Agent versus Non-Agent Motions Influence Language Production: Word Order and Perspective in a VOS language

Manami Sato (manamisato@gmail.com)
Department of British and American Language and Culture, Okinawa International University
2-6-1 Ginowan, Ginowan City, Okinawa 901-2701 JAPAN

Keiyu Niikuni (keiyu@cog.is.tohoku.ac.jp)
Graduate School of Information Sciences, Tohoku University
Aramaki aza Aoba 6-3-09, Aoba-ku, Sendai, Miyagi 980-8579 JAPAN

Amy J. Schafer (aschafer@hawaii.edu)
Department of Linguistics, University of Hawaii
1890 East-West Road, Moore Hall 569, Honolulu, HI 96822 USA

Masatoshi Koizumi (koizumi@sal.tohoku.ac.jp)
Graduate School of Arts and Letters, Tohoku University
Kawauchi 27-1, Aoba-ku, Sendai, Miyagi 980-8576 JAPAN

Abstract

Is language production isolated from our experiences of physical events, or can physical motion affect the conceptual saliency of the components of a to-be-described event, in ways that affect its linguistic description? This study examined the influence of physical motion on the interpretation and description of simple transitive events. More specifically, we investigated whether engagement in non-speech physical actions affects the relative location of verbs versus arguments in sentence production, and the relative location and prominence of Agents, by testing native speakers of Truku, a language that allows flexibility in each of these options and presents under-studied typological patterns.

Keywords: embodiment; conceptual saliency; sentence production; motion; endangered language; Verb-initial language

Introduction

Everyday conversations are often conducted when the interlocutors are simultaneously engaged in another physical activity. We often talk while cooking, cleaning, exercising, and so on. Does such activity influence how language is produced and perceived? While there are long-standing arguments for connections between the comprehension of language about motion and actual motion production (Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006), this question remains underexplored. Moreover, experimental studies of language processing have been heavily dominated by studies of a small sample of the world's languages (i.e., languages in which Subject precedes Object and Verb appears in the medial or final position), under-representing important typological patterns. Here, we evaluated whether engagement in non-speech physical actions affects the relative location of verbs versus arguments in sentence production, and the relative location and prominence of phrases that refer to Agents, by testing native speakers of Truku, a language that allows flexibility in these options. More specifically, we examined how Truku speakers conceptually interpret and linguistically describe transitive events in their selections of voice (Actor voice, Goal voice) and word order (Verb-Object-Subject versus Subject-Verb-Object) in a picture-description task.

Conceptual Saliency in Sentence Production

When people linguistically describe transitive events, they must apprehend who-did-what-to-whom information, assign appropriate semantic roles to each entity, align the relevant phrases in a linear order, and produce an utterance (Bock & Loebell, 1990; Ferreira & Slevc, 2007). For a transitive event in which a girl is kicking a boy, active and passive expressions (i.e., *A girl kicked the boy,* and *A boy was kicked by the girl*, respectively) are both grammatical, so what makes speakers select one particular expression over another? Previous studies have shown that one of the factors that influences the selection of active/passive voice and word order in describing or recalling transitive events is conceptual accessibility (Japanese: Tanaka, Branigan, McLean, & Pickering, 2011; Spanish: Prat-Sala & Branigan, 2000; Tzeltal: Norcliffe et al., 2015).

More specifically, more conceptually salient entities such as animate nouns tend to be mentioned earlier in a sentence than less conceptually salient ones such as inanimate nouns. They also tend to be realized with higher grammatical functions, affecting voice alternations. For example, when Spanish speakers describe the event depicting *A train ran over the woman*, the conceptually salient entity the woman tends to be mentioned earlier in the utterance, triggering a word order alteration from SVO to OVS (e.g., *A la mujer la atropello el tren*, literal translation: woman-ran over-train, 'The woman, the train ran her over') or the selection of the animate entity, the woman, as the sentential Subject, resulting in a passive voice construction (e.g., *La mujer fue* *atropellada por el tren*, literal translation: woman-be run over-by train, 'The woman was run over by the train') (Prat-Sala & Branigan, 2000).

Animacy is well-known as a factor that influences the form of utterances, but what determines relative saliency when the agent and patient entities are both animate? Here, we tested the claim that there are embodiment effects in language production – e.g. voice and word order are affected by conceptual salience/perspective, which are themselves affected by physical motion. To test this idea, we placed participants in a situation where they were the agent of a pulling motion, the patient of one, or not involved in motion, to evaluate whether this differential physical involvement affected the speakers' internal attention to actions, agents, or patients, and their subsequent production of sentences that described transitive events with two human animate participants. We first provide background about the grammar of Truku, the language used for our study, and then turn to the predictions of our experiment.

Truku language

Truku is a Formosan Austronesian language spoken in an area north-east of Puli in Central Taiwan. Truku is recognized as an endangered language, with approximately 20,000 native speakers. Importantly for this study, Truku allows both verb-object-subject (VOS) word order as shown in (1), and subject-verb-object (SVO) order, as shown in (2). The VOS order is considered the basic word order (Tsukida, 2009), while the SVO order is derived by preposing the subject to the sentence initial position. Recent experimental studies have provided empirical evidence supporting the analysis of VOS as the basic word order in Truku. For example, using a sensibility judgment task, Ono, Kim, Tang, & Koizumi (2016) have shown that Truku speakers processed VOS sentences faster than SVO. In an event-related potential (ERP) experiment, comprehending SVO sentences elicited a larger P600 effect compared to VOS ones, indicating that processing VOS sentences required a lower cognitive load compared to SVO sentences (Yano, Niikuni, Ono, Kiyama, Sato, Tang, Yasunaga, & Koizumi, 2017).

In addition to VSO versus SVO word order flexibility, Truku has a symmetrical voice system, allowing alternations in which argument is made syntactically salient. In Tagalog, another verb-initial language that has a symmetrical voice system, the argument associated with the selected voice is known to be more prominent (Sauppe, Norcliffe, Konopka, Van Valin, & Levinson, 2013; Sauppe, 2016).

Here, we consider the alternation between what we will refer to as Agent voice (AV) and Goal voice (GV). Each of these is marked morphologically on the verb, as indicated in (1) and (2). Both (1a) and (1b) are non-derived transitive events, that is, they are both unmarked and equally transitive. Either argument can be assigned as the syntactically salient element and expressed by the relevant voice (Haude & Zuniga, 2016). When Agent voice is used in transitive sentences, the Subject refers to the Agent and the Object refers to the Patient, as shown in (1a) and (2a). However, in Goal voice (GV) sentences, assignments of grammatical functions and semantic roles are switched – the Subject is now linked to the Patient and the Object to the Agent as shown in (1b) and (2b). Notice that this means that the Agent precedes the Patient in Goal voice VOS and Agent voice SVO sentences (1b, 2a), while the Patient is mentioned before the Agent in Agent voice VOS and Goal voice SVO sentences (1a, 2b).

"The girl kicks the boy"

(1a) V_{AV} $O_{Patient}$ S_{Agent}

qmqah	snaw	niyi	ka	kuyuh	niyi
kick.AV	[boy	DET]	[NOM	girl	DET]

(1b) V_{GV} O_{Agent} $S_{Patient}$

qqahan	kuyuh	niyi	ka	snaw	niyi
kick.GV	[girl	DET]	[NOM	boy	DET]

(2a) S_{Agent} V_{AV} $O_{Patient}$

kuyuh	niyi	o	qmqah	snaw	niyi
[girl	DET	FOC]	kick.AV	[boy	DET]

(2b) $S_{Patient}$ V_{GV} O_{Agent}

snaw	niyi	o	qqahan	kuyuh	niyi
[boy	DET	FOC]	kick.GV	[girl	DET]

Predictions

If physical involvement by a speaker affects the conceptual saliency they establish for elements represented in transitive events, we can make two predictions regarding verbal responses of Truku. First, we predict that irrespective of agentivity in motions, physical involvement cognitively highlights the action component of an event (versus the participants in it), making participants more likely to produce the verb as the initial element in their utterances. This predicts that motion engagement triggers more VOS responses compared to when participants are not involved in motion.

Second, if motions modulate the perspective from which participants perceive the transitive event, then their performance as the agent of a pulling motion (our *Pull-Agent* condition) may stimulate or generate a tendency to adopt the perspective of the agent for subsequent events, and thus we predict it will increase their use of the Agent voice (versus the Goal voice) in their linguistic descriptions of those events. On the other hand, the experience of being pulled (our *Pull-Patient* condition) may decrease or hinder their sense of agency, and favor the patient perspective. We predict this will increase responses using a Goal voice to describe transitive events.[2]

[2] One could also predict a preference to mention the Agent earlier in the sentence in Pull-Agent versus Pull-Patient conditions. We consider this possibility in the Discussion section.

Embodiment Experiment

By manipulating the type of motion a participant engaged in (Pull-Agent, Pull-Patient, or Static conditions, described further below), we investigated whether motion unconsciously influences how speakers of Truku perceive, encode, and describe simple transitive events.

Participants

Thirty native speakers of Truku (21 females) participated in this study. They had all been born and raised in the village. Their average age was 60.60 (SD=7.96). All participants reported normal or corrected-to-normal hearing and vision. Although the participants in the sample speak Mandarin Chinese as a second language and have some knowledge of Japanese, the individuals selected for the study were limited to ones whose dominant language is strictly Truku.

Materials

For Target stimuli, we created sixty line-drawings of transitive events in which an agent physically acts on a patient. Twenty similar line-drawings of intransitive events served as Filler stimuli. Half of the Target events represented hand-related actions (e.g., A girl pushed the woman) and another half represented non-hand related actions (e.g., A man kicked the boy) (Figure 1). This distinction between hand-involved and non-hand-involved events was created to assess if hand motions cognitively highlight an action component specifically in hand-oriented events or generally in hand and non-hand oriented events. The picture stimuli involved three female characters (i.e., girl, woman, elderly woman) and three male characters (i.e., boy, man, elderly man). Each character evenly appeared in sixty events as an agent or a patient. To minimize the possibility that the participants would be influenced by the gender of the characters, the two characters were always either two males or two females. The overall size of the agent and patient characters was drawn to be similar, and the left-right positions of the agent and the patient in all transitive events were counterbalanced throughout the six lists. Filler stimuli depicted intransitive events[3] in which a single character conducted non-hand related motions (e.g., A boy is skipping, A woman has fainted). Since these twenty filler events appeared twice, each participant encountered a total of a hundred trials including sixty different transitive events and forty intransitive events.

[3] Intransitive events are most commonly described with Agent voice (Oiwa-Bungard 2017:11, 116).

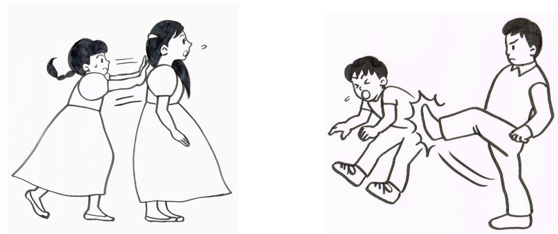

Figure 1: Hand related and non-hand related events

Experimental Design and Procedure

Participants individually took part in the experiment in a quiet room in Jingmei village, Taiwan. They sat in front of a laptop computer holding one side of a 15-inch wooden stick while an experimenter who held the other side of the stick sat across the table (Figure 2). To encourage participants to focus on the task itself, a partition was set up between the two people. The partition had a small hole so that they could hold the stick and move it through the hole. A black mouse pad with a large yellow-star was placed on the right side of the laptop computer. The experiment was programmed using Python (ver. 2.7.13) and some functions of PsychoPy (Peirce, 2007) for stimulus presentation.

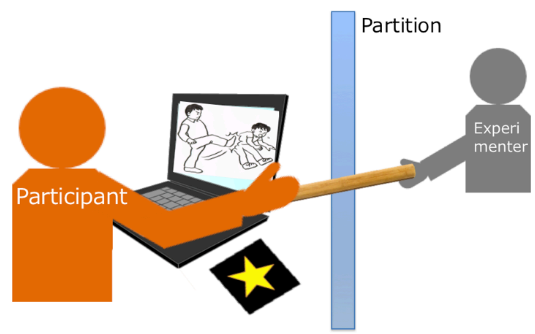

Figure 2: Experimental setup

Participants were asked to grab the stick with their right hand at the beginning of the experiment and hold it during the entire experimental period. The experiment began with a practice session followed by the main experiment. Participants performed twelve practice trials and were allowed to repeat them until they felt comfortable performing the task.

Each trial started with the screen displaying a yellow star for 3000ms. As soon as the star appeared on the screen, the participants placed and rested their right hand on the star-marked mouse pad. Then, the screen became either green or gray for 3000ms. They were instructed to pull the stick when the screen was green (Pull-Agent condition) while to do nothing and remain still when the screen was gray (Static condition). When the screen was gray (and a participant's hand was resting on the mouse pad, the participants were sometimes unexpectedly pulled by the experimenter (Pull-Patient condition). This created three types of motion manipulations (e.g., Pull-Agent, Pull-Patient, Static conditions). A cross then appeared in the center of the computer screen for 1500ms, followed by the pictures. Participants described the picture they saw as quickly and accurately as possible in a simple sentence of Truku language. The experiment then proceeds to the next trial.

The main experiment was composed of two blocks, separated by a brief break. The first block presented twenty Target trials in the Static condition, kept separate from the motion conditions to eliminate any continuing influence from motion that could have occurred in a design that mixed the three conditions. These trials provide a baseline measure of word order and voice preferences when no motion is involved. The second part of the main experiment was composed of forty Target trials, in either the Pull or Pulled condition, and forty Filler trials, in the Static condition. Trials were arranged in a different random order for each participant with the restriction that two Target trials never appeared consecutively. The verbal responses that the participants produced were recorded for subsequent transcription. The entire experiment including the practice session and the main experiment took approximately forty-five minutes.

Results

Response coding

We collected a total of 3000 verbal responses (1800 target and 1200 filler sentences) from the 30 participants. A native speaker of Truku transcribed all of the target sentences. Then two native speakers of Truku, both of whom were blind to the purpose of the present study, individually coded each target sentence for word order (i.e., VOS, SVO, or Other) and the voice morphologically marked on the verb (e.g., Agent voice or Goal voice). The two coders showed a high consistency rate (Cohen's κ = .98): in cases of inconsistency, the coders discussed and decided on a consistent single code for that production. "Other" responses included incomplete and/or ungrammatical sentences, extremely long sentences, and responses composed of multiple sentences. The total number of Other responses among all participants was 161 for the Static condition (27% of the overall 600 target trials), 230 (38%) for the Pull-Agent condition, and 225 (38%) for the Pull-Patient condition. In the following statistical analysis, we eliminated Other responses as well as the data from filler trials. The data from one participant were excluded from the analyses because half of the target responses (30 out of 60 responses) were coded as Other responses.

Statistical analysis

The sentence production data from target trials were analyzed using logistic mixed effects models with participants and items as random factors (Jaeger, 2008)[4]. In these analyses, motion conditions were treatment-coded with the Static condition as the reference level. The *R* programing language (R Core Team, 2017) and the *glmer* function within the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) were used for the analysis.

Motion on VOS vs. SVO word order

We first assessed whether motion involvement influenced the word order that participants produced. Table 1 shows the absolute frequencies (and percentages) of VOS and SVO word order responses in each condition. A logit mixed model in which SVO responses were coded as 0 and VOS responses as 1 revealed that significantly more VOS word order sentences were produced in each of the motion conditions compared to the Static condition: Pull-Agent (β = 2.15, SE = 0.25, z = 8.49, p < .001) and Pull-Patient (β = 1.80, SE = 0.24, z = 7.45, p < .001). A follow-up analysis found no significant difference between Pull-Agent and Pull-Patient conditions (p = .140). These results suggest that the relative frequencies of VOS and SVO word orders when Truku speakers participated in motions (regardless of the agentivity) significantly differ from those when they participated in no motion[5].

Table 1: Frequencies of VOS/SVO responses.

	Frequency	
Condition	VOS	SVO
Static	221 (48%)	228 (52%)
Pull-Agent	252 (69%)	112 (31%)
Pull-Patient	244 (66%)	127 (34%)

Motion on AV vs. GV in VOS responses

To assess voice selections, we conducted two statistical analyses, one for VOS and one for SVO responses since how motion interacts with the selection of the perspective may differ in these two word orders. Table 2 shows the absolute frequencies (and percentages) of AV and GV responses for VOS productions. A logit mixed model in which AV responses were coded as 0 and GV responses as 1 found effects neither for the Pull-Agent condition (β = 0.78, SE = 0.42, z = 1.83, p = .065) nor the Pull-Patient one (β = 0.74, SE = 0.42, z = 1.78, p = .076) compared to the Static condition. A follow-up analysis found no significant difference between Pull-Agent and Pull-Patient conditions (p = .926).

[4] We included only random intercepts because models with random slopes failed to converge.

[5] Because linguistic analyses of Truku are limited, we are unable to assess the baseline relative frequencies of VOS vs. SVO word orders.

Table 2: Frequencies of AV/GV responses within the VOS responses.

	Frequency	
Condition	AV	GV
Static	186 (88%)	25 (12%)
Pull-Agent	214 (85%)	38 (15%)
Pull-Patient	206 (84%)	38 (16%)

Motion on AV vs. GV in SVO responses

Table 3 shows the absolute frequencies (and percentages) of AV and GV responses within the SVO responses. A logit mixed model in which AV responses were coded as 0 and GV responses as 1 revealed that in the Pull-Patient condition, significantly more GV sentences were produced than in the Static condition ($\beta = 1.51$, $SE = 0.40$, $z = 3.77$, $p < .001$), while no significant effect was found in the Pull-Agent condition ($\beta = 0.79$, $SE = 0.43$, $z = 1.82$, $p = .068$). A follow-up analysis found no significant difference between Pull-Agent and Pull-Patient conditions ($p = .077$).

Table 3: Frequencies of AV/GV responses within the SVO responses.

	Frequency	
Condition	AV	GV
Static	207 (91%)	21 (9%)
Pull-Agent	93 (83%)	19 (17%)
Pull-Patient	93 (73%)	34 (27%)

Discussion

We investigated the notion that speakers interpret and describe the events by maintaining perspectives and saliency patterns related to the motions that they were previously engaged in. As we predicted, regardless of Agentive (pulling) or Non-agentive (being pulled) motions, involvement in motion immediately prior to describing a picture appeared to increase the saliency of action information, hence the action component of the sentence (the verb) appeared earlier in the utterances. Therefore, experiencing motion seemed to make participants significantly more likely to produce responses with a VOS word order, compared to when the participants did not experience an immediately preceding motion.[6] This result is compatible with incremental accounts in sentence production where speakers start with a sentence with an easily planned or retrieved word (Bock, 1993; Bock & Warren, 1985; Levelt, 1989; Ferreira & Slevc, 2007).

As for the SVO-word-order responses, while the Pull-Agent motion did not increase the frequency of AV responses (which seems to be due to a ceiling effect), the Pull-Patient motion, corresponding to our prediction, significantly increased GV responses. This indicates that physical motion, and more specifically the participants' role as Agent or Patient of the action, influenced the voice selection. On the other hand, in the VOS-word-order responses, we found no significant effect of motion on the frequency of AV/GV responses. In order to explain this asymmetry between the significant effect of motion in SVO responses and the lack of an effect of motion in VOS responses, let us consider another aspect of the Truku sentences, i.e., Agent/Patient argument order.

In Truku sentences, as we mentioned above, an AV-VOS sentence realizes (verb-)Patient-Agent argument order, while a GV-VOS sentence realizes (verb-)Agent-Patient order. Likewise, an AV-SVO sentence realizes Agent(-verb)-Patient order, while a GV-SVO sentence realizes Patient(-verb)-Agent order. If there is some tendency to not just use the grammatical voice that adds salience to an argument, but also place that argument earlier in linear order, the VOS word order puts the two drives in conflict, while SVO word order allows both to be satisfied simultaneously. Following this reasoning, the Pull-Patient motion should not only induce the participants to adopt a patient perspective and select GV for the verb, but also lead them to mention the patient entity as soon as possible in the utterances. As a result, responses with Goal voice and SVO word order (i.e., Patient-Agent argument order) should be especially favored, but not those with VOS order (i.e., Agent-Patient order). This is indeed the pattern we see in our results.

We found no difference between hand and non-hand events in any analysis, which is more suggestive of a general effect of speakers' sense of agency than a domain-specific one.

In sum, our results show that physical motion has an impact on how Truku speakers interpret and describe transitive events. Motions that speakers are engaged in increase salience to the action component represented in subsequent events, increasing VOS word order. Moreover, at least in the case when voice preferences are compatible with a preference for early mention of a conceptually salient argument, agentive motions align with the generally preferred adoption of an agent perspective while non-agentive motions facilitate a shift to a patient perspective, reflected in subsequent voice selection.

Acknowledgments

This research was supported by Grant-in-Aid for Young Scientists (A) (PI: Manami Sato, #JP16H05939) and Grant-in-Aid for Scientific Research (A) (PI: Masatoshi Koizumi, # JP15H02603).

[6] We acknowledge that the blocked design, in which participants were first exposed to Static trials and then Motion trials, allows for additional interpretations. We plan to conduct a follow-up study that counterbalances the order of Static and Motion blocks.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*, 1–39.

Bock, J. K., and Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47–67.

Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 453–469). Oxford: Oxford University Press.

Fox, J. (2003). Effect displays in *R* for generalised linear models. *Journal of Statistical Software*, *8*, 1–27.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, *9*, 558–565.

Haude, K., & Zuniga, F. (2016). Inverse and symmetrical voice: On languages with two transitive constructions. *Linguistics*, *54*(3), 443–481.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Levelt, W. J. M. (1989). *Speaking from intention to articulation*. Cambridge, MA: M.I.T. Press.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*, 1187–1208.

Peirce, J. W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.

Prat-Sala, M., & Branigan, H. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*, 168–182.

Oiwa-Bungard, M. (2017). Morphology and Syntax of Gerunds in Truku Seediq: A Third Function of Austronesian “Voice” Morphology. PhD dissertation, University of Hawaii, Mānoa.

Ono, H., Kim, J., Tang, A., & Koizumi, M. (2016). *VOS preference in Seediq: A sentence comprehension study*. Poster presented at the annual meeting of the Austronesian Formal Linguistics Association, Tokyo, Japan.

R Core Team (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology*, *7*, 95.

Sauppe, S., Norcliffe, E., Konopka, A. E., Van Valin Jr., R. D., & Levinson, S. C. (2013). Dependencies first: Eye tracking evidence from sentence production in Tagalog. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci 2013)* (pp. 1265–1270). Austin, TX: Cognitive Science Society.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: evidence from Japanese. *Journal of Memory and Language*, *65*, 318–330.

Tsukida, N. (2009). Sedekku-go no bumpou [A grammar of Seediq]. PhD dissertation, Tokyo University.

Yano, M., Niikuni, K., Ono, H., Kiyama, S., Sato, M., Tang, A., Yasunaga, D., & Koizumi, M. (2017). *VOS Preference in Truku sentence processing: Evidence from event-related potentials*. Poster presented at the annual meeting of the Society for the Neurobiology of Language, Baltimore, MD.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*, 1–11.