

## 談話標識語の音韻的研究 (II)

川森 雅仁 川端 豪 島津 明  
NTT 基礎研究所  
〒 243-01 神奈川県厚木市森の里若宮 3-1

### あらまし

本稿では、自然な日本語発話における談話標識語の音韻的特徴づけについて記述する。特に、「はい」「うん」「ええ」などの語の音調特徴を目標とする。音調性質の音韻的特徴のための表現法を提案し、それに基づいて談話標識語の特徴的音調性質を記述する。実際のデータの分析結果は、個々の発話トークンに対しかなり高い割合で、特徴付けが正しいことを示し、この方法が経験的に有望であることを示唆している。談話標識語の部分集合の新しい特徴付けに付いても述べる。

### キーワード

対話理解, 談話標識語, イントネーション, 音韻論

## A Phonological Study on Japanese Discourse Markers (II)

MASAHITO KAWAMORI TEKESHI KAWABATA AKIRA SHIMAZU  
NTT Basic Research Laboratories  
3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 243-01 Japan

### Abstract

In this paper, we describe some of the phonological characteristics of discourse markers in naturally occurring utterances of Japanese. In particular, our description is specifically aimed at a characterization of prosodic features of such utterances as *hai*, *un*, *ee*. For the phonological characterization of prosodic features, we introduce a representational framework, in terms of which we describe the characteristic prosody patterns of discourse markers. Our analysis of actual data shows that our characterization within our domain has promising empirical merit, giving correct prediction for high percentage of the token instances. A new way of characterizing the subset of discourse markers is also touched upon.

### Keywords

Dialogue understanding, discourse markers, intonation, phonology.

## 1 Introduction

A spontaneously spoken, natural Japanese discourse contains many instances of such features as interjections, non-sentential particles, disfluencies, checked utterances, and repairs, which are not part of its written counterpart. Many interjectory expressions, called discourse markers, are also regarded as one of the features of naturally occurring discourse. This paper investigates discourse markers from a phonological perspective, looking specifically into their intonational patterns.

We are interested in discourse markers for various reasons. There is the simple fact that a natural discourse is full of these expressions. As scientists interested in language, we naturally desire to know in what way they are so often used, and for what reason. Such pure curiosity aside, there is another theoretical reason for our interest. As was reported in [16] and [5], 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 some light on intended meaning and other issues concerning pragmatics [14].

There is also a more practical side to our interest. The inability to handle discourse markers would limit the capacity of an expert system [12][20]. It 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. 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.

In this paper, we try to describe some of the phonological characteristics of discourse markers in naturally occurring utterances of Japanese. In particular, our description is specifically aimed at a characterization of prosodic features of such utterances. In [16], and also in [5], was reported surface characteristics of some of the discourse markers: *hai*, *ee*, and *un*. In our previous [6], we extended the domain of that paper by studying the prosodic characteristics of these expressions and some others. Our purpose in this paper is two-fold: one is to augment, especially in a more quantitative way, the findings we reported in [6]; the other is to explore other dimensions of prosody involved in Japanese discourse markers.

Gram. Category	Examples
Interjection	oh, ah
Adverbial	well, now, then, in any case
Conjunction	because, and, but, so
Set phrases	you know, I mean

Table 1: Some English Discourse Markers

After clarifying the notion of discourse markers in Section 2, we introduce in Section 3 our framework for representing phonological features of prosody. In Section 4, we explicate and discuss the result of analysis on data taken from actual speech. In the penultimate section, we remark on some possible consequences of the present study and future directions thereof.

## 2 Discourse Markers

It is generally acknowledged that in English such linguistic expressions as *oh*, *well*, *okay*, *now*, *then*, and *you know* are used to convey important information about the structure of a discourse, if not about its semantics. These expressions are usually referred to as *discourse markers*, although it has become increasingly more common to hear them referred to as *cue phrases*. We here use the term *discourse markers* because we do not wish to suggest that these expressions, except for a few, are in any sense phrases. The English discourse markers are exemplified in Table 1.

Expressions that functionally correspond to these English discourse markers can also be abundantly found in Japanese discourses. In [19] is given a fairly comprehensive list of Japanese discourse markers commonly encountered in Japanese discourses. [7] is also to be noted for samples of Japanese discourse markers. In our previous work [6], we categorized, mostly for convenience and only provisionally, discourse markers into four groups: fillers, responsiveness, conjunctives, and modals. These are exemplified in Table 2.

Fillers are those expressions which have ordinarily been taken as rather 'meaningless' or 'unimportant', something akin to slips of tongue. 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*.

Responsives are what [16] call *interjectory responses* and roughly correspond to what are traditionally referred to as *aizuchi* in Japanese. Formally, 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. Functionally, they are usually used to make 'responses': a simple affirmative response to a question, an accepting response to a request, or just an acknowledging response to a previous utterance.

Conjunctives are, as the name implies, those expressions used to conjoin utterances. But this characterization might not carry much weight in the way of explicating these expressions, for most, if not all, utterances in natural discourse are 'conjoined' in some sense

Type	Examples
Filler	<i>anoo</i> , <i>etto</i> , <i>ee</i> , <i>aa'ma'</i> , <i>sono</i> , <i>n</i>
Responsive	<i>hai</i> , <i>ee</i> , <i>un</i> , <i>haa</i>
Conjunctive	<i>yappari</i> , <i>de</i> , <i>ja</i> , <i>toiruka</i>
Modal	<i>ne</i> , <i>yo</i> , <i>desho</i>

Table 2: Some Japanese Discourse Markers

or other. Such reservation notwithstanding, the expressions belonging to this group seem to share some characteristics. Many of them are apparently derived from conjunctive particles (*setsuzoku-joshi*) or conjunctive forms of verbs. Many others are simply conjunctive adverbs or idiomatic set-phrases used conjunctively. This category is somewhat eclectic as regards the forms of expressions it comprises.

The last group in the table, modals, comprises those expressions traditionally labeled as sentence final particles or end particles. Although it may be controversial whether these are truly discourse markers or not, we group them as such because these expressions play important roles in structuring discourse and dialog [4].

### 3 Framework of Representation

In order to give a qualitative characterization, it is useful to have a systematic method of representing features in terms of which to analyze the phenomena at hand.

As reported in [6], we have developed a framework for describing prosodic features of Japanese, generally in line with Pierrehumbert's theory of intonation[11] but deviating from it in some respects. In Pierrehumbert's theory 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[17][15]. 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.

Unlike Pierrehumbert et al., however, we do not assume the existence of intermediate phrases in Japanese, nor do we believe that Japanese has pitch accents as understood in the sense [2] uses this term.

Although different in such respects from each other, English and Japanese still, we believe, share some prosodic features. A prominent example is the tone that indicates a major break between phrases. This type of tone is typified by Figure 1<sup>1</sup>. Notice that at the end the pitch is lowered, followed by a pause. Such a tone is represented in our notation as L%. % here stands for a boundary tone, a signal that this tone is used at the 'boundary' between phrases. Another example of a case in which Japanese prosody is like that of English is probably the tone that indicates a request for information, as typified by Figures 2 and 3. Figure 2 shows an utterance of the phrase *arimasudeshoo* (isn't there?) which was uttered as part of a larger clause. We can notice the rising of pitch at the end of the phrase. Figure 3, on the other hand, shows a truncated phrase *de, nenrei-wa?* (and your age?), which, although not a full sentence, is perfectly comprehensible as a question. This also has a rising pitch at the end. This pattern of intonation is represented as H%. Again, the % mark indicates that this tone marks a 'boundary' between major units of discourse.

<sup>1</sup>The examples are taken from the data discussed in Section 4. The graph represents the autocorrelation coefficients of the input sound wave. The upper end corresponds to 0 msec, and the lower end corresponds to 15 msec.

Since we do not assume that Japanese has intermediate phrases, our description of the Japanese prosody system is rather simple. In order to account for more varieties that can be found at the end of a phrase, we introduce two new notations: 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 of notation, the length of a vowel may also 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. For example, *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.

We can add, provisionally at least, these 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).

H+H and L+L correspond to what in the traditional orthography are represented by a long bar (—), while H<sup>-</sup> and L<sup>-</sup> are related to the feature represented by a small *tsu* (ゝ).

With such an inventory of symbols for representation, we can describe the prosodic characteristics of many Japanese discourses in general, and discourse markers in particular.

From our pilot study on discourse markers, based on a relatively small number of sample data, we derived the notable features of three groups of discourse markers, as was reported in [6]. We represent these characteristics using the notation described above.

#### Filler

- may sometimes have a slight L tone in the beginning;
- is followed by a flat long H+H or L+L
- does not have a sharp drop L% at the end, but ends with H or H&.

#### Responsive

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

#### Conjunctive (non-adverbial)

- seldom has a low tone at the beginning;
- is followed by a flat, not necessarily long, tone;
- usually lacks a big drop L% at the end, and often ends with H or H&.

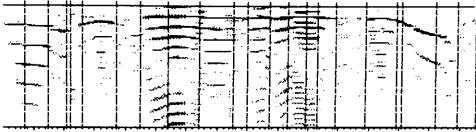


Figure 1: *jishin arimasu L%* (I am positive)

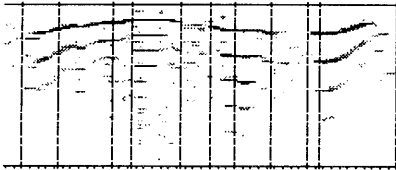


Figure 2: *arimasu-deshou H%* (isn't there?)

The characteristics of responsives can be seen in Figures 4, 5, and 6. Note that these have strikingly similar intonation patterns. This salient prosodic feature of responsives is characterized in our notation as **HL%** tone. Figure 7 exemplifies the typical intonation pattern of a filler. One can notice the flat, long stretch of the vowel in the graph. This prosodic feature is characterized as **H+H** tone. Figure 8 is but one example of a conjunctive discourse marker. Notice the long flat vowel sloping slightly down toward the end. We have yet to see if this pattern applies to all conjunctives; it is likely that many adverbial markers like *yappari* (then again) may turn out to have a very different pattern from non adverbials such as *de* (and) and *toiuka* (or rather). But it still seems to be the case that many conjunctives do carry the pattern exemplified by Figure 8. Other examples of conjunctives can be found in [6].

Whether these features can be generalized to a wider range of data is one of the questions the present paper is intended to address.

## 4 Analysis

In order to ascertain whether the above characterization in fact correlates with reality, we have analyzed some empirical data, taken from actually recorded and almost spontaneous, albeit somewhat controlled, discourses.

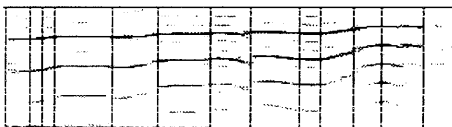


Figure 3: *de, nenrei wa H%* (and, your age?)

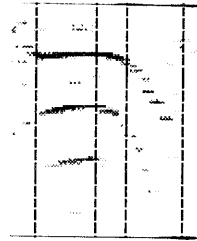


Figure 4: A typical *hai HL%*

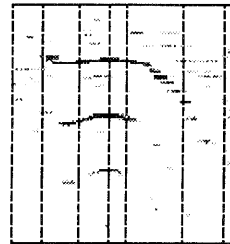


Figure 5: A typical *un HL%*

### 4.1 description of the data.

The data we used for analysis was collected by Osaka [10]. The subjects were instructed to do certain tasks, but were not instructed as to what expressions to use or as to how they should speak. We analyzed six conversations, the total recording time of which is approximately 1200 seconds.

We used a digital signal processor, developed by the second author, to extract waveforms in the vicinity of utterances of *hai* and other responsive discourse markers. We collected each token utterance, and then looked at the pitch patterns.

Analysis was made by observing whether or not the pitch pattern of the token utterance of a discourse marker conformed with the pattern predicted by the theory presented above and in [6]. We did not pay attention to what type of utterance a particular token represented,



Figure 6: A typical *ee HL%*



Figure 7: A typical *ee* H+H

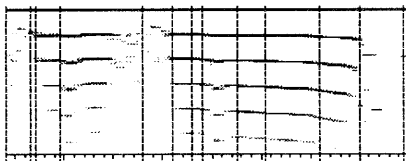


Figure 8: *sorekara* H+H (and then,)

but rather only to the pitch pattern of the token. We specifically looked at whether the typical responsive discourse markers, *hai*, *un*, *ee*, do in fact have the reported characteristics.

## 4.2 results

There were 308 token utterances of the form *hai*. Of these 292 had the form HL%, as predicted by the theory. 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*.

One typical pattern in which the predicted pattern did not occur is the following:

- (1) *hai wakari mashita*  
 HL& L%  
 yes understand PAST  
 'OK, I got it.'

Table 3a for *hai*

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

Table 3b for *un*

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

Table 3c for *ee*

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

Table 3: Summary of the Results

The lowering at the end of *hai* is not as low as the expected form, or the typical one, like that of Figure 4. This type of *hai*, that followed by another phrase, can take various prosodic patterns, and sometimes have such features as H& or even H. Although one would expect the position of *hai* to be crucial for the pattern, yet as was shown by the data, such deviations were rather small in number.

Another case of un-typical *hai* is the following:

- (2) *settei shimasunode hai*  
 L& HH&  
 arrangement do yes  
 'I will arrange it, yes'

Notice that *hai* comes right after the utterance of the same speaker; this is not used as a responsive to the interlocutor's utterance but as a sort of confirmation of one's own utterance. It is not clear whether the position of *hai* is the prominent factor in the determination of the pitch pattern. Many such utterances also had the typical HL% pattern.

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 that was reported in our previous work is not too far from truth. In fact, it is rather tempting to say that the predicted patterns are indeed predominantly common for these two discourse markers.

A possibly more interesting, and perhaps more challenging, case is that of *ee*. As can be seen from Table 2, *ee* is both a filler and a responsive. (Note Figures 6 and 7.) 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, as in Figure 7. The HL% pattern, as in Figure 6, 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 3.

## 4.3 Discussion

Our results suggest some interesting things. 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.

Another thing to be noted is that *un*, though quite stable with the 80 percent occurrence of the same pattern, has a greater variety than does *hai*. Since *un* is used in a more familiar and informal setting while *hai* is more

formal than *un*, this may suggest that the formality level of a discourse plays a role in determining the prosodic features of responsives. This in turn may be used in support of the familiar contention that prosody carries much information that has a lot to do with social and emotional aspects of discourse.

Still another point is that *ee* is more often used as a filler than as a responsive. This may also 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 for natural language understanding, with the ambiguity, or possibly even indeterminacy, of the correlation between its various forms, including prosody, and functions.

## 5 Remarks

### Classification by form, length, and pitch

Our system of notation suggests an interesting way of characterizing discourse markers, especially fillers and responsives. The inventories of discourse markers in literature, such as seen in [19], [7], [18], and Table 2, together with the characterizing features discussed in Section 3, seem to suggest that many of the 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:

**Type A (regular):** 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 ' 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*, *ho*, 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.

### Type B (independent):

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, ecto*

Note also that *to* in *ecto* 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 *ecto* 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** As we described in Section 3, there are usually four lengths: **H+H**, **L+L**, **H<sup>-</sup>**, and **L<sup>-</sup>**, though occasionally one may also need the labels like **H+L** or **L+H**. Any of the sound forms can basically 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 acknowledgement. It is often found in an utterance like the following:

(3) *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*, usually written *aa*, usually signals some kind of hesitation, as in:

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

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

Those of Type B 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 usually short **HL%**, in a situation where one calls out to reply, **H+H** plus **L%** may be used.

**Pitch** The two pitches, **H** and **L**, can also 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?*

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.

In [19] are given similar explanations of the above mentioned differences in meaning among the distinct forms of discourse markers in operational or functional terms, using 'a performance model'. There is also, as was noticed earlier in this paper, a rather comprehensive, though not exhaustive, list of discourse markers, a list that arranges discourse markers into groups defined in terms of operational, or performative, concepts.

It is worth noticing that all the discourse markers listed there are accounted for in the categorization in this section, except for those we call conjunctives, of which there are three examples in [19].

[19] suggests a possibility of such categorization as ours and further comments that each of the features, such as length and pitch, correlates to 'an independent mental function'. Although whether such correlations can be ascertained or not is clearly well beyond the scope of our present work, it seems safe to say that our characterization of fillers and responsives in terms of the three features above shows some promise.

## Two types of fillers

This characterization in turn suggests 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 that which includes  $eeto_{H+H}$  and  $anoo_{LH+H}$ . 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, so to speak, 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.

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. The interesting thing is that there are three prosodic characteristics that correspond to these three categories: the typical prosody of an anaphoric filler is either L<sup>-</sup> or H<sup>-</sup>; that of a cataphoric filler is (L)H+H; and that of a responsive is HL%.

## A 'syntax' of discourse markers

The correlation between the typical prosodic characteristics of fillers and responsives and the three categories discerned among them suggests a possibility of conjecturing a 'syntax' of discourse markers.

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<sup>-</sup> 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:

*a anoo, eeto anoo, a hai, sodesu hai;*

while these are rarely observed:

*hai a, eeto e, sodesu a.*

Summarizing, we get the following rough generalization:

- Fillers with L<sup>-</sup> or H<sup>-</sup>, the anaphoric fillers, such as  $a_{L-}$ ,  $e_{L-}$ , and  $e_{H-}$ , 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  $anoo_{H+H}$ ,  $e_{H+H}$ , and  $e_{L+L}$ , 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 + 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:

- (5) *a anoo tokyoo-ni eeto detaindesuyo hai*  
 L<sup>-</sup> LH+H LH+H HL%  
 Tokyo-TOWARD want-to-go-out  
 'I want to go to Tokyo.'

This rule itself is rather too simple-minded, 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, [9]. 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.

## 6 Conclusion

This paper has discussed the phonology of intonational features of discourse markers, especially of responsives and, to a lesser extent, of fillers. We have reintroduced the representational framework discussed in [6] with slight extension, namely the length features of vowels. Using the notation, we have described the characteristics of three of the four groups of discourse markers. Our analysis of actual data shows that our characterization, within the domain of responsives *hai*, *un*, and

*ee*, has some empirical merit, giving correct prediction for more than 90 percent of the instances in the case of *hai* though much more work remains to be done to truly ascertain the validity of the theory. The coverage of data must next be expanded to conjunctives and modals, along with the other fillers and responsives.

We have also discussed, as part of the remarks, the possibility of characterizing the subset of discourse markers, those of fillers and responsives, in terms of the phonetic form together with the prosodic features, length and pitch, of vowels. The possibility of developing a 'syntax' of discourse markers has also been touched upon.

## Acknowledgements

We would like to express our sincere thanks to our director, Dr. Ken'ichiro Ishii, for his support. We would also like to thank the members of the discourse understanding research group for their valuable comments.

## References

- [1] M. Beckman and J. Pierrehumbert. "Japanese Prosodic Phrasing and Intonation Synthesis." 24th ACL Meeting, (1986) 173-180.
- [2] D. Bolinger. "A theory of pitch accent in English", *Word*, **14**(1958), 109-149.
- [3] J. Hirschberg and D. Litman. "Now Let's Talk about *Now*: identifying cue phrases intonationally", ACL, 1987.
- [4] M. Kawamori. "Japanese Sentence Final Particles and Epistemic Modality" TR IEICE NLC91-12, 1991.
- [5] M. Kawamori, A. Shimazu, and K. Kogure. "Roles of Interjectory Responses in Spoken Discourse", 1994 International Conference on Spoken Language Processing, (1994) 955-963.
- [6] M. Kawamori, T. Kawabata, and A. Shimazu "A Phonological Study on Japanese Discourse Markers." TR IEICE, NLC95-26. (1995), 53-60.
- [7] S. Kobayashi, M. Yamamoto, S. Nakawaga. "Kantōsi, Iinaosi-tō-no Syutugenn-ni Kansuru Onkyōteki Tokuchō" (Acoustic characteristics concerning the occurrences of interjections, repairs, etc). (1993) 93-SLP-1-2.
- [8] S. Maynard. *Kaiwa Bunseki (Discourse Analysis)*. Kuroshio Shoten, 1993.
- [9] M. Nakano, A. Shimazu, and K. Kogure. "A Grammar and a Parser for Spontaneous Speech". COLING '94, 1994.
- [10] N. Osaka. "Conversational Turn-Taking Model Using Petri Net", ICSLP 90, 1990.
- [11] J. Pierrehumbert and J. Hirschberg. "The Meaning of Intonational Contours in the Interpretation of Discourse." in Cohen *et al.* (eds.), *Intentions in Communication*. MIT, 1990.
- [12] M. Pollack, J. Hirschberg, B. Webber. "User Participation in the Reasoning Process of Expert Systems", AAAI82, 1982.
- [13] D. Schiffrin. *Discourse Markers*. Cambridge, 1987.
- [14] T. Sadanobu and Y. Takubo. "The Discourse Management Function of Fillers — a case of *etto* and *ano(o)* —", International Symposium on Spoken Dialogue. (1994)271-274.
- [15] H. Sato. "Purosodii-no Seisei", (Generation of Prosody). *Gengo*, 21(1992) 58-65.
- [16] A. Shimazu, M. Kawamori, and K. Kogure. "Analysis of Interjectory Responses in Dialogue", TR IEICE, SIG-NLC-95-9, 1993.
- [17] M. Sugito. *Nihonjin-no Koe (The Japanese Voice)*. Izumi Shoin, 1994.
- [18] K. Takagi, N. Houra, S. Itahashi. "Characteristics of Utterance Units in Relation to the Topic Flow of Spoken Dialog. 93-SLP-1-3, (1993).
- [19] Y. Takubo. "Towards a Performance Model of Language". TR SLIP 1-3. (1994)15-22.
- [20] M. Walker and S. Whittaker. "Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation", ACL90, 1990.