Journal of East Asian Linguistics (2020) 29:211–236
https://doi.org/10.1007/s10831-020-09205-9

Agentive versus non-agentive motions immediately influence event apprehension and description: an eye-tracking study in a VOS language

Manami Sato[1] · Keiyu Niikuni[2] · Amy J. Schafer[3] · Masatoshi Koizumi[4,5]

Received: 1 August 2018 / Accepted: 5 December 2019 / Published online: 12 March 2020

Abstract The embodied cognition hypothesis postulates that human cognition is fundamentally grounded in our experience of interacting with the physical world (Barsalou in Behav Brain Sci 22:577–609, 1999). Research has shown bi-directional associations between physical action and the processes of understanding language: language comprehension seems to activate implied visual and motor components (Zwaan and Taylor in J Exp Psychol Gen 135(1):1–11, 2006), and action behavior seems to facilitate the comprehension of associated action-language (Beilock et al. in Proc Natl Acad Sci USA 105:13269–13273, 2008). Although numerous research studies have reported a link between action and language comprehension, the exact nature of their association remains subject to debate (Chatterjee in Lang Cognit 2:79–116, 2010). Moreover, the role of action in the production of language is under-explored, as are general language production processes in Austronesian languages. The endangered Austronesian language Truku provides typological patterns that are both under-examined in psycholinguistic research and informative for questions of language production. Truku allows flexibility in the relative location of verbs versus arguments in sentence production, and uses a symmetrical voice system that marks the prominence of different participants in an event. Working with native speakers of Truku, we tested whether performing physical motions

✉ Manami Sato
manamisato@gmail.com

[1] Department of British and American Language and Culture, Okinawa International University, Okinawa, Japan

[2] Department of Clinical Psychology, Niigata Seiryo University, Niigata, Japan

[3] Department of Linguistics, University of Hawai'i, Hawaii, USA

[4] Department of Linguistics, Graduate School of Arts and Letters, Tohoku University, Sendai, Japan

[5] National Institute for Japanese Language and Linguistics, Tokyo, Japan

immediately affects the conceptual saliency of the components represented in a to-be-described event in ways that guide speakers' visual attention and shape their utterance formulation. More specifically, we investigated whether speakers' engagement as an agent or patient in a non-speech physical action affects initial eye-fixations on agent versus patient participants in a visual scene, as well as word order and grammatical voice choices in the speakers' descriptions of simple transitive events. The results revealed significant effects of physical action on the relative location and prominence of agents in subsequent sentence formulation, and on online patterns of eye fixations. These results provide further support for language-action connections in cognitive processing, and shed light on the cross-linguistic patterns of sentence production.

Keywords Embodiment · Conceptual saliency · Sentence production · Motion · Endangered language · Verb-initial language · Visual world paradigm

1 Introduction

Even when people perceive the same event, the way they describe it can vary. On seeing a child chasing an old man, for example, one person might say: *The little girl is chasing the elderly man*, while another person seeing the same event might say: *The elderly man is being chased by the girl*. What makes speakers select one expression over another? The relative order in which speakers attend to visual information may be one factor that influences how they interpret and linguistically describe events. For example, Gleitman et al. (2007) showed that an initial visual position (which they experimentally manipulated with a brief attentional cue) influences the subsequent linguistic encoding of a transitive event. That is, if speakers' eyes are initially guided to the agent, they will tend to describe the event in an active sentence and mention the agent first (e.g., *The guy kicks the boy*). But if their eyes are instead first guided to the patient, they will show an increased tendency to describe the event in a passive sentence beginning with the patient (e.g., *The boy is kicked by the guy*). In the absence of visual cues, however—and when all properties including the conceptual saliency of the agent and the patient are equivalent (Bock and Warren 1985; Prat-Sala and Branigan 2000)—what attracts the speaker's visual attention to either the agent or the patient entity?

In the present study, we test the possibility that physical motion might influence the process of event apprehension. Everyday conversations are often conducted when the interlocutors are simultaneously engaged in another physical activity. We often talk while moving our hands, such as when cooking, cleaning, exercising, and so on, and being jointly engaged in an activity with our interlocutor. Such physical activities might influence how we perceive events and the language we produce. It is known that our comprehension of action-related language activates or perhaps even relies on neural systems used in actual movements (Barsalou 1999; Glenberg and Kaschak 2002; Feldman and Narayanan 2004; Pulvermuller 2013; Zwaan and Taylor 2006). Glenberg et al. (2008) have shown a causal link between the motor activities and the comprehension of action language. In their study, participants first

continuously moved 600 beans one at a time either toward or away from the body, and then judged whether the object-transfer sentences describing toward or away directionality were sensible or not. Results showed participants were slower to respond when the direction of the sentences and previous physical motion matched, indicating that action planning and sentence processing are linked in the cognitive system. It therefore seems possible that embodied information, evoked by motor activities, could affect speakers' internal states and subsequent processes of sentence formulation.

While evidence has been accumulating for many years to support a connection between the comprehension of language about motion and actual motion production, the question of whether motion plays a role in the processes of visually apprehending and linguistically describing an event remains under-explored. Moreover, experimental research on language processing has been heavily dominated by studies on a small sample of the world's languages, leaving important typological patterns unrepresented. Here, we demonstrate that engagement in non-speech physical actions affects the relative location of initial eye fixations on an agent versus a patient character, the relative ordering of verbs versus arguments in sentence production, and the relative ordering and prominence of phrases that refer to agents. We show these effects by testing native speakers of Truku. Unlike commonly studied languages in psycholinguistic research, Truku allows verb-initial sentences. It also employs a symmetrical voice system (described further below), which differs from the active-passive voice alternation of languages such as English, and it allows notable variation in word order. More specifically, we examined how Truku speakers conceptually interpret and linguistically describe transitive events by analyzing the location of their initial eye gaze (agent character, patient character), their selection of voice (Actor voice, Goal voice, described further below), and their selection of word order (Verb-Object-Subject versus Subject-Verb-Object) in a picture-description task using a visual world eyetracking paradigm.

1.1 Language grounded in action

The embodied cognition hypothesis postulates that human cognition is fundamentally grounded in our experience of interacting with the physical world (Barsalou 1999). Much of the work investigating these postulated internal constructions of meaning have come from tests of internal visual imagery. For example, previous studies utilizing a sentence-picture verification task have repeatedly demonstrated that language comprehenders not only activate implied visual components of a mentioned object such as its shape (Stanfield and Zwaan 2001), orientation (Zwaan et al. 2002), and visibility (Yaxley and Zwaan 2007), but also incrementally update the implied shape of the object as they process new words (Sato et al. 2015). Such internal constructions of meaning also involve properties of events such as the presence or absence of perspectives modulated by a grammatical subject (Brunyé et al. 2009; Sato and Bergen 2013) and aspectual information related to event stages (e.g., resulting or ongoing; Madden and Zwaan 2003).

Moreover, a wide range of research has shown bi-directional associations between action and language comprehension. For example, previous studies have revealed that comprehending action-based language generates implied motor details such as direction of hand motion (Glenberg and Kaschak 2002), which is argued to rely on the automatic neural activation of the corresponding motor areas of the brain (e.g., Barsalou 1999; Pulvermuller 2013). Most empirical findings on action-language support stimulus-response congruency effects: sentence comprehension automatically activates representations of particular physical movements, which is reflected in reaction times (RTs) of subsequent motion tasks. For example, Glenberg and Kaschak (2002) asked participants to respond to stimuli using a response key located either nearby (which required a toward-movement of the participant's hand) or far away (which required an away-movement). They found that toward-actions were performed faster after comprehending sentences that denoted a toward-motion (e.g., *open the drawer*) while away-actions were executed faster in response to the away-motion sentences (e.g., *close the drawer*). The overlapping effects of directional information of action-language and physical responses are known as "action compatibility effects" (ACE), and they have been taken to involve a mechanism whereby understanding directional information automatically activates an associated motor area of the brain, which facilitates the execution of congruent physical movements. Conversely, accumulated physical experiences may facilitate associated action-language comprehension processes. For example, Beilock et al. (2008) demonstrated that personal experiences or interests not only change the neural regions activated, but also facilitate the comprehension of related action-language. More specifically, they showed that both fans' and players' experience with ice-hockey changed the neural region, the left premotor cortex, that is responsible for well-learned action planning, which enhanced the participants' comprehension of language about ice-hockey (e.g., *The hockey player passes with his backhand*). Beilock et al. thus demonstrated that non-linguistic behaviors influence action-language comprehension.

Although numerous research studies have reported a link between actions and language comprehension, the exact nature of language–action associations remains subject to debate (Chatterjee 2010; Miller et al. 2018). One of the primary controversies is rooted in the empirical method of previous studies, which rely heavily on behavioral reaction time (RT) data such as ACE effects. Yet RT is known to be sensitive to various factors (e.g., stimulus perception, recognition, comprehension, decision making) that may be involved in the cognitive processes that take place between the stimulus onset and response completion. Thus, it is difficult to ensure that RT is affected solely by (in)compatibility between the motor activation generated by action-language processing and the actual motor execution. Therefore, in addition to behavioral RT-based experiments, Miller et al. (2018) conducted a series of event-related potential (ERP) experiments to measure online motor activation during the processing of action-language. They investigated whether processing the meanings of verbs expressing hand or foot (effector) actions activated the motor areas that are responsible for executing the associated hand or foot actions in various paradigms (e.g., lexical decision, sensibility judgment, Stroop paradigm, and a recognition memory task). Their RT data supported

compatibility effects, indicating that effector-specific semantic processes of action verbs enhance the responses executed by the corresponding effectors. In contrast, the ERP data showed no compatibility effects. These ERP results are seemingly inconsistent with embodied cognition models because they suggest that activation of the motor areas associated with hand or foot movements is neither automatic nor necessary when comprehending or representing action meanings, and raise a challenge to the interpretation of the RT results attested in their study and previous ones. In short, while there is some kind of close association between language and action, the mechanisms underlying it require further explanation. Data from a richer array of tasks could help pinpoint the nature of the connection and its consistency across tasks.

The studies discussed above examined whether and how the comprehension of language induces action-related information. Much less research has investigated whether non-linguistic actions influence how we perceive the world and describe it in language production. Gesture studies have shown that hand motions not only facilitate children's solving of equivalence problems, but also improve knowledge maintenance in long term memory (Cook et al. 2008; Goldin-Meadow et al. 2009; Goldin-Meadow and Wagner 2005). Sato (2010) found that prior physical activity affects the subsequent process of message construction. She reported that producing toward- versus away-hand motions facilitated the construction of sentence content denoting the corresponding direction. She hypothesized that direction-oriented physical activities feed directional information to the subsequent process of constructing an internal message, and conceptually frame the relational meaning of an under-determined message. The current study extends such investigations, and focuses on the role of actions with or without the participant's agentivity on the subsequent conceptual and linguistic processing of transitive events.

1.2 Conceptual saliency in sentence production

When people linguistically describe transitive events, various factors influence how the speakers perceive and encode the events. These factors interact with each other in intriguing ways. A speaker must apprehend who-did-what-to-whom information, assign appropriate semantic roles to each entity, and align the relevant phrases in a linear order to produce an utterance (Bock and Loebell 1990; Ferreira and Slevc 2007).

Previous studies have shown that one influential feature in the selection of active/passive voice and word order when describing or recalling transitive events is conceptual accessibility (Japanese: Tanaka et al. 2011; Spanish: Prat-Sala and Branigan 2000; Tzeltal: Norcliffe et al. 2015). More specifically, conceptually more salient and more accessible entities such as animate nouns tend to be mentioned earlier in a sentence than conceptually less salient and less accessible ones such as inanimate nouns (Bock 1986; Dowty 1991). The former also tend to be realized with higher grammatical functions, affecting voice alternations (such as Active or Passive voices, and the options for Truku reviewed below). For example, when Spanish speakers describe the event that a train ran over a woman, the conceptually salient entity *the woman* tends to be mentioned earlier in the utterance. This triggers

either a word order alternation from the canonical Spanish SVO to OVS (i.e., *A la mujer la atropello el tren*, literal word-order: 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 (i.e., *La mujer fue atropellada por el tren*, literal word-order: woman-be run over-by train, 'The woman was run over by the train') (Prat-Sala and Branigan 2000). Animacy is well known to influence the form of utterances, but what determines relative saliency when the agent and patient entities are both animate? For example, in a transitive event in which a girl is kicking a boy, the participants are equally accessible on animacy grounds, and active and passive expressions (i.e., *the girl kicked the boy* and *the boy was kicked by the girl*, respectively) are both grammatical, so what makes speakers select one particular expression over another is less clear.

In this study, we tested the claim that there are embodiment effects on language production; that is, that linguistic choices such as the selection of 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 a pulling motion, or not involved in motion, and then elicited their production of sentences describing transitive events with two human participants, while also tracking their eye movements. Thus, we were able to evaluate whether this differential physical involvement affected the speakers' internal attention to actions, agents, or patients, and their subsequent linguistic choices. In the following section, we consider a sense of agency as a possible conceptual property that influences how speakers perceive an event and apprehend the relational structure of the event.

1.3 Motion and sense of agency

A sense of agency refers to the feeling or experience of controlling one's own actions (Haggard and Chambon 2012). The degree of one's sense of agency depends on volitionality and predictability; that is, people feel a stronger sense of agency when the motion is conducted easily, smoothly, and as one predicted (Chambon et al. 2014; Wenke et al. 2010). Our research explores the question of whether or not physical movements modulate a sense of agency in the time course of framing an event. Initiating intended actions might stimulate a person's sense of agency, changing the subsequent conceptual processes of event apprehension, thereby influencing linguistic behaviors. More specifically, we hypothesized that engaging in a pulling motion (prior to event perception) would enhance the saliency of the concept of agency in the cognitive system, which would increase (1) attention to finding the agent character in the visual scene, (2) the level of activation of the agent in the mental model of the event depicted in the visual scene and (3) the level of activation of the agent thematic role when linguistically encoding the event depicted in the visual scene. We further hypothesized that each of these enhancements could in turn increase the likelihood of shifting or maintaining eye fixations to an agent character. In the linguistic encoding process, we hypothesized that increased conceptual salience of the agent thematic role would increase the likelihood of selecting an agent perspective to frame the transitive event. This, we reasoned,

would then lead to increased selection of the Agent voice, which makes the agent argument grammatically prominent and supports an agent point-of-view (see Sect. 1.5 for description of the Truku voice system). Likewise, when speakers are being pulled, the passive motor experience may reduce their sense of agency, and reduce the prominence of the agent in the scene and the adoption of an agent perspective. This may in turn decrease the likelihood of looks to the agent character (while increasing those to the patient character) and decrease the selection of Agent voice. To explore these possibilities, we tracked speakers' eye movements in order to examine whether motion influences how speakers conceptually process event representations and recorded their spontaneous descriptions of depicted events.

1.4 Eye-movement and sentence production

Language production experiments using the visual-world paradigm provide fine-grained temporal data that show the order in which speakers access information during their message and utterance formulation (Bock et al. 2004). In the version of the paradigm used here, speakers' eye movements were recorded from the point at which they first viewed a scene to the point when they finished describing the depicted event linguistically. Because people tend to look at the entity they are thinking about and will talk about, data from eye fixations can be used as an indicator of how people access entities represented in an event, identify their relationship, and construct a sentence structure.

The rapid extraction of the gist of a depicted event (e.g., identifying event categories and comprehending the entities represented in the event, their animacy features, and their relationships or roles) occurs within 400 ms of the onset of a picture. This apprehension period (or conceptual encoding period) frames the subsequent process of utterance formulation (Bock et al. 2004; Griffin and Bock 2000). The time window of eye fixations subsequent to the first 400 ms is known to reflect linguistic encoding processes; that is, the patterns of eye fixations reflect the order in which the entities will be uttered. In a test of the verb-initial language Tzeltal, which allows VOS and SVO word orders, Norcliffe et al. (2015) demonstrated that the patterns of eye fixations during conceptual encoding and linguistic encoding periods differ between VOS and SVO sentence production.

The exact relationship between initial eye fixations and the processes involved in sentence formulation is still under debate. One proposal is that attention to a character directly induces lexical encoding, and therefore the character that draws the initial attention is assigned to be the subject. For example, by manipulating speakers' initial eye gaze with brief visual cues (60–75 ms), Gleitman et al. (2007) demonstrated that the character to which the eye is directed first affects which element speakers start their sentence with, as well as their structural choices (active vs. passive sentences). In a depicted event such as a guy kicking a boy, initial fixations directed to a picture of "the guy" led to increased production of an active sentence such as *The guy is kicking the boy*, while those directed to a picture of "the boy" generated an increased rate of production of passive structures such as *The boy is being kicked by the guy.* In other words, perceptual saliency determines the subsequent linguistic encoding.

On the other hand, some studies have suggested that initial attention to a character reflects not perceptual saliency, but the consequences of advance planning of the relational structure of an event and the generation of a structural framework for an upcoming sentence (Konopka and Mayer 2014; Norcliffe et al. 2015). On this view, during the early event apprehension phase, speakers generate a structural frame, which in turn directs their eyes to a particular character. And during a later phase, lexical retrieval takes place so that their eyes are guided to the character they will mention first (Bock et al. 2003; Griffin and Bock 2000; Norcliffe et al. 2015). If initial fixations are not driven by low-level perceptual properties, then low-level perceptual properties may be subordinate to conceptual properties in event apprehension processes. Truku, the language used for our study, allows word order alternations (VOS and SVO) as well as voice alternations (Agent voice and Goal voice), making it ideal for disentangling our understanding of the initial fixations either as determining the first element in the utterance or as a consequence of framing the event from a specific perspective. In the next section, we provide background about the grammar of Truku, and then turn to the predictions of our experiment.

1.5 The 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, and a critical shift toward Mandarin has been observed among its approximately 20,000 native speakers. Tang (2011) assessed Truku linguistic proficiency across age groups between 10 to 65 and found dramatic intergenerational attrition of Truku. More specifically, native speakers of Truku are gradually losing phonological and morphosyntactic aspects of the language.

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). VOS is considered the basic word order (Tsukida 2009), while the SVO order is derived by preposing the subject to the sentence-initial position. As shown in examples (2a) and (2b), this preposed subject is marked with a topic marker *o*, which syntactically expresses which entity holds pragmatic attention/focus (i.e., topicalization) (Tang 2011). Recent experimental studies have provided empirical evidence supporting this analysis. For example, using a sensibility judgment task, Ono et al. (2016) showed that Truku speakers process VOS sentences faster than SVO sentences. In an event-related potential (ERP) experiment, comprehending VOS sentences elicited a smaller P600 effect compared to SVO ones, indicating that processing VOS sentences created a lower cognitive load compared to SVO sentences (Yano et al. 2017).

Truku also has a symmetrical voice system, allowing alternations in which arguments are made syntactically salient. In symmetrical voice languages such as Truku and Tagalog, the selected voice is signaled by an overtly marked voice marker and the argument associated with the selected voice becomes the pivot or prominent argument in the event (Sauppe 2016, 2017; Sauppe et al. 2013). All voices are equally marked so that there is no default voice, which may not be the

case in languages with asymmetrical voice like English (Foley 2008) (e.g., the active voice is the default form while the passive voice is derived).

Here, using the simple transitive sentence of "The girl kicks the boy" in Truku, we consider the alternation between what we will refer to as Agent voice (AV) and Goal voice (GV). Both Agent and Goal voice constructions are equally transitive (1a and 1b, 2a and 2b); constructions with Goal voice are not derived from those with Agent voice. Each of these voices is marked morphologically on the verb, as indicated in (1) and (2). When the Subject refers to the Agent (A) and the Object refers to the Patient (P), the verb is marked as Agent voice in transitive sentences, as shown in (1a) and (2a). However, in Goal voice (GV) sentences, the assignment of grammatical functions and semantic roles is 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 following examples of (1a) to (2b) represent "The girl kicks the boy."

(1a) Agent voice (AV), VOS, Verb-patient-agent (VPA)

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

(1b) Goal voice (GV), VOS, Verb-agent-patient (VAP)

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

(2a) Agent voice (AV), SVO, Agent-verb-patient (AVP)

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

(2b) Goal voice (GV), SVO, Patient-verb-agent (PVA)

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

2 The study

2.1 Purpose of the study

Although some factors that affect message formulation (e.g., perceptual saliency, conceptual accessibility of referents in the event based on animacy and imaginability) have been widely studied, extra-linguistic factors existing prior to event

apprehension have received little attention. Because we cannot be detached from the physical world, however, previous actions may play a role in how we perceive and select the parts of an event to be highlighted and how we construct the relational structure of the event to be described. In this study, we explore three specific questions based on sentence production and eye-movement patterns. First, does the cognitive state of being in control of one's own action (i.e., a sense of agentivity) influence the relational structure of perceived events and thereby predict the selection of Agent-perspective or Goal-perspective sentences in Truku? Second, can motions determine the word order (VOS vs. SVO) of subsequently produced sentences? Third, do motions executed prior to event perception modulate how speakers apprehend the event, and if so, is this reflected in their initial eye-fixation patterns?

2.2 Predictions

If the physical involvement of a speaker affects the conceptual saliency of elements represented in transitive events for that speaker, we can make two predictions regarding verbal production in Truku and one prediction regarding eye fixations.

First, physical involvement (i.e., motion) may cognitively highlight the action component of an event (as opposed to the participant components), making speakers more likely to produce the verb as the initial element of their utterances, regardless of speaker agentivity in the motion. If this is the case, we predict that engaging in motion will trigger more VOS responses compared to when the speakers do not engage in motion.

Second, if motions modulate the perspective from which the speakers perceive the transitive event, then executing a pulling motion (i.e., the *Pull-Agent* condition) should increase the speakers' sense of agency and stimulate their tendency to adopt the agent perspective when they perceive subsequent events. We therefore predict that the Pull-Agent condition will increase the speakers' use of the Agent voice (versus the Goal voice) in their description of transitive events. On the other hand, an experience of being pulled (i.e., the *Pull-Patient* condition) may decrease or reduce the speakers' sense of agency, leading them to prefer the patient perspective. We therefore predict that the Pull-Patient condition will increase the speakers' use of the Goal voice to describe transitive events.[1]

Third, early gaze fixations should reflect the conceptual formulation processes that determine the relational structure of speakers' event descriptions. Therefore, if conducting the pulling action (i.e., the *Pull-Agent* condition) increases speakers' sense of agency, we predict that this condition will guide the speakers' eyes first to the agent character in the event. On the other hand, if the experience of being pulled (i.e., the *Pull-Patient* condition) decreases the speakers' sense of agency, we predict that this condition will direct their attention to the patient character first.[2]

[1] Another possible prediction is that speakers would prefer to mention the Agent earlier in the Pull-Agent condition than in the Pull-Patient condition. We consider this possibility in the Discussion section.

[2] Subsequent to the conceptual formulation process, in the period when the speaker is uttering the sentence, we anticipated that gaze would correspond to the order of mention of characters in the unfolding production.

3 Embodiment experiment

By manipulating the type of motion the participants 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.

3.1 Participants

Thirty native speakers of Truku (21 female) participated in this study in exchange for a culturally appropriate amount of monetary compensation based on the National Dong Hwa University standard. They had all been born and raised in Jingmei village, Taiwan, which is where the study was conducted. Their average age was 60.60 (*SD* = 7.96). All participants were right-handed and reported normal or corrected-to-normal hearing and vision. Although the participants spoke Mandarin Chinese as a second language and had some knowledge of Japanese, they were all fluent in Truku and had self-reported that their dominant language was Truku. Approval for the study was obtained from the Ethics Committee of the Graduate School of Arts and Letters, Tohoku University, Japan.

3.2 Materials

For the 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. These were combined in six lists. Half of the target events represented hand-related actions (e.g., a girl pushing a woman) and the other half represented non-hand related actions (e.g., a man kicking a boy). Figure 1 presents examples. The picture stimuli involved three female characters (i.e., girl, woman, elderly woman) and three male characters (i.e., boy, man, elderly man). All of the characters appeared in an equal number of events (sixty each) as an agent or a patient. People tend to engage with a particular character when interpreting depicted events (i.e., empathetic projection; Decety and Sommerville 2003; Lamm et al. 2007). Therefore, to minimize the possibility that the participants would be influenced by the gender of the characters, the 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 skipping, a woman fainting). Because these twenty filler events appeared twice, each participant encountered a total of one 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).

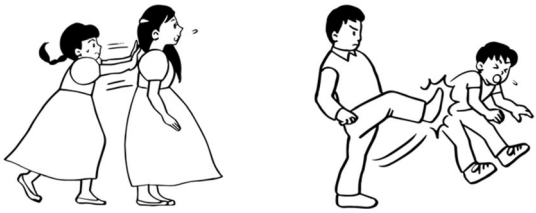

Fig. 1 Hand related and non-hand related events

3.3 Experimental design and procedure

Participants individually took part in the experiment in a quiet room in the village. We used a Tobii X3-120 eye tracker (120 Hz sampling frequency) to track their eye movements during the experiment. Participants sat in front of a laptop computer holding one end of a 15-inch (i.e., 38 cm) wooden stick with their right hand while an experimenter who held the other end of the stick sat across the table (Fig. 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 through which the stick extended on each side. 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.

The experiment began with a practice session. Participants performed twelve practice trials involving all three action conditions, which they were allowed to repeat until they felt comfortable performing the task.

As shown in Fig. 3, each trial started with the screen displaying a yellow star for 3000 ms. As soon as the star appeared on the screen, the participants placed and rested their right hand (still holding the stick) on the star-marked mouse pad. Then, the screen became either green or gray for 3000 ms. They were instructed to pull the stick when the screen was green, and to do nothing and remain still when the screen was gray. When the screen was gray, the participants were sometimes pulled by the experimenter. This created three types of motion manipulations (e.g., Pull-Agent, Pull-Patient, and Static conditions). We used screen color to signal the action instead of action words (e.g., *pull*, *remain still*) to ensure that any motor information

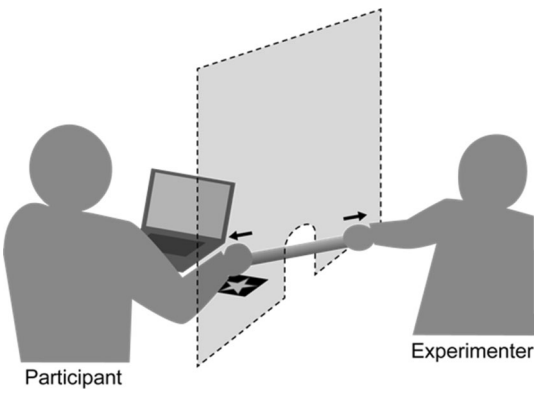

Fig. 2 Experimental setting

came from non-linguistic motor activities rather than from comprehending the action words (Barsalou 1999). A fixation cross then appeared in the center of the computer screen for 1500 ms, followed by a picture. Participants had been instructed to describe each picture they saw as quickly and accurately as possible in a simple sentence in the Truku language. After the participant completed an utterance, the experiment proceeded to the next trial. Participants' verbal responses were recorded by a TASCAM DR-40 voice recorder for subsequent transcription.

The main experiment was composed of two parts, separated by a brief break. In the first part, participants completed twenty target trials in the Static condition, which was 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 also were intended to provide a baseline measure of word order and voice preferences when no motion was involved. The second part of the main experiment was composed of forty target trials, in either the Pull-Agent or Pull-Patient 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 experimenter could not see the participant or the computer screen, but did see a green, red, or yellow light that signaled the condition, so that the experimenter could create or allow the appropriate motion (i.e., green: participant's pulling, red: experimenter's pulling, or yellow: remaining still). The participants could not see these lights, so the participants did not know when they would be pulled by the experimenter. In order to make the cognitive load comparable in all three conditions (Pull-Agent, Pull-Patient, Static), the experimenter gave instructions prior to the practice session although the twenty trials in the first part were all in the Static condition. Therefore, during this first part, the participants may have expected but did not experience pulling and being pulled motions. The entire experiment including the practice session and the main experiment took approximately forty-five minutes.

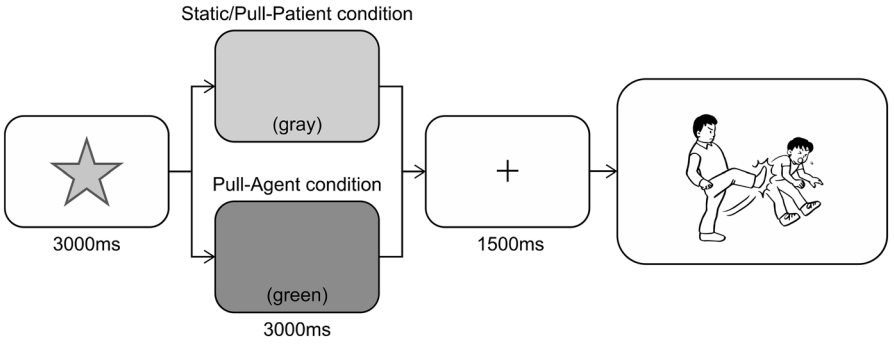

Fig. 3 Flow of the trial

3.4 Results: sentence production

3.4.1 Response coding

We collected a total of 3000 verbal responses (1800 target and 1200 filler) from the thirty participants. A native speaker of Truku transcribed all of the target sentences. Two native speakers of Truku who did not know the purpose of the study individually coded each target sentence for word order (i.e., VOS, SVO, or Other) and for the voice morphologically marked on the verb (i.e., Agent voice or Goal voice). Other responses included the lack of a verbal response, incomplete and/or ungrammatical sentences, semantically inappropriate sentences, extremely long sentences, and responses composed of multiple verbs (Table 1). The total number of Other responses among all participants was 161 for the Static condition (27% of 600 target trials), 230 (38%) for the Pull-Agent condition, and 225 (38%) for the Pull-Patient condition. In the following statistical analysis, following standard practice in sentence production research, we eliminated Other responses as well as the data from filler trials. The distribution of the remaining responses across the three Motion conditions is visualized in Fig. 4.

3.5 Statistical analysis

The sentence production data from the target trials were analyzed using logistic mixed effects models with participants and items as random factors (Jaeger 2008). In these analyses, motion conditions were treatment-coded with the Static condition as the reference level. We conducted backward model comparisons from the maximal model for the design and included random slopes for fixed factors only if they improved model fit at $p < .20$. The R programming language (R Core Team 2017) and the glmer function within the lme4 package (Bates et al. 2015) were used for the analysis.

3.6 Motion and VOS versus SVO word order

We first assessed whether motion involvement influenced the word order that participants produced. Table 2 shows the absolute frequencies (and percentages) of VOS and SVO word order responses in each condition. A logistic regression mixed

Table 1 Distributions of Other responses

Condition	Distribution of other responses					
	No response	Incomplete	Ungrammatical	Semantically inappropriate	Long	Multiple verbs
Static	12 (7.5%)	30 (18.6%)	1 (0.6%)	75 (46.6%)	18 (11.2%)	25 (15.5%)
Pull-Agent	63 (27.4%)	46 (20.0%)	4 (1.7%)	88 (38.3%)	7 (3.0%)	22 (9.6%)
Pull-Patient	51 (22.7%)	41 (18.2%)	4 (1.8%)	103 (45.8%)	4 (1.8%)	22 (9.8%)

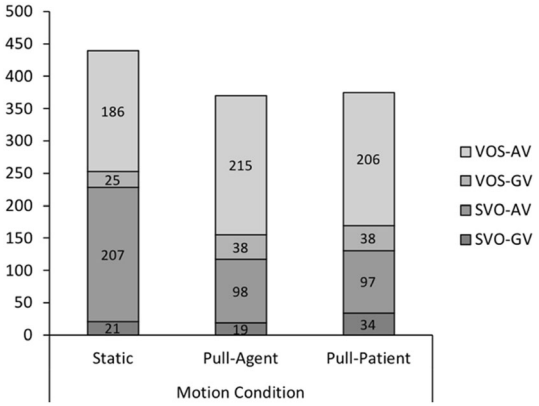

Fig. 4 Counts of production response types for each experimental condition

Table 2 Frequencies of VOS/SVO responses

Condition	Frequency	
	VOS	SVO
Static	221 (48%)	228 (52%)
Pull-Agent	253 (68%)	117 (32%)
Pull-Patient	244 (65%)	131 (35%)

Table 3 Parameters of the logistic mixed effects regression model for the analysis of VOS/SVO responses

	β	*SE*	*z*
(Intercept)	− 0.28	0.77	− 0.37
Motion: pull-agent	2.13	0.41	5.22***
Motion: pull-patient	1.96	0.44	4.48***

***$p < .001$, **$p < .01$, *$p < .05$

effects model in which SVO responses were coded as 0 and VOS responses were coded as 1 revealed that significantly more VOS word order sentences were produced in the two motion conditions compared to the Static condition[4] (Table 3). These results suggest that motion generates a high verb saliency, resulting in an increase of VOS word order.

3.7 Motion and AV versus GV in VOS responses

How motion interacts with the selection of perspective may differ by word order. (Recall that in VOS sentences the argument that matches the voice selection is mentioned in the S position, after the object, while in SVO sentences it is mentioned

[4] As shown in Table 2, the intercept was not significant, indicating that the proportion of VOS/SVO responses in the Static condition is not statistically different from chance.

Table 4 Frequencies of AV/GV responses within the VOS responses

Condition	Frequency	
	AV	GV
Static	186 (88%)	25 (12%)
Pull-agent	215 (85%)	38 (15%)
Pull-patient	206 (84%)	38 (16%)

Table 5 Parameters of the logistic mixed effects regression model for the analysis of AV/GV responses within the VOS responses

	β	*SE*	*z*
(Intercept)	− 3.53	0.64	− 5.53***
Motion: pull-agent	0.77	0.42	1.83
Motion: pull-patient	0.75	0.42	1.78

***$p < .001$, **$p < .01$, *$p < .05$

first.) Therefore, to assess voice selection, we conducted two statistical analyses, one for VOS responses and one for SVO responses. Table 4 shows the absolute frequencies (and percentages) of Agent voice (AV) and Goal voice (GV) responses for VOS productions. A logistic regression mixed effects model in which AV responses were coded as 0 and GV responses were coded as 1 found no effect for either the Pull-Agent condition or the Pull-Patient condition compared to the Static condition (Table 5). There was, however, a significant overall preference for Agent voice.

3.8 Motion and AV versus GV in SVO responses

Table 6 shows the absolute frequencies (and percentages) of AV and GV responses within the SVO responses. A logistic regression mixed effects model in which AV responses were coded as 0 and GV responses were coded as 1 found that significantly more GV sentences were produced in the Pull-Patient condition than in the Static condition, but found no significant effect of the Pull-Agent condition (Table 7). Once again, there was a significant overall preference for Agent voice.

3.9 Discussion: sentence production results

We investigated the notion that the way speakers interpret and describe events is affected by perspectives and saliency related to motions that they were previously engaged in. As we predicted, regardless of motion agency (pulling or being pulled), 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 following motions. Therefore, experiencing motion seemed to make participants significantly more likely to

Table 6 Frequencies of AV/GV responses within the SVO responses

Condition	Frequency	
	AV	GV
Static	207 (91%)	21 (9%)
Pull-agent	98 (84%)	19 (16%)
Pull-patient	97 (74%)	34 (26%)

Table 7 Parameters of the logistic mixed effects regression model for the analysis of AV/GV responses within the SVO responses

	β	*SE*	*z*
(Intercept)	− 3.32	0.61	− 5.40***
Motion: Pull-agent	0.77	0.43	1.77
Motion: Pull-patient	1.49	0.40	3.72***

***$p < .001$, **$p < .01$, *$p < .05$

produce responses with a VOS word order, compared to when the participants did not experience an immediately preceding motion.[5] This result is compatible with incremental accounts of sentence production that suggest speakers prefer to start sentences with an easily planned or retrieved word (Bock and Warren 1985; Ferreira and Slevc 2007; Levelt 1989).

Regarding motion effects on voice selection, we predicted that a pulling motion would highlight the agent role, leading to sentences in which the verbs were marked with Agent voice. On the other hand, when a being-pulled motion highlights the patient role, we predicted the verb would be marked with Goal voice. 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 significantly increased GV responses, in line with our prediction. This indicates that physical motion, and more specifically the participants' role as agent or patient of the action, influenced their voice selection, at least to some extent. 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 Truku sentences: Agent/Patient argument order.

In Truku sentences, as mentioned above, an AV-VOS sentence realizes a (verb-) Patient-Agent argument order (Agent last), while a GV-VOS sentence realizes a (verb-)Agent-Patient order (Patient last). However, an AV-SVO sentence realizes an Agent(-verb)-Patient order (Agent first), while a GV-SVO sentence realizes a Patient(-verb)-Agent order (Patient first). If there is some tendency to use the

[5] We acknowledge that the experimental design in which participants completed first the Static trials and then the Motion trials allows for other interpretations of these results. We plan to conduct a follow-up study that counterbalances the order of the Static and Motion trials.

grammatical voice that adds salience to an argument, and also a tendency to place that argument earlier in linear order, the VOS word order would create conflict between these tendencies, while the 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 their utterances. As a result, Goal voice responses with SVO word order (i.e., Patient-Agent argument order) should be favored, but those with VOS order (i.e., Agent-Patient order) should not. This is indeed the pattern we see in our results.

In sum, our results show that physical motion has an impact on how Truku speakers interpret and describe transitive events. Motions that speakers engage in increase the salience of the action component represented in subsequent events, increasing the production of sentences with the 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, as reflected in subsequent voice selection. We also found a very strong overall preference for Agent voice.

3.10 Results: eye-tracking

3.10.1 Statistical analysis

The time window (0–500 ms) for the eye-tracking data analyses was selected based on Griffin and Bock's (2000) proposal of 0–400 ms as the time window for an event apprehension period (cf. Norcliffe et al. 2015; Sauppe et al. 2013). Because our participants were somewhat elderly and not familiar with computer manipulations or experimental tasks and settings, and their speech onset was relatively slow compared to that of younger participants in previous studies,[6] we extended the time window from 0–400 ms to 0–500 ms (Norcliffe et al. 2015). We analyzed the looks to each entity (agent/patient) in a picture for 0–500 ms after the picture onset to assess (a) if motions modulate speakers' initial eye gazes and (b) if the first entity that speakers look at is incrementally encoded and mentioned in the utterance, as Gleitman et al. (2007) reported.

For the statistical analysis, the gaze data were aggregated into 100 ms time bins and then analyzed using grouped logistic mixed effects models (Donnelly and Verkuilen 2017) with participants and items as random factors.[7] The data from one participant were excluded from the analyses due to excessive amounts of track loss. In addition, trials with greater than 30% track loss out of all samples for the

[6] Average speech onset time for overall target trials was 2424 ms in our study (SD by participant = 687 ms; Mean age of participants = 60.6), while it was 1738 ms (for a grand mean of four sentence types; Mean age of participants = 28) in Norcliffe et al. (2015).

[7] The R code of the analyses was as follows (cf. Donnelly and Verkuilen 2017: 41):
glmer(cbind(C + 0.1, N – C + 0.1) ~ fixed effects…
C: sum of looks to agent/patient AOI
N - C: sum of looks not to agent/patient AOI.

Table 8 Parameters of the logistic mixed effects regression models for the first analysis

	β	*SE*	*z*
AOI: agent			
(Intercept)	− 0.63	0.14	− 4.49***
Time	0.29	0.07	4.28***
Motion1: Static versus others	− 0.33	0.30	− 1.10
Motion2: Pull-agent versus pull-patient	− 0.10	0.18	− 0.54
Time × Motion1	0.44	0.04	11.84***
Time × Motion2	− 0.14	0.02	− 6.64***
AOI: Patient			
(Intercept)	− 1.24	0.14	− 8.73***
Time	0.51	0.05	9.80***
Motion1: Static versus others	0.16	0.24	0.66
Motion2: Pull-agent versus pull-patient	0.11	0.14	0.81
Time × Motion1	− 0.15	0.18	− 0.82
Time × Motion2	0.08	0.09	0.87

***$p < .001$, **$p < .01$, *$p < .05$

0–500 ms time window were excluded from the data (19.8% of the data). We conducted backward model comparisons and included random slopes for fixed factors only if they improved model fit at $p < .20$. The following predictors and interaction terms were included in the mixed models: Time, Motion Condition, and Time × Motion Condition in the first analysis (Table 8) and Time, Agent/Patient Order in verbal responses, and Time × Agent/Patient Order in the second analysis (Table 9). In the second analysis, trials in which the verbal response was coded as Other were eliminated from the data.

In the first analysis, motion conditions were Helmert-coded (Static condition = [1/3, 0]; Pull-Agent condition = [− 1/6, − 1/2], Pull-Patient condition = [− 1/6, 1/2]), so that first the Static condition is compared to the two motion conditions, and then the two motion conditions are compared to each other. In the second analysis, the verbal responses were deviation-coded. The time variable (5 time bins) was standardized (to z-scores) in both analyses.

The analysis was performed with R programming language (R Core Team 2017), using the eyetracking R package (Dink and Ferguson 2015) and the lme4 package (Bates et al. 2015). The eyetrackingR package was also used for creating Figs. 5 and 6.

(a) Effects of physical motion on event apprehension

Figure 5 shows the proportion of looks to the agent/patient objects for each motion condition (0–2000 ms after the picture onset), and Table 8 shows the parameters of the logistic mixed effects models. As Fig. 5 shows, we observed a clear tendency, from approximately 250 ms after the picture onset, for participants to look more at the agent object in the Pull-Agent condition than in the Pull-Patient

Table 9 Parameters of the mixed effects regression models for the second analysis

	β	*SE*	*z*
AOI: Agent			
(Intercept)	− 0.74	0.20	− 3.71***
Time	0.35	0.12	3.05**
Agent/patient order	0.12	0.30	0.39
Time × agent/patient order	− 0.06	0.22	− 0.26
AOI: patient			
(Intercept)	− 1.86	0.22	− 8.30
Time	0.98	0.14	7.08***
Agent/patient order	0.42	0.26	1.62
Time × agent/patient order	0.01	0.21	0.05

***$p < .001$, **$p < .01$, *$p < .05$

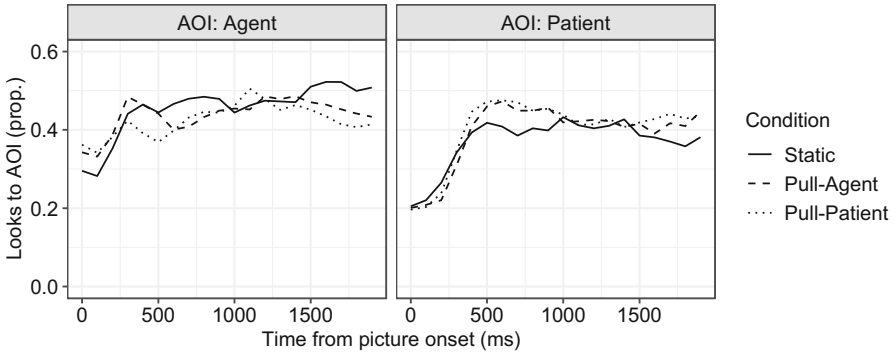

Fig. 5 Proportion of looks to the agent/patient AOI (area of interest) for each motion condition

condition; this resulted in a significant interaction of Motion Condition (Pull-Agent vs. Pull-Patient) and Time (see Table 8). These results correspond to our prediction that motion with agentivity would arouse a sense of agency and thus initially direct a speaker's eyes to the agent entity.

As shown in Fig. 5, it appears that participants looked more at the patient character in the Pull-Patient condition than in the Pull-Agent condition after 250 ms. The mixed effects model, however, found neither a significant interaction of Motion Condition and Time nor a significant main effect of Motion Condition. Therefore, we do not see significant support in data from fixations to the Patient AOI for our prediction that motion without agentivity would make speakers adopt the patient perspective in apprehending an event and thus initially show significantly increased looks to the patient entity.

The effect on fixations to the patient might be weakened by the strong overall agent bias in the early region. We believe that the early bias (starting from the onset of the depicted scene) for looks to the Agent AOI over the Patient AOI is due to an

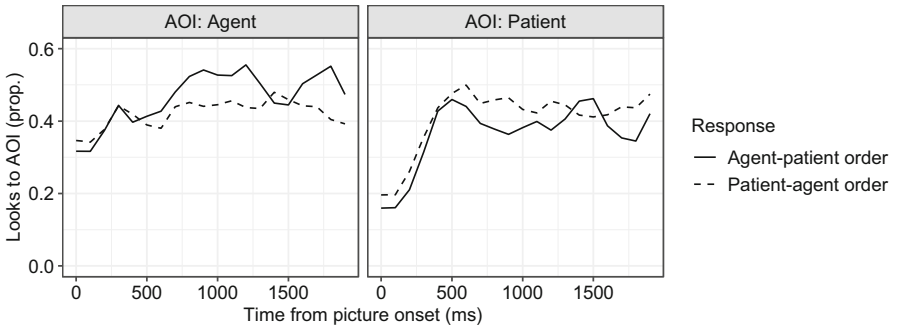

Fig. 6 Proportion of looks to the agent/patient AOI in agent-patient (i.e., AV-SVO and GV-VOS) and patient-agent (AV-VOS and GV-SVO) order sentences

artifact of an unintended relationship between the fixation cross location and the location of the effector portion of the Agent. For verbs like "kick", "punch", and "choke", the AOI for the Agent included the central portion of the screen that had previously displayed the fixation cross. However, because the design was counterbalanced, this overlap was evenly distributed across motion conditions.

(b) Initial looks for verbal responses with agent-patient or patient-agent order

Figure 6 shows the proportion of looks to the agent/patient objects for agent-patient order responses (i.e., AV-SVO and GV-VOS responses) and patient-agent order responses (i.e., AV-VOS and GV-SVO responses), to assess the relationship between the speakers' early gaze patterns and the argument order in the subsequent productions. Table 9 shows the results of the mixed effects model analysis. As shown in Table 9, we found neither a significant effect of Agent/Patient Order nor a significant interaction between Agent/Patient Order and Time. In the later portion of the trial (i.e., after 500 ms from the picture onset), participants looked at the agent or the patient entity as they linguistically mentioned it, replicating the findings reviewed above. For example, the fixation patterns show that participants more often looked at the agent character first and the patient character next when they went on to produce a sentence with an agent-patient order compared to how often they looked at the entities in this order when they went on to produce a sentence with a patient-agent order. Conversely, they clearly looked more often at the patient character first and at the agent character second when they went on to produce a sentence with a patient-agent order compared to when they went on to produce a sentence with an agent-patient order. These results show that the linguistic encoding of events is a temporally dynamic process (Bock et al. 2004).

3.11 Discussion: eye-movement data

We conducted two analyses to examine how non-linguistic motions influence the order of information that speakers' eyes are directed to and whether or not the initially attended entity is retrieved first and mentioned first in the utterance. First, if

motions with or without agentivity change degree of sense of agency, pulling motions should direct speakers' eyes to the agent character during the early time window while being-pulled motions should guide their eyes to the patient character. As we predicted, pulling motions generated more looks to the agent compared to being-pulled motions in the time window associated with event apprehension. Being-pulled motions, however, did not increase looks to the patient. The finding of agent-motion effects suggests that volitional motions increased speakers' sense of agency and made them comprehend the event from the agent perspective. The absence of an effect of the non-agent motion may be related to the nature of the experience of being pulled. That is, the being-pulled motion might be able to reduce one's sense of agency, but this does not necessarily make speakers adopt a patient's perspective. Although it is known that grammatical subjects can flexibly determine which perspective people engage with when interpreting events (Brunyé et al. 2009), experiencing agentivity in a motion might not have the same effect as comprehending a grammatical subject. This may suggest non-equivalent natures or functions of language versus motion in the process of perspective adoption. Moreover, the strong bias in the early window to look at the agent (evident in Figs. 4 and 5 by comparing the two panels within each figure) may have led to competition against looks to the patient that suppressed an agency effect.

Another plausible interpretation for the absence of being-pulled motion effects on the patient relies on how the apprehension process of the event is framed. That is, the event apprehension process may not be about paying attention to either the agent or the patient figures, but about detecting the agent against everything else.[8] In fact, during the phase of event apprehension where speakers extract the gist of an event (Griffin and Bock 2000), they can detect the agent depicted in the picture even when it is presented for a very short time (37 ms) (Hafri et al. 2012). Moreover, people generally allocate more visual attention to the agent presented in the scene due to a stronger cognitive bias toward the agent. Agents who initiate the described events/actions are cognitively more prominent and provide more information about the event than patients, and thus play a role as cognitive attractors in apprehending the event (Cohn and Paczynski 2013; Sauppe 2016). The strong preference for agents over patients in event apprehension may explain why no motion effect on the visual attention to the patient was observed.

We also examined whether the element mentioned first was correlated with what was fixated first. If speakers incrementally encode the initially accessed element in the subsequent utterance (Gleitman et al. 2007), verbal responses with the agent-patient order would show the agent character to be the initially attended element in the first 500 ms. Similarly, verbal responses with the patient-agent order would be accompanied by initial eye fixations on the patient character. Our results, however, show that an initial gaze position to either the agent or the patient character did not correspond to subsequent encoding of agent-patient word order versus patient-agent word order. Importantly, though, this analysis was based on productions for which VOS order dominated the responses. In VOS productions, the linear order and the voice marking lead to opposite predictions for which character will receive the most

[8] We thank an anonymous reviewer for suggesting this possibility.

fixations in the first 500 ms. The VOS responses in the Motion conditions were predominately expressed with Agent voice, and there was indeed a strong preference to look at the agent in the early time window. Because voice is linguistically encoded on the verb as well as through the morphological marking on the subsequent noun phrases, an early effect of voice is consistent with linguistic encoding necessary for the production of the verb. To summarize the effects on linguistic encoding: in their verbal descriptions, the speakers did not show a tendency to mention first whichever character they had looked at first in these data, but they did show a strong association between the voice used in the productions and the character that was looked at first.[9] These patterns are consistent with Sauppe et al. (2013)'s findings for an effect of voice and conceptual planning in Tagalog during an early time window.

More generally, since voice is marked morphologically on the verb and Agent voice was also the preferred voice, the results are compatible with both a strong effect of initial visual attention on formulation choices (i.e., to use Agent voice), and also with an effect of the general bias for Agent voice to influence what gets fixated first (i.e., the Agent). Further research will be necessary to tease these apart.

4 General discussion

The primary theoretical contribution of this study is to provide empirical evidence that non-linguistic, physical motion has a causal impact on subsequent event apprehension, thereby affecting linguistic behaviors. Regarding motion effects on subsequent linguistic encodings, we argue that the Truku speakers unconsciously incorporated an extra-linguistic factor, a sense of agency driven by motions, and exhibited motion effects on their selection of word order (VSO vs. SVO) as well as voice (Agent voice vs. Goal voice). The first of these effects emerged robustly; the second was less reliable. Regarding motion effects on eye fixations, we provide online empirical evidence that further supports the claim that motor experiences influence subsequent conceptual activities in a rather automatic fashion. Taken together, our main findings—that motions influence the event encoding process and subsequent utterances (i.e., word order and voice selection), and that the being the agent of a pulling motion enhances the preference for early looks to the agent—lead us to conclude that non-linguistic motor information can highlight an entity in a speaker's perceiving, framing, and encoding of an event.

Further research on typologically different languages such as Japanese will clarify whether action saliency driven by motor activities is a universal property or a unique property for Truku speakers, whose basic word order is VOS. In Truku, verb-initial utterances might be facilitated by motion involvement because VOS speakers' cognitive system is habitually tuned to the action component. Japanese has a basic word order of SOV (the mirror image of Truku word order), making it a

[9] Due to few GV productions in the Motion conditions (i.e., VOS word order in Pull-Agent vs. Pull-Patient conditions: 38 vs. 38 responses; SVO word order in Pull-Agent vs. Pull-Patient conditions: 19 vs. 34 responses), the data cannot be further divided into Word Order and Voice to tease apart their effects on fixations.

good testing ground for this question. If people are universally sensitive to embodied information and cognitively incorporate it when interpreting an event, motions should also influence Japanese speakers' initial attention in event apprehension processes as well as their selection of active versus passive voice in speech production. On the other hand, if speakers of a verb-final language show no motion effects, this would suggest that speakers of a verb-initial language in particular, rather than all speakers, incorporate motor information when perceiving and describing of events because their routine use of verb-initial language makes them sensitive to action components and shapes their cognition. It would also be valuable to test whether these findings replicate in other verb-initial languages.

Currently available and widely accepted comprehension and production models were proposed primarily on the basis of findings from languages in which the Subject precedes the Object, as well as non-verb initial languages including English. Recent comprehension studies have begun to investigate the processing cost of word order, and voice, and verb-argument relationships in a VOS language (Ono et al. 2016; Yano et al. 2017), but much less research has investigated the psycholinguistic processes of how VOS-language speakers encode a message and transform it into language. We see the current study in Truku as an early step in the investigation of the cognitive systems of VOS-language speakers and hope that it will be followed by many other studies of the online production processes in VOS languages.

Acknowledgements We would like to thank three anonymous JEAL reviewers as well as James Huang, whose comments have led to substantial improvements in the work presented here. We would also like to thank the audience at the 40th Annual Conference of the Cognitive Science Society (CogSci 2018) as well as the International Workshop on Seediq and Related Languages: Grammar, Processing, and Revitalization at the Harvard-Yenching Institute, Harvard University, in May 2018, where an earlier version of this paper was presented. We are indebted to the villagers who participated in our experiment and to Apay Ai-yu Tang, Yudaw Pisaw, Kimi Yudaw, Yuki Kumus, Reykong Yudaw, Muhung Yuki, and Ubing Yuki, for professional as well as heartwarming support. This research was supported by a Grant-in-Aid for Scientific Research (A) (PI: Masatoshi Koizumi, # JP15H02603), a Grant-in-Aid for Scientific Research (S) (PI: Masatoshi Koizumi, # JP19H05589), a Grant-in-Aid for Young Scientists (A) (PI: Manami Sato, #JP16H05939), and a Grant-in-Aid for Scientific Research (B) (PI: Manami Sato, # 19H01263).

References

Barsalou, L.W. 1999. Perceptual symbol system. *Behavioral and Brain Sciences* 22: 577–609.

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

Beilock, S.L., I.M. Lyons, A. Mattarella-Micke, H.C. Nusbaum, and S.L. Small. 2008. Sports experience changes the neural processing of action language. *Proceedings of the National Academy of Sciences, USA* 105: 13269–13273.

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

Bock, K. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18: 355–387.

Bock, K., D.E. Irwin, and D.J.J. Davidson. 2004. Putting first things first. In *The integration of language, vision, and action: Eye movements and the visual world*, ed. F. Ferreira and M. Henderson, 249–278. New York, NY: Psychology Press.

Bock, K., D.E. Irwin, D.J.J. Davidson, and W.J.M. Levelt. 2003. Minding the clock. *Journal of Memory and Language* 48: 653–685.

Bock, K., and H. Loebell. 1990. Framing sentences. *Cognition* 35: 1–39.

Brunyé, T.T., T. Ditman, C.R. Mahoney, J.S. Augustyn, and H.A. Taylor. 2009. When you and I share perspectives: pronouns modulate perspective taking during narrative comprehension. *Psychological Science* 20: 27–32.

Chambon, V., N. Sidarus, and P. Haggard. 2014. From action intentions to action effects: how does the sense of agency come about? *Frontiers in Human Neuroscience* 8(320): 1–9.

Chatterjee, A. 2010. Disembodying cognition. *Language and Cognition* 2: 79–116.

Cohn, N., and M. Paczynski. 2013. Prediction, events, and the advantage of agents: the processing of semantic roles in visual narrative. *Cognitive Psychology* 67(3): 73–97.

Cook, S.W., Z. Mitchell, and S. Goldin-Meadow. 2008. Gesturing makes learning last. *Cognition* 106(2): 1047–1058.

Decety, J., and J.A. Sommerville. 2003. Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Science* 7(12): 527–533.

Dink, J.W., and B. Ferguson. 2015. "*eyetrackingR: An R library for eye-tracking data analysis*." http://www.eyetrackingr.com. Accessed 3 April 2018.

Donnelly, S., and J. Verkuilen. 2017. Empirical logit analysis is not logistic regression. *Journal of Memory and Language* 94: 28–42.

Dowty, D.R. 1991. Thematic proto-roles and argument selection. *Language* 67: 547–619.

Feldman, J., and S. Narayanan. 2004. Embodied meaning in a neural theory of language. *Brain and Language* 89(2): 385–392.

Ferreira, V.S., and L.R. Slevc. 2007. Grammatical encoding. In *The Oxford handbook of psycholinguistics*, ed. M.G. Gaskell, 453–469. Oxford: Oxford University Press.

Foley, W. 2008. The place of Philippine languages in a typology of voice systems. In *Voice and grammatical relations in Austronesian languages*, ed. P. Austin and S. Musgrave, 22–44. Stanford, CA: CSLI.

Gleitman, L.R., D. January, R. Nappa, and J.C. Trueswell. 2007. On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language* 57: 544–569.

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

Glenberg, A.M., M. Sato, and L. Cattaneo. 2008. Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology* 18(7): 290–291.

Goldin-Meadow, S., S.W. Cook, and Z.A. Mitchell. 2009. Gesturing gives children new ideas about math. *Psychological Science* 20(3): 267–272.

Goldin-Meadow, S., and S.M. Wagner. 2005. How our hands help us learn. *Trends in Cognitive Sciences* 9(5): 234–241.

Griffin, Z.M., and K. Bock. 2000. What the eyes say about speaking. *Psychological Science* 11: 247–279.

Hafri, A., A. Papafragou, and J.C. Trueswell. 2012. Getting the gist of events: recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General* 142(3): 880–905.

Haggard, P., and V. Chambon. 2012. Sense of agency. *Current Biology* 22: R390–R392.

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.

Konopka, A.E., and A.S. Meyer. 2014. Priming sentence planning. *Cognitive Psychology* 73: 1–40.

Lamm, C., C.D. Batson, and J. Decety. 2007. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience* 19(1): 42–58.

Levelt, W.J.M. 1989. *Speaking from intention to articulation*. Cambridge, MA: MIT Press.

Madden, C.J., and R.A. Zwaan. 2003. How does verb aspect constrain event representations? *Memory and Cognition* 31(5): 663–672.

Miller, J., K. Brookie, S. Wales, S. Wallace, and B. Kaup. 2018. Embodied cognition: is activation of the motor cortex essential for understanding action verbs? *Journal of Experimental Psychology. Learning, Memory, and Cognition* 44(3): 335–370.

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

Oiwa-Bungard, M. 2017. "*Morphology and syntax of gerunds in Truku Seediq: A third function of Austronesian "voice" morphology.*" PhD dissertation, University of Hawai'i, Honolulu.

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

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

Prat-Sala, M., and H. Branigan. 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.

Pulvermuller, F. 2013. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences* 17: 458–470.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 3 April 2018.

Sato, M. 2010. *Message in the 'body': Effects of simulation in sentence production.* PhD dissertation, University of Hawai'i, Honolulu.

Sato, M., and B.K. Bergen. 2013. The case of missing pronouns: Does mentally simulated perspective play a functional role in the comprehension of person? *Cognition* 127: 361–374.

Sato, M., A.J. Schafer, and B.K. Bergen. 2015. Metaphor priming in sentence production: concrete pictures affect abstract language production. *Acta Psychologica* 156: 136–142.

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

Sauppe, S. 2017. Symmetrical and asymmetorical voice systems and processing load: pupillometric evidence from sentence production in Tagalog and German. *Language* 93: 288–313.

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

Stanfield, R.A., and R.A. Zwaan. 2001. The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science* 12: 153–156.

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

Tang, A. 2011. "*From diagnosis to remedial plan: A psycholinguistic assessment of language shift, L1 proficiency, and language planning in Truku Seediq.*" PhD dissertation, University of Hawai'i, Honolulu.

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

Wenke, D., S.M. Fleming, and P. Haggard. 2010. Subliminal priming of actions influences sense of control over effects of action. *Cognition* 115: 26–38.

Yano, M., K. Niikuni, H. Ono, S. Kiyama, M. Sato, A. Tang, D. Yasunaga, and M. Koizumi. 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.

Yaxley, R.H., and R.A. Zwaan. 2007. Simulating visibility during language comprehension. *Cognition* 105: 229–238.

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

Zwaan, R.A., R.A. Stanfield, and R.H. Yaxley. 2002. Language comprehenders mentally represent the shape of objects. *Psychological Science* 13: 168–171.