Title

Spontaneous co-speech gestures with prompt phrases reflect linguistic structures

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Kimura, Hina
Yasuda, Tetsuya
Kobayashi, Harumi

Publication Date

2023

Peer reviewed

Spontaneous co-speech gestures with prompt phrases reflect linguistic structures

Hina Kimura (23rmd14@ms.dendai.ac.jp)
Graduate School of Tokyo Denki University, Ishizaka, Hatoyama-machi, Hiki-gun
Saitama 350-0394, Japan

Tetsuya Yasuda (t-yasuda@g.ecc.u-tokyo.ac.jp)
The University of Tokyo, 3-8-1, Komaba, Meguro-ku,
Tokyo 153-8902, Japan

Harumi Kobayashi (h-koba@mail.dendai.ac.jp)
Tokyo Denki University, Ishizaka, Hatoyama-machi, Hiki-gun
Saitama 350-0394, Japan

Abstract

This study aimed to investigate whether people produce spontaneous co-speech gestures that reflect the underlying linguistic branching structures and additional information when speech is restricted by prompt phrases. Participants were asked to convey information about an animated movie using a three-word phrase in Japanese that could be interpreted in two different ways depending on their underlying branching structures. The animated movie included or did not include important information that was not described by the prompt phrases. The results showed that most participants produced gestures while uttering the phrase, and the onset of gesture reflected the underlying branching structures. A time-series analysis revealed that the occurrence of object-action gestures that depicted a noun's movement tended to reflect the associated branching structures and semantic elements. People spontaneously produce gestures, which may syntactically and semantically help to disambiguate ambiguous phrases.

Keywords: spontaneous gesture, syntactic gesture, linguistic structure, speech-gesture onset, branching structure

Introduction

Why do people often use gesture in addition to speech? Previous studies have suggested that people use co-speech gestures to emphasize a part of speech (Bull & Connelly, 1985), compensate for the content of speech (Iverson & Goldin-Meadow, 2005; Kita & Özyürek, 2003), enhance language comprehension (Kelly, Özyürek, & Maris., 2010), and for their own thinking (Goldin-Meadow & Beilock, 2010). Another important reason for this finding has been proposed: People use co-speech gestures to disambiguate the inherently ambiguous linguistic structures of their utterances.

The fact that language has hierarchical structures can be illustrated using a three-word phrase (Fujita & Fujita, 2022). For example, the phrase "green tea cup" can be interpreted in two different ways: as either a cup for green tea or a green-colored tea cup. In left branching (LB), the adjective "green" first makes a branch with the noun "tea" and then "green tea" makes a branch with "cup," resulting a structure whose meaning is "a cup for green tea." On the other hand, in right branching (RB), the noun "tea" first makes a branch with "cup" and then "tea cup" makes a branch with the adjective "green," resulting a phrase that can be interpreted "a green-colored tea cup." Thus, language has a hierarchical structure at the deep structure level, whereas at the surface structure level, utterances are sequentially produced words and thus do not show a deep structure. Therefore, it may be difficult to determine the correct meaning. Nonetheless, we do not usually feel that the meaning of our language is ambiguous. Previous studies have shown that prosody plays a role in disambiguation (Hirose & Mazuka, 2015; Ito, Arai, & Hirose, 2015; Okahisa & Shirase, 2018). For example, Hirose (2020) showed that when participants heard an adjective + noun + noun (i.e., a three-word) phrase in which a change in pitch accent did not occur, they interpreted the phrase with left branching. However, when participant heard the same three-word phrase with "metrical boost" in which changing of pitch accent in the first noun occurred, the participants interpreted the phrase with right branching. In addition to prosody, there may be other means of disambiguation.

Some recent studies have examined the role of co-speech gestures in syntactic disambiguation (Kashiwadate, Yasuda, Fujita, Kita, & Kobayashi, 2020; Kita, Özyürek, Allen, Brown, Furman, & Ishizuka, 2007; Özyürek, 2014). Regarding deep structures, in particular branching structures, Kashiwadate et al. (2019, 2020) suggested that gestures can be used to identify syntactically ambiguous structures in a study of Japanese ambiguous phrases. They focused on the ambiguity of four-word Japanese phrase structures that can be interpreted as having at least two different meanings, and investigated the possible role of gestures in the disambiguation of meanings. The results showed that the onsets of gestures differed to reflect different Japanese linguistic structures, and a detailed time-course analysis showed that the gestures tended to be synchronized according to linguistic chunks. They suggested that gestures could be used to disambiguate syntactically ambiguous structures.

Handa et al. (2021) examined the contribution of gestures to ambiguous clause structures consisting of three words with verbs. They focused on branching structures (i.e., right or left branching) that may be found in the same utterance. The stimulus was ambiguous clause structures with a verb like Verb + Noun-1 + Noun-2, for example, "*Rakka-shiteiru* (fall + ing) *otoko-no* (man + particle) *keitai* (cell-phone)." The participants were asked to utter the prompt phrases while

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

gesturing. The participants' gestures were analyzed in terms of their timing. The results showed that the timing of the onset of the first gesture was slower in the right branching (RB) than the left branching (LB) clause structure. They reported that this slowness corresponded to chunks in the clause structure.

Thus, recent investigations of the role of co-speech gestures suggested that when participants were asked to perform gestures while uttering phrases, the onset and combination of their gestures tended to reflect the underlying linguistic structures. However, the participants' gestures in these studies were forced and not spontaneous; therefore, gesture production might have included some unnaturalness. It is not known whether people spontaneously produce syntactically informative gestures.

Do people spontaneously use gestures to disambiguate syntactically ambiguous utterances? In this study, the speaker was not asked to produce gestures but was simply asked to convey information to a listener about an animated movie in which a walker encountered various scenes. The utterance the speaker could use was controlled; it was a three-word phrase in Japanese such as "*Rakka-shiteiru* (fall + ing) *neko-no* (cat + particle) *shashin* (photo)." This phrase can be interpreted either "The photo that depicts a falling cat" ("Falling cat" construction) or "The falling photo that depicts a cat ("Falling photo" construction)." [1] In addition, for each phrase construction, there were two situations regarding the informativeness of the depicted scenes. For example, for the "Falling cat" construction in a more informative situation, the falling cat was falling into a crocodile's mouth. In a less informative situation, the cat fell onto grass. We expected most participants to spontaneously produce gestures that reflected the underlying linguistic structures, and additionally hypothesized that the participants would spontaneously produce additional gestures that were not described in the controlled utterance, such as the crocodile's mouth in the more informative situation.

We also investigated the effect of verb types because a previous study (Handa et al.) reported that the onset and frequency of gestures vary between gestures that depict "fall" and "fly." Additionally, the meanings of these verbs can be expressed relatively easily in a specific manner and path, such as falling straight down or flying in an arc. The timing of the participants' language and the content of their gestures were analyzed using a time-series analysis to determine the moments when participants used gestures while speaking the prompt phrase.

Method

Participants

Thirty-one Japanese monolingual students who spoke Japanese as their first language and one Chinese student participated in the study (M_{age} = 21.6; *SD* = 0.78). The analysis excluded groups for which the response sheets were misaligned or incomplete or that contained invalid responses. Therefore, we used data from 12 groups that utilized spontaneous cospeech gestures. The experiment was conducted in accordance with the university's code of confidentiality and ethical treatment of human subjects.

Condition and Stimuli

The experimental conditions consisted of branching (2: left-, right-branching), informativeness (2: more-, less-informative), and verbs (2: fall, fly).

The branching condition concerned the ambiguous phrases presented in the stimulus video, in addition to related scenes. Under this condition, there were two levels: left or right branching. An example of an ambiguous three-word phrase structure used in this study was "*Rakka-shiteiru* (Fall + ing) *neko-no* (Cat + particle) *shashin* (Photo)," Verb (V: verb) + Noun-1 (N1: first noun) + Noun-2 (N2: second noun) (Figure 1). In left branching, the Verb and Noun-1 are chunked first. This phrase can be interpreted as "a photo that depicts a falling cat ({{falling, cat}, photo})" In contrast, in right branching, Noun-1 and Noun-2 are chunked first. This phrase can be interpreted as "a falling photo that depicts a cat" ({falling, {cat, photo}}).

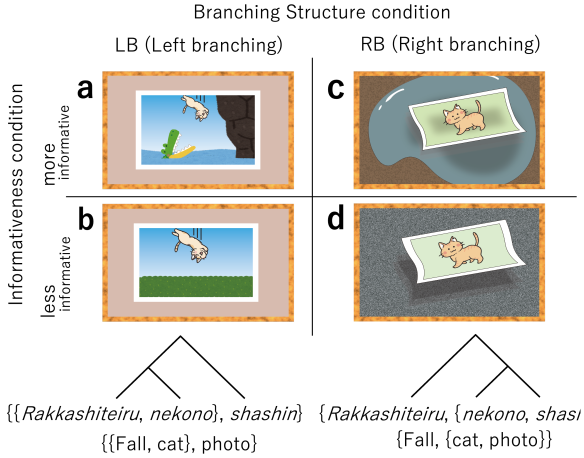
Branching Structure condition
LB (Left branching)
RB (Right branching)
Informativeness condition
more informative
less informative
a
b
c
d
{{Rakkashiteiru, nekono}, shashin}
{{Fall, cat}, photo}
{Rakkashiteiru, {nekono, shasi
{Fall, {cat, photo}}

Figure 1: An example of a stimulus "Falling cat photo" which can be interpreted differently based on deep structures. Two informativeness situations were used for each structure: left branching (more or less informative with crocodile or grass) and right branching (more or less informative with puddle or ground). Listeners were then tested on selecting the correct picture out of four options.

[1] Japanese is generally described as a left-branching language where modifiers come before the syntactic head of the sentence (Makino & Tsutsui, 1989). In the stimuli in this experiment, the head was at the end of the phrases and the modifiers came before the head in either the left-branching or the right-branching phrase.

The informativeness condition had two levels: more informative and less informative. In the "more informative situation," the photo included more information about the falling cat, that is, the cat was falling into a crocodile's mouth. The situation involved an unusual risky element and the participants were expected to talk about the crocodile. In the "less informative situation," the photo included less information about the falling cat because it was falling onto grass, so the situation did not involve something unusual and the participant was not expected to talk about the grass. Likewise, in the "more informative situation" in RB, the cat was falling onto a puddle, so the picture depicted an unusual risky situation (the photo may be wet and destroyed), but in the "less informative situation" the photo was falling onto the ground and the situation did not involve an unusual element of risk.

The verb condition consisted of two verbs: "falling" and "flying." "Fall" implies a strong directionality like gravity (determined by the simple physical reality of the world), whereas "fly" does not. These two verbs were selected to reflect two major movements.

The experimental stimuli consisted of 12 different phrases constructed with 2 verbs (i.e., V), 6 first nouns (i.e., N1), and 3 second nouns (i.e., N2). Twelve pictures were created to represent left- and right-branching structures with background information of varying informativeness, for example, crocodile (high informativeness) /grass (low informativeness).

The stimuli consisted of two sets of animated movies. One set (more informative) included 12 trials with pictures representing left- and right-branching structures under more informative conditions. Another set (less informative) included 12 trials of stimuli with pictures representing the left- and right-branching structures with less informative conditions.

With respect to the factorial design, the branching structure and verb conditions were within-participants factors, while the informativeness condition was a between-participants factor.

Procedure

The participants in each pair were randomly assigned to either the role of the speaker, who was tasked with delivering the content of a stimulus, or the listener, who was responsible for observing the speaker's utterances and gestures.

In the initial phase, the speaker (Participant A) and listener (Participant B) sat at a table. Each participant was positioned such that they could only see their own monitor and not the other participant's. The experimenter then provided instructions about the stimuli to be used and explained the speaker's and listener's roles, as well as the general flow of the experiment. Additionally, the experimenter informed the participants that they would receive a reward based on the percentage of correct responses after the experiment was completed.

After receiving the instructions, the speaker watched the stimulus and freely conveyed its content to the listener. The researcher asked the speaker to say a specific prompt phrase displayed during the stimulus. The listener was then asked to complete a test in which they had to choose a picture from a list (four pictures), as shown in Figure 1. Once all the trials were completed, Participants A and B switched roles and locations and repeated the entire trial. Additionally, participants saw each scene only once because the informativeness condition was a between-participant factor.

A digital video camera (FDR-AX40, Sony) was used to record the entire session. Additionally, the camera captured the listener's upper body, including arms and facial expressions. However, for this analysis, we did not examine facial expressions.

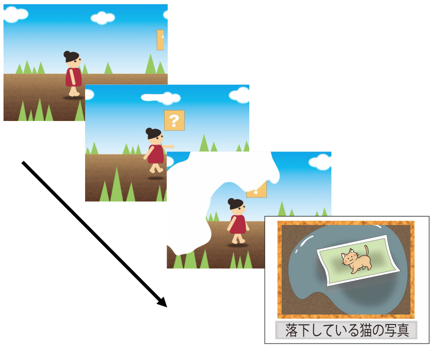

Figure 2: Flow of the animated movie presented for the speaker.

Coding

ELAN (ver. 6.4) was used for analysis. Speech and gestures were coded for each branching structure based on the video data (29.97 frames/sec).

We analyzed the gestures according to Kendon's (2004) gesture phases, which capture the movement dynamics of the gesture, including the stroke itself, the preparatory movements leading up to the stroke, the recovery phase when the gesture relaxes or withdraws, and the post-stroke hold phase when the gesture sustains its position at the end of the stroke. Each gesture was classified into one of three categories.

We analyzed the uttered words (e.g., verb, noun-1, and noun-2) using ELAN. Since the stimulus consisted of "Verb (e.g., *rakka-shiteiru*; fall + ing) Noun-1 (e.g., *neko-no*; cat + particle) Noun-2 (e.g., *shashin*; photo)," we classified the words into three categories. To determine the start of the uttered word, the sound wave function in ELAN was used to identify the point at which the word could be clearly heard. To determine the end of an uttered word, we identified the point at which the vowels became difficult to discern.

To examine the relationship between co-speech gestures and their meanings, the gesture time-series data were aligned with the speech time-series. The time points of the speech time-series ranged from 5 s before the start of the utterance to 2 s after the end of the utterance. We analyzed the difference in the onset time of the gesture "stroke" relative to the onset time of the utterance "N1 (first noun)." According to Handa et al. (2021) who used similar phrases as stimuli, in left branching, the "N1" utterance (first noun) was chunked with the "verb" utterance (i.e., falling cat); contrariwise, in right branching, the "N1" utterance was chunked with the "N2" utterance (second noun; i.e., cat's photo). Thus, the timing of the N1 utterance is important to specify the branching structure.

We analyzed the relationship between verbs and specific manner-and-path gestures, such as object action gestures, using a time-course analysis. Object-action gestures depict an action and related objects simultaneously or sequentially. For instance, when the verb "fall" was used in the left branching structure, participants tended to produce object-action gestures that depicted both the verb and the N1 object (e.g., a falling cat), with gestures flowing quickly and straight. These gestures were coded as object-action gestures. Conversely, in the right-branching structure, the participants produced object-action gestures that reflected the verb and the N2 object (e.g., a falling photo), with gestures depicting a slow and fluttering motion. We also coded these gestures as object-action gestures.

Analysis

To examine the synchronicity between speech and gesture, we investigated whether participants' coordination of speech and gesture differed among conditions such as context (informativeness) and ambiguous phrases (left- and right-branching) that included verb information. Linear mixed models (LMMs) were constructed using the *lmer* function in the lme4 package to fit each time point using the restricted maximum likelihood.[2] In addition, each condition was coded using dummy coding and centered for effect coding (e.g., −0.5, 0.5).

First, a maximum model was constructed that included experimental conditions and their interactions as fixed effects and individual and item differences as random effects. Subsequently, a forward stepwise method was used to examine the candidate models that would fit the data obtained. Model selection suggested that Informativeness, Branching, Verb, and their interactions (excluding the interaction between Informativeness and Verb) be applied as fixed effects (Formula = lmer(Onset ~ Branching * (Informativeness + Verb) + (1|participants) + (1|item)), *df* = 9, *AIC* = 791.7, *weight* = 0.157).

To examine gesture onsets in a time series, we analyzed gesture production in a time series using cluster-based permutation analysis[3] (CPA), which has been utilized in studies of electroencephalography (EEG) and the visual world paradigm. First, the time course of gesture data, including conditions, was computed by binning at 100 msec intervals for the CPA, which requires high-density data in the time series. We then specified that the target data be analyzed to compare with the branching conditions. In addition, branching conditions were coded as effect coding.

We computed the CPA via the generalized linear mixed model (GLMM) to use the "*clusterperm.glmer*" function with a binomial distribution. This GLMM applied the branching condition as a fixed effect and individual and item differences as random factors, with the number of permutations of 1000 times. CPA is a nonparametric statistical technique capable of identifying significant patterns of two factors in a time series.

Results

Synchronicity between speech and gesture

To compare the synchronicity between speech and gestures, the LMM was computed using the *lmer* function in lme4:

The LMM fit between speech-gesture onset time and each condition via lme4 revealed that the intercept (β = 1.01, *df* = 18.53, *t* = 5.34, *p* < .001), Informativeness (β = 0.797, *df* = 19.20, *t* = 2.10, *p* = .049), and Verb (β = 0.458, *df* = 9.76, *t* = 3.07, *p* = .012) were significant. The LMM also revealed that Informativeness × Branching structure (β = −0.708, *df* = 206.81, *t* = −2.43, *p* = .016) and Branching structure × Verb (β = −0.722, *df* = 211.92, *t* = −2.48, *p* = .014) interactions are significant.

To reveal the simple main effects of the interactions, post hoc tests were conducted pairwise using emmeans. Regarding the interaction between branching structure and informativeness, there were significant differences in informativeness between the RB structure (*df* = 24.9, *t.ratio* = −2.83, *p* = .009) and branching structure in the more informative condition (*df* = 210.0, *t.ratio* = 1.98, *p* = .049). In the RB structure, the onset time of the gesture stroke relative to the onset time of the Noun-1 utterance was significantly later with more informative (*EMM* = 1.62, 95%*CI*[1.06, 2.17]) than less informative (*EMM* = 0.47, 95%*CI*[−0.11, 1.04]; Figure 3). However, in the LB structure, the gesture onset time did not differ between the informativeness situations (*df* = 26.2, *t.ratio* = −1.017, *p* = .319).

Regarding the interaction between branching structure and verb, there were significant differences between the verb in the RB structure (*df* = 35.0, *t.ratio* = −3.92, *p* = .0004) and the branching structure in the verb "Fall" (*df* = 214.0, *t.ratio* = 2.01, *p* = .046). In the RB structure, the onset time of gesture stroke relative to the onset time of the Noun-1 utterance was significantly later when the participants uttered the verb "Fall" (*EMM* = 1.45, 95%*CI* [1.00, 1.90]) than when they uttered the verb "Fly" (*EMM* = 0.63, 95%*CI* [0.19,1.08]; Figure 4). However, in the LB structure, the gesture onset

[2] R software (R core team, 2022), lme4 (ver. 1.1-26; Bates et al., 2015), ImeTest (ver. 3.1-3; Kuznetsova et al., 2017), emmeans (ver. 1.6.3; Lenth et al., 2018), ggeffects (ver. 1.0.2; Lüdecke, 2018), and MuMIn (ver. 1.43.17; Bartoń, 2020) were used for statistical modeling. If a statistical model found an interaction effect in their factor, a post-hoc test we used the contrast function in the emmeans (ver. 1.80; Length et al., 2022) We also used tidyverse (Wickham et al., 2019), dplyr (Wickham et., 2022), and stringr (Wickham, 2022) for data handling.

[3] clusterperm and permutes (Voeten, 2022) was used for cluster-based permutation analysis.

time did not differ between these verbs (*df* = 30.7, *t.ratio* = −0.774, *p* = .463).

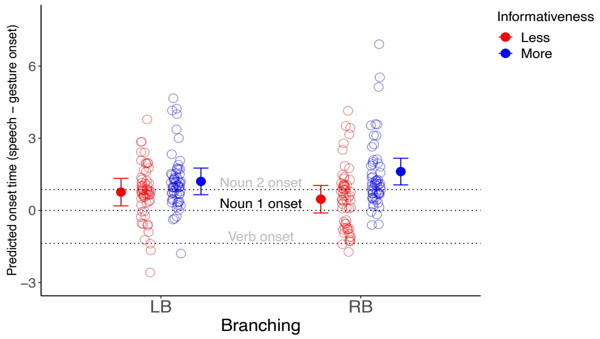

Figure 3: Gesture onset times based on the N1 utterance onsets regarding interaction effect between branching structure and informativeness.

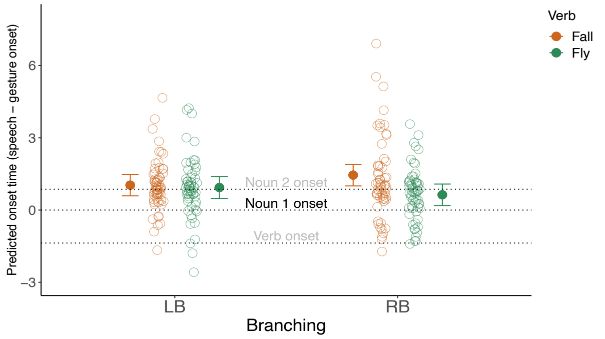

Figure 4: Gesture onset times based on the N1 utterance onsets regarding interaction effect between branching structure and verb.

Gesture onset in time series

To compare gestural expressions and their time courses, CPAs were computed using the *clusterperm.glmer* function in permutes. Figure 5 shows the time course of the results of the CPA based on GLMMs.

Regarding the verb "Fall (Figure 5a)," CPA revealed that object action gestures were confirmed by the significant positive cluster in the less informative condition (−1100 – 100 ms, cluster mass statistic = 331.8, *p* < .05), indicating that more occurrences of the left branching than the right branching was observed. CPAs also revealed that the object action gestures were confirmed by the significant negative cluster in less informative (1800 – 2900 ms, cluster mass statistic = 528.9, *p* < .05), indicating that more occurrences of the right branching than the left branching was observed.

In addition, in the more informative condition, positive and negative clusters were confirmed to resemble cluster ranges in the less informative condition (pos. 0 – 700 ms, cluster mass statistic = 53.7; neg. 2400 – 5000 ms, cluster mass statistic = 1142.4, *p* < .05).

Regarding the verb "Fly" (Figure 5b)," CPAs revealed that, the object action gestures were confirmed by the negative cluster in less informative (−1100 – 400 ms, cluster mass statistic = 692.4; 2500 – 4000 ms, cluster mass statistic = 484.4, *p* < .05), indicating that more occurrences of the right branching than the left branching was observed, whereas in more informative, the object action gestures were confirmed by a positive cluster (−800 – −200 ms, cluster mass statistic = 218.7; 3800 – 4500 ms, cluster mass statistic = 152.0, *p* < .05), indicating that more occurrences of the left branching than the right branching was observed.

In addition, we observed that participants tended to add more gesture in the "more informative" situations, such as a crocodile opening its mouth or a puddle appearing. In both the more and less informative situations, the path and manner of falling or frying actions were observed by gestures.

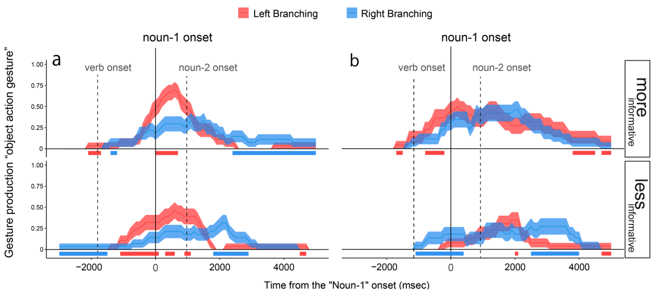

Figure 5: Time-course in the object-action gestures with the verbs "Fall" (a) and "Fly" (b). The error bars show standard errors. The red line at the bottom (around *y* = 0) indicates the significant positive cluster, and the blue line indicates the significant negative cluster. The *x*-axis shows the time course based on the onset time of Noun-1, and the *y*-axis represents the gesture production, with a value close to 0 indicating that few participants produced the gesture and a value close to 1 indicating that almost all participants produced the gesture.

Discussion

We examined whether participants produced spontaneous gestures to reflect branching structures when they uttered Japanese prompt phrases. The results for the gesture onset time showed that the onset of the gesture stroke was delayed compared to the onset of the utterance of Noun-1 (the first noun) when the participants watched animated that consisted of a right branching structure and a more informative situation (i.e., the picture included additional information but the prompt phrase did not). Interestingly, the onset of the gesture stroke was not delayed when participants watched animated movie involving a left-branching structure and a more informative situation. In right branching, it is important to convey the chunk of the first and second nouns ([cat] [photo]) that must be chunked with the verb ([fall]), so that the participants pay attention to the chunk of the two nouns, which comes relatively late in the utterance. As this delay was observed only with a more informative picture, participants were likely to find it difficult to add more information. This may explain why the participants later produced gestures. This speculation seems to indicate that the timing of emerging gestures is influenced by the branching structure of linguistic structures.

The findings of these analyses and a visual inspection of gesture time series demonstrated that the participants gestured using syntactic chunking when they said the verb "fall," but not much when they said the verb "fly." This may imply that gestures can be used to convey both the movement depicted by the verb and the related syntactic structure about the verb "fall," whose suggested the movement and the path are simply determined by the gravity concept. However, as the verb "fly" itself does not depict a determined path, participants may use gestures to express semantic information according to the agent of the verb "fly," therefore the participants have paid more attention to describe the movement of the verb. These findings seem to suggest that in this study, gestures play an important role to express both the syntactic structure and verb meanings.

In conclusion, this study showed that participants produced spontaneous gestures while uttering phrases that can be interpreted with two different meanings. The results suggest that people spontaneously produce co-speech gestures, which may syntactically and semantically help disambiguate ambiguous phrases.

Acknowledgments

We would like to thank all the participants in this study. This study was supported by JSPS/MEXT KAKENHI JP17H06382 (H.K.), JSPS KAKENHI JP20H01763 (H.K.), and JP20K03375 (T.Y.).

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bull, P., & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior*, *9*(3), 169–187.

ELAN (Version 6.4) [Computer software]. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fujita, H., & Fujita, K. (2022). Human language evolution: A view from theoretical linguistics on how syntax and the lexicon first came into being. *Primates*, *63*(5), 403–415.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674.

Handa, Y., Yasuda, T., & Kobayashi, H. (2021). The use of co-speech gestures in conveying Japanese phrases with verbs. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 1555–1559.

Hirose, Y., & Mazuka, R. (2015) Anticipatory processing of novel compounds: Evidence from Japanese. *Cognition*, *136*, 350–358.

Ito, K., Arai, M., & Hirose, Y. (2015). The interpretation of phrase-medial prosodic prominence in Japanese: Is it sensitive to context? *Language, Cognition and Neuroscience*, *30*, 167–196.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371.

Kashiwadate, K., Yasuda, T., Fujita, K., Kita, S., & Kobayashi, H. (2020). Syntactic structure influences speech-gesture synchronization. *Letters on Evolutionary Behavioral Science*, *11*(1), 10–14.

Kashiwadate, K., Yasuda, T., & Kobayashi, H. (2019). Do people use gestures differently to disambiguate the meanings of Japanese compounds? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*, 527–531.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(1), 1–26.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Estimated Marginal Means, aka Least-Squares Means. Retrieved from https://CRAN.R-project.org/package=emmeans

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Makino, S., & Tsutsui, M. (1989). *Dictionary of basic Japanese grammar*. The Japan Times.

Okahisa, T., & Shirose, A. (2018). Influence of hand gestures on prosodic disambiguation of syntactically ambiguous phrases. *Acoustical Science and Technology*, *39*, 171–17.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Voeten, C. C. (2018). permutes: Permutation tests for time series data. R package version 0.1. Available online at: https://CRAN.R-project.org/package=permutes

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., ... & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.