

**VOS 語順話者と SO 語順話者が捉える事象認知の違い：
VOS 語順であるタロコ語話者を対象としたジェスチャー実験を通して**

**Speakers of VOS and SO word order languages interpret the
world differently: A gesture study with Truku (VOS) speakers**

**里 麻奈美
Manami Sato**

要旨

本稿の目的は、ヒトは言語に関わらず普遍的に「動作主-対象者/物-動作」の順番で事象を認知している、という近年のジェスチャー研究の主張を実験的に再検証することである。これまでの研究では SO 言語 (i.e., 主語(Subject)-目的語 (Object)を基本語順に持つ言語) のみが対象となっているため、「動作主-対象者/対象者-動作」という認知選好が、(i)言語に依らない人間の普遍的認知選好であるのか、または(ii)言語における SO 語順が事象の認知順序に影響した結果なのかは不明である。そこで、本研究では目的語が主語に先行する VOS という基本語順を持つタロコ語を研究対象とした。タロコ語は台湾で話されている先住民族言語の一つで、絶滅に瀕している言語でもある。タロコ語母語話者は、単純な線画事象 (e.g., 女の子がボールを投げている) を見て、ジェスチャーだけで描写するように指示された。24名のジェスチャー産出の順序を分析した結果、先行研究とは異なり「動作主-対象者/物-動作」の産出はほぼ皆無に近く、動作を表すジェスチャーを最初または中間に位置づけるものが大多数であった。これは、VOS語順であるタロコ語話者が、SO言語話者とは異なる順序で事象を認知している事を示唆する結果だと言える。常に動詞先行の語順に触れている VOS 語順話者は、事象を構成する「動詞の要素」に敏感な認知機能を持っていることを意味しているのではないだろうか。

ABSTRACT

This study explores a fundamental claim of recent gesture studies that all humans cognitively interpret the world in Subject-Object-Action order, no matter what language they speak. The claim is based on research with speakers of SO languages (i.e., languages in which Subject-Object is the basic word order), and therefore could reflect either (i) a natural human cognitive preference, as claimed, or (ii) the influence of the basic word order of SO languages. To disentangle these two possibilities, this study tested speakers of the Truku language, in which Object precedes Subject in the basic word order (VOS). Truku is an endangered language spoken in Taiwan. The participants, 24 native speakers of Truku, were shown line drawings depicting simple transitive events (e.g., a girl throwing a ball), and were asked to describe the events in the drawings in pantomime. They produced almost no SOV gestures, instead showing a preference for verb-initial and verb-medial sequences. The study's results indicate that these speakers of Truku, a VOS language, see or interpret the world differently from the speakers of SO languages examined in previous studies. More specifically, this study sheds light on the cognitive systems of VOS speakers and suggests that VOS language speakers' constant engagement in verb-initial word order deeply affects their cognitive mechanisms in such a way as to highlight the action component of events.

1. Introduction

1.1 Language and thought

Research has demonstrated that language influences cognition in a wide variety of ways. Everyday linguistic activities such as reading, writing, listening, and speaking generally involve a linearization of words associated with specific meanings or thoughts (Bock & Loebell, 1990; Bock, Loebell, & Morey, 1992; Levelt, 1989). For instance, when reading, writing, and listening, people perceive, form, and receive the words one at a time. Likewise, when speaking, thoughts are chunked and packed into appropriate words, aligned in a specific linear order, and verbally delivered one word at a time.

The basic or most frequent word order varies among languages. In all their linguistic activities in daily life, people are exposed to and become accustomed to a certain word order – the basic one in their own language. For English speakers, the basic word order is Subject-Verb-Object (SVO; e.g., *A girl pushed a boy*), while for Japanese speakers, it is Subject-Object-Verb (SOV; e.g., *Onnanoko-ga otokonoko-wo oshita* [Literally: girl-NOM boy-ACC pushed], “A girl pushed a boy”). For speakers of the Truku language, an endangered language spoken in Taiwan, the basic order is Verb-Object-Subject (VOS; *Hnliqan Rabay ka kipu tgami niyi* [Literally: peel off Rabay NOM stamp the], “Rabay peels off the stamp”).

If we, as humans, cannot be detached from our thinking and speaking, then it is reasonable to assume that the constant experience of a particular word order in our own language influences and establishes a pattern of thinking or recognizing (perceiving) the world (Slobin, 1991, 2003, 2006). The tight connection between one’s language and one’s cognition suggests that speakers of languages with different word orders conceptually perceive events differently. Let us take the example of the event of a girl pushing a boy. If a word order or pattern of a language that is routinely used over years can influence one’s conceptual behaviors, speakers of SVO languages such as English may automatically pay attention first to the girl (Subject) and then to the boy (Object), in accord with the word order of their native language. On the other hand, speakers of VOS languages such as Truku might focus first on the boy (Object) who has been pushed and then on the girl (Subject).

1.2 Elicited pantomime as a reflection of thought

In the field of psycholinguistics, elicited pantomime production has recently come to be widely used and accepted as an empirical method to investigate whether language influences how people perceive a pictured event and externally deliver it through manual modes (Futrell et al., 2015; Gibson et al., 2013; Goldin-Meadow, So, Ozyurek, & Mylander, 2008; Hall, Ferreira, & Mayberry, 2014; Langus & Nesper, 2010; Schouwstra & de Swart, 2014). This method has become popular in part because, as manually describing what one sees in a picture is usually quite simple, it can be easily employed regardless of participants’ age, gender, educational background, and language.

In previous studies employing the elicited pantomime paradigm, speakers of languages with various word orders, including English, Chinese, Spanish, Italian, Turkish, and Korean, have been instructed to manually describe simple transitive events (e.g., a girl throwing a ball) depicted in line drawings or video clips. The crucial output in this paradigm is the constituent order in the elicited pantomime of the subjects, objects, and actions in the transitive event (e.g., *girl* as a subject, *ball* as an object, and *throw* as an action in the event of a girl throwing a ball).

Although people might see an event holistically instead of perceiving subject, object, and action components separately or individually, the participants are required to select one component at a time when they describe the event in pantomime. In this way, the elicited pantomime task makes people temporally align the constituents (subject, object, and action). The temporal alignment they choose is taken to reflect how they conceptually perceive and represent the world in their minds.

Such gesture research investigating the effects of language on thought has consistently shown that human beings cognitively access and represent transitive events in the Subject–Object–Action order, no matter what word orders their native languages have (Gibson et al., 2013; Goldin-Meadow et al., 2008; Hall, Ferreira, & Mayberry, 2014; Langus & Nespors, 2010). In other words, these results conflicted with the Slobin’s claim that language influences the way we think.

For example, in the study conducted by Goldin-Meadow et al. (2008), they examined participants with Subject-Verb-Object (SVO) order languages including English, Chinese, and Spanish, as well as participants with Subject-Object-Verb (SOV) order languages such as Turkish. The participants saw short animations that depicted simple transitive events (e.g., a girl twisting a knob; a boy opening a box). After watching each animation, the participants described what they had just seen by using their hands only. The results showed that the primary order of gestural production for all the participants was Subject-Object-Action (e.g., girl-knob-twist, boy-box-open). The researchers concluded that regardless of one’s native language, the

Subject-Object-Action order is the conceptual representation that is the most natural and universal sequence for all human beings. However, this is not an uncontroversial claim, as discussed in the next section.

1.3 Is the way we perceive and interpret the world “language specific” or “language universal”?

Regarding the influence of language on thought, two viewpoints make opposing claims. Slobin’s thinking-for-speaking hypothesis (1991, 1996) primarily asserts that linguistic structures influence the way we perceive the world; this is the language-specific view. On the other hand, the gesture studies discussed in section 1.2 claim that Subject-Object-Action is the most natural order, so humans cognitively prefer it regardless of their native language; this is the language-universal view. How, then, can we examine the question of whether or not the language we speak unconsciously influences the way we perceive the world?

Although the previous studies using pantomime production have claimed that the SOV sequence is a universal preference, they have only assessed languages in which the subject precedes the object (SO languages) in the basic word order (e.g., SVO in English, Chinese, and Spanish and SOV in Turkish in Goldin-Meadow et al., 2008; SVO in Italian in Langus & Nespors, 2010). But evidence from only SO languages with their structurally limited orders is not sufficient to conclude that all humans universally perceive the world in the SOV sequence. Solely assessing SO language speakers' choices cannot disentangle the two possibilities, which are that the apparent preference for Subject-Object-Action sequences is the result of (i) a natural, universal human cognitive preference or (ii) the influence of the word order of SO languages.

In order to disentangle these two possibilities – the language-specific view or the language-universal view – it is necessary to examine a language in which the object precedes the subject (OS) in the basic word order. It has been difficult, however, for researchers to access speakers of OS languages, for several reasons. First, only a small number of OS languages are spoken. According to Dryer (2005), some languages lack a dominant word order, but 82.8% of the world's languages are SO (including SOV, SVO, and VSO) languages, while 3.2% are OS (including VOS, OVS, and OSV).

Furthermore, physical and technological obstacles exist as well. OS languages tend to be spoken by limited numbers of people, in remote locations, where the lab facilities needed by cognitive and psycholinguistic researchers are not available. Hence, OS languages have not been experimentally tested in the fields of psycholinguistics and cognitive linguistics. Nevertheless, as discussed above, a VOS language like Truku is ideal to assess the language-specific and language-universal views; this study therefore conducted experiments in Truku, whose basic word order is VOS.

1.4 The Truku language

Truku is one of the three dialects of Seediq (with Toda and Tgdaya), an Austronesian language spoken in the north-eastern part of Puli in Central Taiwan, past Wushe and the valleys to the east and north-east, and across the Central Mountain Range to Truku Gorge and the Pacific coast (Holmer, 1996, p. 9). Truku's basic word order is VOS, as shown in the example below, but it also permits SVO.

Truku morphology tends to be agglutinating, that is, one word is usually composed of one lexical morpheme and some optional affixations such as prefixes, infixes, suffixes, or circumfixes.

k-em-rut	qesurux	sapah	ka	Lowking.
AV-cut	fish	house	NOM	Lowking
'Lowking cuts (the) fish at home.'				

It is reported that there are now about 30,074 Truku Seediq people, but the exact number of people who can speak the Truku language is unknown. Research shows a severe and sudden decline of fluent speakers in recent years. Only the grandparental generation can now speak fluent Truku, which probably amounts to only a couple thousand people. According to Krauss's classification of degrees of language endangerment (2007), Truku is categorized as a definitely endangered language.

2. Elicited pantomime experiment in the Truku language

2.1 Aims of the current study

This study addresses a question derived from Slobin's thinking-for-speaking hypothesis (1991, 1996), whose primary assertion is that linguistic structure may influence the way we think and the way we perceive the world. If this hypothesis holds true, then speakers of the Truku language would not only position linguistic entities (e.g., Subject, Object, Verb) in the Verb-Object-Subject order, but also perceive and interpret events in the same sequence. On the other hand, if, as previous elicited pantomime studies claim, all people share a coherent way of perceiving and describing events nonverbally, then Truku speakers should also conceptually prefer the Subject-Object-Verb order, although it is different from the word order of the language they speak.

2.2 Method

2.2.1 Participants

All participants were born, raised, and live in Jingmei village. They were all native speakers of Truku Seediq who acquired it as their first language from birth. Because the language has not been successfully transmitted to the younger generations, the age of eligible participants was quite advanced. Their approximate mean age was 70 years old. Each participant completed a demographic form after finishing the elicited pantomime experiment. Although the participants were carefully selected by a native Truku speaker who was born and grew up in the village, the demographic form was used to ensure the participants' Truku proficiency in comprehension and production. In addition, it included questions regarding participants' name, age, address, and whether they had normal hearing and eyesight via self-evaluations using a 5-point scale (in which 1 represents "very poor" and 5 represents "excellent, native"). Their proficiency ratings in both comprehension and production were all above 4, and close to 5.

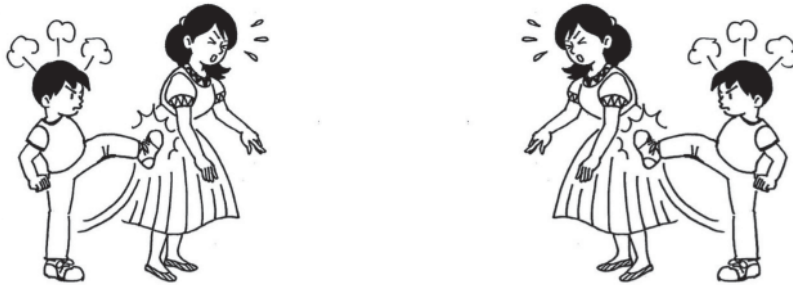
2.2.2 Design and picture materials

Thirty line-drawing pictures were created. Each picture depicted a simple action-related event in which a person acted on something (e.g., a man pushes a box) or somebody (e.g., a girl hits a boy). The pictures included three types: (1) Person-Person (e.g., a boy is kicking a woman), (2) Person-Object (e.g., a girl is throwing a ball), and (3) Object-Person (e.g., a police car is chasing a woman)

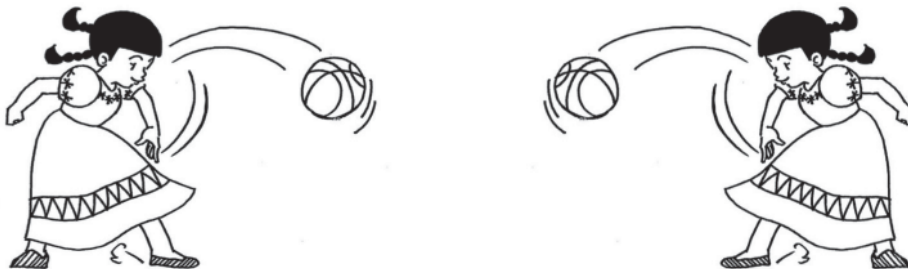
(Figure 1). The first two types, Person-Person and Person-Object, were critical for the experimental purposes while the Object-Person type was prepared for the sake of event variation. The Person-Person and Person-Object materials were created in order to see if the orders of conceptual access would be altered depending on whether the entities involved were animate (Person) or inanimate (Object). Ten events used for critical conditions included *chase*, *kick*, *poke*, *pull*, *punch*, *push*, *roll*, *run into*, *shoot*, *throw*. Each of the three types had ten pictures, resulting in thirty pictures in total.

Since humans have a tendency to first pay attention to the left side of a scene they are viewing and then shift to the right, a mirror version of each of the thirty pictures was created. The mirror versions were expected to enable us to assess which sequence the participants conceptually preferred when thinking about the event and describing it manually.

(1) Person-Person type



(2) Person-Object type



(3) Object-Person type



Figure 1. Three types of events in the original and the mirror versions

2.2.3 Setting and procedure

The participants completed the elicited pantomime experiment individually in a quiet room located in their home village. The participants were seated in front of a laptop computer, on which they viewed the experimental materials. A video camera was set up in a corner of the room to record the experiment. To minimize any influences from other languages (e.g., Chinese, English, Japanese) on the participants, the two research assistants (RAs), who were Truku native speakers, used Truku exclusively when giving the participants the experimental instructions.

The RAs explained the participants' task in the elicited pantomime experiment by emphasizing that not language, but hand gestures were required when describing the depicted events. Each participant first completed a practice session followed by the main experiment. In the practice session, a picture depicting either a single object (e.g., girl, ball) or a single action (e.g., throw) was presented on a screen (Figure 2). The purpose of this session was to make participants come up with their own creative gestures for each item that appeared as part of the events depicted in the main experiment. Note that one item included an action involving a body part other than the hands (i.e., kick), and the participants, who were seated throughout the experiment, imitated the kicking action using their hands.



Figure 2. Individual pictures used in the practice session

Let us take as an example the event where a girl is throwing a ball. In this case, a picture of a “girl” appeared on the screen, and the participants created their own gestures to describe the “girl,” without employing any words. Likewise, when a picture of a “ball” or “throw” was shown on the screen, they created their own gestures accordingly. Through this practice session, all gestural components that were a part of the events used in the main experiment were expected to become equally easily accessible for the participants.

After the participants had created and used gestures for each character and object, they moved

on to the main experiment. The main experiment was composed of thirty trials. In each trial, a cross first appeared in the center of the computer screen for 1000 ms. The trial's picture then automatically appeared on the screen. The three types of pictures (Person-Person, Person-Object, Object-Person) were arranged in pseudo-random order. Participants had been instructed to describe the picture they saw, using their hands, as quickly and accurately as possible. In addition, participants were carefully reminded that all three components in the picture (i.e., Subject, Object, and Action components) should be expressed by their pantomime. To move to the next trial, they pressed the space bar.

The practice and main experiment sessions together took approximately 20 to 25 minutes. Each participant's pantomime was videotaped for the purpose of the later analyses. After each participant completed the main experiment, the RA administered the demographic form in Truku.

2.2.4 Differences between previous studies and the current study

This study differed from previous studies in terms of how participants generated the gestures. That is, the previous studies (e.g., Gennari, Sloman, Malt, & Fitch, 2002; Goldin-Meadow et al., 2008; Papafragou, Hulbert, & Trueswell, 2008) had no practice session, so their participants were required to produce spontaneous pantomimes during the experiment, while the participants in this study were given the opportunity to create and practice individual items in advance.

The reason for the current study's procedure was to avoid the need for the participants to create their pantomime gestures and describe the event at the same time, which could lead participants to make the easiest gesture first. In this case, their pantomime would not reflect the sequential order of the event as they thought of it in their minds. This study's practice session was intended to make all of the components represented in the event (i.e., people, objects, and actions) equally accessible to the participants when they were perceiving the events in the main experiment. Hence, the order of their gestures should purely reflect the order of their conceptual alignment (i.e., thought).

2.2.5 Data coding

A total of 720 pantomime productions (480 critical and 240 filler) were collected from the 24 participants in the main experiment. One RA and the author separately watched the 24 participants' video clips and coded the sequence of the participants' production in each trial by the three components of Subject (S), Object (O), and Verb (V). For example, if a participant pantomimed a kicking action first, a tree second, and a boy last for the event of "a boy kicking a tree," this sequence of pantomimes would be coded as "VOS" (Figure 3).

Kick (Verb)



Tree (Object)



Boy (Subject)



Figure 3. A gesture production of VOS order

The two coders then compared the results of their coding; in the case of inconsistency, they reviewed the video clip in question again, discussed it, and decided on a consistent single code sequence for that production. Coding was inconsistent between the two coders in less than 15 of the 480 critical items, and they were able to resolve each case by this process.

Although the participants were repeatedly reminded to produce all three components when describing the events, they sometimes produced either less or more. Because the experimental purpose was to assess in which order the participants nonverbally accessed and described the three components (Subject, Object, and Action) when perceiving the event, all pantomime productions with less than three (e.g., VO, V) or more than three (e.g., SOSV, OSVSO) components were categorized as “Other” and excluded from the data analysis. The excluded data amounted to 5.6% (27 of 480); the data remaining for the analysis were therefore 94.4% (453) of the productions made in response to the critical items.

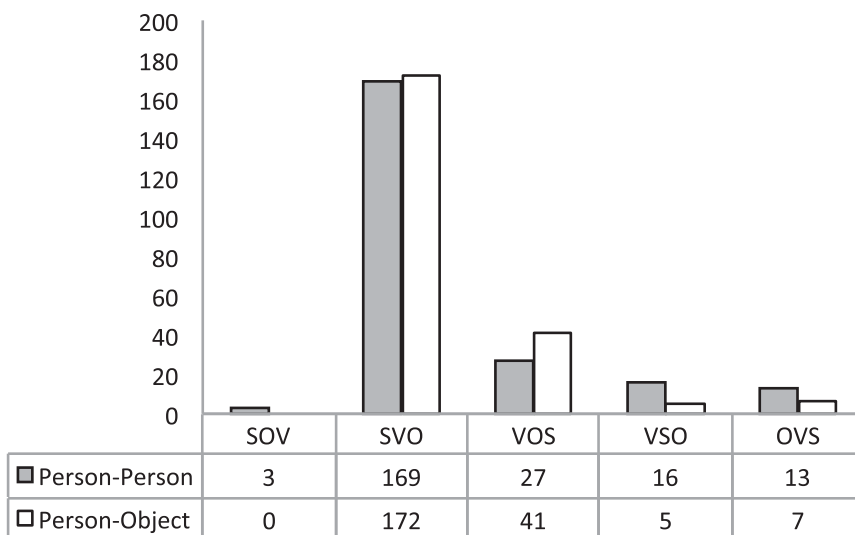
2.3 Results

The coded data were categorized and analyzed in four ways to examine four different points: (i) the order of the three components (i.e., Subject, Object, Verb), (ii) the order of the Subject and Object components (i.e., SO or OS), (iii) the position of the Action (verb) component (i.e., verb-initial, verb-medial, or verb-final), and (iv) the order of the Action and Object components (i.e., VO or OV).

The primary reason for categorizing the responses as in (ii) and (iii) was to individually highlight the order of the conceptually equivalent entities, that is, the tangible entities of the Subject and Object in (ii) and the physical action entity of the Action (the verb) in (iii). The reason for considering the relationship between Object and Action as in (iv) was to respond to Dryer’s (2005) argument for the importance of distinguishing between OV and VO types of languages in predicting language features and properties in nature. For the analyses, a chi-squared test was employed.

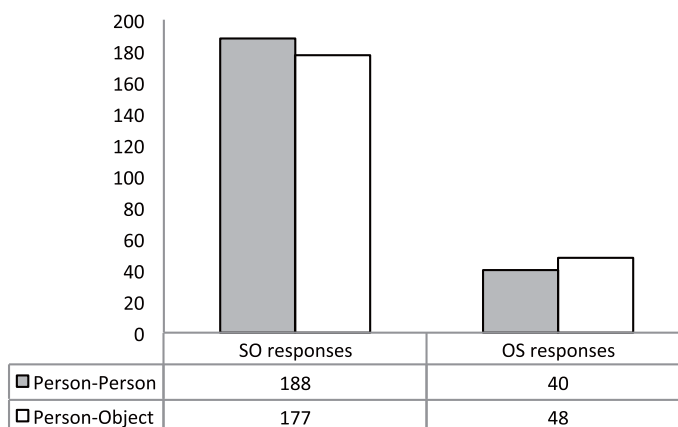
First, Graph 1 below shows the general tendencies of how the VOS language speaking participants manually described the two types of events with three components (i.e., Subject, Object, Verb). As shown in Graph 1, the participants produced five sequences. The dominant order was

SVO (341 out of 453 responses, 75.3%) followed by VOS (68 responses, 15%), VSO (21 responses, 4.6%), OVS (20 responses, 4.4%), and SOV (3 responses, 0.7%). If SOV is the most natural order for human beings, it should be the most dominant order. However, it was not the most dominant sequence ($\chi^2(1)=466, p<0.001$). No significant differences in Person-Person and Person-Object types of events were observed in any of these gestural orders.



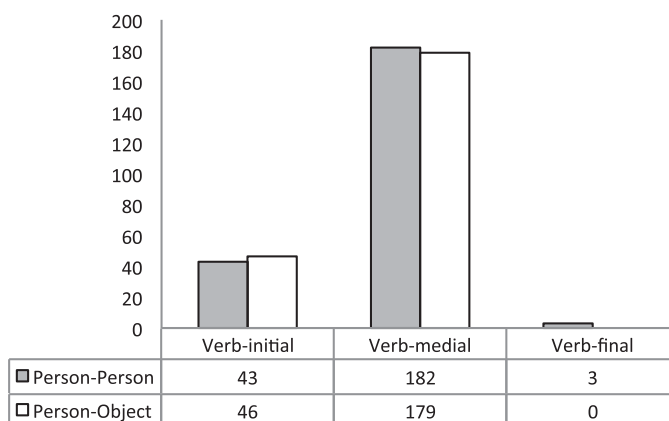
Graph 1. Order of the three components in pantomimes of the two event types

Second, Graph 2 shows the responses recategorized according to the order of the Subject and Object components (i.e., SO or OS order). As the graph indicates, Truku speakers gestured more SO sequences than OS ones (i.e., 365 responses (80.6%) and 88 responses (19.4%), respectively). Distributions of SO and OS sequences were not significantly different between the two event types ($\chi^2(1)=1.04, p=0.3$).



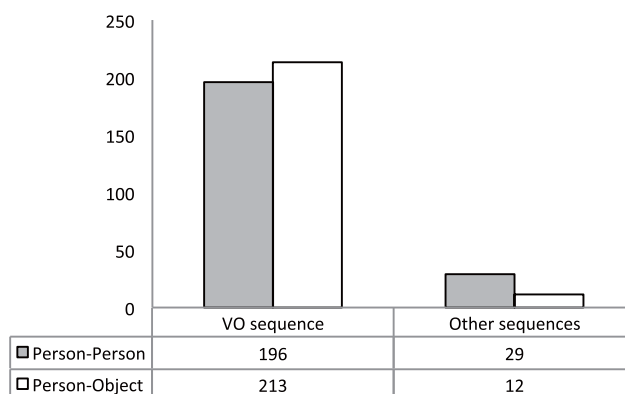
Graph 2: Gestural order of SO and OS responses in Person-Person and Person-Object types

Third, Graph 3 below shows the data again recategorized based on the position of the Action (verb) component (i.e., verb-initial, verb-medial, or verb-final). As Graph 3 represents, the predominate sequence was verb-medial (361 responses, 79.7%) followed by verb-initial (89 responses, 19.6%). In contrast to previous research, in which almost all sequences produced were verb-final, the participants in this study produced very few verb-final sequences (3 responses, 0.7%). There were no differences based on the positions of the action in the two types of event, Person-Person and Person-Object ($\chi^2=3.1, p=0.21$).



Graph 3: Gestural order of verb-initial, verb-medial, or verb-final in Person-Person and Person-Object types

Finally, the relationship between Object and Action (verb) was considered. As shown in Graph 4, there was a clear preference for the VO order. There were 430 responses with the VO order (94.9%) and only 23 responses with the OV order (5.1%). This study's observation of a dominant preference for the VO sequence displayed by speakers of the VOS word order Truku language is the first such observation to be reported; this preference has never been observed in speakers of the SVO or SOV languages that were examined in previous studies.



Graph 4: Gestural order of the VO and Other sequences in Person-Person and Person-Object types

3. Discussion

This study explores the fundamental claim of recent gesture studies that all humans cognitively interpret the world and process the entities that constitute the world in an identical Subject-Object-Action order, no matter what language they speak. In other words, linguistic structures have no influence on the cognitive activities of humans perceiving events. However, this reportedly universal preference could reflect (i) a natural cognitive preference for human beings, as previous studies have claimed, or (ii) an influence rooted in the word order of SO languages (i.e., languages in which Subject precedes Object in the basic word order).

To disentangle these two possibilities, the study examined the Truku language, in which Object precedes Subject in the basic word order. If Subject-Object-Action is a universal preference, Truku speakers should also prefer the Subject-Object-Action sequence when thinking about an event. However, if such preferences reflect the influence of word order, then Truku speakers would be expected to prefer a different sequence.

The gestural responses of the Truku speaker participants, which were examined in four ways as shown in section 2.3, shed light on various aspects of the conceptual systems of Truku speakers. First, in contrast to the findings of previous studies, this study observed almost no sequences of Subject-Object-Verb gestures. This result suggests that SOV could be the preferred sequence for SO speakers, but it is not a universal human preference. Second, if the SO order is a unified representational system for the nonlinguistic expression of simple events, only the SO order should have been observed in the current study. However, approximately 20% of the responses were in the OS constituent order. This result suggests the possibility that OS language speakers conceptually perceive events differently from SO language speakers.

Third, regarding the position of the verb in nonverbal descriptions, cognitive complexity could be a factor that influences verb position. It has been said that object entities (i.e., subject and object) and actions (i.e., verb) differ in terms of their qualities as well as their properties: Object entities are cognitively simpler and less relational than actions (Bickerton, 1992; Gentner & Boroditsky, 2001; Goldin-Meadow et al., 2008; Hall et al., 2014). If this is the case, then people might tend to access the tangible object-entity components earlier than the action information represented in an event, leading to a cognitive preference for the SOV sequence. However, the dominant verb-medial responses observed in this study imply that there is no clear qualitative distinction between object and action information in people's cognitive representations. In addition, the finding that 20% of the responses were verb-initial sequences, while less than 1% were verb-final sequences, indicates that action entities are not universally perceived as cognitively more complex than object entities. In other words, the cognitive complexity of entities could vary depending on speakers' native languages. If the constant experience of word order unconsciously influences and establishes a pattern in people's thinking and perception of the world, then the experience of speakers who normally produce a verb as

the initial component of their utterances in their daily linguistic activities should shape their cognitive systems in such a way as to incline these speakers to attend to action entities before object entities. This might result in preventing Truku speakers from placing the verb in the final position.

Finally, this study's observation of a strong preference for VO gesture sequences is unique in the available research: The speakers of the VOS Truku language who participated in this study showed a VO preference that has never been observed in the speakers of SVO or SOV languages tested in previous studies. The implication is that the VO sequence has a special cognitive status for VOS language speakers. That is, Truku speakers might pack the Action and the Object together as a cognitive unit in their minds.

Taking this study's results together with the results of previous studies, speakers of the VOS language Truku apparently see or interpret the world differently from speakers of SO languages. Thus, the claim of universality in how people think about and describe events, which was based solely on SO languages, does not hold true when we test it with an OS language. This assessment of Truku language therefore suggests a possible impact of word order on thinking about the world, supporting the view that the use of language indeed influences cognition (Gordon, 2004). However, what this study has observed is not a superficial effect in which word order is directly projected onto the speakers' conceptual orderings. Rather, the word order of Truku seems to modulate the speakers' cognitive mechanisms deeply, so that the action component of events is highlighted. Speakers of a verb-initial language may conceive of the action of an event as playing a fundamental role, framing the event and assigning the roles of the object entities, such as who did what to whom.

It is worth noting that while the current study and the previous studies used similar simple events involving persons and objects, they used different materials and event variations. To make a direct comparison with this study's results, it is necessary to conduct the same experiment with speakers of SO languages using identical materials and settings. For this reason, this experiment is next being conducted with English speakers.

References

- Bickerton, D. (1992). *Language and species*. Chicago, IL: Chicago University Press.
- Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.
- Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99, 150–171.
- Dryer, M. S. (2005). The order of subject, object and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *World atlas of language structures* (pp. 330–333). Oxford: Oxford University Press.
- Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, 136, 215–221.

- Gennari, S., Sloman, S., Malt, B., & Fitch, T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79.
- Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). New York: Cambridge University Press.
- Gibson, E., Piantadosi, T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*, *24*(7), 1079–1088.
- Goldin-Meadow, S., So, W. C., Ozyurek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*, 9163–9168.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499.
- Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating constituent order change with elicited pantomime: A functional account of SVO emergence. *Cognitive Science*, *38*(5), 943–972.
- Holmer, A. J. (1996). *A parametric grammar of Seediq*. Lund, Sweden: Lund University Press.
- Krauss, M. (2007). Classification and terminology for degrees of language endangerment. In M. Brenzinger (Ed.), *Language diversity endangered* (pp. 1–8). Berlin: Mouton de Gruyter.
- Langus A., & Nespors, M. (2010). Cognitive systems struggling for order. *Cognitive Psychology*, *60*, 291–318.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Papafragou, A., Hulbert, J., & Trueswell, J. C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 155–184.
- Schouwstra, M., & de Swart, H. (2014) The semantic origins of word order. *Cognition*, *131*, 431–436.
- Slobin, D. I. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics*, *1*, 7–26.
- Slobin, D. I. (1996). From “thought and language” to “thinking to speaking.” In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.
- Slobin D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157–91). Cambridge, MA: MIT Press.
- Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam/Philadelphia: John Benjamins.

Funding

This study was supported in part by JSPS KAKENHI Grant Numbers 15H02603, 16H05939.