Language, Information and Computation(PACLIC 11), 1996, 297-306

A Phonological Study on Japanese Discourse Markers

Masahito KAWAMORI Akira SHIMAZU Takeshi KAWABATA

{kawamori, shimazu}@atom.brl.ntt.co.jp, kaw@idea.brl.ntt.co.jp

NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 243-01 Japan

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. We show that these utterances are characterizable as discourse markers, and that they comprise a well-defined category, characterizable in a regular manner by their phonologico-prosodic properties.

Our report is based on an experiment involving spontaneously spoken conversations, recorded in a laboratory environment and analyzed using digital devices. Prosodic patterns of discourse markers occurring in the recorded conversations have been analyzed. Several pitch patterns have been found that characterize the most frequently used Japanese discourse markers.

1 Introduction

A spontaneously spoken, natural Japanese discourse contains many features which are not part of its written counterpart: disfluencies, interjections, repairs, non-sentential particles, and checked utterances.

Aizuchi, or back-channel utterances, and other so-called redundant utterances, such as *hai*, *un*, *anoo* and *ee*, which are counterparts of the English "uh-huh", "yes", or "ok", are especially abundant in Japanese discourse.

These are often taken as the manifestation of the irregularity and non-systematic nature of spoken language. They are traditionally considered to be spurious or meaningless, not especially contributing to language. From this point of view, these utterances would be nothing but 'disfluencies', some kind of 'noise'.

Yet we believe these utterances have important functions in discourse, comparable in many respects to what is called *discourse markers*. In this paper, we show that these utterances are indeed characterizable as discourse markers, and that they comprise a well-defined category, characterizable in a regular manner by their phonologico-prosodic properties.

These expressions are important because, among other things, they often provide information about the structure of discourse they occur in and about the speaker's intention or plan. Sidner 1985 states that discourse markers are necessary for recognizing the relations between the intended acts and the overall plans of the speaker. A number of researchers have noticed the relation between discourse structure and intonational, or prosodic, characteristics. Grosz and Hirschberg 1992, using an independently motivated theory of discourse model, show that there are significant associations between intonational features and discourse structure. Similarly, Nakajima and Allen 1992 discuss the correlation between prosodic information and the topic structure of discourse. One of the first works to discuss the relation between discourse markers and prosody, in the context of discourse structure, is Hirschberg and Litman 1987. Based on a study of "now" in natural recorded discourse, Hirschberg and Litman propose that, in speech, intonational characteristics play a crucial role in distinguishing between cue and non-cue uses, helping to disambiguate the structure of a discourse.

We are interested in these Japanese discourse markers for various reasons. First of all, in order to understand and explain the language in actual use, one cannot avoid treating these phenomena. Secondly, they may have particular functions, not required in written language but specifically called for in its spoken counterpart. There are also 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 (Kawamori *et al.* 1994). 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).

In addition to these theoretical interests, clarifying these phenomena may serve more practical purposes. For example, constructing a truly friendly human-machine interface would most likely require a systematic knowledge of these features. Conversely, the inability to handle these utterances would limit the capacity of an expert system (Pollack *et al.* 1982), (Whittaker and Stenton 1988). A system without such an ability may fail to allow the user to participate in the reasoning process by not letting her think while the system is giving answers or questions, or to give the exact answer the user wants by not noticing her hesitation or surprise.

Attempts at clarifying Japanese discourse markers, however, have not so far been a major enterprise. Even their status as discourse markers itself has not been widely accepted, nor is there a general agreement as to what constitute the category. Moreover, much of what little effort made has been exerted to the acoustic aspects of these expressions, (Kobayashi *et al.* 1993), and the qualitative aspects of these Japanese discourse markers have not hitherto received much attention. But phonological study is essential in clarifying the exact phonologico-prosodic nature of these expressions.

We investigate this aspect of Japanese discourse markers, looking specifically into their intonational patterns. We analyze the intonational features of discourse markers in naturally occurring utterances of Japanese. The assessment of our analysis and characterization is made using empirical data.

The paper is organized as follows. The first section introduces the Japanese tone features represented in terms of a Japanese variant of Tobi system. In the second section, we give a general introduction to Japanese discourse markers. There we discuss the phonological characteristics of responsives and fillers, our main concern in this paper. The third section describes the nature of the experiment we conducted, and the results we obtained.

2 Japanese Tone Features

We use as a framework for describing prosodic features of Japanese a system of notation essentially the same, with some modifications, as what is used by Pierrehumbert and Hirschberg 1990.

In this system, intonational contours are described as sequences of low and high tones in the fundamental frequency (f_0) contour, viewed as the physical correlate of pitch. This manner of description is basically the same as what has been widely practiced in representing lexical accent patterns of Japanese (Sugito 1994). Thus our description of the Japanese prosody is essentially based on the two tone features, H and L, which correspond to a higher tone and a lower tone, respectively.

Although different in some respects from each other, English and Japanese share certain prosodic features (Beckman and Hirschberg 1986). A prominent example is the tone that indicates a major break between phrases. Such a tone is represented in our notation as L%. Another example of a case in which Japanese prosody is like that of English is probably the tone that indicates a request for information. This pattern of intonation is represented as H%. In addition to these two boundary tone symbols, we introduce two new tone symbols: H& and L&. These symbols represent tones that are 'intermediate' in the sense that H& is not quite as high as H% and L& not as low as L%.

In addition to the above features, 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).

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),
- $\mathtt{H}^-$ (for a short vowel with an abrupt stop at higher pitch),
- $\mathtt{L}^-$ (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 by Pierrehumbert and Hirschberg 1990. With such an inventory of symbols for representation, we may describe the prosodic characteristics of Japanese discourse markers.

3 Japanese Discourse Markers

In this section, we give a general introduction to Japanese discourse markers.

As was mentioned in the introduction, what are often called redundant, interjectory utterances, such as *aizuchi*, or back-channels, and hesitations are here regarded as part of the category of expressions generally called *discourse markers*.

Schiffrin 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".

There are other terms used for the expressions denoted by Schiffrin's discourse markers. *A cue phrase* is one of the most recent, and probably the most frequently used, ones. Other terms include 'clue word' and 'discourse particles'. They seem to refer to roughly the same set of linguistic expressions (in English). According to Hirschberg and Litman 1987, "cue phrases are linguistic expressions — such as *okay, but, now, anyway, by the way, in any case, that reminds me* — which 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."

As is clear from the above definitions, discourse markers, or cue phrases, are expressions that are used to convey explicit information about the structure of a discourse. Notice that Schiffrin's idea of discourse markers is more general than Hirschberg's and Litman's of cue phrases, in that the former is not necessarily limited to phrases or words and that units of talk are not directly related to explicit information about the structure of a discourse. Because the expressions in which we are interested in this paper are not phrases, we refer to them as discourse markers.

Unlike their English counterparts, Japanese discourse markers are mostly non-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 inherent status as full-fledged lexical items in dictionary. They are easily recognized as 'words'. Their likely Japanese counterparts, *eeto* and *hai*, on the other hand, usually have no other functions than as discourse markers.

This fact may explain why there exists little consensus among researchers as to which 'words' constitute Japanese discourse markers. They are conventionally grouped into interjections, that wastebasket category. The traditional view on these expressions is succinctly summarized in the words of Martin 1975, who says that "these elements stand outside the domain of the well-formed sentence itself."

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. On the other hand, the question of distinguishing among the different functions born by discourse markers becomes important.

3.1 Types of Japanese Discourse Markers

Although there is no received categorization, Japanese discourse markers can be roughly 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, while the remaining categories comprise full-fledged lexical items. Sentence final particles are words like *yo* and *ne* that are usually attached to the end of a sentence to express speaker's attitudes (Kawamori 1991). Conjunctives are expressions like *ja* (then), *sorekara* (and then), and *toiuka* (but rather), which are mostly derived from conjunctions and conjunctive particles, that are used to express various relations between sentences.

As our main focus in the present work is on the so-called redundant expressions and interjections, analyses of conjunctive discourse markers and sentence final particles are deferred to elsewhere since their treatment calls for more

	Formal Situation	*Informal Situation*
hai	95.7%	15.0%
un	0 %	84.6%
ee	4.3 %	0.4%
total	100 %	100 %

Table 1: Distribution of Responsive Markers

thorough and elaborate research of its own. In the following we look at responsives and fillers a little more closely.

3.1.1 Responsive Discourse Markers

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.

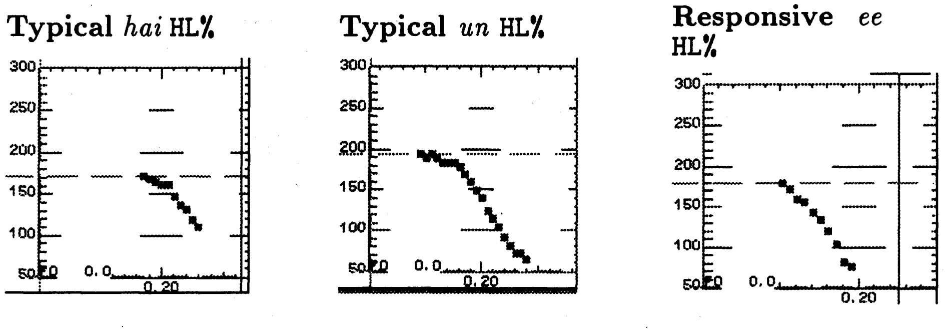
An example of a responsive discourse marker is *hai*, one of the most frequently used words in spoken Japanese as well as one of the most complex and difficult to analyze. If used in response to a question, *hai* means a simple "yes", while if it is in reply to a request, it means an accepting "OK". 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.

In our corpora, the distribution of *hai*,*un*, and *ee* is as in Table 1. Here the *formal situation* means a situation in which the participants in the conversation are not acquainted with each other, while *informal situation* means one in which the participants are acquainted with each other to some extent. The data are taken from a body of conversations we collected. The formal ones altogether comprise 1713 seconds, involving 10 people. One participant in the conversation was instructed to tell the other participant the route to a place to which the latter was to go, which the former was not told until the conversation started. The informal ones altogether comprise 788 seconds, involving again 10 people. The task was essentially the same as that of the formal ones. It should also be noted that there appeared no other responsives in the corpora. As can be seen from the table, *hai* is by far the most common in a rather formal conversation, while it can also be used in an informal situation. On the other hand, *un* is the most common in an informal situation. The fact that there are no other responsives frequently encountered in usual conversation justifies our especial emphasis on these responsive markers.

Our observation shows that a responsive discourse marker typically has the following intonational features:

- it seldom has `L` at the beginning;
- it generally ends with a short `HL%`.

These features are exemplified by the three responsives, *hai*, *un*, and *ee*, in Figures 1[1].


Typical *hai* HL%

Typical *un* HL%

Responsive *ee* HL%

Figures 1:

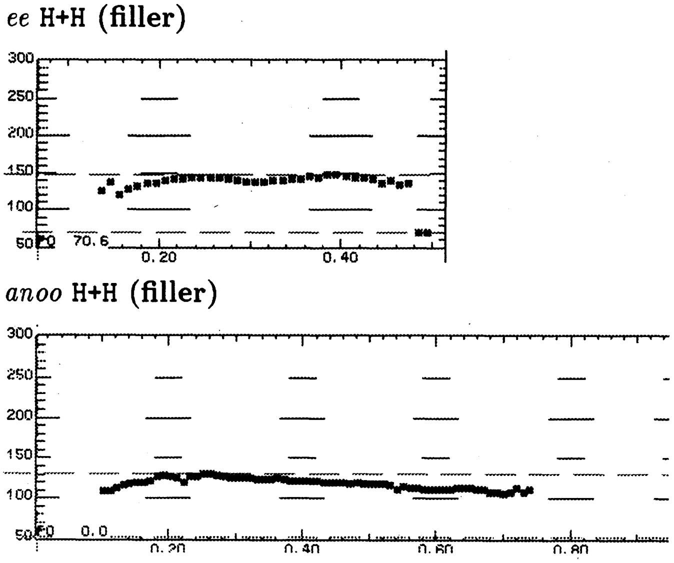
3.1.2 Fillers

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 *eeto*.

A filler typically has the following intonational features[2]

- it may sometimes have a slight L tone at the beginning;
- it is generally followed by a flat long H+H or L+L
- it usually has no sharp drop L% at the end, but ends with H or H&.

These features are exemplified by the two fillers, *ee*H+H and *anoo*, in Figures 2.


ee H+H (filler)

anoo H+H (filler)

Figures 2:

Whether these intonational features correctly characterize these types of discourse markers is not self-evident, as this is an empirical question. On the other hand, if these features do characterize the above three types of discourse markers, then it is likely that intonational patterns of discourse markers are not arbitrary or idiosyncratic, dependent on each individual marker, but rather systematically categorizable according to a general categorization of discourse markers, such as the above.

We conjecture that these features somehow demarcate the three types of Japanese discourse markers and that these discourse markers are more systematically recognizable than was previously thought possible.

4 Experiment

In order to discern the characteristic features of the discourse markers in question, we have analyzed empirical data, taken from actually recorded, spontaneous discourses.

The subjects were instructed to do certain tasks, but were not instructed as to what expressions to use or as to how they speak. We analyzed six conversations, the total recording time of which is approximately 1200 seconds[3]. The recordings were done in two-track mode so that the utterances of one speaker can be clearly distinguished from those of the other.

We used a digital signal processor to extract waveforms of the utterances of a discourse marker. We collected each token utterance, calculated its f_0 pattern (Secrest and Doddington 1983), and labeled them with the pitch pattern.

Analysis was made by observing the pitch pattern of the token utterance of a discourse marker. We paid particular attention to the pitch pattern of the token, as well as to what function each utterance of a token conveyed.

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 less 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%.

These two cases show that our characterization of these two discourse markers is probable.

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 occurences 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.

Our results suggest some interesting facts. First of all, they show that our characterization of the responsives is quite effective and grasps some, if not many, prosodic characteristics inherent in such discourse markers. As with *hai*, our characterization can be said to be correct for more than 90 percent. This shows that *hai*, as a responsive, has a rather stable character, and might suggest that such a stable character could be put to some practical purposes. This latter awaits still a future research to be made more concrete.

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: Summary of the Results

Another thing to be noted is that *ee* is more often used as a filler than as a responsive. This may have something to do with style; *ee* used as a responsive does seem to be somewhat restricted in its possible contexts of usage, and it may be taken as more 'affected' than *hai*, which is more neutral. The exact nature and origin of the relatively infrequent use of *ee* as a responsive aside, it is certainly clear that a discourse marker like *ee* poses a greater challenge to natural language understanding, with the ambiguity, or possibly even indeterminacy, of the correlation between its various forms, including prosody, and functions.

A more important thing to be noted may be that the results of our analysis suggest that the intonational characteristics of these markers are category-dependent; markers of the same category share similar intonational patterns, and, conversely, a set of specific intonational features defines, so to speak, a type of discourse markers. Within the confine of what is reported in this abstract, it is quite clear that responsive discourse markers share strikingly similar intonational features while they have distinct features from what are shared by fillers.

5 Concluding Remarks

We have discussed the intonational characteristics of some of the Japanese discourse markers.

Our analysis has 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.

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 (Kawamori *et al.* 1994).

Notes

[1] The vertical lines in those figures represent f_0 in Hz. The horizontal lines represent time in second.

[2] In fact, there is another type of fillers in Japanese. These fillers are shorter than ordinary 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.

[3] These conversations are from a different set of conversations than the ones already mentioned above.

References

BECKMAN, MARY, and JULIA HIRSCHBERG. 1986. Japanese prosodic phrasing and intonation synthesis. In *the Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 173–180.

GROSZ, BARBARA, and JULIA HIRSCHBERG. 1992. Some intonational characteristics of discourse structure. In *theProceedings of the 2nd International Conference on Spoken Language Processing*, 429–432.

HIRSCHBERG, JULIA, and DIANE LITMAN. 1987. Now let's talk about *now*: identifying cue phrases intonationally. In *the Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 163–171.

KAWAMORI, MASAHITO. 1991. Japanese sentence final particles and epistemic modality. *The Technical Report of the Institute of Electronics, Information and Communication Engineers* NLC 91-12,- 41–48.

——, AKIRA SHIMAZU, and KIYOSHI KOGURE. 1994. Roles of interjectory responses in spoken discourse. In *theProceedings of the 3rd International Conference on Spoken Language Processing.*

KOBAYASHI, S, M. YAMAMOTO, and S. NAKAGAWA. 1993. Accoustic characteristics concerning the occurrences of interjections, repairs etc. *The Technical Report of the Institute of Electronics, Information and Communication Engineers* SLP 93-1-2 7–10.

MARTIN, SAMUEL. 1975. *A Reference Grammar of Japanese.* New Haven: Yale University Press.

NAKAJIMA, SHIN'YA, and JAMES ALLEN. 1992. Prosody as a cue for discourse structure. In *theProceedings of the 2nd International Conference on Spoken Language Processing*, 425–428.

POLLACK, MARTHA, JULIA HIRSCHBERG, and BONNY WEBBER. 1982. User participation in the reasoning process of expert systems. In *the Proceedings of the American Association for Artificial Intelligence*.

SCHIFFRIN, DEBORAH. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

SECREST, BRUCE, and GEORGE DODDINGTON. 1983. An integrated pitch tracking algorithm for speech systems. In *International Conference on Speech and Signal Processing*.

SIDNER, CANDACE. 1985. Plan parsing for intended response recognition in discourse. *Computational Intelligence* 1–10.

SUGITO, MIYOKO. 1994. *Nihonjin-no Koe (The Japanese Voice)*. Tokyo: Izumi Shoin.

TAKUBO, YUKINORI. 1994. Towards a performance model of language. *TR SLIP* 15–22.

WHITTAKER, S., and P. STENTON. 1988. Cues and control in expert client dialogues. In *the Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, 123–130.