

# Discourse Markers in Spontaneous Dialogue: A Corpus based study of Japanese and English

Masahito KAWAMORI<sup>†</sup>

Nippon Telegraph and Telephone,  
NTT Basic Research Laboratories<sup>†</sup>

kawamori@atom.brl.ntt.co.jp

Takeshi KAWABATA<sup>†</sup>

kaw@idea.brl.ntt.co.jp

Akira SHIMAZU<sup>\*</sup>

Japan Advanced Institute  
of Science and Technology<sup>\*</sup>

shimazu@jaist.ac.jp

## Abstract

A spontaneously spoken, natural Japanese discourse contains many instances of the so-called redundant interjections and of back-channel utterances. These expressions have not hitherto received much attention and few systematic analyses have been made, since they were regarded as useless, spurious expressions.

On the basis of the analysis of spoken dialogue corpus, we claim that these utterances have the characteristics of discourse markers, which delimit and define units of discourse. Our corpus consists of task-oriented dialogues conducted both in Japanese and in English. The analysis of the Japanese corpus shows that about half of the turns are started with these so-called redundant utterances, while the English corpus shows that about 25 % of the turns start with corresponding English expressions. This suggests that at least in the case of Japanese, these so-called redundant utterances have much to do with units of discourse, the building blocks of discourse relations, and that they do indeed function as discourse markers.

We show that these utterances comprise a well-defined category, characterizable in a regular manner by their prosodic properties. Prosodic patterns of discourse markers occurring in the recorded corpus have been analyzed. Several pitch patterns have been found that characterize the most frequently used Japanese discourse markers. Based on these characteristic pitch patterns, a system of discourse markers can be designed so that a relatively small number of basic forms are shown to give rise to most of the discourse markers, with intonation patterns corresponding to the functions of discourse markers.

## Introduction

Japanese is an agglutinative language with predominantly verb final sentence patterns. It is to be expected, therefore, that Japanese sentence units are rather easy to recognize, and textbooks and written forms of Japanese indeed vindicate such an expectation.

A spontaneously spoken Japanese discourse, however,

is full of features which are not part of its written counterpart: phrases ending with conjunctions, interjections with verbs attached, unfinished utterances and other disfluencies.

*Aizuchi*, or back-channel utterances, and other so-called redundant utterances, such as *hai*, *un*, *anoo* and *ee*, are especially abundant in Japanese discourse. These are often taken as the manifestation of the irregularity and non-systematic nature of spoken language. Especially when one wants to recognize 'sentences' in discourse, these utterances inevitably stand in one's way, destroying the sentential structures expected by grammar. Because of these utterances, from the standpoint of the conventional grammar, clarifying such matters as which utterances constitute one sentence and which utterances belong to two different sentence units becomes extremely difficult. If the task of discerning sentence units, the building blocks of discourse, is difficult, the difficulty of discerning discourse structure is enormous.

If these utterances are really just noise, then they are hindrance to the understanding of discourse, and spoken discourse would be far more difficult for people to understand than written texts. But this flies in the face of the fact that people do communicate and understand each other in spoken language, and quite easily for that.

Given that spoken language is full of such so-called redundant utterances and that people can nevertheless communicate rather easily, we have to conclude that these utterances are not just noise after all; their function in discourse is effective, and they contribute somewhat to the understanding of spoken discourse. We believe that these utterances have important functions in discourse, as what is called *discourse markers*.

As was reported in [Kawamori *et al.*, 1994], there are fair indications that these expressions play crucial roles in determining discourse structures, especially with respect to units of surface discourse as well as of speech acts and planning. Elucidating such roles can not only clarify syntactically relevant features of discourse but may shed light on intended meaning and other issues concerning pragmatics [Takubo, 1994]. Aside from the

fact that since spoken discourse is full of these expressions, one cannot avoid treating these phenomena in order to understand and explain the language in actual use, we are interested in these utterances because they have particular functions, not required in written language but specifically called for in its spoken counterpart.

In addition to these theoretical interests, clarifying these phenomena serves more practical purposes also. For example, constructing a truly friendly human-machine interface would require a systematic knowledge of these features. Moreover, since written language generally lacks many discourse markers abundant in its spoken counterpart, interfaces between these two media would invariably have to be able to handle discourse markers. For example, a machine designed to take the dictation of its human interlocutor could not do without the ability to discern these discourse markers, unless correct dictation means interspersing every phrase with *ah's* and *uh's*. The inability to handle these features would limit the capacity of an expert system [Pollack *et al.*, 1982], [Whittaker and Stenton, 1988].

### Corpus

In order to analyze spontaneously spoken discourse, we have been collecting task-oriented dialogues. Although we are mainly concerned with spoken Japanese, we are also collecting English dialogues for comparison. The corpus is still growing.

The Japanese corpus comprises 312 dialogues, whose total recording time is about 18 hours (1081 minutes). The recorded conversations are transcribed and the size of the resulting text is about 1 megabyte, which roughly corresponds to 500,000 characters in Japanese orthography and contains approximately 29,000 turns. The transcription is made using a tool that automatically segments speech data into units separated by pauses. Hence a turn is defined prosodically rather than syntactically or semantically.

The English corpus comprises 77 dialogues. The recording time is about 4 hours, and the text size is about 400 kilobytes with approximately 5000 turns. The same tool is used for transcription. The tasks of the English dialogues are the same as those of Japanese. The subjects are all native speakers of the North American variety of English, with varying degrees of exposure to Japanese culture and language.

### Japanese Discourse Markers

Schiffirin [Schiffirin, 1987] gives the operational definition of discourse markers as "sequentially dependent elements which bracket units of talk", units that include such entities as sentences, propositions, speech acts, and tone units, the exact nature of which she deliberately leaves vague. She also suggests that, conversely, discourse markers themselves may define "some yet undiscovered units of talk". [Hirschberg and Litman,

1987] state that they "may, instead of making a 'semantic' contribution to an utterance (i.e., affecting its truth conditions), be used to convey explicit information about the structure of a discourse." Other terms used to designate discourse markers are 'cue phrase', 'clue word', and 'discourse particles'. Examples of English discourse markers are: *okay, but, now, anyway, by the way, in any case, that reminds me*.

Unlike their English counterparts, many Japanese discourse markers are not lexical; they are not generally regarded as comprising a well-defined category. For example, in English, such words as "well" and "now" not only function as discourse markers, but also have their inherent senses like "in a good manner" and "the present time". Their status as 'words' are unquestionable. Their Japanese counterparts, *eeto* and *hai*, on the other hand, have no other functions than as discourse markers. They have no 'meaning' by themselves, and they are never used in written language.

This fact explains why little consensus exists among researchers as to which 'words' constitute Japanese discourse markers and why these utterances have conventionally been bundled into interjections, given up as standing "outside the domain of the well-formed sentence itself," [Martin, 1975].

Another consequence of this fact is that the disambiguation between lexical and non-lexical uses, or that between non-cue and cue usages, of words used as discourse markers does not constitute a particularly urgent problem in Japanese, because most of the time there is nothing to be disambiguated. On the other hand, the question of distinguishing among the different functions born by discourse markers becomes important, since one discourse marker can have different functions depending on the context or its prosodic features.

Although there is no received categorization, Japanese discourse markers can naturally be divided into two main groups: phrasal and non-phrasal markers.

Phrasal markers are those phrases that are routinely used for conveying information about the structure of discourse. They are lexical phrases, like *sore-kara-desune* 'and then' and *sou-shimasu-to* 'the case being so', or words, like *toiuka* 'rather' and *yappari* 'as expected', usually complex in form, with inherent meaning directly related to the discourse.

Non-phrasal (particle) discourse markers can be further grouped into four categories: fillers, responsives, (sentence) final particles, and conjunctives and other adverbial expressions. The so-called redundant expressions and interjections belong to the first two categories. As was noted above, the expressions belonging to these categories are usually not recognized as part of the language proper, and therefore not given the status of morpheme.

Sentence final particles are morphemes like *-yo* and *-ne* that are attached to the end of a sentence or a phrase

to express speaker's attitudes [Kawamori, 1991]. Their role in conversation is somewhat similar to that played by French *quoi* and German *doch* and *jo*.

Other particles, mostly conjunctive in nature and expressing various discourse relations, like *-node* 'because', *-kedomo* 'though', *-to* 'then', *-ga* 'but', also function as discourse markers. Their functions are similar to those of ordinary conjunctives, but particles are more common in spoken language. They are not only used to connect sentences but also to end sentences with, indicating that in conversation they are more like sentence final particles than pure conjunctives.

Conjunctive adverbials are expressions like *ja* (then), *de* 'then', *ato* 'in addition' - many of which are derived from conjunctions and conjunctive particles - that are used to express various relations between sentences.

### Discourse Markers and Units of Discourse

That these non-phrasal markers are indeed discourse markers as defined by Schiffrin can be shown by the analysis of our corpus. We selected 59 dialogues from our Japanese corpus, and found out that 3,210, or 49.3%, of the total turn of 6,507 start with either a filler or a responsive. On the other hand, 4,308, or 66.2%, of the turns end with a particle, filler, or responsive. This seems to suggest that these expressions are indeed used to "bracket units of talk".

These units are not necessarily sentences in the conventional grammatical sense, but are closer to what the traditional Japanese linguistics call *bunsetsu*, or sentential clause, though strictly speaking, they are not what are usually regarded as well-formed *bunsetsu*.

This also shows how precarious the status of sentence is in spoken Japanese discourse, as mentioned at the outset, since it is possible that in at least half of the time, a turn does not consist of a sentence. This suggests that the discourse structure of dialogue cannot be understood in terms of the traditional grammar based on sentence, or well-formed *bunsetsu*, [Dohsaka and Shimazu, 1996].

For comparison, we looked at 13 of the English dialogues on a similar task, and found that 930, or 25.9%, of the 3,584 turns start with *uh*, *ah*, *OK*, *um*, or *right*. We did not specifically look at what comes at the end of each turn, but it appears that an English dialogue tends to have a longer turn with a finished sentence.

### Responsives and Fillers

Since responsives and fillers are those expressions that have hitherto been categorized as *jouchou-go* (redundant words) or *fuyou-go* (unnecessary words) and hence regarded as redundant or needless [Takagi *et al.*, 1993], [Kobayashi *et al.*, 1993], little attention has been paid to give a systematic account of their forms, much less their functions. But, as we said in the introduction, the elucidation of their exact status is more urgent, in that they occur so frequently in spoken language. In

the following we look more closely at responsives and fillers and try to characterize their nature.

**Responsives** are what [Kawamori *et al.*, 1994] call *interjectory responses* and roughly correspond to what in Japanese are traditionally referred to as *aizuchi*, or back-channel utterances in English. These expressions are rather limited in their realization: there are only a few expressions belonging to this class in the Standard Japanese. Their forms seem to be restricted to expressions with two morae.

An example of a responsive discourse marker is *hai*, one of the most frequently used words in spoken Japanese as well as one of the most complex and difficult to analyze. If used in response to a question, *hai* means a simple "yes", while if it is in reply to a request, it means an accepting "OK". When used by itself, at the beginning of a sentence, it usually means, "now" or "well". In addition to all these functions, it also has its most common use as an expression of acknowledgment, as does "uh-huh" in English. Of the responsives, *hai* is by far the most commonly used, and most neutral in style.

**Fillers** are those expressions which have ordinarily been taken as rather 'meaningless' or 'unimportant'. They are usually characterizable as consisting of one or two vowels, with or without consonants. Their syllable compositions are generally very simple. The expressions of this class have functions, and forms, rather similar to the fillers in English, like *mm* and *ah*. Typical examples of fillers are *anoo* and *etto*.

Another type of fillers are shorter than the fillers discussed above, and often amount to no more than a catch of voice. Examples are *a* and *e*. The phonological characteristics of this type of fillers seem to be somewhat similar to those of responsives, but the exact clarification is rather difficult because these expressions are uttered in extremely short duration, usually less than 100 milliseconds, making it almost impossible to detect them as voiced sounds.

### Taxonomy of Responsives and Fillers

The inventories of these discourse markers in literature, such as listed by [Takubo, 1994], [Kobayashi *et al.*, 1993], and [Takagi *et al.*, 1993], together with the characterizing features discussed above, seem to suggest that these discourse markers can be classified in a simple way in terms of the three dimensions: the phonetic form of the marker, the tone of a vowel, and the length of a vowel. More concretely, the following three can be offered as important aspects in classifying filler and responsive discourse markers:

#### Sound forms

These are the basic phonetic forms that constitute the realizations of discourse markers. These forms can be further divided into two groups in the following way:

**Regular Type: (Type R)** This group is composed of those markers that are regularly constructed from two elements: the initial sound plus the main vowel. These discourse markers can be summarized in the following table.

initial sound	vowels			
	a	e	un	o
/ʔ/	a	e	un	o
/h/	ha	he	hun	ho
/m/	ma	-	-	mo

In the table, the mark ʔ stands for a glottal stop; in practice it shows that there is no noticeable consonant. The 'vowel' *un* is so-called because it is not /u/ plus /n/, but rather a nasalised *u*. Note that the basic phonetic forms of both fillers and responsives are included in this group. The table indicates that there are 10 basic phonetic forms of discourse markers, including *o*, *ha*, and *mo*. Those forms in the *o* column may not actually be fillers but simple interjections, for the usages of their actual instances seem to differ from those of the fillers. We include them in the table for the sake of completeness.

**Independent Type: (Type I)** The other group of sound forms is made up of those fillers/responsives with more than one mora that do not belong to the above regular type. These are small in number, and can be summarized as follows:

*hai, ano, kono, sono, eeto*

Note that *to* in *eeto* may be regarded as a separate particle, for *unto* and *ot-to* are also possible. In fact, in familiar occasions, people often use such expressions as *hai-tto* and *a-tto*. If this is true then *eeto* should be regarded as *ee* plus *to*, and not as a single form.

It is further to be noted that, again, responsives and fillers share a type. Consideration of the sound forms, therefore, leads us to believe that fillers and responsives somehow constitute a larger, more general category which comprises phonetically related, but prosodically and functionally distinct, elements.

### Length

In addition, the length of a vowel is also to be considered. In Japanese, the long vowel and its short counterpart are phonemically distinct so that a word with a short vowel is distinguished lexically from a corresponding word with a long vowel, as can be seen from *obasan* (aunt) and *obaasan* (grand mother). Similarly, a discourse marker with a short vowel often has different functional/semantical features than does its long vowel counterpart. The filler *anoo* with the long vowel *oo* is usually thought to be different, lexically and functionally, from *ano* with the short *o*.

Hence we add the following four features concerning the lengths of vowels in our inventory:

- H+H (for a lengthened vowel at higher pitch),
- L+L (for a lengthened vowel at lower pitch),
- H<sup>-</sup> (for a short vowel with an abrupt stop at higher pitch),
- L<sup>-</sup> (for a short vowel with an abrupt stop at lower pitch).

Notice that + and - are used in a different manner from the way they are meant in [Pierrehumbert and Hirschberg, 1990]. With such an inventory of symbols for representation, we can describe the prosodic characteristics of Japanese discourse markers [Kawamori et al., 1996].

### H+H F0 pattern (*ee*)



Figure 1:

### HL% F0 pattern (*hai*)

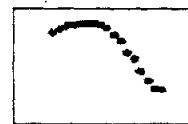


Figure 2:

Any of the sound forms can take on one of these lengths, although some of them may not actually be used in discourse. For example, the sound form *a* often takes L<sup>-</sup>; this usually shows awareness or acknowledgment. It is often found in an utterance like the following:

- (1) *a, wakari mashita*  
 L<sup>-</sup> L%  
 FILLER understand PAST  
 'uh, OK.'

*a* with the higher H<sup>-</sup> would usually imply some element of surprise.

The long H+H accompanying *a*, often written as *aa*, signals some kind of hesitation, as in:

- (2) *aa, hai*  
 H+H HL%  
 FILLER yes  
 'well, yes.'

Similarly, the other markers in Type R above may have either of the long or short tones.

Those of Type I also can be realized with either a long vowel or a short one. *ano* as a filler is usually uttered with H+H, but it may also have L<sup>-</sup>. Although *hai* is

Form	Tone				
	L <sup>-</sup>	H <sup>-</sup>	H+H	HL%	(L)H%
<i>a</i>	Recognition	Recognition	Hesitation	Response	-
<i>e</i>	Recognition	Recognition	Hesitation	Response	Question
<i>hai</i>	-	-	-	Response	Question
<i>ano</i>	-	-	Hesitation	-	-

Table 1: Function and Prosodic Patterns of Markers

usually short HL%, in a situation where one calls out to reply, H+H plus L% may be used. Table 1 summarizes these points succinctly.

### Pitch

The two pitches, H and L, can be used to characterize different markers. As noted above, *a* with H+H can signal hesitation, but the same *a* with HL%, also written *aa*, is usually taken as a responsive, albeit an impolite, or arrogant, kind. The sound form *e* with L<sup>-</sup> is usually interpreted in a similar way to what *a* with L<sup>-</sup> is interpreted. But *e* with H<sup>-</sup>% is usually used to express a question, rather like English *Pardon?* or *huh?*

We analyzed another set of recorded dialogues and observed the pitch pattern of the token utterance of a responsive discourse marker. There were 308 token utterances of the form *hai*. Of these 292 had the form HL%, as predicted. This is approximately 95 percent of the cases. There are 16 instances in which *hai* did not have HL% pattern. Of these, 8 were either immediately following or immediately followed by some other utterances, including two instances of *hai hai*.

Single utterances of *hai* that do not have HL% pattern comprise fewer than 3 percent. These had the pattern HL+L% or HL+L with the lengthening of the last vowel. There were 60 token utterances of *un*, which may be considered an informal counterpart of *hai*. Of these, 49, or approximately 82 percent, were of the pitch pattern HL%.

A possibly more interesting, and perhaps more challenging, case is that of *ee*, for *ee* is both a filler and a responsive. Our result shows that there were 96 occurrences of the token form *ee*, of which 76, or about 80 percent, were of (L)H+H pattern. The HL% pattern comprised fewer than 10 percent of the total 96, while other patterns counted 11, or slightly more than 10 percent. As it stands, this result does not refute our characterization, but it only shows that *ee* may be used more often as a filler than as a responsive. The results are succinctly summarized in Table 2.

This way of characterizing discourse markers, more precisely fillers and responsives, is interesting because it gives a good idea of the domain of discourse markers by systematically offering us a large sample of these expressions.

As was noticed earlier in this paper, [Takubo, 1994] gives a rather comprehensive, though not exhaustive,

Table 2a for *hai*

pattern	HL%	others	total
number	292	16	308
percentage	94.8	5.2	100

Table 2b for *un*

pattern	HL%	others	total
number	49	11	60
percentage	81.7	18.3	100

Table 2c for *ee*

pattern	HL%	(L)H+H	others	total
number	9	76	11	96
percentage	9.3	79.2	11.5	100

Table 2: Pitch Patterns of Fillers and Responsives

list of discourse markers, a list that arranges discourse markers into groups defined in terms of operational, or performative, concepts. [Takagi *et al.*, 1993] and [Kobayashi *et al.*, 1993] also give similar lists. It is worth noticing that all the discourse markers listed in these are accounted for in the categorization in this section.

### Phoricity

Our taxonomy of discourse markers shows that there is a dividing line between two types of fillers. One of them is the type that is composed of such fillers as  $a_{L-}$ ,  $e_{L-}$ , and  $e_{H-}$ ; the other is the long type, which includes  $ee_{(L)H+H}$  and  $ano_{(L)H+H}$ . The former may be called *anaphoric* because a member of this class is generally uttered when there is an antecedent situation, either linguistic or otherwise, that triggers its utterance: a situation that is surprising, outstanding, or simply salient for some reason. The latter class, on the other hand, may be called *cataphoric*, for its member is usually used to prepare the listener, and possibly the speaker also, for imminent continuance of speech by signaling, for example, hesitation. Responsives are by definition anaphoric, because they always presuppose something to which they are used to reply. We call the property of being anaphoric or cataphoric *phoricity* of

that expression.

This demarcation in fillers, along with the remark above about the more general, comprehensive category comprising both fillers and responsives, shows that there are in fact three distinctive categories among fillers and responsives: anaphoric fillers, cataphoric fillers, and responsives. Moreover there are three prosodic characteristics that correspond to these three categories: the typical prosody of an anaphoric filler is either  $L^-$  or  $H^-$ ; that of a cataphoric filler is  $(L)H+H$ ; and that of a responsive is  $HL\%$ . This situation is summarized in Table 3.

Whether the correspondence between the phoricity of an expression and its prosodic pattern is universal or not remains to be seen. It would be interesting to pursue this in comparison with other languages other than English and Japanese.

The correlation between the typical prosodic characteristics of fillers and responsives and the three categories discerned among them also defines the possibility of linear precedence relation among these discourse markers and discourse units.

As noted above there are cases in which the speaker refers back to his own utterance with a *hai*. There are many examples of *a* with  $L^-$  preceding *hai*. But it is rare to find *ee* with the  $H+H$  prosody followed by *hai*. The following are frequently observed sequences of discourse markers:

$\langle a, hai \rangle, \langle a, anoo \rangle, \langle souden, hai \rangle;$

while these are rarely observed:

$\langle hai, a \rangle, \langle ee, e \rangle, \langle souden, a \rangle.$

Summarizing, we get the following rough generalization:

- Fillers with  $L^-$  or  $H^-$ , the anaphoric fillers, such as  $aL^-, eL^-,$  and  $eH^-$ , come at the beginning of a 'unit';
- *hai* and other responsives with  $HL\%$  come at the end of a 'unit'. (Sometimes it may be the only unit.)
- The cataphoric fillers, such as  $anoH+H, eH+H,$  and  $eL+L$ , come in between some 'units'.

Such generalization helps us to formulate a rule that possibly characterizes the way discourse markers occur in daily discourses. One possible way of doing so would be to posit a rule like the following:

$$(\tau^-)((\{L\}\tau + \tau)^*, \text{UNIT})^+(\text{HL}\%),$$

where UNIT is some antecedently defined unit of discourse, possibly *bunsetsu*,  $\tau$  is either  $H$  or  $L$ , the raised

	Phoricity	Tone	Examples
Filler	Cataphoric	$(L)H+H$	<i>ee, anoo</i>
	Anaphoric	$L^-, H^-$	<i>a, e, ma</i>
Responsive	Anaphoric	$HL\%$	<i>hai, ee, un</i>

Table 3:

+ indicates the possibility of iterating the same element once or more times, and the raised \* indicates the possibility of iterating the same element zero or more times.

While imperfect as it stands, this rule 'defines' many surface strings as well-formed discourses. For example, the following is well-formed according to this rule:

- (3)  $a$  *anoo tokyoo-ni* *ee to detaindesuyo* *hai*  
 $L^-$   $LH+H$   $LH+H$   $HL\%$   
Tokyo-TOWARD want-to-go-out  
'I want to go to Tokyo.'

This characterization itself is rather too simple, and more work has to be done to improve upon it, a task which is beyond the scope of this paper and is to be taken up in a later work. But such a rule will be useful in devising a 'grammar' of the spontaneously spoken language taken in its totality, an attempt made by, for example, [Nakano *et al.*, 1994]. Indeed, once a truly natural discourse is taken seriously, an attempt at such a grammar and a formulation such as in this section, seems not only indispensable but also inevitable.

Another thing to be noted is that such a rule underscores the fact that these expressions are not to be regarded as redundancies, but rather as legitimate discourse markers, as has been claimed at the outset.

## Concluding Remarks

We looked at Japanese discourse markers, comparing them with their English counterparts. Although many English discourse markers are phrasal or lexical, with independent status as words, many Japanese discourse markers are non-phrasal and some of them have been denied the status of being discourse markers, dismissed as redundant or spurious.

We claimed, on the basis of the analysis of spoken dialogue corpus, that these utterances have the characteristics of discourse markers, which delimit and define units of discourse. Our corpus consists of task-oriented dialogues conducted both in Japanese and in English. The analysis of the Japanese corpus shows that about half of the turns are started with these so-called redundant utterances, while the English corpus shows that about 25 % of the turns start with corresponding English expressions. This suggests that at least in the case of Japanese, these so-called redundant utterances have much to do with units of discourse, the building blocks of discourse relations, and that they do indeed function as discourse markers.

We also examined the formal properties of these non-phrasal discourse markers, fillers and responsives, specifically looking at their prosodic features. Our analysis suggested that the intonational characteristics of these markers are category-dependent, in that markers of a category share similar intonational patterns. The existence of natural phonological demarcations among

the discourse markers suggests a systematic categorization of these expressions, a taxonomy of discourse markers that may enable us to systematize the seemingly chaotic, ad-hoc way these expressions are currently treated. We noted that the phoricity of a discourse marker corresponds to a certain prosodic pattern. Although whether such a correspondence is universal or not remains to be seen, it would be interesting to pursue this in comparison with other languages other than English and Japanese.

Our taxonomy would in turn suggest the possibility of developing a system of linear precedence relations among discourse markers and discourse units, with rules governing when and where a discourse marker may occur in an utterance. Such a 'syntax' of discourse markers naturally has an important bearing on constructing a grammar of natural, spoken discourse [Nakano *et al.*, 1994].

On the other hand, the presumed categoricity does not seem to be so fine-grained as to provide clear-cut phonological telltales distinguishing among the "functional meanings" of a member of one category: the different functional meanings of *hai*, for example, does not seem to be disambiguated solely by the differences in pitch patterns. Such finer-grained distinctions could only be made with a help of context; one has to take into account what type of expression or speech act precedes the discourse marker, and in what position of a phrase the marker occurs [Walker and Whittaker, 1990].

### Acknowledgments

The authors would like to thank Dr. Ken'ichiro Ishii, the director of Information Science Laboratory, NTT Basic Research Labs, and the other members of Dialogue Understanding Group for their support.

### References

- [Dohsaka and Shimazu, 1996] Kohji Dohsaka and Akira Shimazu. A computational model of incremental utterance production in task-oriented dialogues. In *Proceedings of the 16th International Conference on Computational Linguistics*, 1996.
- [Hirschberg and Litman, 1987] Julia Hirschberg and Diane Litman. Now let's talk about *now*: identifying cue phrases intonationally. In *the Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 163-171, 1987.
- [Kawamori *et al.*, 1994] Masahito Kawamori, Akira Shimazu, and Kiyoshi Kogure. Roles of interjectory responses in spoken discourse. In *the Proceedings of the 3rd International Conference on Spoken Language Processing*, 1994.
- [Kawamori *et al.*, 1996] Masahito Kawamori, Akira Shimazu, and Takeshi Kawabata. A phonological study on Japanese discourse markers. In *the Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation*, 1996.
- [Kawamori, 1991] Masahito Kawamori. Japanese sentence final particles and epistemic modality. *The Technical Report of the Institute of Electronics, Information and Communication Engineers NLC 91-12*, pages 41-48, 1991.
- [Kobayashi *et al.*, 1993] S. Kobayashi, M. Yamamoto, and S. Nakagawa. Acoustic characteristics concerning the occurrences of interjections, repairs etc. *The Technical Report of the Institute of Electronics, Information and Communication Engineers SLP 93-1-2*, pages 7-10, 1993.
- [Martin, 1975] Samuel Martin. *A Reference Grammar of Japanese*. Yale University Press, New Haven, 1975.
- [Nakano *et al.*, 1994] Miki Nakano, Akira Shimazu, and Kiyoshi Kogure. A grammar and a parser for spontaneous speech. In *Proceedings of the 15th International Conference on Computational Linguistics*, 1994.
- [Pierrehumbert and Hirschberg, 1990] Janet Pierrehumbert and Julia Hirschberg. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 271-311. MIT Press, 1990.
- [Pollack *et al.*, 1982] Martha Pollack, Julia Hirschberg, and Bonny Webber. User participation in the reasoning process of expert systems. In *the Proceedings of the American Association for Artificial Intelligence*, 1982.
- [Schiffirin, 1987] Deborah Schiffirin. *Discourse Markers*. Cambridge University Press, Cambridge, 1987.
- [Takagi *et al.*, 1993] K. Takagi, N. Houra, and S. Itahashi. Characteristics of utterance units in relation to the topic flow of spoken dialog. *The Technical Report of the Institute of Electronics, Information and Communication Engineers SLP 93-1-3*, 1993.
- [Takubo, 1994] Yukinori Takubo. Towards a performance model of language. *TR SLIP*, pages 15-22, 1994.
- [Walker and Whittaker, 1990] Marilyn Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 1990.
- [Whittaker and Stenton, 1988] Steve Whittaker and Phil Stenton. Cues and control in expert client dialogues. In *the Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 123-130, 1988.