Processing loads related to word order preference during sentence production in Japanese: An NIRS and eye tracking study

Tohoku Psychologica Folia
2014, *73*, 36–45

Processing loads related to word order preference during sentence production in Japanese: An NIRS and eye tracking study

Yasuhiro TAKESHIMA（竹島康博）[1], Godai SAITO（齋藤五大）[1],
Ryo TACHIBANA（立花　良）[1], Riku ASAOKA（朝岡　陸）[1],
Jiro GYOBA（行場次朗）[1] and Masatoshi KOIZUMI（小泉政利）[1]

Word order influences sentence comprehension and production processes. Previous studies on Japanese have shown that sentences with subject-object (SO) word order induce lower processing loads for comprehension than those with object-subject (OS) word order. Furthermore, sentences with SO word order are found to be constructed more frequently than those with OS word orders. Therefore, SO word order is thought to be preferred to the OS word order in both sentence comprehension and production. However, as indices of word order preference, most of previous studies measured the processing loads in sentence comprehension tasks, while measuring the frequency in the sentence production tasks. In this study, we directly compared the processing loads between different word orders during sentence production by measuring utterance latency (i.e., the time required to initiate the utterance), brain activity, and eye movement. Results showed that these three measures indicated higher processing loads for the OS word order compared to the SO word order. In basic Japanese syntactic word ordering, the subject precedes the object. Therefore, the processing load can be expected to increase in sentences with the OS word order compared to those with the SO word order. Based on the appropriate measurements of the processing load, the present study confirmed this prediction also for sentence production.

Key words: NIRS, Eye movement, Processing loads, Sentence production, Word order preference

Introduction

Word order affects the sentence comprehension process. For example, Japanese readers take less time to judge whether a sentence makes sense when it has subject-object (SO) word order (i.e., SOV sentence) than when it has object-subject (OS) word order (i.e., OSV sentence) (Tamaoka, Sakai, Kawahara, Miyaoka, Lim, & Koizumi, 2005). Moreover, longer reading times for OSV sentences in Japanese were investigated in self-paced reading and eye-tracking studies (Mazuka, Itoh, & Koizumi, 2002; Imamura & Koizumi, 2008). In functional magnetic resonance imaging (fMRI) studies, the left inferior frontal gyrus (IFG) was activated more during the processing of sentences with the OS word orders compared to those with the SO word order (Kim, Koizumi, Ikuta, Fukumitsu, Kimura, Iwata, Watanabe, Yokoyama, Sato, Horie, & Kawashima, 2009; Kinno, Kawamura, Shioda, & Sakai, 2008). Therefore, SO

1. Graduate School of Arts and Letters, Tohoku University, Japan (E-mail: yasuhiro.takeshima@gmail.com)
Corresponding Author to: Kawauchi 27-1, Aoba-ku, Sendai 980-8576, Japan

sentences are assumed to induce a lower processing load for comprehension than OS sentences.

The effects of word order have been reported in sentence production studies as well. Japanese sentences with SO word orders are produced more frequently than those with OS word orders (Imamura & Koizumi, 2011). In a sentence recall experiment, the recall rate was higher for SO than for OS word order sentences (Tanaka, Branigan, McLean, & Pickering, 2011). Therefore, SO word order is preferred to OS word orders in the sentence production process.

These findings indicate preferences for SO word order in both sentence comprehension and production processes. However, the index for word order preference differed between sentence comprehension and production in most previous studies (e.g., Tamaoka et al., 2005; Imamura & Koizumi, 2011). In sentence comprehension, the index was the processing load calculated from response time and brain activity; on the other hand, the sentence production index was the production frequency. The effects of word order preference may appear to differ between sentence comprehension and production simply because the preference index differs. Therefore, the word order preference of production needs to be investigated using the same index as the one used for comprehension.

In the present study, we attempted to compare the processing load between SOV and OSV sentence production, targeting Japanese speakers. We simultaneously measured utterance latency, brain activity, and eye movement. The utterance latency was assumed to reflect the processing load in sentence production (Lindsley, 1975; Smith & Wheeldon, 1999). We also focused on activity in the lateral prefrontal cortex (LPFC) detected by near-infrared spectroscopy (NIRS). The left IFG, especially the Broca's area located in the LPFC, plays an important role in sentence production (Horwitz, Amunts, Bhattacharyya, Patkin, Jeffries, Zilles, & Braun, 2003). Eye movement was also predicted to differ in participants while producing SOV and OSV sentences (cf. Griffin & Bock, 2000; Gleitman, January, Nappa, & Trueswell, 2007).

Methods

Twenty right-handed individuals (13 women and 7 men) participated in this experiment. They reported normal or corrected-to-normal vision. Handedness was assessed with the Edinburgh Handedness Inventory (Oldfield, 1971). In this test, a positive laterality quotient (LQ) score indicates that the participant is right-handed, while a negative LQ is indicative of left-handedness; those with LQ scores of 0 were considered mixed-handed. The LQ was positive for each of the participants (Mean LQ = 0.88, SD = 0.14). The present sudy was approved by the ethics committee of the Graduate School of Arts and Letters, Tohoku University. All participants gave written informed consent prior to their participation.

During the experiment, participants sat in front of a LCD-display (Tobii: Tobii TX300 screen unit; resolution: 1024 × 768 pixels; refresh rate: 60 Hz). The viewing distance was about 60 cm. The experimental stimuli were controlled by E-Prime 2.0 with E-Prime Extensions for Tobii (Psychology Software Tools, Inc.) and a PC (Dell: Precision T7500).

We used a voice recorder (Olympus: Voice Trek DS-850), an eye tracker (Tobii: Tobii TX300; sampling rate: 300Hz), and a multi-channel NIRS system (Shimadzu: FOIRE-3000) to record the participants' utterances, eye movements, and the relative concentration changes in oxygenated, deoxygenated, and total hemoglobin (Coxy-Hb, Cdeoxy-Hb, and Ctotal-Hb). A whole-head probe cap was used. The probes were placed over the prefrontal cortex of each hemisphere, each consisting of a 4 × 3 array with six emitters and six detectors, constituting 17 channels per hemisphere (see Fig. 1a). We arranged these probes in reference to the international 10-20 system: the locations of the 5th receiver and 15th emitter were arranged at T8 and T7, respectively. The sampling rate of each channel was 10 Hz.

We prepared twenty pictures (45 × 32°). On half of these pictures, an agent person and a patient person were depicted, and on the other half of the pictures, an agent person and a patient object were depicted (Appendix). The agent person and patient person/object were depicted at the left and right sides of each other. In order to counterbalance the locations of the agent and patient, we also prepared 20 pictures that were mirror images of the original pictures. The sizes of the agent person and patient person/object were nearly identical. Before the experiment, participants observed these pictures in a random order and described the event depicted in the picture using Japanese transitive sentences. If the participants could not utter or incorrectly interpreted the content, the experimenter explained the picture. After this session, participants were fitted with NIRS probes and then performed nine-point eye tracking calibration. Each trial initiated with the presentation of a fixation cross for 8 s, followed by an instructional display for word order for 5 s (Fig. 1b). A short beeping sound was presented for 0.5 s simultaneously with the onset of this instructional display to use in utterance latency analysis. In the instructional display, the Japanese cursive syllabary "*ga (ha)*" or "*wo (ni)*" was displayed. Then, the picture was presented for 8 s. Participants were instructed to utter as concisely as possible the contents of the picture on subject-initial order for the "*ga (ha)*" display and on object-initial order for the "*wo (ni)*" display. Thus, it was expected that participants uttered the contents in SOV and OSV sentences after "*ga (ha)*" and "*wo (ni)*" displays, respectively. The fixation cross appeared again after the picture display (Figure 1b). Participants performed a total of 40 trials: 2 (word order; SOV or OSV) × 20 pictures. Of the 20 pictures each participant saw, the agent was depicted on the left side in 10 pictures, and on the right side in the other 10 pictures.

Analysis

One participant whose NIRS data were not correctly acquired was excluded from further analysis. Moreover, we excluded data from trials when participants could not utter the instructed word order, where we judged the sentence was grammatically strange, and from trials when participants corrected themselves during the utterance. These led to the removal of 1.18% of all the data.

Utterance latency. We calculated the utterance latency for the two word order conditions

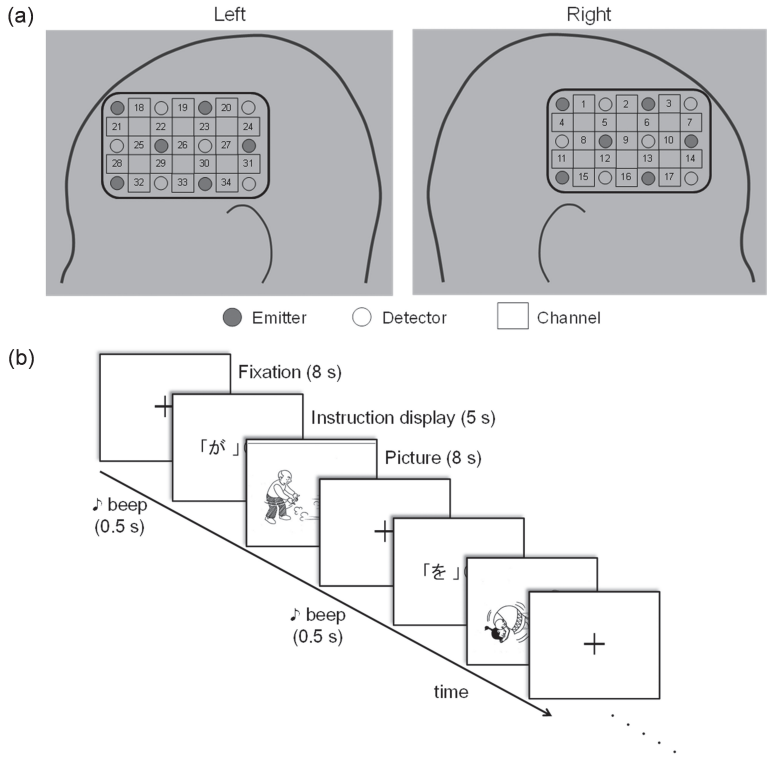


Figure 1. (a) Locations of the probes used in near infrared spectroscopy measurements. (b) Schematic representation of the procedure in the present experiment.

using Sound Engine (http://soundengine.jp/). The latency was defined as the interval between the onset of the picture presentation and the starting point of the participant's utterance. Interjections, for example "Let's see," were not included in the utterance.

Brain activity. We focused on Coxy-Hb as an index of neural activation because it is more sensitive to changes in regional cerebral blood flow than Cdeoxy-Hb and Ctotal-Hb concentrations (Hoshi, 2003). The raw Coxy-Hb data were high-pass filtered at 0.02 Hz to remove signal drift (Taga, Asakawa, Maki, Konishi, & Koizumi, 2003). Data were averaged for each participant and each condition and converted into z-scores (Otsuka, Nakato, Kanazawa, Yamaguchi, Watanabe, & Kakigi, 2007; Schroeter, Zysset, Kruggel, & von Cramon, 2003), so that the mean values and standard deviations for the base line period (5 s before picture presentation) were 0 and 1, respectively. We averaged the Coxy-Hb values during the picture presentation period (8 s).

The cortical regions corresponding to each channel were estimated by measuring three-dimensional (3D) coordinate data of each probe from two participants (one woman and one man) using a 3D digitizer (Polhemus: FASTRAK). The 3D coordinate data were averaged

across participants. Channel coordinates were calculated as the midpoint between the nearest two probes, and channel positions were registered to the Talairach coordinate space (Talairach & Tournoux, 1988) using NIRS-SPM software.

Eye movement. We calculated the mean fixation time on the agent and patient during the time of each picture (8 s) using area of interest (AOI) analysis. We used two AOIs of comparable size (but not necessarily the same shape) to properly cover the agent and patient (Fig. 2).

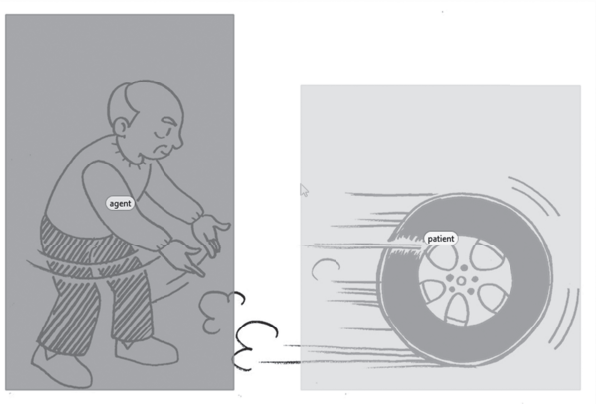


Figure 2. An example of AOI setting. The left AOI covers the agent area. The right AOI covers the patient area. In this case, while a larger AOI (14 × 23°) is applied to the agent because the agent area is vertically long, another AOI (17 × 19°) is to the patient because its area is comparatively square.

Results

Utterance latency. The results of utterance latency analyses are shown in Table 1. We conducted a paired *t*-test to compare the latencies between word orders. The analyses revealed that the latency was significantly shorter in the SOV condition than in the OSV condition (t (18) = 4.48, $p < .001$). These results indicate that SO word order has a lower processing load in sentence production.

Brain activity. We focused channels 29, 30, 32, and 33 over regions we thought were centered on Broca's area (channel 29: Brodmann area [BA] 45, Talairach coordinates, [x, y, z = -57, 29, 17]; channel 30: BA 45, Talairach coordinates, [x, y, z = -62, 6, 32]; channel 32: BA 45, Talairach coordinates, [x, y, z = -54, 39, 6]; channel 33: BA 44, Talairach coordinates, [x, y, z = -60, 16, 14]). The results of each channel are shown in Fig. 3. We conducted a paired *t*-test with the Coxy-Hb peak values on each channel to compare brain activity between word order conditions. On channel 32, the participants showed significantly higher peak values in the OSV condition than in the SOV condition (t (18) = 1.72, $p < .05$). However, significant differences

were not observed between word order conditions on the other channels (channel 29: t (18) = 0.28, p = .39; channel 30: t (18) = 0.70, p = .25; channel 33: t (18) = 0.53, p = .30).

Eye movement. The results from analyses of eye movement data are shown in Fig. 4. We conducted paired t-tests for every 100 ms to compare fixation time between the agent and patient in the SOV and OSV conditions. In the SOV condition, while there was no significance from 1,800 to 2,100 ms and 2,900 to 3,400 ms (ts (18) < 1.83, ps > .05), all other time periods were significant (ts (18) > 2.14, ps < .05). The relative fixation time was higher for the patient than for the agent in periods from 2,100 to 2,900 ms, and vice versa in other periods. In the OSV condition, while there were no significance from 700 to 800 ms and 1,500 to 1,800 ms (ts (18) < 1.57, ps > .05), all other time periods were significant (ts (18) > 2.28, ps < .05). Relative fixation time was higher for the patient than for the agent in the period from 800 to 1,500 ms, and vice versa for all other periods. In the case of significance in both the SOV and OSV conditions, the relative fixation times were higher for the agent than for the patient immediately following picture presentation.

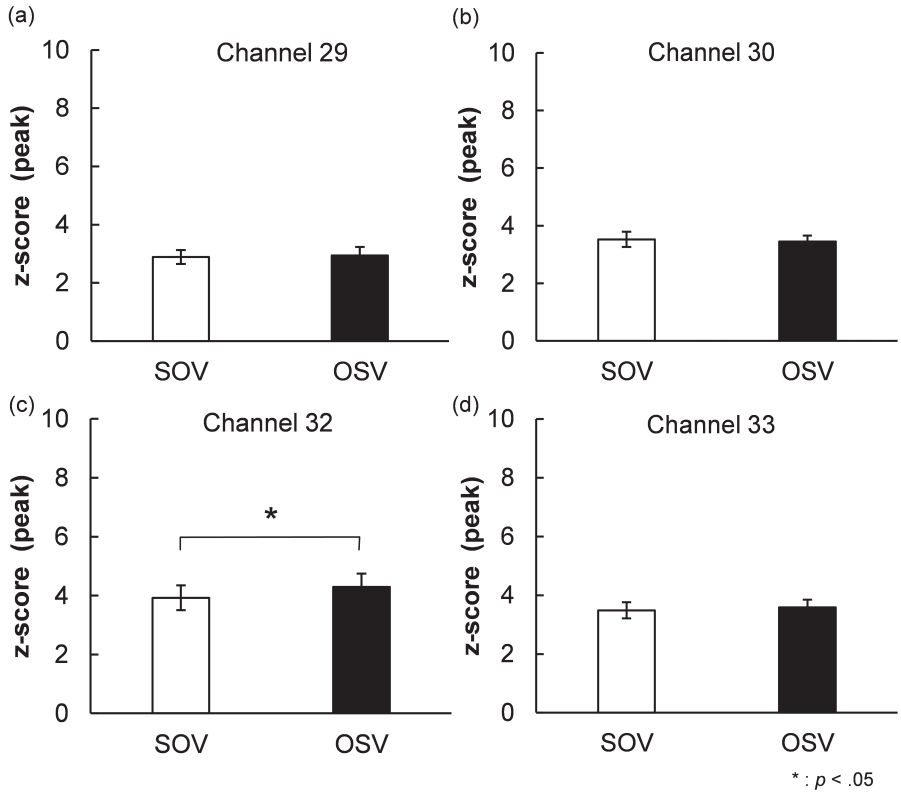


Figure 3. Average peak concentration changes of oxyhemoglobin under the word order conditions in the four channels estimated as Broca's area. (a) The results for channel 29. (b) The results for channel 30. (c) The results for channel 32. (d) The results for channel 33. Error bars represent standard error of the mean (n = 19).

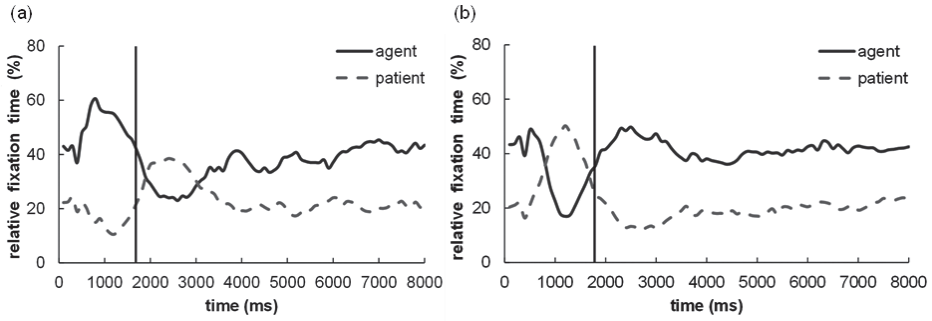


Figure 4. The results of eye movement data. (a) The relative fixation time in the SOV condition during the picture presentation. (b) The relative fixation time in the OSV condition during the picture presentation. The black vertical line represents the start of utterance.

Table 1. The means and standard deviations of utterance latency for word order conditions.

Word order	Utterance latency (ms)	
	M	*SD*
SOV	1,592	216
OSV	1,771	321

Note: $n = 19$, M = mean, SD = standard deviation

Discussion

The present study examined the processing load during sentence production using utterance latency, brain activity, and eye movement.

The utterance latency was shorter in the SOV than the OSV condition. This indicated that the processing load in sentence production was higher in the OSV condition than the SOV condition, since the utterance latency was assumed to reflect the processing load (Lindsley, 1975; Smith & Wheeldon, 1999). This result was in line with the previous findings that the OSV sentences needed higher processing loads than the SOV sentences during sentence comprehension (Tamaoka et al., 2005).

The peak Coxy-Hb on channel 32 was higher in the OSV condition compared with the SOV condition. The estimation of NIRS channel location indicated that channel 32 recorded activity in the Broca's area, which plays an important role in speech production. The activity in the Broca's area was assumed to increase with increasing processing load in sentence production. However, the brain activity differences might simply reflect differences in the number of words in the sentences between word order conditions. Thus, we compared the word counts between the SOV and OSV conditions. The results indicated that the word counts

were nearly identical between the SOV (M = 5.49) and OSV (M = 5.46) conditions (t (18) = 1.58, p = .13). Therefore, the results of the Coxy-Hb indicated that the processing load of the OSV condition was higher than that of the SOV condition during utterance.

Eye movements also differed between the SOV and OSV conditions. In the SOV condition, the agent-directed fixation bias reached a peak 700–800 ms after picture presentation, and then the relative fixation times were approximately equivalent between the agents and the patients during the 1800–1900-ms period. The utterance was initiated between this agent-peak period and the equivalent fixation time period. After the onset of utterance, the relative fixation time for the patient was higher than for the agent (2100–2900 ms), and then the agent-directed fixation bias was observed again until picture presentation offset. On the other hand, in the OSV condition, the patient-directed fixation bias appeared faster compared with the SVO condition (800–900 ms). The utterance was initiated during the equivalent fixation time period between the agents and the patients (1500–1800 ms).

The syntactically basic word order is SOV in Japanese. Therefore, the processing load can be expected to increase in OSV sentences compared with SOV sentences in both the sentence comprehension and production processes. The present results clearly indicated higher processing loads during sentence production for OSV sentences based on more appropriate indices, the utterance latency, brain activity, and eye movement data. Future studies will need to conduct the same investigation in other languages to confirm the generality of the present findings.

Acknowledgments

We are most grateful to the participants. We are also very grateful to Dr. Hiroshi Shibata for his kind cooperation to estimate the cortical regions corresponding to each NIRS channel. This research was supported by a Japanese Society for the Promotion of Science KAKENHI to MK (Grant-in-Aid for Scientific Research (S): No. 22222001 and Challenging Exploratory Research: No. 26580069).

References

Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544-569. doi:10.1016/j.jml.2007.01.007

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274-279. doi: 10.1111/1467-9280.00255

Horwitz, B., Amunts, K., Bhattacharyya, R., Patkin, D., Jeffries, K., Zilles, K., & Braun, A. (2003). Activation of Broca's area during the production of spoken and signed language: A combined cytoarchitectonic mapping and PET analysis. *Neuropsychologia*, *41*(14), 18681876. doi:10.1016/S0028-3932(03)00125-8

Hoshi, Y. (2003). Functional near - infrared optical imaging: Utility and limitations in human brain mapping. *Psychophysiology*, *40*(4), 511-520. doi:10.1111/1469-8986.00053

Imamura, S., & Koizumi, M. (2008). Bunrikai ni okeru jyouhou kouzou to tougo kouzou no kougosayou ga syoujiru

taimingu ni tuite [On the time course of the interaction between information structure and syntactic structure in sentence comprehension]. *Proceedings of the 137th Conference of the Linguistics Society of Japan*, 92-97. (in Japanese)

Imamura, S., & Koizumi, M. (2011). A centering analysis of word order in Japanese. *Tohoku Studies in Linguistics*, 20 59-74.

Kim, J., Koizumi, M., Ikuta, N., Fukumitsu, Y., Kimura, N., Iwata, K., Watanabe, J., Yokoyama, S., Shigeru, S., & Kawashima, R. (2009). Scrambling effects on the processing of Japanese sentences: An fMRI study. *Journal of Neurolinguistics*, *22*(2), 151166. doi:10.1016/j.jneuroling.2008.07.005

Kinno, R., Kawamura, M., Shioda, S., & Sakai, K. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Human Brain Mapping*, *29*(9), 1015-1027. doi:10.1002/hbm.20441

Lindsley, J. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, *7*(1), 1-19. doi:10.1016/0010-0285(75)90002-X

MacWhinney, B. (1977). Starting points. *Language*, *53*, 152-168.

Mazuka, R., Itoh, K., & Kondo, T. (2002). Costs of scrambling in Japanese sentence processing. In Nakayama, M. (Ed.) *Sentence processing in East Asian languages* (pp. 131-166). Stanford, CA: CSLI Publications.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113. doi:10.1016/0028-3932(71)90067-4

Otuska, Y., Nakato, E., Kanazawa, S., Yamaguchi, M. K., Watanabe, S., & Kakigi, R. (2007). Neural activation to upright and inverted faces in infants measured by near infrared spectroscopy. *NeuroImage*, *34*(1), 399-406. doi:10.1016/j.neuroimage.2006.08.013

Schroeter, M., Zysset, S., Kruggel, F., & Cramon, D. (2003). Age dependency of the hemodynamic response as measured by functional near-infrared spectroscopy. *NeuroImage*, *19*(3), 555-564. doi:10.1016/S1053-8119(03)00155-1

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*(3), 205246. doi:10.1016/S0010-0277(99)00053-0

Taga, G., Asakawa, K., Maki, A., Konishi, Y., & Koizumi, H. (2003). Brain imaging in awake infants by near-infrared optical topography. *Proceedings of National Academy of Sciences of the United States of America*, *100*(19), 10722-10727. doi:10.1073/pnas.1932552100

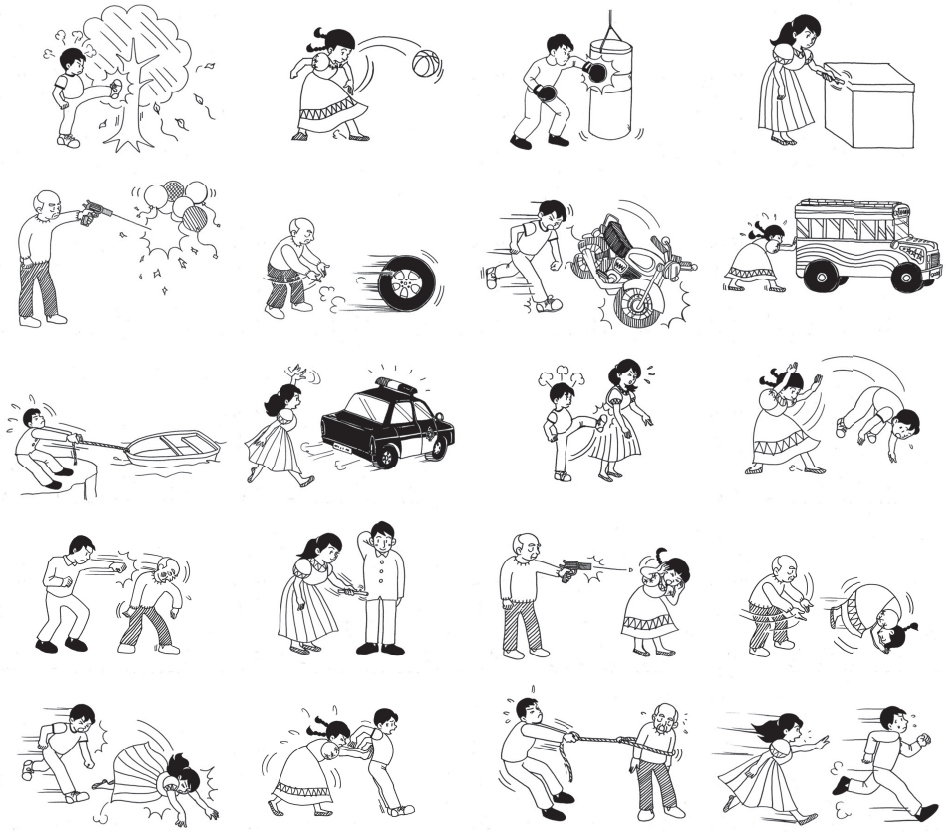
Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Image*. Stuttgart: Thieme Medical Publishers.

Tamaoka, K., Sakai, H., Kawahara, J., Miyaoka, Y., Lim, H., & Koizumi, M. (2005). Priority information used for the processing of japanese sentences: thematic roles, case particles or grammatical functions? *Journal of Psycholinguistic Research*, *34*(3), 281-332. doi:10.1007/s10936-005-3641-6

Tanaka, M., Branigan, H., McLean, J., & Pickering, M. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*(3), 318330. doi:10.1016/j.jml.2011.04.009

(Received December 12, 2014)
(Accepted January 5, 2015)

Appendix



All picture samples were used in the present study (In this case, the agent and patient are always depicted on the left and right sides, respectively).