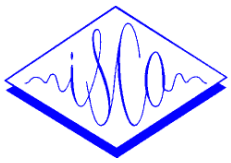

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

PROSODIC INFORMATION AND PROCESSING OF TEMPORARILY AMBIGUOUS CONSTRUCTIONS IN JAPANESE

Jennifer J. Venditti[1]
Department of Linguistics, Ohio State University
222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210 USA

Hiroko Yamashita[2]
Department of East Asian Languages & Literatures, Ohio State University
204 Cunz Hall, 1841 Millikin Rd., Columbus, OH 43210 USA

ABSTRACT

This study examines the contribution of acoustic information to distinguishing 'temporarily ambiguous' structures in Japanese. In the written language, since there is no morphological marking on relative clauses, there is a consistent ambiguity (until the head noun is presented) as to whether a clause is a simplex sentence or a subordinate clause of a complex NP. Experiment 1 shows that, in the spoken language, there are differences in fundamental frequency, duration and amplitude in native speakers' productions which distinguish these two constructions. Furthermore, experiments 2 and 3 show that this information is indeed used by listeners in identifying the structure of the utterance prior to the point of disambiguation (i.e., presentation of the head noun), and even prior to the completion of the verb in the relative clause.

1. INTRODUCTION

Processing of Japanese has received much attention recently in both experimental and theoretical psycholinguistics (e.g. [1, 2, 3]). The strictly head-final, SOV character of the language, as well as the fact that any verbal argument may be a phonologically null pronoun (*pro*), predict that there will be frequent initial misanalyses in parsing Japanese sentences. However, previous studies have focused on the processing of the written language, and little attention has been paid to the possible on-line contributions to processing from acoustic information such as fundamental frequency (F0), duration and amplitude. The present study gives evidence that prosodic information is indeed utilized by speakers and listeners in differentiating segmentally identical but structurally distinct constructions in Japanese.

The study asks the following questions: What kind of acoustic information is present in the stream of speech to cue the intended structure of an ambiguous string of segments? Is this information salient; can it be used on-line by listeners in disambiguation?

In order to address these questions, simplex sentences and complex NPs such as those shown in (1) were compared in production and perception experiments.

(1) Simplex construction:
a. Mari-ga yonda.
Mari-NOM read
'Mari read (it).'

Complex NP construction:
b. Mari-ga yonda hakusyo-wa omokatta.
Mari-NOM read report-TOP was heavy
'The report which Mari read was heavy.'

Since there is no morphological marking on relative clauses in Japanese, it is impossible to tell just from the string /mari-ga yonda/ whether it is a simplex sentence or a subordinate clause of a complex NP. That is, these constructions are 'temporarily ambiguous' (see [4]) in written Japanese: the structure of the sentence does not become clear until the information after the verb becomes available, as discussed by Inoue [2]. However, it is possible that there is acoustic information available in the spoken language prior to this point which will serve to cue the intended structure. If such information is present and used by listeners, then they should be able to distinguish between the simplex sentences and complex NP constructions with little difficulty.

2. EXPERIMENT 1 — PRODUCTION

In order to determine the prosodic organization and acoustic information which is characteristic of each construction, a production experiment was conducted.

2.1. Experimental Design

Three native speakers of Tokyo Japanese read 16 different types of simplex and complex sentences 10 times each, making a total of 160 utterances. Sentences containing a lexically accented verb /yo'nda/ 'read' as well as an unaccented verb /yonda/ 'call' were elicited. All sentences were presented in Japanese orthography. The speech data was digitized at 8KHz with 8 bit resolution and analyzed using ESPS *waves*™ speech software. The following acoustic measurements were made: 1) *Duration* of each mora in the ambiguous string /mari-ga yonda/ was measured from a wide-band spectrogram. 2) *Average amplitude* measurements were taken of each mora in the verb. Amplitude measurements were normalized by dividing them by the average amplitude of the first mora in the sentence (/ma/) in order to eliminate any variation due to differences in overall amplitude among sentences. The average amplitudes of /ma/ were not different across the two sentence types. 3) *Fundamental frequency* values of the peaks (accentual H* for accented words or phrasal H tones for unaccented words) and valleys (L% phrasal boundary tones) in the contour were recorded for the ambiguous phrase. Attention was directed to these three acoustic parameters since previous studies have suggested that they play a major role in cueing the syntactic structure of an utterance (e.g. [5, 6, 7]).

2.2. Results

Figure 1 shows the results of the duration measurements for one speaker in the accented verb condition. The fifth (/n/) and sixth (/da/) morae of the ambiguous string /mari-ga yonda/ are significantly longer

in the simplex sentence than in the complex NP construction. This result was true for the other speakers, in both verb accentuations. The lengthening of the simplex structure is most likely due to the phenomenon of final lengthening which is described in the literature for English (e.g. [5, 8, 9]) and Japanese (e.g. [10, 11]). Here, the ambiguous string in the simplex case is in itself a single *utterance* (the highest level of the prosodic hierarchy as described by Pierrehumbert & Beckman [12]), and thus the segments final to this prosodic unit are lengthened. This contrasts with the complex construction, which contains no major boundary after this string. Notice that it is not only the final mora which is lengthened, as Kaiki et al. [11] describe, but the penultimate mora is significantly longer as well. This long range effect of final lengthening for Japanese has also been reported by Campbell (see [13]).

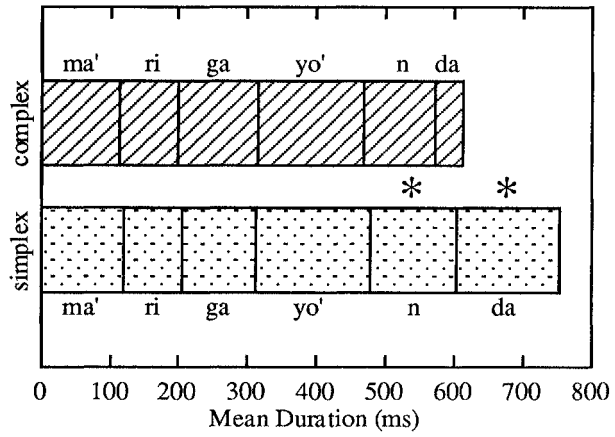

Figure 1. Mean durations of each mora in the ambiguous string in both simplex and complex constructions (speaker RS, accented verb). Asterisks mark segments which are significantly different at the 99% confidence level in a one way ANOVA.

Figure 2 shows the data of the average amplitude measurements of each mora in the verb (the same speaker, accented verb). At each mora, the amplitude of the simplex construction is significantly lower than that in the complex construction. This result was found for the other speakers as well, in both verb accentuations, though in the case of one speaker in the accented verb condition, the difference between the two structures was not significant for the first mora (/yo/) of the verb. In all cases the last two morae were significantly different between constructions. Again, since the ambiguous string in the simplex construction forms its own *utterance*, this observed lowering in amplitude in the verb is most likely a final lowering effect which operates within this prosodic domain.

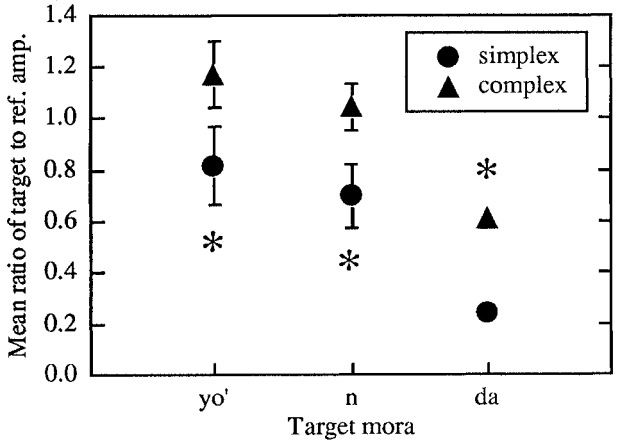

Figure 2. Mean average amplitudes of each mora of the verb in both simplex and complex constructions (speaker RS, accented verb, standard deviation error bars). Asterisks mark segments which are significantly different at the 99% confidence level in a one way ANOVA.

Schematizations of the F0 contours of each construction are shown in Figure 3 (the same speaker, accented verb). Here, the mean fundamental frequency values and standard deviation error bars are given for each peak (P) and valley (L) in the ambiguous phrase. Each rise and fall in the contours represents the tonal pattern of a single *accentual phrase*, a unit of the prosodic hierarchy of Japanese which is tonally defined by a H phrasal tone, optional H*L accent, and L% boundary tone (see [14, 12]). The prosodic phrasing at this level in both the simplex and complex contours is the same: in each case the subject and the verb of the ambiguous string form their own *accentual phrase*. The difference between these two structures is then not the accentual phrasing, but rather the overall height of the contours. The simplex contour is in general lower than the complex one, and this divergence becomes significant around the accent peak of the subject /mari/. The same results were found for the other speakers and verb accentuations, though the exact location at which the contours began to diverge varied between the first and third measurement points depending on speaker. The lower F0 values for the simplex contour can be explained by the phenomenon of final lowering (the lowering of F0 near a prosodic phrase boundary), which has been shown for English (e.g. [15]) and Japanese [12]. According to Pierrehumbert & Beckman, the domain of final lowering in Japanese is the *utterance*, and lowering effects are seen in the last half second of this prosodic unit. The results of the present study support this view, though according to the duration data the lowering seems to begin a bit earlier (600 ~ 650ms from end).

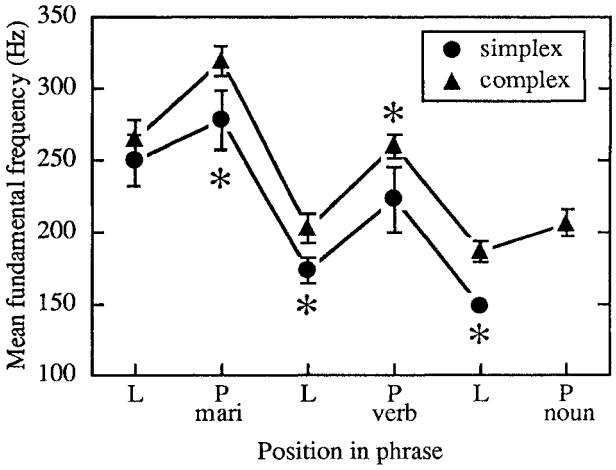

Figure 3. Schematized F0 contours of the ambiguous phrase in both simplex and complex constructions (speaker RS, accented verb, standard deviation bars). Asterisks mark segments which are significantly different at the 99% confidence level in a one way ANOVA.

While there is a 'temporary ambiguity' present in the written language between simplex sentences and subordinate clauses of complex NPs in Japanese, results

from this production experiment show that in fact in the spoken language these two structures are distinct. This is shown by the significant differences in duration, amplitude, and F0.

3. EXPERIMENT 2 — PERCEPTION

It is useful to know if these differences found in speakers' productions are perceptually salient; if listeners can use them in disambiguation. A perception test was conducted to determine the usefulness of these acoustic cues in identifying the syntactic structure.

3.1. Experimental design

The ambiguous string of each utterance from the production experiment was digitally spliced from both simplex and complex sentence types and recorded onto cassette tape in random order. Twenty milliseconds of silence was appended to each end of the extracted portion, and Gaussian noise was overlaid onto the signal to make identification more difficult. Three levels of noise were used: no noise, soft noise (speech-to-noise ratio of +12dB), and loud noise (speech-to-noise ratio of +6dB). Thirteen native listeners were asked to make a forced choice decision whether the stimulus was 'going to end' (simplex case) or 'going to continue' (complex NP case). Since all stimuli ended immediately after the ambiguous phrase, there was no resolution of the ambiguity after hearing the entire phrase. Also, listeners were not given any feedback about their response.

3.2. Results

The identification rate in this perception experiment was near perfect in all noise conditions. All of the listeners scored between 94% and 99% correct. This suggests that native listeners are able to use acoustic cues to reliably identify the syntactic structure of an utterance even though the segmental content is ambiguous.

4. EXPERIMENT 3 — GATING PERCEPTION

Results of experiment 2 demonstrated that listeners can identify the intended syntactic structure of a segmentally ambiguous string. However, since in this experiment the whole ambiguous phrase was presented to the listeners, it is not clear at which point (i.e. while hearing the phrase or after the entire phrase was revealed) listeners were able to identify the structure. Therefore, another perception experiment was conducted to investigate whether listeners use these cues on-line during the presentation of the phrase.

4.1. Experimental design

In order to determine the point within the ambiguous phrase at which listeners can identify the structure of the utterance, the gating paradigm (as described in [16]) was used. The ambiguous string was divided into its six component morae (6 gates), and this region was then presented to listeners in intervals which progressively increased in length by one mora. That is, first listeners heard just /ma/, then /mari/, then /mariga/, and so on until the whole ambiguous phrase was revealed. At each gate, listeners were again asked to judge whether the stimulus would end (after the verb) or continue. They also gave a rating between 0% and 100% of how confident they were in their answer at each gate. As with experiment 2, there was no ambiguity resolution or feedback about responses. Twenty one native listeners participated in this experiment.

4.2. Results

In this experiment, the identification rate (by at least the last gate) ranged from 59% to 100% correct, indicating that this task was more difficult than experiment 2. Figure 4 shows the confidence rating and isolation point results for the correct responses only (80% of the data). The graph indicates that, not surprisingly, as the ambiguous phrase is revealed, listeners become more confident in their responses.

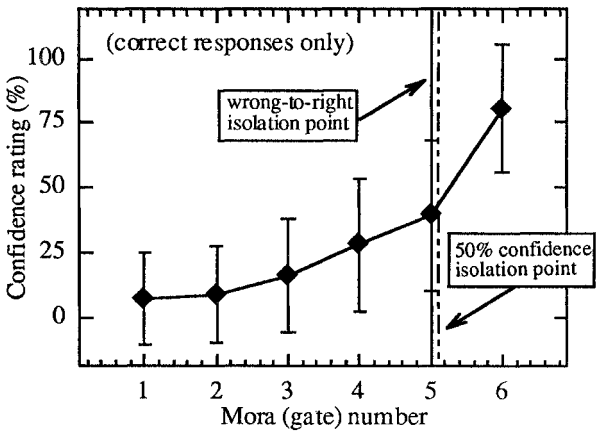

Figure 4. Mean confidence ratings and isolation points in the gating perception experiment (based on 21 listeners, standard deviation bars).

More important are the isolation point measures. The first isolation point is the *wrong-to-right isolation point*; the point at which listeners changed their answer from an incorrect to correct response. We take this measure to indicate the point in the ambiguous string at which listeners heard acoustic information which cued them to the correct structure. This, by definition, does not include cases in which listeners chose the correct response from the beginning and did not change their answer. The second isolation point measure is the *50% confidence isolation point*; the point at which listeners had the correct answer and were at least 50% confident of their response. Both of these measures fall almost at the same point: near the fifth mora of the ambiguous string (/n/ of /yonda/). This indicates that acoustic information is available to listeners to use on-line, before the entire ambiguous phrase is revealed. It suggests that in spoken Japanese listeners need not wait until the information after the verb becomes available, and not even until the completion of the verb, to know whether the clause will be a simplex sentence or a subordinate clause of a complex NP.

5. SUMMARY

The current study has shown that there are robust acoustic differences between simplex sentences and subordinate clauses of complex NPs in Japanese. These differences are a result of differing prosodic phrasing of the two constructions. Furthermore, the study has demonstrated that the differences play a large role in cueing the intended structure of ambiguous phrases. This suggests that acoustic information is used in on-line processing of Japanese, a language which syntactically has many ambiguities and opportunities for potential initial misanalyses.

ACKNOWLEDGMENTS

A full version of this paper will appear in the *Proceedings of the Conference on Formal Approaches to Japanese Linguistics*, published by MIT Working Papers in Linguistics. This paper has benefited from comments by Mary Beckman, Julie Boland, Stefanie Jannedy, Keith

Johnson, Sun-Ah Jun, and audiences of the OSU Cognitive Science Brownbag Series, 7th Annual CUNY Conference on Human Sentence Processing, and MIT Conference on Formal Approaches to Japanese Linguistics. All errors, however, are our own. This research was supported in part by the Ohio State University Department of Linguistics and Department of East Asian Languages and Literatures, and by Title VI FLAS fellowships to both authors.

REFERENCES

[1] Yamashita, H. 1994. *Processing of Japanese and Korean.* Unpublished doctoral dissertation, Ohio State University, Columbus, OH.

[2] Inoue, A. 1991. *A Comparative Study of Parsing in English and Japanese.* Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.

[3] Mazuka, R. 1991. Processing of empty categories in Japanese. *Journal of Psycholinguistic Research* 20:215-232.

[4] Beach, C. 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language* 30:644-663.

[5] Lehiste, I., J.P. Olive, & L.A. Streeter. 1976. Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America* 60:1199-1202.

[6] Streeter, L.A. 1978. Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America* 64:1582-1592.

[7] Uyeno, T., H. Hayashibe, K. Imai, H. Imagawa, & S. Kiritani. 1980. Comprehension of relative clause construction and pitch contours in Japanese. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 14:225-236.

[8] Klatt, D. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59:1208-1221.

[9] Cooper, W.E., J.M. Paccia, & S.G. Lapointe. 1978. Hierarchical coding in speech timing. *Cognitive Psychology* 10:154-177.

[10] Takeda, K., Y. Sagisaka, & H. Kuwabara. 1989. On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America* 86:2081-2087.

[11] Kaiki, N., K. Takeda, & Y. Sagisaka. 1990. Statistical analysis for segmental duration rules in Japanese speech synthesis. *Proceedings of ICSLP* 2:17-20.

[12] Pierrehumbert, J.B., & M.E. Beckman. 1988. *Japanese Tone Structure.* Cambridge, MA: MIT Press.

[13] Sagisaka, Y. 1992. On the modeling of segmental duration control. In *Speech Perception, Production, and Linguistic Structure,* Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka, eds. Tokyo: IOS Press.

[14] Poser, W.J. 1984. *The Phonetics and Phonology of Tone and Intonation in Japanese.* Unpublished doctoral dissertation, MIT, Cambridge, MA.

[15] Liberman, M.Y. & J.B. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In *Language and Sound Structure,* M. Aronoff & R.T. Oehrle, eds. Cambridge, MA: MIT Press.

[16] Grosjean, F. 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28:267-283.

[1] venditti@ling.ohio-state.edu

[2] hbutler@magnus.acs.ohio-state.edu