviated IEE) for that IR. For example in the following example:

- (1) *Ee-to desu-ne <hai> jaa*  
Let's see be-SFP <yes> then  
'Well, let's see, <yes>, in that case...'

the word *ne* is the IEE that elicits *hai*. Table 1 shows the expressions that elicit *hai* responses and the number of times these expressions appear in the dialogues.

Although we would expect that an IEE occurs immediately *before* an IR, accurate measurements reveal that an IEE and an IR can actually overlap [7].

Generally speaking, we can conclude that the sentence final particle *ne* and the end forms of auxiliary verbs are by far the most common expressions that elicit *hai*.

Sentence-ending expressions, such as verbs in their end forms, *desho(o)*, the sentence final particles *ka* and *yone* are among the top 10 expressions in the dialogues that elicit *hai*. Usually it is easy to understand the use of *hai* at the end of a sentence, but in the dialogues we analyzed, we found that in more than half the cases, *hai* is used in other places in a sentence.

Sentence final particles also often elicit IRs other than *hai*. Table 3 shows the top 10 expressions that elicit *un* responses. About 75% of all *un* elicited are in response to one of these 10 expressions.

Rnk	1	2	3	4	5
IEE	<i>ne</i>	<i>ka</i>	<i>te</i>	<i>desho</i>	<i>yo, kara</i>
%	40.66%	5.73%	4.32%	4.23%	2.53%
Rnk	7	8	9	10	
IEE	nominal	<i>ga</i>	<i>wa</i>	<i>moshimoshi</i>	
%	2.35%	2.25%	2.16%	2.07%	

Table 2: Ten most common IEE for *hai* in radio hotline dialogues

Rnk	1	2	3	4	5
IEE	nominal	<i>ka</i>	<i>te</i>	<i>kedo</i>	<i>ni</i>
%	18.63%	11.51%	10.41%	7.12%	6.58%
Rnk	6	7	8	9	10
IEE	<i>de</i>	<i>yone</i>	<i>no</i>	AUX-F	<i>ne</i>
%	5.75%	4.11%	3.84%	3.29%	3.01%

Table 3: Ten most common IEEs for *un*

Seven of these 10 expressions are among the top 10 expressions that elicit *hai*. Table 4 shows the top 10 expressions that elicit *ee* responses. About 68% of the *ee* observed are in response to one of these 10 expressions. Seven of these 10 expressions are again among the top 10 expressions that elicit *hai*.

Below, we focus on the main features of the expressions that elicit *hai* in the L<sub>A</sub>T<sub>E</sub>X dialogues.

### Sentence endings

About 49% of the expressions that elicit IRs are sentence-ending expressions. IRs are elicited in about 90% of the cases where the sentence-ending expressions such as *ne*, *desho(o)*, *ka*, and *yone* are used. The most common IR these expressions elicit is *hai*. There were few cases, however, where IRs are elicited when verbs are used in their end forms. In these cases, responses other than IRs are uttered, or there are no responses at all.

Sentence final particles such as *ne* are closely related to the sense of mutual understanding between the speaker and the addressee[3], thus *ne* usually elicits a response from the hearer. This fact implies that the uses of IRs involve intricate control strategies governed by inferences about the interlocutor's beliefs and intentions.

- (2) *Getsuyoubi dekimasu yo-ne* <Hai>  
 Monday can SFP <Yes>  
 'You can do it on Monday, right? <Yes>'

*Hai* immediately following a sentence-ending expression generally expresses a response or an answer, for a sentence-ending expression shows that some kind of speech act is completed, and *hai* uttered can be interpreted as in response to that speech act. Thus, the exact nature and interpretation of the IR depends on the type of speech act that elicits it. For example, if the utterance at whose end comes IEE expresses the speech act of *request*, *hai* is equivalent to a statement of *compliance*. If, on the other hand, the utterance expresses a *yes-no question*, then *hai* is a statement of *affirmative answer* while *hai* following a *declarative* speech act is interpreted as expressing *acknowledgment*.

Note that *hai* is not always in response to a speech act. In the following example, in which it follows a WH-question, *hai* expresses the hearer's acknowledgment that he has in fact heard the speaker's utterance and is used as a signal urging the speaker to go on.

Rnk	1	2	3	4	5
IEE	<i>ne</i>	nominal	AUX-F	<i>te</i>	<i>ka</i>
%	9.25%	8.22%	7.53%	6.85%	6.14%
Rnk	6	7	8	9	10
IEE	<i>yone</i>	<i>kedo</i>	<i>de</i>	<i>ni</i>	<i>no,wa</i>
%	5.82%	5.82%	5.48%	4.79%	4.11%

Table 4: Ten most common IEEs for *ee*

- (3) *Anpaasanto-wa dou shiteru-no? <Hai?>*  
 Ampersand -TOPIC how doing-SFP <Yes?>  
 'What about the ampersand? <Pardon me?>'

### Interjections

In the dialogues analyzed, there are 245 cases where interjections or interjection-like words elicit IRs, comprising about 11% of all expressions that elicit IRs. These interjections or interjection-like words include *hai*, *moshi moshi*, or *un*. The example below demonstrates this use of *hai*.

- (4) *Ja, sotchi-no machigai dattan -desu-ne? <Hai> hai*  
 Well, then your side-GEN mistake was be SFP? <Yes> yes  
 So, it was your mistake, wasn't it? <Yes> yes.

### Nominals

Table 1 seems to indicate that there are many cases where *hai* is uttered after a nominal. However, compared with the number of times nominals appear, the number of times they elicit a *hai* response is relatively small. In general, *hai* is uttered after a nominal when the nominal is at the head of a noun phrase and the object can be identified.

As the example below illustrates, there are many cases where a tone of inquiry is used when the sentence ends with a nominal. There are also cases where *hai* is uttered after a person's name is called.

- (5) *E-ga tsuita-mono? <Hai>*  
 picture-SUBJ attached-one? <Yes>  
 'The one with the picture? <Yes>'
- (6) *Maruchi-coramu? <Hai>*  
 Multi-column? <Yes>  
 'Multi-column? <Yes>'

### Case particles and adverbials

In the dialogues, there are 251 cases where case particles and adverbials elicit IRs, making up about 11% of all expressions that elicit IRs. Case particles and adverbials include *de*(at, by), *ni*(to, in), *no*(of), *wa*(expressing TOPIC), and *wo*(expressing OBJECT). *Hai* is a common IR at the end of a clause.

- (7) *Ano... sakki-no tokoro-de <Hai>*  
 Well, the other time-GEN place-at <Yes>  
 'Well, what we just talked about <Yes>'

### The particle *no*

When *no* generates an IR, the words immediately before *no* are formative nouns such as *ato*, *tokoro*, *hoo* and *toki*. As in the example below:

- (8) *Purinto-auto shita-no* <*hai*> *arimasu-yone*  
 Print-out done-thing <yes> there are-SFP  
 'You do have a printout, <yes> don't you?'

there are also cases where the quasi-nominal particle is regarded as the terminus of the noun clause.

- (9) *Begin-tabbyura-no ato-no* <*hai*>  
**begin-tabular**-GEN after-thing<yes>  
*agyumento -tte iu-no-wa R deshou?*  
 argument so-to-speak-TOPIC R -be ?  
 'The argument after **begin tabular** <yes> is R, right?'

Generally, when *hai* is uttered after *no* and *no* after a noun phrase, the object can be identified by the noun phrase.

For example, in (9) the addressee can recognize what "*begin tabyurra no ato*" (after **begin tabular**) indicates, and acknowledges this fact with his *hai*. In the dialogue, both dialogue participants mutually understand which *begin tabyurra* is intended.

Based on the assertion that *hai* can be inserted after a noun, one would expect it possible to insert *hai* after *begin tabyurra*, but such is not the case here. An explanation for this fact may be that *begin tabyurra* itself functions as the response to the other participant's utterance and that the adding of *hai* would be redundant.

### Conjunctive particles

In the dialogues, there are 243 cases where conjunctive particles elicit IRs, which is about 11% of all expressions that generate IRs. Conjunctive particles include *te*, *kara*, *ga*, *to*, and *de*.

- (10) *Zenkai-no tsuzuki -nan-desu-ga* <*hai*>  
 previous-time -GEN continuation being -be -but <yes>  
 'To continue from last time <yes>'

### Declinable words in the participial adjective form

Declinable words in the participial adjective form are few, but an example is included below.

- (11) ... *te-iu* <*hai*> *fairu arimashita-yone*  
 ... so-called <yes> file there was -SFP  
 'there was a file called... <yes>, right?'

### In the middle of a word

Below is an example.

- (12) *tabi=* <=*hai hai*=> *=tabi*  
 times <yes yes> times  
 'some=<yes, yes=>=times'

M	(1) <i>Eeto, konaida, ano osshatte ita fairu ni tsuite nan desu keredomo</i> (Well, about the file, that we talked about the other day...)	$G(M, MB(M, S, G(M, P(\text{Talk\_with\_on}(M, S, \text{file}))))))$
S	(2) < <i>Un</i> > ( < <i>Mm</i> >)	
M	(3) <i>Mittsu atte,</i> (There are three of them,)	$G(M, MB(M, S, (\exists x, y, z)[\text{File}(x) \wedge \text{File}(y) \wedge \text{File}(z)]))$
	(4) <i>hitotsu wa K-san to I-san no kaiwa to iu koto de,</i> (one of them is a conversation between K and I,)	$G(M, MB(M, S, \text{File\_of}(x, \text{Conv\_K\&I})))$
	(5) <i>ima yatterun desu kedomo,</i> (that I'm doing now)	$G(M, MB(M, S, \text{Do\_thing}(M, x)))$
S	(6) < <i>Un</i> > ( < <i>Mm</i> >)	
M	(7) <i>ato no futatsu wo kakunin shitai no desu ga,</i>  (but I want to check up on the other two.)	$G(M, MB(M, S, G(M, \text{Knowref}(M, y, \text{File}(y)) \wedge \text{Knowref}(M, z, \text{File}(z))))$

Figure 1: Speech Acts and Interjectory Responses

## 4 Discussion

### 4.1 Interjectory Responses

As we mentioned before, IRs such as *hai* have conventionally been classified as interjections. Interjections include certain words that express excitement, calls, responses, and greetings. Our analysis shows that this conventional classification, however, does not fully convey the functions of these words.

One important function of these IRs is their function as *back channel utterances* (or *aizuchi* in Japanese) intended to keep the conversation flowing. These short responses are utterances that are inserted at suitable places in the other dialogue participant's utterances. These words often indicate acceptance, reception, or acknowledgment of the other participant's utterances.

IRs are also used as answers directed at the other participant's requests or yes-no questions. When used in response to requests, IRs express compliance to the requests while in response to yes-no questions they express affirmation of the other participant's proposition.

One could thus say that short responses intended to keep the dialogue flowing are a type of response directed at an utterance, but it is difficult to associate the word IR simply with back channel utterances. Given the above perspective, what we call IRs in this paper include both short responses for facilitating dialogue flow and responses.

### 4.2 Dialogue Coordination

Many researchers on dialogue understanding have focused their study on *cooperative* dialogues. Although we acknowledge the importance of cooperation in dialogue, we suspect if this emphasis on cooperative dialogues had not led us to neglect the more fundamental aspect of dialogue that is naturally presupposed by cooperation, namely the *coordination* in dialogue.

Cooperation in dialogue entails helpfulness with respect to the domain and tasks involved. But

Coordination level	Characteristic of dialogue
Utterance Coordination	Following the (formal) rules minimally required to establish a conversation
Discourse Coordination	Observing linguistic codes necessary for a dialogue to be semantically coherent
Intention Coordination	Carrying out a dialogue with a view to understand, not only the literal, but also the intended interpretation
Helpful Coordination	Helpful to the interlocutor's intentions

Table 5: Levels of Coordination

we recognize as legitimately constructed dialogues many dialogues in which helpfulness of the participants is not necessarily upheld; frankly, many of the dialogues we actually encounter in everyday life are not as cooperative as we might like them to be.

As we saw before, IRs embody various semantic/pragmatic functions according to their position and the types of IEEs that elicit them. In a yes-no question, for example, *hai* expresses affirmation of the presented proposition. *Hai* may also indicate simply that the addressee has heard the speaker's utterance and is encouraging the speaker to continue. When used in response to a request, it expresses compliance to the request.

But if the addressee, failing to respond affirmatively, does not comply with the speaker's request and rejects it, the participants would still be interpreted as coordinating insofar as they are participating in the same conversation, though the addressee may not be cooperative with the speaker with respect to the latter's intentions. This can be regarded as an example of coordination without cooperation.

In general, we can posit at least four different layers or aspects in dialogue according to the level at which coordination takes place. (1) The first or the lowest layer is the level of coordination at which implicitly encoded formal rules of basic conversation turns, such as those of using IRs (as back channel) and turn-taking, that are minimally required to establish a conversation are followed. (2) The second layer is the level at which coordination is secured in such a way that applications of linguistic codes necessary for a dialogue to be semantically coherent (*e.g.* presuppositions) are ensured. (3) The third layer is the level of coordination at which dialogue is carried out with a view to understand, not only the literal, but also the intended interpretation, including such pragmatically induced information as indirect speech act and implicatures. (4) The fourth layer is the level at which the dialogue participant, understanding his interlocutor's intentions, is being truly helpful in the dialogue. Although there may be more layers than are listed here, this classification of levels seems reasonable enough to merit consideration. (Table 5)

Given such a hierarchical structure of coordination, the functions of an IR such as *hai* are interpreted as ranging over distinct levels according as the speech act type of its IEE varies. Thus, contrary to what has often been suggested about *aizuchi*, IRs do not simply function as a way of demonstrating understanding or acknowledgment, but rather constitute part of a complex coordination mechanism that encompasses different hierarchical levels of coordination.

The IRs in Figure 1 are examples of coordination. In this dialogue, when M says '*yatterun-desu kedo-mo*' (I'm doing it, but...), S gives the short response *un*. With this response, S demonstrates that he has heard M's utterance and encourages her to continue. The IR *un* is used to coordinate the dialogue, but whether S, by using *un*, is actually cooperating with respect to the domain and task cannot be determined by the IR alone.



### 4.3 Relationship with Acceptance

Clark and Schaefer [1] proposed a model of discourse based on the concept they call *contribution*. A unit of contribution is composed of 'presentation' and 'acceptance.' The mutual beliefs of the participants of a dialogue are the basis of its advancement. Although the set of mutual beliefs usually increases as a dialogue advances, uttering expressions do not automatically make the contents of a dialogue converted into mutual beliefs. Clark and Schaefer emphasize that mutual beliefs are formed by the joint activity of the speaker's presentation of utterances together with the demonstration of acceptance by his interlocutor.

Within the framework of such a theory, IRs may be considered to function as signs of acceptance. We have to take care to note, however, that a sign of acceptance (of an utterance) does not automatically guarantee the establishment of a mutual belief. With this respect, the exact nature of the relationship between a purportedly accepted utterance and a mutually believed proposition is not fully discussed by Clark and Schaefer.

Although it is evident that mutual belief is necessary for dialogue, we emphasize the fact that a mutual belief formed through acceptance is a mutual belief only in the sense that each dialogue participant assumes that it is indeed mutually believed and that there is no guarantee that the interpretations of the two participants do in fact coincide.

Although it is generally held that acceptance is considered to have occurred when evidence of misunderstanding is not openly recognized or when evidence of understanding is apparent, we often observe cases in which the supposed mutual beliefs of each of the participants in fact diverge and a certain type of revision is needed to reestablish a truer set of mutual beliefs. Thus, we can see that it would be appropriate to view the mutual belief created through acceptance as a *supposition* [8] that can be inserted in the discourse record of the speaker and the addressee.

### 4.4 Relationship with Propositional Attitude

Expressions that elicit responses are linked to propositional attitudes and speech acts. In general, it is possible to insert IRs in places where a specific propositional attitude is expressed.

To confirm this trend, we conducted an experiment. We presented analysts with dialogue data and had them divide the utterances according to their actions and label utterances with propositional attitudes. The data presented to the analysts was not divided, except for the short responses by the addressee, such as *ee*, within a speaker's utterance. The three analysts consulted are language processing researchers with some knowledge of speech act theory. We did not ask the researchers to classify specific acts within the utterances. As Figure 1 shows, all three analysts divided the utterances in the same way. They also labeled acts within the utterances in quite similar ways.

One of the labeling results is shown on the right side of Figure 1. Here,  $\mathbf{G}(S, P)$  indicates that 'speaker *S* has proposition *P* as his goal', and  $\mathbf{MB}(S, H, P)$  indicates that '*S* and *H* have the mutual belief *P*'.  $\mathbf{G}(S, \mathbf{MB}(S, H, P))$  indicates that 'the goal of *S* is to have proposition *P* develop into a mutual belief held by *S* and *H*'.

Cohen [2] points out that telephone conversations have many referring acts in which the referents are specified. Our dialogue data also has many utterances containing referring acts. However, ways of making reference vary with the domain of the dialogue. The dialogue data in Cohen's study is from conversations about the assembly of a pump, and possibly because the referent is something concrete that can be visualized, relatively simple descriptions are used. Since the dialogues in our study deal with computer files, it is more difficult for the participants to describe the referents, and the participants use many utterances to do so.

For example, in the dialogue in Figure 1 the speaker uses (1), (3), (4), (5), and (7) to determine the referents, which are the *ato no futatsu* (the other two) files. In this section of the dialogue, the addressee utters responses after each block of the speaker's utterances.

## 4.5 Processing Unit of Dialogue

In spoken dialogue, deciding on the unit of analysis is often problematic, as sentences are usually incomplete. As can be seen from Table 1, IRs can be useful in determining what constitutes a processing unit. An experiment on processing [6] in fact suggests that noun phrases and conjunctive modification phrases are considered appropriate processing units. That responses often overlap with the utterances that immediately precede them provides insights on the mechanisms of language understanding. Given that an important constraint in language processing is real-time response capability, it may be doubted if the off-line model, though theoretically lucid, sacrifices some important aspects of language processing.

## 5 Conclusion

We have discussed the role played in dialogue by IRs such as *hai*, *un*, and *ee*. Our analysis data have been telephone dialogues concerning the use of the  $\text{\LaTeX}$  system and telephone dialogues on other topics. Our analysis has shown that IRs do not simply function as a way of demonstrating understanding or acknowledgment, but rather as part of a complex coordination mechanism that encompasses different hierarchical levels of coordination.

We have claimed that the effective use of IRs in discourse entails that a discourse cannot be construed simply as a string of sentences and that the interpretation of speech acts should be associated with more finely-grained discourse units rather than with conventional sentences. Since IRs are closely related to the contents and functions of discourse particles, especially of sentence final particles such as *ne* and *yo*, the uses of IRs involve intricate control strategies governed by inferences about the interlocutor's beliefs and intentions.

## References

- [1] H. Clark and E. F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989.
- [2] P. R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics*, 10:97–146, 1989.
- [3] M. Kawamori. Japanese sentence final particles and epistemic modality. *SIG-NL-84, Information Processing Society of Japan*, pp. 41–48, 1991. (in Japanese).
- [4] L. Lamport.  $\text{\LaTeX}$ : *A Document Preparation System*. Addison-Wesley Publishing Company, 1986.
- [5] S. K. Maynard. *Kaiwa-Bunseki (Conversation Analysis)*. Kuroshio, Tokyo, 1993.
- [6] M. Nakano, A. Shimazu, and K. Kogure. A grammar and a parser for spontaneous speech. In *Proceedings of the 15th International Conference on Computational Linguistics*, 1994.
- [7] N. Osaka. Conversational turn-taking model using petri net. In *Proceedings of ICSLP 1990*, pp. 1297–1300. ICSLP, 1990.
- [8] R. H. Thomason. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pp. 325–364. MIT Press, 1990.