

[公開]

TR-M-0048

Prosodic Marking of Information Status
in Tokyo Japanese

マーク シュワーツ
Marc Swerts

1999.12.10

ATR 知能映像通信研究所

PROSODIC MARKING OF INFORMATION STATUS IN TOKYO JAPANESE

Marc Swerts

Abstract

This paper reports the results of a pilot study on the prosodic marking of information status, i.e., whether information is 'new' or 'given' in the discourse, in Tokyo Japanese noun phrases (NPs). Previous investigations have brought to light that such discursual information is reflected in accent distribution in Dutch NPs, and in degree of accentuation in Italian NPs. First results on Japanese, in which accent is lexically determined, reveal that information status, both for accented and unaccented words, is mainly reflected in pitch range, with higher range used to signal new information, and lower range used to signal given information. In addition to the effect of discourse, there is 'downstep', which lowers the pitch of words following an accented word inside a prosodic phrase.

1 Introduction

This report describes research on prosodic correlates of information status in Tokyo Japanese noun phrases (NPs), which fits in a larger study on similarities and differences between languages in their use of prosody to signal contextual information. In Dutch and English, discourse context is reflected in the distribution of pitch accents in that such accents are used as pointers to information which is 'new' or 'contrastive' in the discourse, whereas 'given' information is typically deaccented. Support for such a correspondence comes from acoustic (Terken, 1984), perceptual (Nooteboom and Kruyt, 1987) and cognitive (Terken and Nooteboom, 1984) studies. These types of evidence may lead to the expectation that the relation between information status and (de-)accentuation is universal. Yet, there have been a number of studies showing that this is definitely not the case. In particular, Romance languages such as Catalan, Spanish or Italian, strongly resist deaccentuation, especially at the end of intonation groups. For instance, while Ladd (1996) acknowledges that Italian allows deaccentuation on sentence level (e.g. repeated full NPs may be deaccented), he argues that Italian strongly disfavours deaccentuation *within* noun phrases (NPs) or other syntactic constituents. That there are such syntactic restrictions appears from the following Italian example (after Ladd, 1996):

Italian Corriere è come camminare in FRETТА, soltanto si deve andare
 molto più in FRETТА
English Running is like walking in HASTE, only you have to go much
 MORE in haste (literally)

According to Ladd, a pattern where the second *fretta* would be deaccented and the accent would 'move' to *più* (like in the English transliteration) is rejected by Italian native speakers, because it only deaccents a part of the adverbial phrase *molto più in fretta*. His claims have been supported by experimental evidence from Swerts, Avesani and Krahmer (1998), who did a contrastive analysis of

accentuation patterns in Dutch and Italian. Their finding was that these two languages differ in the way they mark information status inside NPs: in Dutch, accent *distribution* is the main discriminative factor, whereas in Italian, there is a *gradient* difference between accents, with less prominent accents being reserved for given information, and more prominent ones for new and contrastive information, this difference being reflected in pitch range variation. In other words: while it is true that under certain conditions Italian resists deaccentuation, it was found that given and new information is still encoded in prosody.

Coming back to Japanese, it is of course well-known that accent is a lexical feature in this language, so that words are either accented or not irrespective of discourse context. For instance, the word “kiroi” (yellow) is always unaccented, whereas “aoi” (blue) is accented. This implies that in Japanese discursive information cannot be signalled by accent distribution. This does not entail, however, that such information cannot be cued by prosody at all. The literature (Ladd 1996; Pierrehumbert and Beckman, 1988) mentions that differences in information status in Japanese are reflected in prosody in basically two ways, i.e., in phrasing and in prominence. On the one hand, there is the finding that information put in focus is produced as a separate prosodic phrase, whereas unfocused material is “dephrased” (as opposed to deaccented), meaning it is prosodically integrated in a larger phrase. On the other hand, it is reported, in line with observations for Italian (see above), that focused material is given a higher pitch than its unfocused counterpart.

There are however a couple of interfering factors which make the picture for Japanese more complex. First, there is the complicating issue of catathesis or downstep, defined as the phenomenon that “the H tone of a phonological word, whether accented or unaccented, will be lower if it follows an accented word than if it follows an unaccented word” (Ladd 1996, p. 93; see also Pierrehumbert and Beckman, 1988). In principle, downstep is obligatorily triggered in such tonal sequences, but it is blocked by the occurrence of an intermediate phrase boundary¹. Second, there are words that are intrinsically more likely to become dephrased, and words which are more likely to become separate phrases. For instance, looking at a number of colour names, there appears to be a difference between words that represent “real” adjectives, such as “aoi” or “kiroi”, and nouns that can become adjectives when the genitive marker -no is added, like “mizuirono” (light blue) or “midorino” (green). While this morpheme is obligatory for the latter type, it is optional for the former types of adjectives, so that “kiirono” and “aono” are also possible. Intuitively, words which lack a -no marker are more likely to be dephrased, whereas words with an obligatory -no are more likely to form separate phrases.

Based on previous research, there are still a couple of unanswered questions regarding the prosodic marking of information status in Tokyo Japanese. First, given the two possible effects of discourse structure, dephrasing and pitch range, it is not entirely clear which of these is thought to be the primary correlate of information status. Second, most previous studies dealing with this issue are based on analyses of read-aloud data, the results of which do not necessarily generalize to spontaneous data. The goal of this paper is therefore to check whether some of the findings are also true for semi-spontaneous speech data.

¹One gets the impression that there is some circularity in the definition of intermediate phrase boundaries and catathesis (downstep) in that criteria are mixed: that is, if Pierrehumbert and Beckman find counterevidence against catathesis, they sometimes claim that this is due to the occurrence of an intermediate phrase boundary, without always giving evidence, on independent grounds (perception test, clear pitch or pausal cue), that there is indeed such a boundary. See for instance their statement on p. 105: “In many other tokens [...] where glottalization and pausing were not evident, Beckman still felt strongly that there was an intermediate phrase boundary”.

Sofar, the research presented below has been limited to the analysis of pitch range in order to answer the following questions: (1) Is it the case that information status is reflected in pitch range differences?; (2) If so, is there a difference between accented vs. unaccented words?; (3) If so, is there a difference between words which need a genitive marker (no) and those where this marker is optional?; (4) Is there also an effect of downstep on pitch range? Given the fact that the author is not a native speaker of Japanese, the issue of phrasing will be left for future work, in addition to other measurements.

2 Method

The experimental materials were elicited by means of a variant of a paradigm used in previous research on Italian (Swerts, Avesani and Krahmer, 1999) and Dutch (Krahmer and Swerts, 1998). Accent patterns for Tokyo Japanese² were obtained in a (semi-) spontaneous way via a simple elicitation task. Subjects, eleven speakers in total (2 female, 9 male), were asked to describe stacks of differently coloured geometrical figures in an individual task, in such a way that it would be possible for a listener to reconstruct the sequence of figures from an available set on the basis of the recorded speech. The participants were somewhat constrained in the way they had to describe the figures in that they were not allowed to use ellipsis (which may also cue information status; see e.g. Walker, Iida and Cote, 1994) and had to describe the colour and figure names without uttering words like the Japanese equivalent of “and then” or “next” and without adding extra verbs like “put” or “take”. The data thus obtained allow an unambiguous operationalization of the relevant contexts. A property is defined to be *new* (N) to the discourse if it was not mentioned in the immediately preceding description, and it is *given* (G) if it was mentioned in the previous turn. This becomes clear by looking at the following example which gives a sequence of figure descriptions:

	(figure description)	(translation)	discourse context
1	kiiroi maru	<i>yellow circle</i>	
2	kiiroi sankaku	<i>yellow square</i>	GN
3	midorino sankaku	<i>green square</i>	NG
4	aoi shikaku	<i>blue triangle</i>	NN
	...		

The geometrical figures and their respective colours were chosen such that the following conditions were met: (1) there had to be unaccented and accented versions of the figure terms and the colour descriptions and (2) the colour terms had to consist of nouns that need the genitive marker -no to become an adjective and ‘real’ adjectives where the use of -no is optional for which the accent status remains the same irrespective of the presence or absence of -no. Also, the figure and colour names had to receive unanimous identification, which excluded the use of a colour name like “orange” since this can be translated into Japanese in two ways (“daidaiiro” or “orenzi”). Based on these criteria, the colour and figure descriptions given in figure 1 were selected for the experiment.

With these materials, different combinations of accented and unaccented colour (with -no optional or obligatory) and figure names were elicited in the

²Pierrehumbert and Beckman (1988) and Haraguchi (1977) have shown that the different Japanese dialects may differ considerably in terms of the accent status of words and in terms of prosodic behaviour so that it was chosen in this study to restrict the analysis to one dialect, Tokyo Japanese, which has been studied quite extensively.

	COLOUR		FIGURE
	no optional	no obligatory	
unaccented	kiroi (kiirono) "yellow"	mizuirono "light blue"	shikaku "square"
accented	aoi (aono) "blue"	midorino "green"	sankaku "triangle"

Figure 1: Colour and figure names elicited in the experiment

description task, yielding for instance "midorino shikaku" (accented-accented) or "kiroi sankaku" (unaccented-unaccented). By varying the sequential order of the figures, all possible combinations were collected for the eleven speakers in three different discourse contexts: (1) all new (NN), (2) only colour new (NG) and (3) only figure new (GN). The fourth logical possibility (GG; both colour and figure are given) was excluded.

Speech utterances were recorded in an isolated studio, and then fed into the computer with a 16 kHz sampling frequency. The digitized audio data were analysed with the xwaves software package, including measurements of fundamental frequency (F_0). In the following section, results will be given on the difference between F_0 maxima in the "colour" and "figure" names to see whether the discourse context influenced pitch range. In addition, the average F_0 maxima in the "figure" names will be given as a function of the accent status of the preceding word to test the effect of downstep. Other measurements, the interpretation of which is not completely finished and will therefore not be discussed below, are: the duration (in milliseconds) of the colour and figure terms, the realization of colour names with or without the -no marker, the realization of the figure names in full (shikaku, sankaku) or reduced (shikak, sankak) form and the presence or absence of a clear pause between the colour and figure names.

3 Results

Figure 2 gives the difference between the F_0 maxima in the first and second word of elicited adjective-noun sequences for various discourse contexts: negative values below 0 mean that the maximum in the second word is lower than that of the preceding word, and vice versa. Not surprisingly, it appears that these F_0 values in general reflect the accent status of the word sequences, with large negative values for accented-unaccented combinations (aoi/midorino sankaku), with large positive values for unaccented-accented combinations (kiroi/mizuirono shikaku), with small negative values for accented-accented combinations (aoi/midorino shikaku), and with small positive values for unaccented-unaccented combinations (kiroi/mizuirono sankaku). Concentrating on the discourse effect on pitch range, it appears that the F_0 differences in the leftmost (NG) and rightmost (GN) bars in the different figures, are always in the expected direction: that is, the maximum is relatively high with respect to the preceding maximum in the GN condition, and relatively low in the NG condition, meaning that pitch range reflects the information status in

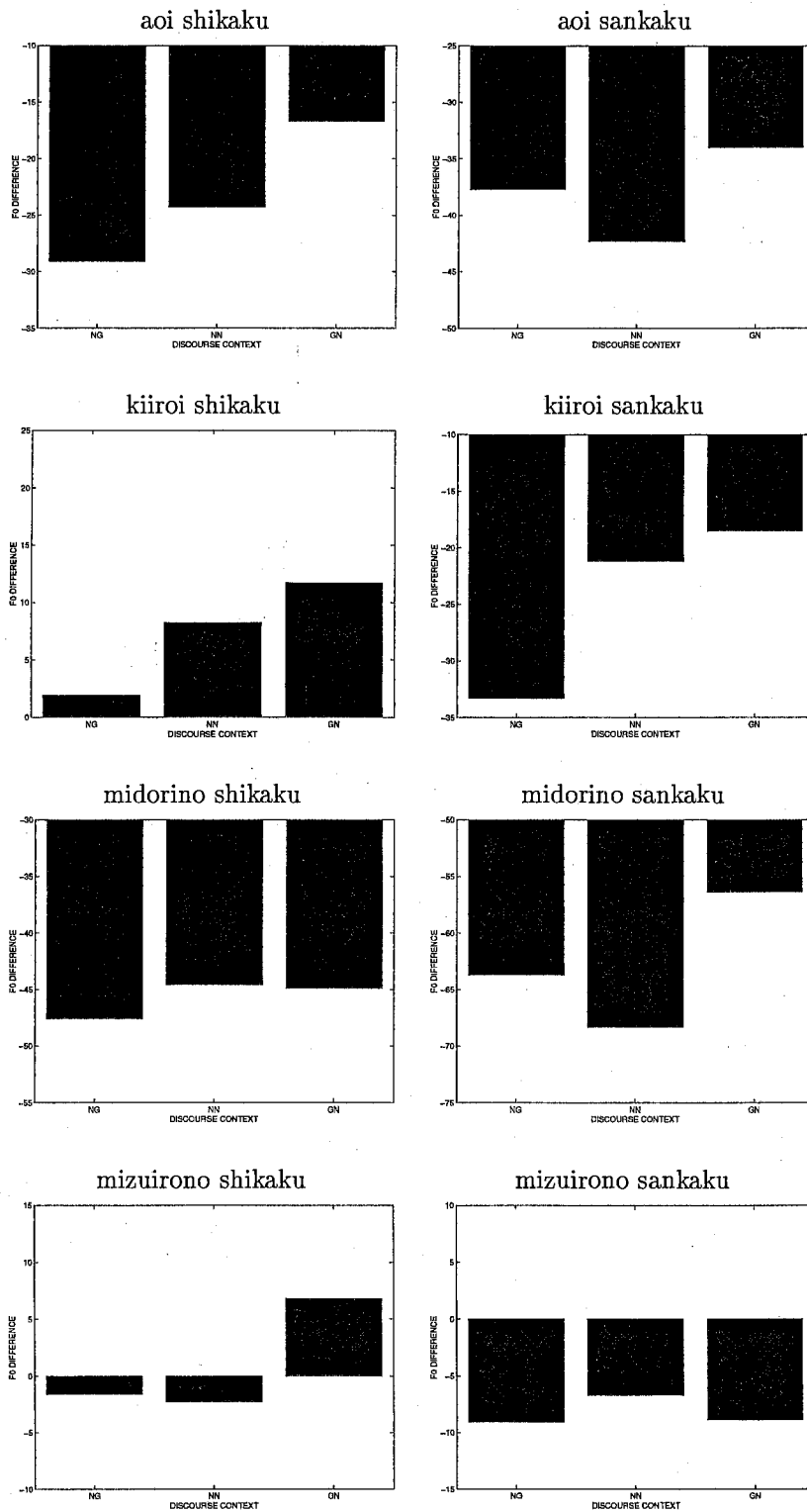


Figure 2: Difference between the F_0 maxima in the first and second word of elicited adjective-noun sequences for various discourse contexts.

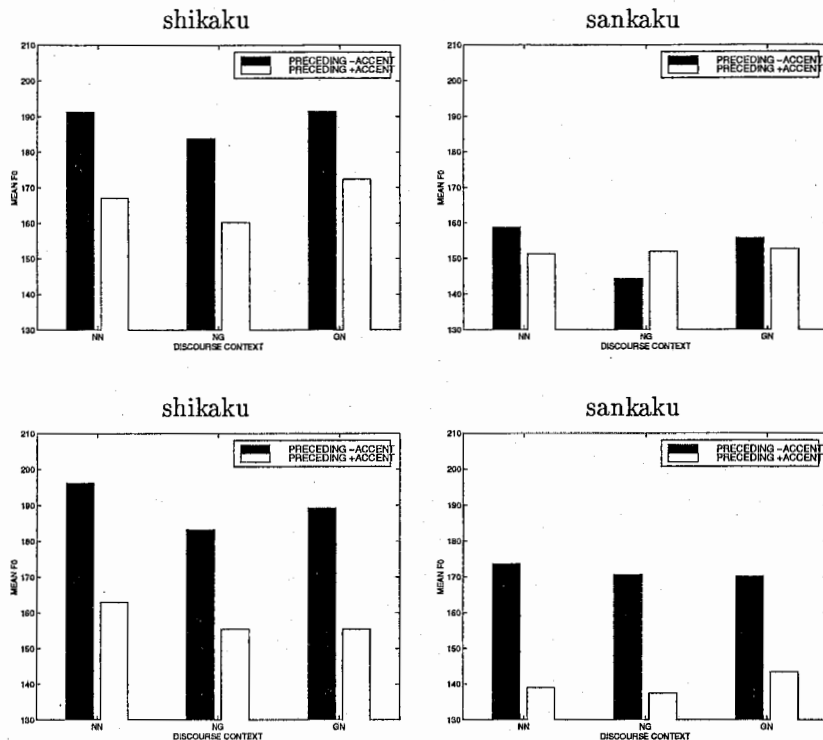


Figure 3: Mean values of the F_0 maxima in “shikaku” and “sankaku” as a function of the accent status of preceding words. In the two top figures the preceding word is “kiiroi” (black bar) or “aoi” (white bar); in the two bottom figures the preceding word is “mizuirono” (black bar) or “midorino” (white bar).

these cases. Note that this is true for all combinations of accented and unaccented words. What is puzzling, however, is that the NN cases do not behave in a uniform way: one would have expected that the values for this condition would lie in between those of the NG and the GN conditions (like is indeed the case in e.g. “aoi shikaku” or “kiiroi sankaku”), but this is not always true (e.g. “midorino sankaku” or “aoi sankaku”).

Figure 3 visualizes the effect of downstep or catathesis on the F_0 maxima in “shikaku” or “sankaku”. In line with what one would expect on the basis of previous literature, it appears that these maxima are relatively low when the preceding word is accented, and relatively high when it is unaccented. This effect is true both for accented and unaccented nouns, with one exception: the maxima for the NG condition when the figure-name is “sankaku”, preceded by “kiiroi” or “aoi”.

4 Discussion and conclusion

Summarizing the main results of the analyses that have been carried out so far, it can be seen that pitch range is influenced both by the discourse context (ignoring the somewhat puzzling outcome for the NN cases) and downstep. These two factors appear to have an orthogonal, independent effect on the F_0 maxima. In this way, the findings seem to confirm earlier observations made by e.g. Pierrehumbert and Beckman (1988) based on read-aloud data. The discourse effect

on pitch range resembles that of Italian from an earlier study (Swerts, Avesani and Krahmer, 1999), in which it was found that Italian speakers could not easily deaccent given words inside NPs: in a similar task as the one presented above, the Italian speakers always accented both the colour and figure terms irrespective of the discourse context; yet, the difference between new and given information was encoded prosodically by means of gradient differences in accents, with smaller pitch accents on given information, and larger accents on new or contrastive information. The current study finds similar degree differences reflecting information status in Tokyo Japanese, both for accented and unaccented words.

These findings remind one of the remarks of Ladd (1996) on universal tendencies in intonation and, more specifically, on the principles-and-parameters view on prosody: "The basic idea is that there are a number of 'parameters' or dimensions of variation on which languages can vary in specified ways; in many cases these parameters are said to have a 'default setting' which can be overridden in a particular language" (p. 295). In this way, one is able to distinguish surface symptoms (that may be language-specific) from deep structural effects (that are universal). Based on the observations on languages that are quite different from an intonational point of view, one could argue that they are similar in that they treat focused utterance materials different from unfocused ones, in that the former are realized more emphatically than the latter. Yet, the actual parameter settings between the languages differ considerably: Dutch distinguishes accented from unaccented words to mark the difference, Italian cues the difference by means of degree differences between accented words, and Japanese uses pitch range to distinguish both unaccented and accented words that are in focus or not. Obviously, a truly universal view on prosodic correlates of information structure should relate these findings to other linguistic phenomena that may distinguish languages, and that may be used to signal discourse context, such as word order variation, morphemic markers and ellipsis.

There are a number of possible future studies that could be done. First, a number of further acoustic measurements could be carried out to see whether information status is reflected in still other prosodic features (like duration, pausing and loudness), also to investigate the effect of "dephrasing". Second, given that the results reported above represent average speaker behaviour, one could envisage a perceptual evaluation to separate "good" from "bad" speakers to see whether the former give clearer discourse cues than the latter. Such a test could, for instance, consist of a simple task, given to an independent group of listeners who have to try and reconstruct the discourse history: e.g. if there are clear prosodic signals to information status, it should be possible for listeners, when presented with a particular figure description, to tell whether the colour term or the figure term is new respect to a prior description, or whether both are new.

Acknowledgments

The research reported upon in this paper was carried out while Marc Swerts was a visiting researcher at ATR-MIC during a three weeks stay in Fall 1999. Thanks are due to Yasuhiro Katagiri, Bob Ladd, Haruo Kubozono, Mits Ota, Emiel Krahmer, Noriko Suzuki and Miki Taniguchi for discussions and various kinds of assistance. Marc Swerts is sponsored as a postdoctoral researcher by the Flemish Fund for Scientific Research (FWO-Flanders), and is affiliated with IPO, Center for User-System Interaction (Eindhoven, The Netherlands) and with Antwerp University (Antwerp, Belgium).

Bibliography

- S. Haraguchi (1977): *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.
- E. Krahmer and M. Swerts (1998): Reconciling two competing views on contrastiveness. *Proc. ICSLP*, November-December 1998, Sydney, Australia, pp. 1299-1302.
- D.R. Ladd (1996): *Intonational Phonology*. Cambridge: Cambridge University Press.
- S.G. Nootboom and J.G. Kruyt (1987). Accent, focus distribution and the perceived distribution of Given and New information: an experiment. *Journal of the Acoustical Society of America* 82: 1512-1524.
- J. Pierrehumbert and M. Beckman (1988): *Japanese Tone Structure*. Cambridge, London: MIT Press.
- M. Swerts, C. Avesani and E. Krahmer (1999): Reaccentuation or deaccentuation: a contrastive analysis of prosodic marking of information structure in Italian and Dutch. *Proc. ICPHS*, August 1999, San Francisco (CA), USA.
- J. Terken (1984): The distribution of pitch accents as a function of discourse structure. *Language and Speech* 27: 269-289.
- J. Terken and S.G. Nootboom (1987): Opposite effects of accentuation and deaccentuation on verification latencies for 'given' and 'new' information. *Language and Cognitive Processes* 2: 145-163.
- M. Walker, M. Iida and S. Cote (1994): Japanese discourse and the processing of centering. *Computational Linguistics* 20 (2): 193-231.