Phonological factors in the shaping of semantic prominence in Japanese: Evidence from gaze fixations

Yusuke Sugaya
Chukyo University

Keywords: prominence, eye-tracking, pitch, duration, intensity

This study examined how phonological factors influence semantic prominence in Japanese by performing an eye-tracking experiment. Semantic prominence may be shaped by a wide range of factors in syntax (e.g., grammatical status and word order), semantics (e.g., semantic roles), and pragmatics (e.g., information structure); however, it is also likely to be substantially influenced by phonological factors (Langacker 2008). Nonetheless, not all forms of acoustic emphasis are directly linked to semantic prominence, and the effects of certain phonological cues on semantic prominence may vary across languages. For the Japanese language, Kori's (1989a, 1989b) studies showed that duration-related factors, such as speech rate and vowel length, have limited impact on the strength of focus, whereas pitch, especially high pitch, exhibits a clear correspondence with focus. In other words, focus in Japanese appears to be marked by pitch more than by duration or intensity.

The present study extends this line of inquiry by employing a method not previously used in this line of research—the visual world paradigm (Tanenhaus et al., 1995)—which tracks participants' gazes during the presentation of visual stimuli. Based on the assumption that longer fixation durations indicate greater cognitive salience, we investigated the role of phonological factors in the formation of semantic prominence. In the experiment, speech stimuli were created to describe AI-generated images, and the noun phrases in sentences were acoustically manipulated using Praat with respect to (v) volume (intensity), (l) duration, and (p) pitch. For example, for an image of a man holding a blue box (see Figure 1), the sentence *Otoko no hito ga aoi hako o mochiageta* (The man lifted the blue box) was presented, and the noun phrase *otoko no hito ga* (the man-NOM) was manipulated in one of the following ways: (v) its volume was increased so that it was pronounced more loudly (see Figure 2-i), (l) it was lengthened so that it was pronounced more slowly and for a longer duration (see Figure 2-ii), or (p) its pitch was raised (see Figure 2-iii). Additionally, stimuli in which two or three of these manipulations were combined

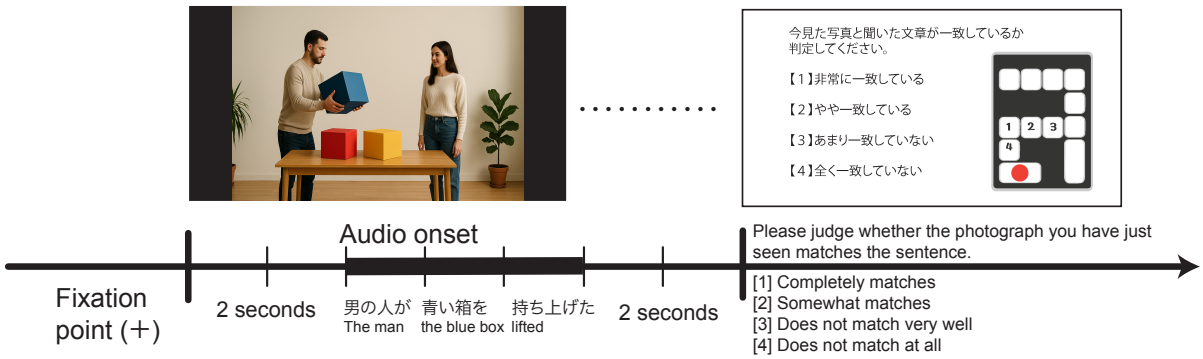
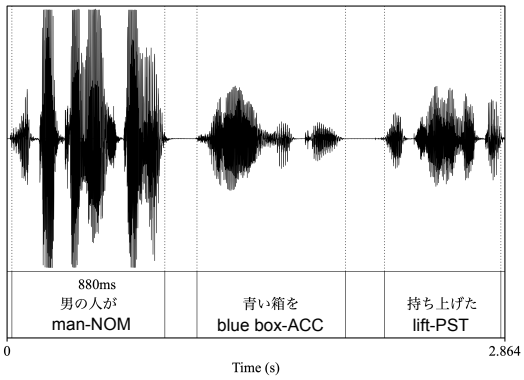


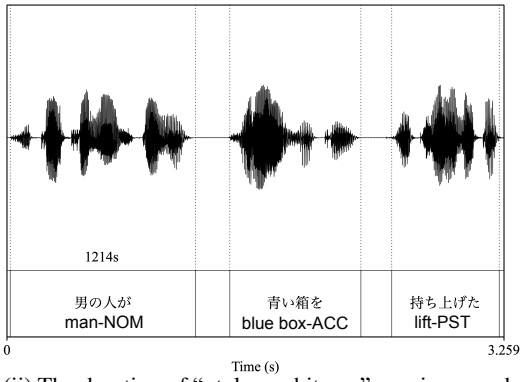
Figure 1: Procedure.

were created and then presented to the participants. As shown in Figure 1, each trial followed a repeated sequence comprising instructions, fixation point presentation, visual stimulus presentation, auditory stimulus presentation, a question (asking participants to rate the degree of correspondence between the image and sentence), and a response.

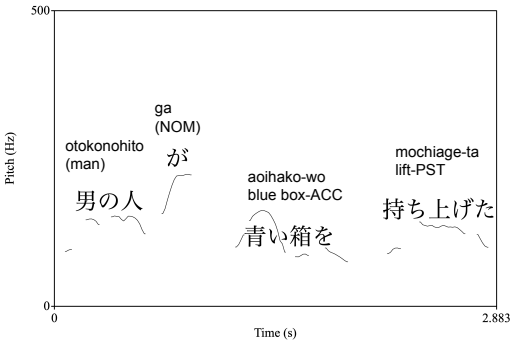
Areas of interest (AOIs) were first defined for each element of the visual stimuli to determine which elements attracted greater visual attention and were therefore more strongly focused on by participants. Overall, total fixation duration for the AOIs corresponding to the referents of noun phrases differed across the experimental conditions compared to the control condition. Under all relevant conditions, fixation duration generally increased relative to that of the control. In particular, the high-pitch condition and the combined conditions that included high pitch produced significant focusing effects. Notably, these results are broadly consistent with those of previous studies. In other words, the pitch height affects focus, and this effect is further strengthened when multiple factors are combined. The methodology adopted in the present experiment is novel in that it captures not only formal prominence at the phonological level but also its connection to semantic aspects and is therefore considered a useful approach. The consistency of the results with those of previous studies further validates the experimental methodology. Moreover, despite the findings' similarity to those of prior studies, obtaining converging evidence using a different methodological approach is in itself meaningful.



(i) The volume of "otoko no hito ga" was increased.



(ii) The duration of "otoko no hito ga" was increased.



(iii) The pitch of "otokono niho ga" was raised.

Figure 2: Auditory stimuli

References

Langacker, Ronald W. 2008. Cognitive Grammar: A Basic Introduction. Oxford: Oxford University Press.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217): 1632–1634.

Kori, Shiro. 1989a. The Roles of Intensity, Duration, and F0 in the Realization of Focus (Fokasu jitsugen ni okeru onsei no tsuyosa, jizoku jikan, F0 no yakuwari). *Speech and Language (Onsei Gengo) III*: 29–38.

Kori, Shiro. 1989b. Emphasis and Intonation (Kyocho to intoneshon). *Japanese Language and Japanese Language Education (Koza Nihongo to Nihongo Kyoiku) 2, Japanese Phonetics and Phonology (Nihongo no Onsei Onin)* 1: 316–342. Tokyo: Meiji Shoin.