LEBS
Vol. 11 No. 1 (2020) 10–14
Letters on Evolutionary Behavioral Science

Syntactic Structure Influences Speech-Gesture Synchronization

Kei Kashiwadate[1], Tetsuya Yasuda[1], Koji Fujita[2], Sotaro Kita[3], Harumi Kobayashi[1,*]

[1]Tokyo Denki University

[2]Kyoto University

[3]University of Warwick

[*]Author for correspondence (h-koba@mail.dendai.ac.jp)

It is known that a phrase may have multiple meanings. Phrases such as "green tea cup" may be interpreted with two different meanings—a "green-colored tea cup" or a "cup of green tea." Then how people know the exact meanings of apparently syntactically ambiguous linguistic expressions? We propose that gesture that accompanies speech may help disambiguate syntactically ambiguous structures. The present study investigated whether the difference in phrase structures influences the production of gestures. Participants produced gestures as they produced a Japanese four-word phrases. We examined all possible synchronization patterns of speech and gestures. We found, for the first time, gestures tended to synchronize with the chunks of words that form a constituent in syntactic structures. Our study suggests that gestures may play an important role in disambiguating syntactically ambiguous phrases. This could be a reason why humans have continuously used gestures even after they acquired a powerful tool of language and why today, they still produce language-redundant gestures.

Keywords
Gestures, Disambiguation, Branching, Iconic gesture, Syntactically ambiguous phrases

Introduction

In human face-to-face communication, people frequently use gestures as well as spoken language (Kita, 2002, 2003; Kita & Saito, 2002). People use gestures for both communication and thinking (Goldin-Meadow & Beilock, 2010). Gestures and language are not necessarily redundant, but these often express different aspects of intended meanings (McNeill, 2005). It has been suggested that such speech-gesture coordination has arisen because gesture is the origin of human language (Corballis, 2009; McNeill, 2005; Tomasello, 2008). Nonetheless, it is generally accepted that humans developed both language and gesture for communication and thinking. However, the exact role of gestures is still unclear, especially when gestures do not complement information expressed in speech, but rather are performed redundantly with speech. People often produce apparently redundant gestures; in fact, adults actually produce more redundant gestures than children (Alibali et al., 2009). The present study investigates the question: "Why do people sometimes produce such redundant gestures when language itself already expresses enough information?" We propose that we produce them because they help disambiguate syntactically ambiguous structures and are actually not redundant.

It is known that phrases can sometimes have multiple meanings. For example, a "green tea cup" can be interpreted either as a "green-colored tea cup" or as a "cup of green tea." The phrase structure with the meaning "green-colored teacup" can be classified as *Pot-Merge*. In this meaning, "tea" and "cup" are first grouped together, and the word "green" plays an adjective role in "tea-cup." In contrast, the phrase structure with the meaning "green tea cup" can be classified as Sub-Merge (Fujita, 2014). Here, "green" and "tea" are first grouped into "green-tea" and they then together play an adjective role for the word "cup" (Figure 1). In instances where the surface structure suggests multiple underlying meanings, we are left with the following question: "how could the recipient then eventually know the intended meaning?"

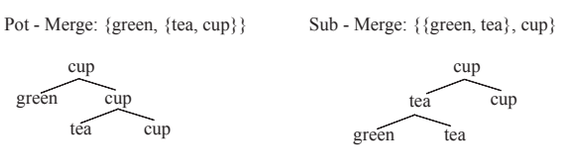

Figure 1. Two different syntactic structures, Pot-Merge and Sub-Merge, in the phrase "green tea cup."

Non-verbal cues have been studied as the means for disambiguation of apparently syntactically ambiguous expressions in addition to verbal information. For example, listeners often interpret meanings of sentences utilizing a downstepping prosodic pattern as one of sequential processing strategies (Venditti, 2006). Some researches of sentence processing have further focused on prosodic cues and pauses to disambiguate interpretation of sentences (Hirose & Mazuka, 2015; Ito et al., 2015; Misono et al., 1997; Venditti, 1994). Typically, native Japanese speakers are found to prefer the left branching interpretation over the right branching interpretation for slightly simpler Japanese compound constructions (Venditti, 1994, 2006). For example, *akai kingyo-no kasa* (red goldfish umbrella) can be interpreted either as an umbrella with a red goldfish or red umbrella with a goldfish. A clear prosodic demarcation that raises the pitch range of the second word (e.g., "*kingyo* (goldfish)" of *akai kingyo-no kasa* (red goldfish umbrella)) promotes the right branching interpretation, i.e., red um-

doi: 10.5178/lebs.2020.73
Received 14 January 2020.
Accepted 19 February 2020.
Published online 15 March 2020.

brella with a goldfish, rather than the left branching interpretation (Hirose & Mazuka, 2015; Ito et al., 2015; Venditti, 1994). However, other types of nonverbal cues have not been explored in sentence processing.

In the present study, we focused on production of nonverbal cues, in particular, gestures. Gestures have not enough yet been examined the disambiguation mechanism in syntactic structures. Humans often simultaneously use gestures to convey information to others. Gestures are commonly produced slightly earlier than its associated speech (McNeill, 1987). Moreover, this can make the listener anticipate the information from the speaker's upcoming speech (McNeill, 1987). Iconic gestures (e.g., depicting actions, motion and shape of objects) and deictic gestures (i.e., pointing gestures) may reflect aspects of the speaker's nonlinguistic spatial representations (Majid et al., 2004). Representational gestures (i.e., iconic and deictic gestures) can express spatial contents or even metaphorically express temporal concepts (Kita et al., 2001). Additionally, gestures express information even when it is difficult to express information in a spoken language (Alibali et al., 2009). Various functions are known about gestures, but the topic of whether gestures can contribute to the disambiguation mechanisms in processing syntactic structures has been largely unexplored. Okahisa and Shirose (2018) recently reported that when people expressed ambiguous phrases using gestures, their use of prosodic cues as a disambiguating cue was reduced. This study suggests the possibility that gestures may have a role in the disambiguation of ambiguous linguistic structures. Using four-word ambiguous phrases (e.g., *kuroi shippo-no ookina neko* (black tail-GEN big cat)), Kashiwadate et al. (2019) reported that the onset of gestures in relation to the target words (the second word "tail") of accompanying speech was different between different structures. Participants started a gesture earlier when they intended to convey the meaning, "a black cat with a big tail" than they intended to convey the meaning, "a big cat with a black tail." This suggested the possibility that speakers used gestures differently to indicate the underlying structure to disambiguate intended meanings. However, this study did not investigate exactly why such difference occurred because gesture and speech were not examined at the syntactic chunk level. Rather, these were examined in holistic level of whole phrases.

To investigate more precisely the difference of gesture co-occurrence timing in ambiguous phrases, we focused on the correspondence between gestures and phrase structures at the syntactic chunk level. The present study investigated whether the difference in chunking between different structures influence the production of gestures using the same production data as Kashiwadate et al (2019). Participants produced gestures as they produced a Japanese four-word phrases. The phrase consisted of Adjective1 + Noun1 + Adjective2 + Noun2 that were syntactically ambiguous. For example, the phrase *kuroi shippo-no ookina neko* in Japanese (i.e., [Black] [tail-GEN] [big] [cat]) can have three different syntactic structures. The three structures have three different meanings (Figure 2): 1) the "Big-tail structure": a black cat with a big tail, 2) the "Big-cat structure": a big cat with a black tail, and 3) the "Black-big-tail structure": a cat with a big black tail.

Which interpretation of this ambiguous phrase do listeners spontaneously choose? In a separate study (see supplementary material 1), we asked university students to interpret the meaning of ambiguous phrases of the type discussed above. The Big-cat structure was the most preferred interpretation. A few people interpreted the phrase as the Big-tail structure, but hardly anybody interpreted the phrase as the Black-big-tail structure. Therefore, we used the two relatively dominant interpretations, the Big-cat structure and the Big-tail structure, in our manipulation of syntactic structure for the production experiment.

We predicted that when speakers produce these phrases, they should produce gestures that are co-expressive with each syntactic "chunk." A chunk refers to a content word sequence that forms a constituent and modifies the head noun in Figure 2. Gestures should be co-expressive to the chunks in the sense that 1) each gesture (stroke)

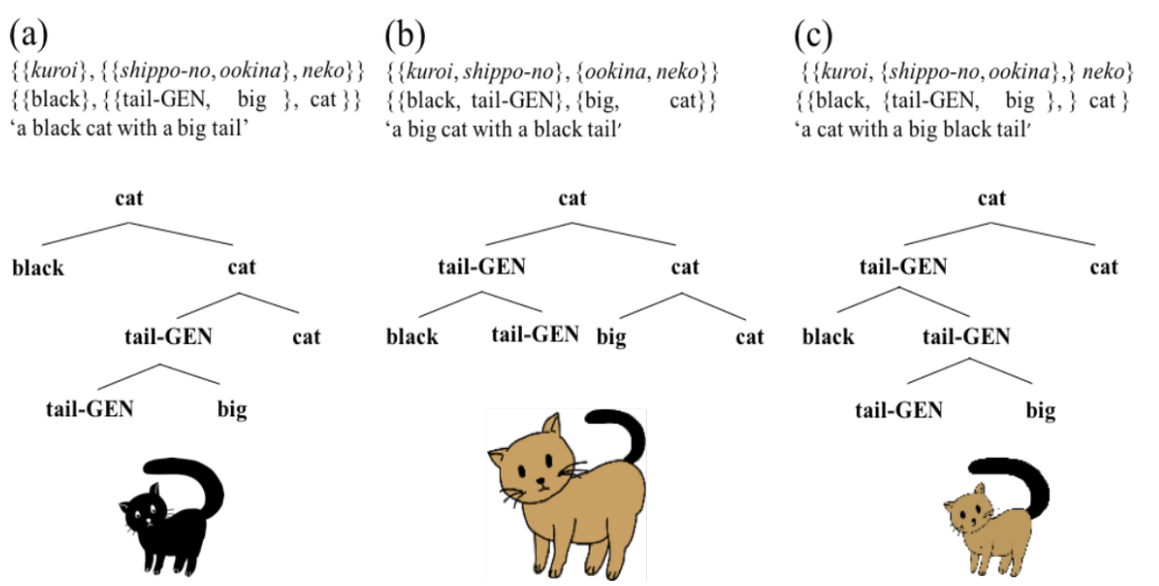

Figure 2. Three branching structures and related pictures; (a) Big-tail structure, (b) Big-cat structure, and (c) Black-tail-big structure.

should be co-temporal with a chunk in speech; that is, the onset and offset of each gesture should roughly coincide with those of a chunk, 2) each gesture should express the content equivalent to the meaning of the concurrent chunk. Thus, for example, they should produce more gestures co-expressive with "*ookina neko*" (big cat) when they express the meaning of the Big-cat structure rather than the meaning of the Big-tail structure. Likewise, they should produce more gestures co-expressive with "*shippo-no ookina*" (big tailed) when they express the meaning of the Big-tail structure rather than the meaning of the Big-cat structure.

Method

Participants

Sixteen monolingual university students who spoke Japanese as a first language participated in this study (M_{age} = 21.6, *SD* = 1.32; 1 female). This study was approved by the ethics committee of the participants' university.

Conditions

There was one independent variable – the phrase structure type (Big-tail structure, Big-cat structure). The stimulus was *Kuroi Shippo-no Ookina Neko* in Japanese (i.e., [Black] [Tail-GEN] [Big] [Cat]). The Big-tail structure was {Black, {{Tail-GEN, Big}, Cat}} with its supposed meaning as "a black cat with a big tail." The Big-cat structure was {{Black, Tail-GEN}, {Big, Cat}} and its supposed meaning was "a big cat with black tail."

Stimuli

The stimuli consisted of 4 types of illustrated animals (cat, elephant, tiger, deer) that can be interpreted either with the Big-tail structure or the Big-cat structure. Thus, there were a total of 8 pictures. For example, the elephant picture in Big-cat structure "*Shiroi Hana-no Ookina Zou* ([White] [Trunk-GEN] [Big] [Elephant])" was illustrated as a big elephant with a white trunk. In order to construct balanced stimuli with a substantial variety of words and phrases, we used various size adjectives such as "big" and "long." For example, "big" may be more readily applicable for "cat," possibly suggesting the Big cat structure ("big cat"), whereas "long" may be more readily applicable for "tail" suggesting the Big tail structure ("long tail"). In addition, we also prepared stimuli using two presentation styles (Side-by-Side, Alone). In the side-by-side presentation, participants looked at two pictures of the Big-cat structure and the Big-tail structure. On the other hand, in the alone presentation, participants look at only one of these pictures. Adjective type and presentation style factors were counterbalanced for all participants.

Procedure

The experiment consisted of two parts – a memorizing phase and a production phase. After the participant filled the consent form, the participant was seated in front of a monitor. First, the experimenter said, "*Please say the language shown using gestures so that others can understand what you mean.*" (*Teijisareta jibutsu wo ku wo iinagara jesuchaa wo tsukatte aiteni wakariyasuiyooni setsumeishitekudasai*, in Japanese). After a familiarization trial, the experiment started. Before each trial, a fixation cross appeared in the center of the monitor. After the cross was fixated for 1 second, a stimulus appeared for 10 seconds for the memorizing phase. Then, the production phase started. In the production phase, the participants made gestures to depict the picture while verbally describing it (Figure 3). The participants were instructed to produce speech and gesture when the picture goes away. There were four Big-tail structure trials and four Big-cat structure trials. Participants' gestures and utterances were recorded by a video recorder (Microsoft LifeCam). Later, we coded the participants' gestures and utterances (see supplementary material 2). Eight participants out of 16 were assigned to the alone presentation, while the other 8 participants were assigned to the side-by-side presentation. In the side-by-side presentation, one of the two objects was presented with a surrounding red frame, and the participants were asked to describe the indicated object.

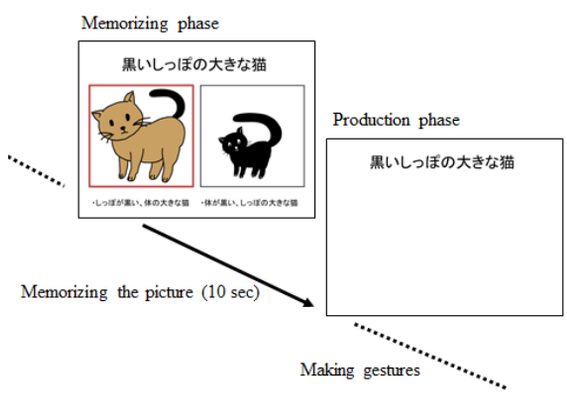

Figure 3. Experimental flow in side-by-side condition. The picture in the memorizing phase consisted of one target phrase (e.g., 黒いしっぽの大きな猫 (Black tail-GEN big cat)) on the top, illustrations in the middle, and explanatory notes for the illustration on the bottom. In this example, there are two explanatory notes. Left explanatory note (i.e., しっぽが黒い、体の大きな猫) indicates a big cat with black tail. Right explanatory note (i.e., 体が黒い、しっぽの大きな猫) indicates a black cat with big tail. Participants made gestures as soon as the memorizing phase was over. In production phase, only the phrase which is on the top was shown.

Results

The total number of gestures that co-occurred in each constituent was measured (Figure 4). To calculate a significance level in each constituent, we compared the number of gestures that participants produced in the Big-cat structure and in the Big-tail structure, using a binomial test. There were more gestures that synchronized with the "first word - second word" chunk, in the Big-cat structure (32/33) than the Big-tail structure (1/33) ($p < .01$, $g = 0.47$). There were more gestures that synchronized with the "third word - fourth word" chunk in the Big-cat structure (39/52) than the Big-tail structure (13/52) ($p < .01$, $g = 0.25$). There were more gestures that synchronized with the first word (only) in the Big-tail structure (28/30) than the Big-cat structure (2/30) ($p < .01$, $g = 0.43$). There were more gestures that synchronized with the "second word - third word" chunk in the Big-tail structure (16/17) than the Big-cat structure (1/17) ($p < .01$, $g = 0.44$). There were more gestures that synchronized with the "second word - third word - fourth word" chunk in the Big-tail structure (20/28) than the Big-cat structure (8/28) ($p < .05$, $g = 0.21$).

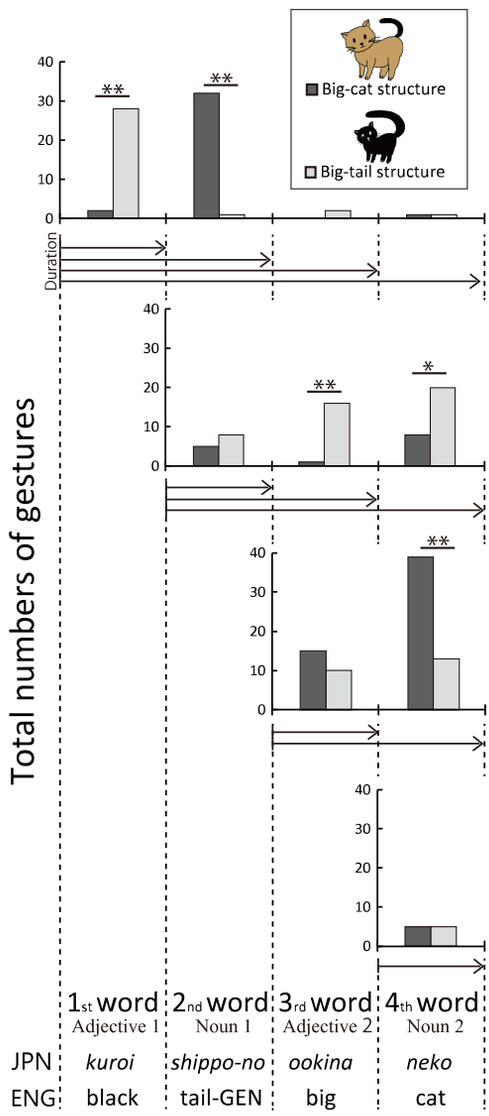

Figure 4. The number of gestures that synchronized with 1st – 4th words were counted. The top panel shows the data for gestures which overlapped with 1st word. The second panel shows the data for gestures that did not overlap with 1st word, but overlapped with 2nd word. The third panel shows the data for gestures that did not overlap with 1st-2nd words, but overlapped with 3rd word. The bottom panel shows the data for gestures that overlapped only with 4th word. Arrows denote the onset and duration of gestures, e.g., for a gesture starting with the 1st word, there were four possible durations: the first word only (the first cluster of bars), the first word and the second word (the second cluster of bars), the first through the third word (the third cluster of bars), the first through the fourth word (fourth cluster of bars). Asterisks indicate statistical significance between the two structures, * $p < .05$, ** $p < .01$.

Discussion

This study investigated whether patterns of speech-gesture synchronization changes, depending on the syntactic structure in syntactically ambiguous phrases. We examined all (10) possible synchronization patterns of speech and gestures as follows: *synchronization with the first word only, the first word and the second word, the first through the third word, the first through the fourth word, the second word only, the second word and the third word, the second through the fourth word, the third word only, the third through the fourth word, and the fourth word only.* We found, for the first time, that gestures synchronized with different words in syntactically ambiguous sentences, depending on the intended syntactic structure. More specifically, gestures tended to synchronize with the chunks of words that form a constituent in the Big-cat structure or the Big-tail structure. As Figure 4 shows, when people produced a sentence with the Big-tail structure, they often produced gestures that spanned *the second word and the third word* (e.g., *shippo-no ookina*, "big tailed"), and also gestures that spanned *the second through the fourth word* (e.g., *shippo-no ookina neko*, "big-tailed cat"). When people produced a sentence with the Big-cat structure, they often produced gestures that spanned *the first word and the second word* (e.g., *kuroi shippo-no*, "black tailed") and also gestures that spanned *the third through the fourth word* (e.g., *ookina neko*, "big cat"). These results mean that the gestures tended to overlap with the chunks of words that corresponded to a constituent in each structure. From the recipient's point of view, a pattern of speech-gesture synchronization is a cue for the syntactic structure of a phrase, which can be used to disambiguate syntactically ambiguous phrases.

One limitation of the current study is that we have instructed participants to produce gestures with speech. Thus, further studies are needed to establish whether the current findings can extend to spontaneously produced gestures. Instructed gestures have been frequently used in recent studies of gestures (e.g., Argyriou et al., 2017, Broaders et al., 2007; Chu & Hagoort, 2014). Such studies can lay an important foundation for our understanding of spontaneous co-speech gestures. Further research to engage in the observation and analysis of gestures in natural speech situations is needed.

What are the implications of the current finding on human evolution? Did the gesture's disambiguation function appear before the emergence of hierarchical syntax or after it in human evolution? Even when there was no hierarchical syntax, people could still utter concatenations of two words. They might have used gestures to disambiguate the meanings of those primitive, linearly syntactic constructions. Ontogenetic evidence shows that children start using gestures early before using syntactic structures, and they use gestures plus word combinations before using two-word utterances. Based on longitudinal study, Iverson and Goldin-Meadow (2005) showed that children who were first to produce gesture-plus-word combinations (e.g., point at bird and say "nap") were first to produce two-word combinations ("bird nap"). Later, children's speech at two-word stage is sometimes produced with gestures.

A seminal example is a child named Kathryn's utterance "mommy sock" reported by Bloom (1970). Kathryn was twenty-one month old, and her mean length of utterances (i.e., the mean number of morphemes per utterance) was 1.32. She said this expression in two completely different situations. One was when Kathryn pointed to her mother's socks. The other was when her mother was putting Kathryn's socks on her. Thus, this noun plus noun construction meant a genitive construction of possessor-possessed relation (mommy's sock), or

Subject-object construction where mommy as a subject and sock as an object. Thus, Kathryn used identical surface structure forms to express two different meanings. Her pointing gesture could disambiguate the meaning of her utterance in the first situation. We therefore guess that ancient human gestures already had a disambiguation function for concatenations of words before the emergence of hierarchical syntax. The function was later on extended for the disambiguation of hierarchically organized syntactic structures.

In conclusion, we found that when people produced speech with gesture, gestures synchronized with a chunk of words that correspond to a constituent in syntactic structures. This finding may be extended to natural speech situation where gestures are spontaneously produced. Our study suggests that gestures may play an important role in disambiguating syntactically ambiguous phrases. This could be a reason why humans have continuously used gestures even after they acquired a powerful tool of language and why today, they still produce language-redundant gestures.

Acknowledgments

We would like to thank all our participants in this study. We would also like to thank Editage (www.editage.jp) for English language editing. This study was supported by JSPS/MEXT KAKEN JP17H06382 (H.K.) and JP16K04318 (H.K.)

Supplementary Material

Electronic supplementary materials are available online.

References

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., & Mainela-Arnold, E. (2009). Gesture-speech integration in narrative: are children less redundant than adults? Gesture, 9, 290-311. https://doi.org/10.1075/gest.9.3.02ali

Argyriou, P., Mohr, C., & Kita, S. (2017). Hand matters: left-hand gestures enhance metaphor explanation. Journal of Experimental Psychology: Learning Memory and Cognition, 43, 874-886. https://doi.org/10.1037/xlm0000337

Bloom, L. (1970). Language development: form and function in emerging grammars. MIT Press.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. Journal of Experimental Psychology: General, 136, 539-550. https://doi.org/10.1037/0096-3445.136.4.539

Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: evidence for interaction in action. Journal of Experimental Psychology: General, 143, 1726-1741. https://doi.org/10.1037/a0036281

Corballis, M. C. (2009). The evolution of language. Annals of the New York Academy of Sciences, 1156, 19-43. https://doi.org/10.1111/j.1749-6632.2009.04423.x

Fujita, K. (2014). Recursive merge and human language evolution. In T. Roeper, & M. Speas (Eds.), Recursion: complexity in cognition. Studies in theoretical psycholinguistics Vol 43 (pp. 243-264). Springer, Cham. https://doi.org/10.1007/978-3-319-05086-7_11

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: the case of gesture. Perspectives on Psychological Science, 5, 664-674. https://doi.org/10.1177/1745691610388764

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: evidence from Japanese. Cognition, 136, 350-358. https://doi.org/10.1016/j.cognition.2014.11.033

Ito, K., Arai, M., & Hirose, Y. (2015). The interpretation of phrase-medial prosodic prominence in Japanese: is it sensitive to visual and discourse context? Language, Cognition and Neuroscience, 30, 167-196. https://doi.org/10.1080/01690965.2013.864778

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. Psychological Science, 16, 367-371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Kashiwadate, K., Yasuda, T., & Kobayashi, H. (2019). Do people use gestures differently to disambiguate the meanings of Japanese compounds? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), Proceedings of the 41st annual conference of the Cognitive Science Society (pp. 527-531). Cognitive Science Society.

Kita, S. (2002). Jesuchaa: Kangaeru karada [Gesture: the body that thinks]. Kaneko Shobo. (in Japanese)

Kita, S. (Ed.). (2003). Pointing: where language, culture and cognition meet. Psychology Press.

Kita, S., Danziger, E., & Stolz, C. (2001). Cultural specificity of spatial schemas, as manifested in spontaneous gestures. In M. Gattis (Ed.), Spatial schemas and abstract thought (pp. 115-146). MIT Press.

Kita, S., & Saito, H. (2002). Jesuchaa: Koui Imi [Gesture, action, and meaning]. Kyoritsu Shuppan. (in Japanese)

McNeill, D. (1987). Psycholinguistics: a new approach. Harper & Row Publishers.

McNeill, D. (2005). Gesture and thought. University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case of space. Trends in Cognitive Sciences, 8, 108-114. https://doi.org/10.1016/j.tics.2004.01.003

Misono, Y., Mazuka, R., Kondo, T., & Kiritani, S. (1997). Effects and limitations of prosodic and semantic biases on syntactic disambiguation. Journal of Psycholinguistic Research, 26, 229-245. https://doi.org/10.1023/A:1025065700451

Okahisa, T., & Shirose, A. (2018). Influence of hand gestures on prosodic disambiguation of syntactically ambiguous phrases. Acoustical Science and Technology, 39, 171-174. https://doi.org/10.1250/ast.39.171

Tomasello, M. (2008). Origins of human communication. MIT Press. https://doi.org/10.7551/mitpress/7551.001.0001

Venditti, J. J. (1994). The influence of syntax on prosodic structure in Japanese. OSU Working Papers in Linguistics, 44, 191-223.

Venditti, J. J. (2006). Prosody in sentence processing. In M. Nakayama, R. Mazuka, & Y. Shirai (Eds.) The Handbook of East Asian Psycholinguistics, Volume II: Japanese (pp. 208-217). Cambridge University Press.