One word at a time: Mental representations of object shape change incrementally during sentence processing

Manami Sato (Hiroshima University), Amy J. Schafer (University of Hawai‘i, Mānoa), and Benjamin K. Bergen (University of California, San Diego)

We report on two experiments that ask when and under what linguistic conditions comprehenders construct detailed shape representations of mentioned objects, and whether these can change over the course of a sentence when new information contradicts earlier expectations. We used Japanese because the verb-final word order of Japanese presented a revealing test case where information about objects can radically change with a subsequent verb. The results show that language understanders consistently generate a distinct and detailed shape for an object by integrating the semantic contributions of different sentential elements. These results first confirm that the tendency to generate specific shape information about objects that are involved in described events is not limited to English, but is also present in Japanese, a typologically and genetically distinct language. But more importantly, they shed light on the processing mechanism of object representation, showing that mental representations are initiated sentence medially, and are rapidly revised if followed by a verb that implies a change to an object shape. This work contributes to ongoing research on incremental language processing—comprehenders appear to construct extremely detailed semantic representations early in a sentence, and modify them as needed.

Key words: sentence processing, incremental processing, Japanese, mental representation

## 1. Introduction

Language comprehenders rapidly and incrementally construct and revise syntactic and semantic representations of ongoing utterances, which allows them to dynamically anticipate upcoming information and respond with changes in expectations (e.g., Altmann and Kamide 1999; Aoshima 2003; Aoshima et al. 2004; Frazier 1987; Mazuka and Itoh 1995; Sedivy et al. 1999). Beyond syntactic and semantic incrementality, it has recently been suggested that deeper comprehension processes—detailed interpretations of sentence meaning—are also incremental. Self-paced reading experiments have shown that comprehenders activate representations of actions implied by sentences, and they do so as soon as or immediately after discriminating information has been provided within the sentence (Taylor and Zwaan 2008; Zwaan and Taylor 2006). The idea that comprehenders may be incrementally constructing detailed representations of what they think described scenes would be like has only limited experimental support as of yet, but bears on several key issues in language processing. In this paper, we report on two experiments that seek to clarify how early within a sentence comprehenders construct detailed object representations for described events and whether representations can change over the course of a sentence when new information contradicts earlier information.

### 1.1. *Incremental processing of syntax and semantics*

Language comprehension is incremental; comprehenders integrate words that they see or hear into ongoing representations of the current utterance. There is rich evidence to support the idea of both semantic and syntactic incrementality (e.g., Altmann and Steedman 1988; Clifton 1993; Frazier and Rayner 1982; Gennari and MacDonald 2008; Levy 2008; Sedivy et al. 1999; Tanenhaus at al. 1995; Traxler et al. 1997). This is true in both head-initial languages like English and in head-final languages like Japanese (e.g., Aoshima 2003; Aoshima et al. 2009; Mazuka and Itoh 1995). For instance, recent eye-tracking studies in which participants process sentences while looking at highly constrained visual representations show that language comprehenders use linguistic information as well as visual and stored semantic knowledge that has been gained from world knowledge to anticipate upcoming non-verb entities in both English and Japanese, among other languages (Kamide et al. 2003; Knoeferle et al. 2005). Sentence processing involves a tight relationship among incremental syntactic and semantic operations and anticipation mechanisms that are able to draw on language understanders' knowledge and experiences.

### 1.2. *Mental representations*

In capturing the meaning of linguistic material, language comprehenders internally construct mental representations of described scenes—they activate mental representations of entities and events mentioned in the sentence (e.g., Barsalou 1999; MacWhinney 2005). Zwaan and colleagues (Stanfield and Zwaan 2001; Yaxley and Zwaan 2007; Zwaan et al. 2002) have presented evidence suggesting that by the time comprehenders complete the processing of an English sentence, they have produced detailed mental representations of the implied orientation, shape, and visibility of mentioned objects. However, these studies used a picture-probe task administered at the end of a sentence, which can produce results consistent with two possible states of affairs. It's possible that shape, orientation, and resolution information is constructed only at the end of the sentence (e.g., as part of a "wrap-up" process involving additional inferencing), or only when induced by the picture probes. On the other hand, it's equally possible that comprehenders activate mental representations of perceptual details as soon as an object is mentioned. Such a processing strategy would allow comprehenders to build richly detailed representations incrementally, but would also risk early commitment to details that could later be proven to be incorrect.

Glenberg and Kaschak (2002) have claimed that understanding language automatically involves activating knowledge about the potential actions that are associated with a mentioned object (e.g., the object affordances; Chao and Martin 2000; Tucker and Ellis 1988, 2004). For example, in the sentence *Diana opened the chocolate box*, the comprehender understands that various opening-movements are possible solely from understanding the verb *open*. However, an appropriate type of opening action will be selectively determined on the basis of the patient argument—opening a *refrigerator* typically indicates the retraction of the arm by an agent who is holding the handle of the refrigerator with one hand with a grasping hand-shape. On the other hand, opening *the bag of chips* implies an outward-horizontal movement with both hands. Previous studies on mental representations have argued that both perceptual and motor experiences underlie our cognitive ability to delineate the implied visual properties of objects or the appropriate motor properties of events described in utterances, and that such mental representations are driven by the tight link between linguistic knowledge and associated experiences in everyday life (Barsalou 1999, 2008; MacWhinney 1999; Zwaan 2004; Zwaan and Radvansky 1998).

Syntactic and semantic representations, then, are processed incrementally. But how much perceptual and motor detail is associated with such incrementality? And does the degree of detail vary with the specific linguistic content? In self-paced reading experiments, Zwaan and Taylor (2006) employed a manual knob-turning response where participants repeatedly turn a knob either clockwise or counter-clockwise in order to display the next word region of sentences presented on a computer screen (e.g., as in “He / realized / that / the music / was / too loud / so he / turned down / the / volume”). They found that native English speakers immediately activated direction-specific motor information when reading a verb region that implied a specific manual rotation (i.e., as soon as encountering “turned down,” counter-clockwise manual rotation is activated). This indicates that participants may be incrementally building internal representations of the manual event at the point where verb information is integrated. However, this evidence is relatively limited, in that it comes solely from English, has only been demonstrated for one type of motor representation, and, because it comes from looking at a verb region preceded by rich semantic context, focuses on whether people start activating such detailed motor event information when they have all of the relevant information. This leaves open the question of whether the verb, which plays an influential role in constructing semantic and syntactic representations, is an obligatory component for the activation of this level of detail in mental representations.

### 1.3. *The verb*

The verb is a major player in sentence meaning. In languages like English, in which verbs typically occur early in the sentence, verb-based information has a considerable impact on predictions of subsequent arguments, because it syntactically and semantically constrains the number of arguments and semantically specifies plausible and suitable referents (e.g., Altmann and Kamide 1999; Dahan and Tanenhaus 2004). Verb choice also may highlight some perceptual details of an object over others. For example, “carry the piano” may activate details related to the size and weight of the piano, while “play the piano” may activate details of sound properties or key texture (Barclay et al. 1974; Tabossi 1998; Zeelenberg et al. 2003).

Because of typological word order differences, languages differ in which syntactic and semantic information can be used as a cue to anticipate subsequent information. For example, in contrast with a head-initial language such as English, in a head-final language like Japanese, the verb typically occurs in sentence-final position, and cannot therefore be used to anticipate its arguments. Instead, arguments marked with different case markers, such as accusative or locative markers, are used to construct an argument structure consistent with the upcoming verb (e.g., Mazuka and Itoh 1995; Nakatani and Gibson 2010; Vasishth and Lewis 2006; Yamashita 1997). However, it isn't known whether comprehenders of head-final languages like Japanese construct detailed perceptual and motor representations of described objects even before they reach the verb. So on the one hand, it seems that, if incrementality is in fact a general and deep-reaching principle of comprehension, then just as these comprehenders build up early expectations about the syntactic and semantic structure of the sentence, so too should they activate detailed information of mentioned entities, even before they reach the verb. On the other hand, it's possible that the verb carries so much information that comprehenders wait until they've reached it to commit resources to a mental representation of the described entities. For example, Japanese speakers might not commit to weight versus musical features for a piano until they knew whether it was being played or moved. In this paper, we focus on the representation of the shapes of direct objects in Japanese sentences. Experiment 1 tests whether comprehenders activate mental representations prior to receiving the relevant verb, and finds that they do. Experiment 2 examines the revision of this shape information when the verb constrains the shape of an earlier mentioned object to be different from what is typically expected.

## 2. Experiment 1: Do comprehenders activate object representations with incomplete sentential information?

This experiment asks whether language comprehenders produce detailed mental representations of object shape with only partial information—even before reaching the verb. We presented native Japanese-speaking participants with Japanese sentences containing direct objects whose most plausible shape would change depending on its location, as in (1). After receiving the direct object and a location phrase, but before the verb, participants saw and responded to picture probes, which we used to test for the presence of a shape-specific representation of the direct object.

## 2.1. *Method*

2.1.1. *Participants.* Forty-eight native speakers of Japanese who were undergraduate or graduate students at the University of Hawai‘i participated in the main experiment. A different group of 28 native Japanese speakers participated in a series of norming studies. All participated in exchange for course credit in an introductory linguistics class, five dollars, or a small gift. All reported normal or corrected-to-normal hearing and vision.

2.1.2. *Sentence materials.* Twenty-eight pairs of critical sentences were created in Japanese, along with 56 filler sentences and 6 practice sentences. Critical sentences were constructed in pairs, in which each sentence mentioned a concrete object and implied one of two shapes for it, as in (1a, b). The mentioned concrete objects were ones that have a relatively canonical (or preferred) shape (as verified in a pretest, described in Section 2.1.4). In each pair, one sentence described an event consistent with the object taking its canonical shape (e.g., an unfolded shirt in 1a), while the other sentence described an event with the same object taking a non-canonical shape (e.g., a folded shirt in 1b). The string of asterisks (***) in these examples indicates the position where a picture probe (described in Section 2.1.3) appeared.

(1a) Canonical object shape (an unfolded shirt):

| | | | | | |
|---|---|---|---|---|---|
| *Haha-ga* | *shatsu-o* | *soto-ni* | *kitinto* | *** | *hoshita.* |
| Mother-NOM | shirt-ACC | outdoors-LOC | neatly | | dried |

‘Mother dried the shirt outdoors neatly.’

(1b) Non-canonical object shape (a folded shirt):

| | | | | | |
|---|---|---|---|---|---|
| *Haha-ga* | *shatsu-o* | *hikidashi-ni* | *kitinto* | *** | *ireta.* |
| Mother-NOM | shirt-ACC | drawer-LOC | neatly | | put |

'Mother put the shirt in the drawer neatly.'

The precise shapes of the objects were never explicitly mentioned, so that if comprehenders were indeed accessing shapes, they would have to be doing so by combining the information in the direct object and location phrases, such as by activating location-appropriate detail for the direct object. For example, in (1a) processing *shirt* and *outdoors* might lead the comprehender to access an appropriate contextual situation (i.e., drying the laundry outside, which is quite common in Japanese culture), and to produce a mental representation of an unfolded shirt; while in (1b) the combination of *shirt* and *in the drawer* provides a context that might evoke a mental representation of a folded shirt.

In order to prevent participants from strategically paying attention to one specific word in a particular position (e.g., the second item) instead of integrating the meaning of all the preverbal phrases, we varied the word order of critical stimuli. Two different word orders were created for critical items by varying the position of the direct object (DO) and locative phrase (Loc), between items (i.e., Subject DO Loc Adverb *** Verb and Subject Loc DO Adverb *** Verb). Both phrase orders are frequent and natural in Japanese (Mazuka et al. 2002; Miyagawa 1997; Miyamoto and Takahashi 2002). In addition, in order to reduce the likelihood that participants would notice that the critical items were constructed with similar sentence structures, the locative phrases were marked by a variety of locational markers such as *-ni* 'in', *-e* 'to', *-de* 'in', *-nouede* 'on', and *-nonakade* 'inside' (direct object phrases exclusively allow the accusative marker, *-o*). This variation in locative markers was expected to make it hard for participants to notice similarities among critical sentences using superficial morphological cues. Furthermore, in order to increase sentential variability, a range of sentence structures was used in the fillers (e.g., NP PP DO PP *** V and NP $_{CP}$[PP DO V] *** V). This prevented participants from anticipating that the pictures would always come after the adverb, as they do in critical items.

In order to have an equal number—42 each—of "yes" and "no" responses to the picture probes, 28 pairs of critical sentences and 14 filler sentences mentioned the following pictured object to elicit "yes" responses, and 42 filler sentences were constructed that were followed by unrelated pictures to elicit "no" responses. Among the sentences constructed to elicit "yes" responses, all pictures for the critical sentences were of objects referred to by direct objects, while filler pictures were of objects referred to by locative or prepositional phrases. This was intended to avoid the possibility that participants would learn that the direct object was more likely to describe the pictured items, and pay more attention to sentential constituents marked with the accusative case marker *-o* for the purpose of the image-matching task, or strategically activate detailed object information just for direct objects.

2.1.3. *Picture materials.* Twenty-eight pairs of critical images that matched the shape of the described objects in the 28 pairs of critical sentences were obtained from a popular clip-art program. Specifically, each pair of critical pictures depicted two different states of an object, one in the canonical (i.e., the most naturally or preferably accessed and frequently observed) shape of the object, and the other in a non-canonical (i.e., recognizable or identifiable, but not the most typical) shape of the object, resulting in a total of 56 critical pictures. For example, for an apple, a round, un-cut apple would be a canonical shape, while a sliced apple would be in a non-canonical shape. Although previous studies using a similar paradigm (for instance, by Zwaan and colleagues) generally do not look specifically at canonicality, we separated the conditions by canonical shape so we could evaluate whether eventual object representation effects would apply to both types of images, or whether they were driven only by preferred or dispreferred shapes or image-meaning pairings. An additional 56 pictures of format similar to the criticals were selected as probes for filler items.

2.1.4. *Three norming studies.* A group of six native Japanese speakers who did not participate in the main experiment verified the quality of each pair of critical pictures through a picture naming study. Participants freely labeled randomly presented pictures, one at a time, to ensure that each picture clearly depicted the intended object regardless of its shape canonicity. There was no effect of shape canonicity in object naming accuracy (canonical, 94.0%, non-canonical, 91.7%, $t_1$ (1,5) = 0.79, $p$ = 0.47, $t_2$ (1,27) = 0.8, $p$ = 0.42). That is, non-canonical shapes were as likely to generate the intended label as canonical shapes. Then, another eight native Japanese speakers read the sentences we created up to but not including the sentence-final verb, and were asked to select which they considered to be the most appropriate of the two selected images, one of which showed the canonical shape of a single object (e.g., an unfolded shirt) and the other a non-canonical shape for that object (e.g., a folded shirt). This ensured that participants would be able to selectively generate the implied canonical or non-canonical shapes of an object by comprehending the sentence with no verb-based information. Participants chose the intended image for 95.5% of the canonical sentences and 97.3% of the non-canonical sentences, resulting in no difference in image selection accuracy between conditions: $t_1$ (1,7) = 1.0, $p$ = 0.35, $t_2$ (1,27) = 0.81, $p$ = 0.42. Finally, another eight native Japanese speakers who did not participate in the main study or either of the other norming studies read isolated target words (e.g., *shirt*) and selected the best matching image, again from the paired canonical and non-canonical images of a single object. Participants chose the canonical image 90.2% of the time and the non-canonical image 9.8% of the time.

2.1.5. *Procedure.* This study employed self-paced reading in Japanese with an interleaved picture verification task. Each participant first experienced six trials in a practice session. After becoming familiar with the task, participants started the experimental session, which consisted of 28 critical trials and 56 fillers, randomly ordered.

Each trial started with a fixation cross that appeared for 250 milliseconds in the center of the screen. Then a sentence was presented visually in a self-paced fashion. After the offset of the fixation cross, the first word of the sentence automatically appeared on the left side of the screen. When participants pressed the space bar, the first word disappeared and the next word appeared directly to the right of the position where the previous word had appeared. Experimental software automatically measured length of reading time (RT) from each press of the space bar to the next. After reading the phrase that immediately preceded the final verb, participants automatically received another centered fixation cross that appeared for 250ms, in order to redirect their attention to the center of the screen and prepare them for a picture stimulus. The cross was then replaced by the picture, and participants had to indicate whether or not the pictured item was mentioned in the sentence that they had just been reading, by pressing one of two buttons, for either “yes” or “no”, as quickly and accurately as possible. As soon as participants pressed the button in order to respond to the picture, the final verb automatically displayed, resuming the moving-window display of the sentence (i.e., it was immediately to the right of the most recently displayed phrase). When participants finished reading the final verb, they pressed the space bar. This initiated the beginning of the next trial.

Participants were expected to respond “yes” to all 28 critical sentences, disregarding objects’ specific shapes (e.g., folded or unfolded), since in all cases the pictured item had been mentioned in the preceding sentential material. The practice session ensured that participants understood this intent. Simple comprehension questions regarding the direct objects (e.g., “Did mother put the *carrot* there?”), prepositional phrases (e.g., "Did mother put them in the *basket*?"), and verbs (e.g., "Did mother *see* the shirt?") randomly appeared after the completion of half of the criticals and half of the filler sentences. These comprehension questions ensured that participants attended to the meaning of the entire sentence, and prevented them from simply trying to remember a particular linguistic element in the sentence in order to respond to the picture stimuli. No feedback was provided for the responses to the pictures or comprehension questions in the main session. Picture verification response choices, picture reaction times, and reading times for each linguistic region of the sentence were recorded by the experimental software for subsequent analysis. Participants took approximately 15 to 20 minutes to complete the entire experiment.

2.1.6. *Design.* Critical items were tested in four conditions, which crossed two sentence types, describing either a canonical or a non-canonical shape of an object, with two picture types, depicting either the canonical or a non-canonical shape. The conditions for each item were distributed across four lists in a counterbalanced design (a Latin Square) so that each participant received each item in only one of the four conditions, and each condition was presented to each participant in equal proportions. Each participant responded to 28 critical items (i.e., 14 sentences that matched the pictures in object shape and 14 sentences that were mismatched with the pictures in object shape) randomly mixed with 56 fillers.

2.1.7. *Predictions.* We predicted that if language comprehenders incrementally construct detailed mental representations of the objects' shapes, then they should already be representing an object with a contextually appropriate shape in sentence medial position (i.e., preverbally) if sufficiently compelling contextual information is provided. Therefore, the canonical shape of a pictured object should be verified faster after reading a sentence that contextually evokes the object's canonical shape than after reading the sentence that evokes a non-canonical shape, while the reverse should be true for the non-canonical pictures.

2.2. *Results*

No subjects were excluded, but four critical items were eliminated from the results because of low accuracy (mean lower than 85%) in the responses to the pictures or the comprehension questions, or because mean picture responses were extremely slow for the item (i.e., they were more than 3 *SD* above the grand mean). In addition, incorrect responses and extremely slow responses (those over 5000 ms) were removed, and responses that were more than 3 *SD* above or below each participant's mean were replaced by the mean $\pm$ 3 *SD* for each participant. Removed data constituted less than 2% of all data, and replaced data constituted less than 1% of all data. Accuracy in picture responses for the remaining items averaged 97.5%; accuracy for each condition is given in Figure 1.

In order to assess whether there could be a speed-accuracy tradeoff present in these data, an error analysis was conducted. We ran Repeated-Measures ANOVAs with picture accuracy as the dependent measure and Sentence and Picture as the independent measures. These revealed no significant main effects of Sentence or Picture, nor an interaction effect (all *F*s < 1). Since we are primarily concerned with the interaction of Sentence and Picture type on reaction time, this analysis confirms that verification accuracy does not limit the interpretability of the reaction time measure.

Turning to the key measure of reaction time to the picture, two-way Repeated-Measures ANOVAs revealed no main effect of Sentence in the subject analysis ($F_1(1,47) = 1.5$, $p = 0.23$, $\eta^2_p = 0.03$) or in the item analysis ($F_2(1,23) = 0.7$, $p = 0.4$, $\eta^2_p = 0.03$). There was a main effect of Picture in the subject analysis ($F_1(1,47) = 7.7$, $p = 0.008$, $\eta^2_p = 0.14$), which was not reliable in the item analysis ($F_2(1,23) = 1.9$, $p = 0.19$, $\eta^2_p = 0.08$). Crucially, however, there was a large, significant interaction between Sentence and Picture both in subject and item analyses ($F_1(1,47) = 12.7$, $p = 0.001$, $\eta^2_p = 0.21$; $F_2(1,23) = 31.2$, $p < 0.001$, $\eta^2_p = 0.58$). That is, when the shape of the pictured object matched the one implied in the prior sentence, the verification times for pictures were faster than when they mismatched (Mean = 1032 ms for Match, Mean = 1195 ms for Mismatch; $t_1 = 3.6$, $p = 0.001$, $t_2 = 5.6$, $p = 0.00001$). Furthermore, as shown in Figure 1 below, subsequent pairwise *t*-tests revealed that participants verified canonical pictures significantly faster while reading a sentence inducing a canonical shape of an object than when reading a sentence implying a non-canonical shape (Mean = 1002 ms for Match (sd = 42; accuracy = 97.8), Mean = 1123 ms for Mismatch (sd = 53; accuracy = 97.4); $t_1 = 2.8$, $p = 0.008$; $t_2 = 2.3$, $p = 0.03$). Likewise, non-canonical pictures were processed significantly faster during sentences that evoked a non-canonical object shape than during their counterparts that evoked canonical shapes, in both subject ($t_1 = 3.0$, $p = 0.004$) and item analyses ($t_2 = 3.4$, $p = 0.003$; Mean = 1061 ms for Match (sd = 49; accuracy = 98.0), Mean = 1266 ms for Mismatch (sd = 77; accuracy = 97.4)).

Overall, these results show that responses were significantly faster when the picture and sentence matched in object shape than when they did not match. Further, extending previous results for match-mismatch effects, the pattern occurred sentence-medially, and held both for canonical shapes and for non-canonical shapes.

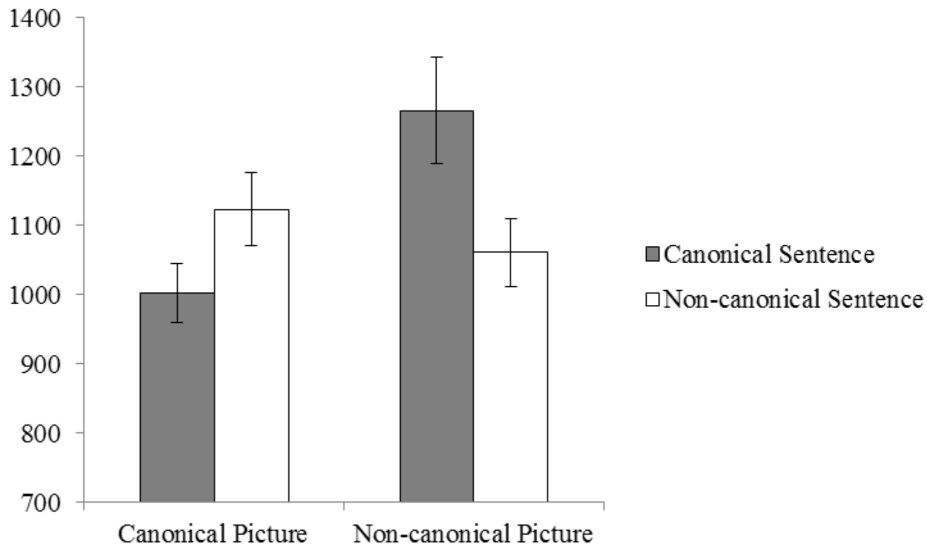


Figure 1. Mean response times (milliseconds) and standard error for verification of pictures that appear *before* the final verb in Experiment 1. Error bars indicate standard error of the difference between the means of canonical and non-canonical sentences in each picture type.

2.3. *Discussion*

The results from Experiment 1 demonstrate that comprehenders mentally represent the implied shape of a mentioned object in the middle of a sentence, and they do so even before being presented with a verb. To do this, they use the direct object and a locative phrase to construct a contextually appropriate shape, which then facilitates their response times to matching picture probes. These match-advantage results are consistent with previous research in English using similar methods, but in which the picture verification took place after the end of the sentence and could potentially reflect sentence wrap-up effects (Stanfield and Zwaan 2001; Yaxley and Zwaan 2007; Zwaan et al. 2002). They are also consistent with other work showing that comprehenders construct detailed mental representations on a word-by-word basis (e.g., Taylor and Zwaan 2008; Zwaan and Taylor 2006). Uniquely, this experiment provides empirical evidence that language comprehenders do not wait until they receive all critically informative sentence material, but instead decide upon the implied shape of the critical object prior to receiving the information contributed by the verb. This is a strong test of incrementality, since the verb can dramatically influence the semantic content; very different object shapes are suggested by “Mother dried the shirt outdoors” and “Mother wrung out the shirt outdoors.” Thus, it is striking that comprehenders anticipate a likely shape even before the verb.

The results from Experiment 1 show that language comprehenders construct shape-specific mental representations of objects prior to the verb, in the middle of a sentence. But what happens next? How do these representations interact with information from subsequent sentence constituents, like verbs? A verb need not always be consistent with the mental representation a comprehender constructs prior to that verb. To this end, in Experiment 2, we investigated whether the verb can radically alter the content of existing shape representations.

## 3. Experiment 2: Do representations of object shape immediately change when a sentence ends with an unexpected verb?

Experiment 1 demonstrated that comprehenders construct object representations corresponding to inferred visual properties of mentioned objects, over the course of sentence processing from partial linguistic information. To do this, they retrieve semantic information triggered by different sentential constituents (e.g., a combination of a direct object and location phrase) that provide implicit information about the contextually appropriate shape of an object (Sedivy et al. 1999) and combine them. Our results suggest that there are temporal dynamics in the process of constructing a detailed mental representation of a described scene, just as there are temporal dynamics in the incremental processing of syntax and semantics.

While anticipation might be an important factor for successful communication in real time, sentences often take unexpected turns. For example, take the Japanese sentence shown in (2).

(2) *Hanabi-taikai-notameni kat-ta yukata-o kirisai-ta.*

fireworks-festival-for buy-PAST *yukata*-ACC tear apart-PAST

'I tore apart the *yukata* (a cotton kimono) that I bought for the fireworks festival.'

As a native speaker reads a sentence like this one, he or she anticipates the upcoming verb. For this sentence, several verbs could be contextually appropriate (e.g., *nagameta* 'looked at', *kashita* 'lent out', *misebirakashita* 'showed off'), but the most likely verb that can be expected from the preverbal information may be *kita* 'wore'. However, confronted instead with the actual verb, *kirisaita* 'tore apart', comprehenders are able to abandon the initially anticipated verb and incorporate the meaning contributed by the unexpected verb. The question is whether the process of constructing a mental representation of the described scene follows this same trajectory. That is, in their mental representations, do comprehenders initially generate a mental representation of a brand-new, whole *yukata* that is most plausible prior to the verb, but immediately change its shape to torn-up pieces of *yukata* as soon as they process the verb implying this modification? If each semantic component is incrementally integrated to build a representation of the scene, then the final verb should change the developing shape representation of the object, making it fit the final, full context.

Alternatively, it could be difficult to revise existing object representations, so that the original ones tend to linger, as has been argued for thematic roles (Christianson et al. 2001). If this is the case, then the original representation could tend to dominate, or both the original representation and the revised representation could show some degree of activation, in much the same way that both factual and counterfactual scenes are accessed in understanding counterfactual and negated sentences (Byrne and Johnson-Laird 2002, 2009; Kaup et al. 2006). Simultaneous activation of both canonical and implied colors of a single object has also been shown (Connell 2007; Connell and Lynott 2007), and it is plausible that shape could work the same way, especially when strongly canonical and clearly implied shapes are both in play.

### 3.1. *Method*

The design of this experiment was the same as Experiment 1, except that unexpected verbs were used and the picture stimuli appeared after the verb, instead of before it.

3.1.1. *Participants*. Fifty-six native Japanese speakers participated in the experiment. Twenty participants were undergraduates at the University of Hawai‘i, who participated in exchange for course credit in an introductory linguistics class, and the remainder were undergraduates at Hiroshima University in Japan, who received a small monetary compensation. In addition, 16 native Japanese speakers at the University of Hawai‘i participated in two norming studies. All participants reported normal or corrected-to-normal hearing and vision.

3.1.2. *Sentence materials*. We constructed 16 pairs of critical sentences that resembled those in Experiment 1 as closely as possible up to the verb, using the same objects as in Experiment 1. Each pair of critical sentences described the same single object, but each implied a different shape. Experiment 2 was different from Experiment 1 in that the verbs in these sentences were selected to cause an unexpected modification in the implied shape of the object. For example, based on Experiment 1, upon reading the sentence in (3a), participants were expected to incrementally generate an object with a specific shape (i.e., a whole egg) prior to the verb. However, in Experiment 2, participants received an unexpected verb, such as *dropped* in (3a), that might immediately force them to reconsider the egg's implied shape to that of a broken egg. Each pair of critical sentences was created to include two types of sentences. One sentence induced canonical shapes of object representations (e.g., an unbroken egg) prior to the verb, to be ultimately revised to non-canonical shapes (e.g., a broken egg) when the verb appeared, as in (3a); the other sentence in each pair motivated the opposite change, from non-canonical to canonical shapes, as in (3b).

The verb changes the implied shape from an unbroken egg to a broken egg:

| | | | | | | |
|---|---|---|---|---|---|---|
| (3a) | *Nana-ga* | *reezooko-nonakani* | *tamago-o* | *subayaku* | *otoshita* | *** |
| | Nana-NOM | refrigerator-LOC | egg-ACC | instantly | dropped | |

'Nana dropped an egg in the refrigerator instantly.'

The verb changes the implied shape from a broken egg to an unbroken egg:

| | | | | | | |
|---|---|---|---|---|---|---|
| (3b) | *Nana-ga* | *huraipan-nouede* | *tamago-o* | *subayaku* | *korogashita* | *** |
| | Nana-NOM | pan-LOC | egg-ACC | instantly | rolled | |

'Nana rolled an egg in the pan instantly.'

For the same reasons discussed in Experiment 1, the positions of the direct objects (DO) and the locative phrases were counterbalanced between items so that half of the critical sentences were constructed with a DO followed by a locative phrase while the other half were structured with a locative phrase followed by a DO. Thirty-two filler sentences were newly constructed with a variety of syntactic structures that involved expected as well as unexpected verbs, so that participants would not recognize the distinctiveness of the critical items. For example, in (4) below, Hajime's expected action toward the glove that his mother knitted for him would be *wearing* it, rather than *stepping* on it. Note that the final verb in fillers might unexpectedly change the event as a whole, but none of the filler sentences required dramatic revision of an object's shape. Therefore, the object's shape remained consistent before and after integrating the verb information.

(4) *Hajime-wa* *haha-ga* *annda* *tebukuro-o* *hunda.*
Hajime-TOP mother-NOM knit glove-ACC stepped
'Hajime stepped on the glove that his mother knitted for him.'

3.1.3. *Picture materials.* Sixteen pairs of critical images and 32 filler images that were similar to the criticals were obtained from a popular clip-art program, as in Experiment 1. Each participant saw a total of 48 stimuli that were expected to produce 24 "yes" (i.e., 16 criticals and 8 fillers) and 24 "no" (i.e., all remaining fillers) responses. All filler sentences that anticipated "yes" responses involved images of objects that were referred to in locative phrases.

3.1.4. *Two norming studies.* We conducted two norming studies: a picture selection task and a sentence completion task. First, in the picture selection study, eight native Japanese speakers who did not participate in any other main or norming task read sentence fragments for each critical sentence without a final verb and then selected one of two shapes (e.g., a broken vs. an unbroken egg) for an object that was most appropriate for what they had read. Participants selected intended canonical images with 93.8% accuracy and intended non-canonical images with 96.9% accuracy. The results confirmed that all sentence fragments prior to the final verbs evoked the intended shapes of objects. This indicated that sentence portions prior to the verbs implicitly designated specific shapes of objects. Second, another group of eight native Japanese speakers read complete sentences that were composed of the sentence fragments used in the first norming study and the unexpected verbs that were supposed to cause the image revision. As in the first norming study, participants were instructed to select the most suitable shape of the object. The results showed that they chose different shapes from those selected in the first norming study (canonical, 93.8%, non-canonical, 95.3% accuracy: no difference between conditions: $t_1$ (1,7) = 0.42, $p$ = 0.69, $t_2$ (1,15) = 0.37, $p$ = 0.72). This suggested that the verbs successfully created scenes where the implied objects' shapes were different from those generated prior to the verbs.

3.1.5. *Procedure*. The procedure was identical to that in Experiment 1, except that the picture appeared after the verb, instead of before it.

3.1.6. *Predictions*. If each linguistic component contributes to the ongoing development of visual representations, then encountering a verb that changes the implied shape of an object should result in a modified object representation. Further, if constructing a mental representation is a rapid and dynamic process, this change should result in an immediate shift to the object representation that fits the verbal information. Therefore, regardless of the initial mental representation, the response times to the images are expected to be faster when the pictured shape matches the meaning denoted not by the preverbal (i.e., Subject, Direct object, and Locative phrase) information (alone), but by that information as filtered through the meaning of the verb. Below, we will use "Canonical Sentence" to refer to sentences in which the verb implies the canonical shape of the critical object, and "Non-canonical Sentence" to refer to the condition in which it implies the non-canonical shape.

### 3.2. *Results*

The data were trimmed as in Experiment 1. The removed data constituted less than 3% of all critical data. No subjects or items were eliminated.

Repeated-Measures ANOVAs with picture accuracy as the dependent measure and Sentence and Picture as the independent measures showed no significant main effects of Sentence ($F_1 < 1$; $F_2(1,15) = 1.26$, $p = 0.27$, $\eta^2_p = 0.02$) or Picture (*Fs* < 1), nor an interaction effect (*Fs* < 1). Overall accuracy rate was 97.5%.

Two-way Repeated-Measures ANOVAs with reaction time as the dependent measure revealed no main effect of Sentence across participants ($F_1(1,55) = 3.6$, $p = 0.062$, $\eta^2_p = 0.06$) or items ($F_2(1,15) = 1.8$, $p = 0.2$, $\eta^2_p = 0.11$), although there was a marginal effect in the by-subjects analysis, where we observed shorter reaction times for Non-canonical sentences (apparently due to the shorter responses for Non-canonical sentences with Canonical pictures, as shown in Figure 2). A significant main effect of Picture was found both in the subject ($F_1(1,55) = 24.1$, $p < 0.001$, $\eta^2_p = 0.3$) and in the item analyses ($F_2(1,15) = 8.8$, $p = 0.01$, $\eta^2_p = 0.37$), with shorter times for Canonical pictures. Of particular interest in this experiment was whether the pictured object was verified faster when it matched the object shape implied by the verb or by the pre-verbal material. The Sentence-by-Picture interaction was a large, significant effect ($F_1(1,55) = 13.4$, $p = 0.001$, $\eta^2_p = 0.2$; $F_2(1,15) = 19.9$, $p < 0.001$, $\eta^2_p = 0.6$), as shown in Figure 2 below. Subsequent pairwise *t*-tests confirmed that participants verified canonical pictures significantly faster after sentences whose verbs implied objects in their canonical shape than after sentences whose verbs implied objects in a non-canonical shape (Mean = 789 ms for Match (sd = 28; accuracy = 96.9), Mean = 909 ms for Mismatch (sd = 51; accuracy = 99.1); $t_1 = 2.5$, $p = 0.016$; $t_2 = 2.6$, $p = 0.021$). Non-canonical pictures were processed significantly faster after sentences whose verbs implicitly denoted non-canonical shapes of objects than after sentences whose verbs implied the objects' canonical shapes (Mean = 859 ms for Match (sd = 40; accuracy = 96.0), Mean = 1082 ms for Mismatch (sd = 57; accuracy = 98.2); $t_1 = 3.75$, $p = 0.0004$; $t_2 = 3.8$, $p = 0.0017$).

In sum, the patterns of the data in this experiment were in a way very similar to those found in Experiment 1: responses were significantly faster when the picture matched the object shape implied by the sentence than when it did not. But the cause for this match-advantage was different. The same sentence onsets that produced faster responses to canonical images in Experiment 1 produced faster responses in Experiment 2 to non-canonical images (and conversely for those that produced faster responses to non-canonical images in Experiment 1), solely due to the verb. This pattern of responses thus provides no compelling evidence that pre-verbal mental representations lingered after the verb was read, but instead is completely consistent with mental representation of object shape being highly dynamic and quickly updateable. As in Experiment 1, match-mismatch effects occurred for canonical pictures as well as non-canonical pictures.

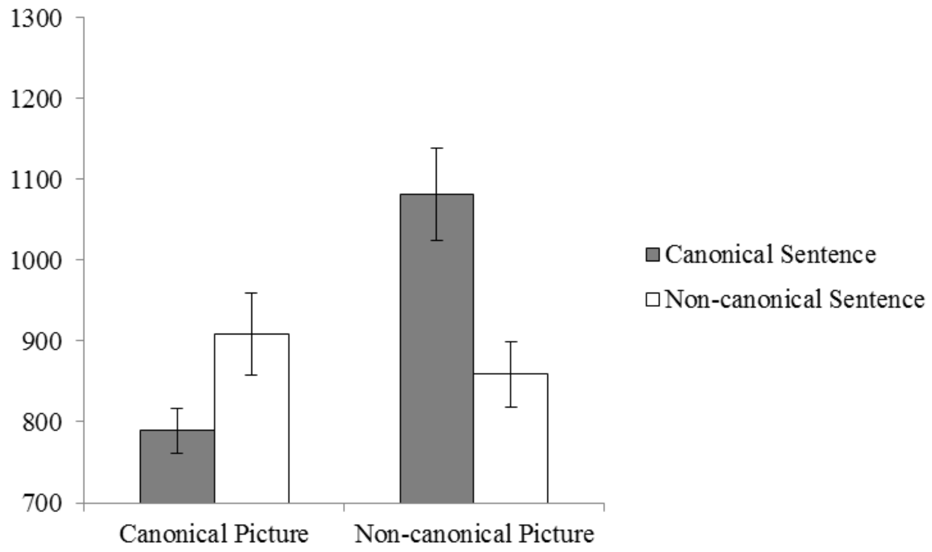


Figure 2: Mean response times (milliseconds) and standard error for verification of pictures that appear *after* the unexpected verb in Experiment 2. Error bars indicate standard error of the difference between the means of canonical and non-canonical sentences in each picture type.

3.3. *Discussion*

The results from Experiment 2 suggest that an object's detailed shape, initially constructed prior to the verb, can be immediately changed to fit the final, full sentential meaning as soon as the final verb is processed. This interpretation rests on a fundamental assumption: that comprehenders produce initial shape representations in sentence-medial position and then revise them as the verb requires. This assumption is supported by the results of Experiment 1, which showed that people evoke detailed representations of shape before receiving verb-based information.

However, there is another plausible interpretation of the effects found in Experiment 1, which is that the effects arose not because people routinely construct detailed mental representations of object shape, but because when they received pictures in the task they were forced to induce mental representations. In other words, it is possible that shape information was only accessed at sentence-medial position because participants had to respond to the picture verification task. If the effects were observed due to this specific task requirement, it therefore might be possible that, in Experiment 2, where a picture appeared at the sentence-final position, people had not in fact already developed mental representations of object shape, even though enough information to build the detailed shape of an object had been provided. Instead, they may have waited until they had received the target picture in the post-verbal position, and at that point, constructed an appropriate object representation for the first time. From such a perspective, the combination of results from Experiment 1 and Experiment 2 is still illuminating, because it suggests that even in such a probe-induced situation, image processing is evaluated against a sophisticated linguistic analysis that reflects the contributions of the verb to the most plausible object shape, and not a more superficial process that would also give strong weight to the information in the locative phrase. To explore the question of whether, in Experiment 2, initial visual representations were developed in sentence-medial position and eventually modified in order to fit the contextual situation, we looked at self-paced reading times (RTs) at two critical linguistic regions (and not the reaction times in response to the picture probes that we've already considered) from Experiments 1 and 2. These are presented in Section 4 below.

## 4. Analysis

As noted above, the linguistic stimuli for all experiments were presented in a self-paced moving-window display. This allowed us to collect reading times for each region of the sentences, which are informative to the question of whether comprehenders were only constructing mental representations in response to pictures, or whether they were constructing them online during sentence processing. We began by comparing the RTs for the first critical object-shape component (either DO or locative phrase) in Experiments 1 and 2. At this point, participants should generate the most accessible representation of the entity mentioned in the phrase. Because comprehenders have only received one of the two critical object-shape phrases, the first region does not contain sufficient information to establish a sentence-specific shape. For the half of the stimuli in which this region presented the DO, comprehenders should generate a canonical shape, whether the sentence goes on in the next phrase to support a canonical DO shape or a non-canonical one. For the other half of the stimuli, they should generate a mental representation of the mentioned location. Note that for this half of the items, lexical content varied between the canonical and non-canonical conditions, and so reading times might reflect simple lexical effects such as word frequency. But on the assumption that all other factors are balanced for the locative phrases, the RTs at this first region should be similar in canonical shape sentences and in non-canonical shape sentences, in both experiments. As predicted, we found no significant difference between the two categories of sentences in either experiment (Experiment 1: Mean = 666 ms for canonical shape sentences, Mean = 709 ms for non-canonical shape sentences, $t_1 = 1.7$, $p = 0.1$; $t_2 = 1.5$, $p = 0.2$. Experiment 2: Mean = 556 ms for canonical shape sentences, Mean = 562 ms for non-canonical shape sentences, $t_1 = 0.5$, $p = 0.6$; $t_2 = 0.1$, $p = 0.9$).[1]

Second, we compared the RTs for the last DO or locative phrase preceding the verb in Experiments 1 and 2. This position is the first point where we expect processing

[1] Reading times are globally different between the two experiments, perhaps because of the different participant samples, or perhaps reflecting the simpler task in Experiment 2, in which pictures follow the sentence, compared to Experiment 1, in which pictures are interleaved with the sentential material. These global differences in response time were also seen in the responses to picture probes (mean = 1113 ms in Experiment 1; 910 ms in Experiment 2).

effects due to combinatorially constructed shape representations, because it is here that participants should have received enough linguistic information to produce the implied shape of the direct object in their mental representations. In Figure 3, reading times for this region are shown separately for sentences that induced canonical shapes prior to the verb and sentences that induced non-canonical shapes of direct objects at the preverbal position. If people incrementally constructed a mental representation of the described object before pictures appeared in the experimental task, and if canonical shapes take less time to mentally represent than non-canonical ones, then reading times at this region for sentences that evoked objects' canonical shapes (according to the norming studies) should be faster than those for sentences that evoked non-canonical shapes in both Experiments 1 and 2.

As Figure 3 shows, we found similar advantages at this position for canonical-shape inducing fragments in the two experiments. In Experiment 1, RTs were significantly faster for sentence fragments evoking canonical shapes than for those evoking non-canonical shapes (Mean = 687 ms for canonical shape sentences (sd = 25), Mean = 729 ms for non-canonical shape sentences (sd = 26)) in the subject analysis ($t_1 = 2.4$, $p = 0.02$), although the difference did not reach significance in the item analysis ($t_2 = 1.0$, $p = 0.3$). The lack of effect in the item analysis may have been due to differences in the lexical content of the locative phrases, which could appear both as initial processing effects for DO-Loc order items or spill-over effects for Loc-DO order items. Crucially, although no picture verification was required sentence-medially in Experiment 2, reading times for phrases that generated canonical shapes were also significantly faster than their counterparts (Mean = 522 ms for canonical shape fragments (sd = 17), Mean = 552 ms for non-canonical shape fragments (sd = 21); $t_1 = 2.2$, $p = 0.04$; $t_2 = 2.8$, $p = 0.01$). This supports the assumption that people processed sentential material faster when it was consistent with preferred, canonical shapes, because comprehenders generated object shape representations incrementally during sentence processing, regardless of the task differences, and prior to presentation of the picture.

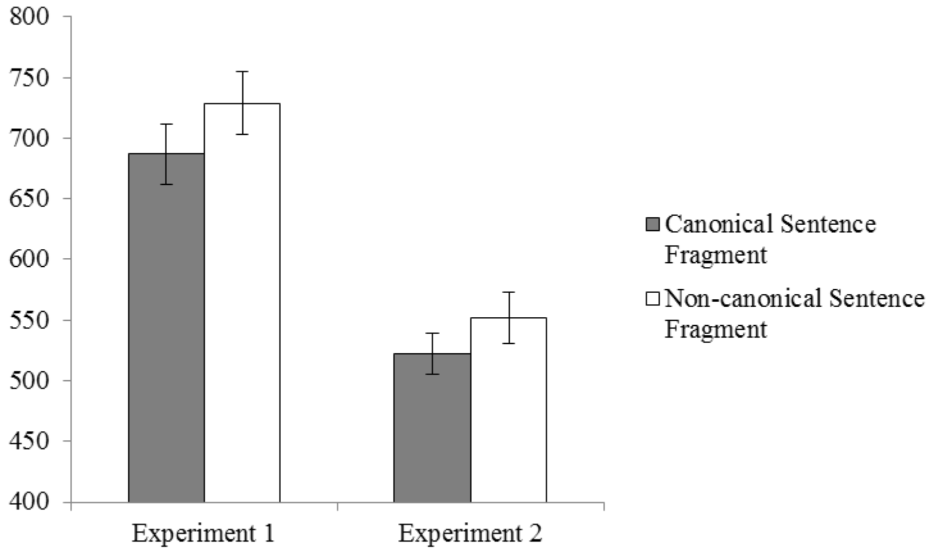


Figure 3: Mean reading times for the *critical object-shape region preceding the verb* (i.e., either the DO or the PP) in Experiments 1 and 2. Error bars indicate standard error.

Other authors (e.g. Ditman et al. 2010) have also argued using different empirical demonstrations that mental representation generation during language comprehension occurs spontaneously and is independent of specific task demands, such as those imposed by picture verification. Considering together all of the evidence for rapid and dynamic sentence comprehension, we believe that the simplest explanation for the match-mismatch effects found sentence-medially in Experiment 1 and both sentence-medially (from reading times) and sentence-finally (from picture decisions) in Experiment 2 is that language comprehenders routinely and incrementally activate representations of object shape detail during sentence comprehension, as an integral part of the sentence comprehension process.

## 5. General discussion and conclusion

Results from two sentence-processing experiments suggest that language understanders consistently generate distinct and detailed mental representations of object shapes by integrating the semantic contributions of different sentential elements. This demonstration in Japanese expands on the findings from English and confirms that the tendency to activate detailed information about objects that are involved in described events is not limited to head-initial, object-centered, and satellite-framed languages (Choi and Gopnik 1995; Gentner 1982; Talmy 1991; Zwaan et al. 2002), but is also present in a head-final, verb-centered, and verb-framed language, Japanese.

More importantly, the word order of Japanese allowed us to test an unexplored temporal aspect of the mental representation processes. Specifically, this study has shed light on the processing mechanism of mental representations by investigating it at both preverbal (or sentence-medial) and postverbal (or sentence-final) positions in order to learn more about the dynamics of mental representation construction.

What kinds of cognitive processes underlie the generation of detailed mental representations of objects during the ongoing processing of a sentence? One possible mechanism would be the accrual of shape probability with each part of the sentence that the comprehender encounters. The direct object noun might lead to activation of a representation of a canonically shaped object, or of both a canonical and non-canonical one, perhaps with different strengths. A locative phrase might influence the strength of activation of these shape representations, such that the one appropriate to that location gains increased strength. All of this could happen with or without information provided by a verb, and if information from the locative and accusative phrases is sufficiently constraining, people would reliably generate a consistent mental representation. A second possibility is that, during sentence processing, participants do not directly construct shape information from non-verbal constituents, but only do so through the intermediate step of anticipating a contextually plausible verb, based on the semantic information in the already-processed preverbal words or sentence fragments. The anticipated verb would then be used to help generate a detailed, appropriate shape of the object before the comprehender encounters the verb.

Previous research conducted in English has successfully revealed some of the details of language-based mental representations, but they have predominantly been tested at the sentence-final position. Overall, the experiments reported above confirm and extend previous findings about mental representations during language comprehension. Our results demonstrate that mental representations are initiated sentence medially, and are rapidly revised if followed by a verb that implies a change to an object shape. Strikingly, revisions to mental representations appear to be easier than some other sorts of reanalysis decisions in sentence comprehension, which can be quite slow or costly (e.g., Christianson et al. 2001). The fact that these detailed mental representations are modified over the course of sentence comprehension highlights the degree to which they are incremental and dynamic.

**Appendix A**: Critical stimuli that depict canonical/non-canonical shape in Experiment 1

(1) 勉は　空に/滑走路に　飛行機を　はっきりと　見た

Tsutomu clearly saw an airplane in the sky/at the runway.

(2) 母は　買い物かごに/フルーツポンチに　りんごを　きちんと　入れた

Mother neatly put an apple in the shopping basket/in the fruit salad.

(3) 幸子は　買い物かごに/シリアルに　バナナを　あわてて　足した

In a hurry, Sachiko added a banana to the shopping basket/to the cereal bowl.

(4) 千恵は　本棚に/コピー機に　本を　丁寧に　のせた

Chie carefully placed the book on the bookshelf/in the copy machine.

(5) 母は　トースターで/パン製造機で　食パンを　手際良く　焼いた

Mother skillfully toasted/baked (both are translated into a single Japanese verb) bread in the toaster/in the bread maker.

(6) 健は　レジに/BBQ グリルに　にんじんを　ぎこちなく　置いた

Ken clumsily put a carrot on the check out counter/on the BBQ grill.

(7) 剛は　シガレットケースの中に/灰皿の中に　たばこを　思いがけず　みつけた

Tsuyoshi unexpectedly found a cigarette in the cigarette case/in the ashtray.

(8) 由紀は　とうもろこし畑で/鍋から　とうもろこしを　できるだけ　取った

Yuki took corn from the corn field/from the pot as much as possible.

(9) 翼は　冷蔵庫に/フライパンに　卵を　適当に　入れた

Tsubasa uncertainly put an egg in the refrigerator/in the pan.

(10) 藍は　かごに/皿に　グレープフルーツを　きちんと　置いた

Ai precisely placed a grapefruit in the basket/on the plate.

(11) 卓美は　リンクで/ベンチで　ホッケー選手を　ちょうど　見かけた

Takumi just happened to see a hockey player on the ice rink/on the bench.

(12) 会社員は　机の上に/車のトランクに　ラップトップを　うっかり　忘れた

A company employee carelessly left a laptop computer on the desk/in the car trunk.

(13) 里美は　畑で/ざるから　レタスを　上手に　取った

Satomi skillfully took lettuce from the field/from the colander.

(14) 店員は　買い物かごに/コロナビールに　ライムを　ちょうど　入れた

Just now, a clerk put a lime in the shopping bag/into the corona beer.

(15) 叔父が　山で/フライパンに　きのこを　偶然　みつけた

The uncle accidentally found a mushroom in the mountain/in the pan.

(16) 純一が　たんすに/洗濯かごに　ズボンを　さりげなく　入れた

Junichi casually put pants in the drawer/in the laundry basket.

(17) 友達が　ショッピングカートから/皿から　パイナップルを　要領よく　横取りした

A friend professionally snatched a pineapple from the shopping cart/from the plate.

(18) 船長が　風で/パドルで　船を　着実に　進めた

The captain steadily moved the boat forward with the wind/with the paddles.

(19) 静香が　庭に/たんすに　シャツを　偶然　みつけた

Shizuka accidentally found a shirt in the garden/in the drawer.

(20) 一郎が　鍋から/箱から　スパゲティーを　きちんと　出した

Ichiro took spaghetti neatly from the pot/from the box.

(21) 美紀が　鉢の中に/ピザの上に　トマトを　確かに　見た

Miki surely saw a tomato in the garden pot/on the pizza.

(22) 隆が　タオル掛けに/床に　タオルを　たまたま　見た

Takashi unexpectedly saw a towel on the towel hanger/on the floor.

(23) 信二が　雨の中で／傘立ての中に 傘を　必ず　見た

Shinji always saw an umbrella in the rain/in the umbrella stand.

(24) 父が　空に/木の枝に 鷹を　はっきりと　見た

Father clearly saw an eagle in the sky/on the branch.

**Appendix B:** Critical items for Experiment 2, labeled by object canonicity for the complete sentence.

| Non-canonical sentences (with canonical onsets) | Canonical sentences (with non-canonical onsets) |
|---|---|
| (1) 翔は　果樹園で　りんごを　一生懸命　切った | (1) 翔は　まな板の上で　りんごを　一生懸命　磨いた |
| Sho cut an apple in the orchard as hard as he could. | Sho polished an apple on the cutting board as hard as he could. |
| (2) 幸子は　バナナの木の上で　バナナを　今さっき　切った | (2) 幸子は　フルーツポンチの中に　バナナを　今さっき　落とした |
| Sachiko cut a banana on the banana tree a short time ago. | Sachiko dropped a banana into the fruit salad a short time ago. |
| (3) 千恵は　本棚に　本を　慎重に　広げた | (3) 千恵は　机の上に　本を　慎重に　立てた |
| Chie carefully opened a book on the bookshelf. | Chie carefully stood the book on the desk. |
| (4) 母が　朝ごはんに　パンを　いつもの様に　まるかじりした | (4) 母が　パン屋さんで　パンを　いつもの様に　切ってもらった |
| Mother bit (something that is in the original shape) bread as usual at breakfast. | Mother asked someone to cut a bread in the bakery as usual. |
| (5) 剛は　たばこの箱に　たばこを　いったん　もみ消した | (5) 剛は　灰皿に　たばこを　いったん　並べてみた |
| Tsuyoshi temporarily extinguished a cigarette in the cigarette case. | Tsuyoshi temporarily arranged cigarettes in the ashtray. |

| | |
|---|---|
| (6) 由紀は　とうもろこし畑で　とうもろこしを　あえて　むいた<br>Yuki purposefully took off (the skin from) a corn in the corn field. | (6) 由紀は　鍋の中に　とうもろこしを　あえて　収穫した<br>Yuki purposefully harvested (and collected) corn in the pot. |
| (7) 奈々が　冷蔵庫の中に　卵を　とっさに　落とした<br>Nana quickly dropped an egg in the refrigerator. | (7) 奈々が　フライパンの上で　卵を　とっさに　転がした<br>Nana quickly rolled an egg in the pan. |
| (8) シェフは　八百屋で　グレープフルーツを　嬉しそうに　試食した<br>The chef happily tasted a grapefruit at the fruit and vegetable shop. | (8) シェフは　スプーンで　グレープ フルーツを　嬉しそうに　転がした<br>The chef happily rolled a grapefruit with the spoon. |
| (9) 里美は　ざるで　レタスを　言われた通り　叩いた<br>Satomi tore a lettuce to pieces in the field as she was asked. | (9) 里美は　畑で　レタスを　言われた通り　ちぎった<br>Satomi hit a lettuce with the colander as she was asked. |
| (10) 酔っ払いが　商品棚に　ライムを　なぜか　絞りつけた<br>A drunken man somehow squeezed and put a (juice of) lime on the merchandise shelf. | (10) 酔っ払いが　コロナに　ライムを　なぜか　投げつけた<br>A drunken man somehow threw a lime at the corona beer. |
| (11) 叔父が　山で　きのこを | (11) 叔父が　フライパンで　きのこを |

| | |
|---|---|
| 手際よく　料理した | 手際よく　運んだ |
| Uncle skillfully cooked mushrooms in the mountain. | Uncle skillfully carried mushrooms in the pan. |
| (12) 純一が　たんすに　ズボンを　いつもの様に　ねじこんだ | (12) 純一が　洗濯かごに　ズボンを　いつもの様に　たたんだ |
| Junichi squeezed pants into the drawer as usual. | Junichi folded (and placed) pants in the laundry basket as usual. |
| (13) 男が　片手で　パイナップルを　楽々と　取り分けた | (13) 男が　包丁で　パイナップルを　楽々と　脅し取った |
| Man easily distributed a pineapple with one hand. | Man threatened (someone) and easily took a pineapple with a knife. |
| (14) 静香が　裏庭で　シャツを　いったん　たたんだ | (14) 静香が　たんすに　シャツを　いったん　かけた |
| Shizuka temporarily folded a shirt in the backyard. | Shizuka temporarily hung a shirt on the drawer. |
| (15) 母が　洗面所に　タオルを　さっき　投げ込んだ | (15) 母が　洗濯かごに　タオルを　さっき　干した |
| Mother threw the towel into the bathroom a short time ago. | Mother dried the towel on the laundry basket a short time ago. |
| (16) 香織が　雑貨屋で　ラッピングペーパーを　なぜか | (16) 香織が　ゴミ箱に　ラッピングペーパーを　なぜか |

| | |
|---|---|
| ぐしゃっと丸めた | 立てかけた |
| Kaori somehow crumpled a (piece of) wrapping paper into a ball in the miscellaneous goods shop. | Kaori somehow stood a (roll of) wrapping paper against the trash can. |

References


Altmann, G. T. M. & Y. Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73. 247–264.

Altmann, G. T. M. & M. J. Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30. 191–238.

Aoshima, S. 2003. *The grammar and parsing of wh-dependencies.* College Park, MD: University of Maryland dissertation.

Aoshima, S., C. Phillips & A. Weinberg. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51. 23–54.

Aoshima, S., M. Yoshida & C. Phillips. 2009. Incremental processing of coreference and binding in Japanese. *Syntax* 12. 93–134.

Barclay, J. R., J. D. Bransford, J. J. Franks, N. S. McCarrell & K. Nitsch. 1974. Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior* 13. 471–481.

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22. 577–660.

Barsalou, L. W. 2008. Grounded cognition. *Annual Review of Psychology* 59. 617–645.

Byrne, R. M. J. & P. N. Johnson-Laird. 2002. Conditionals: A theory of meaning, inference, and pragmatics. *Psychological Review* 109. 646–678.

Byrne, R. M. J. & P. N. Johnson-Laird. 2009. ‘If’ and problems of conditional reasoning. *Trends in Cognitive Sciences* 13. 282–287.

Chao, L. L. & A. Martin. 2000. Representation of manipulable man-made objects in the dorsal stream. *NeuroImage* 12. 478–484.

Choi, S. & A. Gopnik. 1995. Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language* 22. 497–529

Christianson, K., A. Hollingworth, J. F. Halliwell & F. Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42. 368–407

Clifton, C., Jr. 1993. Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology* 47. 222–246.

Connell, L. 2007. Representing object colour in language comprehension. *Cognition* 102. 476–485.

Connell, L. & D. Lynott. 2007. Is a bear white in the woods? Parallel representations of implied object color during language comprehension. *Psychonomic Bulletin & Review* 16(3). 573–577.

Dahan, D. & M. K. Tanenhaus. 2004. Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2). 498–513.

Ditman, T., T. T. Brunyé, C. R. Mahoney & H. A. Taylor. 2010. Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition* 115. 172–178.

Frazier, L. 1982. Sentence processing: A tutorial review. In M. Coltheart (ed.), *The psychology of reading* (Attention and Performance 7), 559–586. London: Erlbaum.

Frazier, L. & K. Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14. 178–210.

Gennari, S. P. & M. C. MacDonald. 2008. Semantic indeterminacy and relative clause comprehension. *Journal of Memory and Language* 58. 161–187.

Gentner, D. 1982. Why nouns are learned before verbs: Linguistic relativity and natural partitioning. In S. A. Kuczaj (ed.), *Language, thought and culture* (Language Development 2), 301–334. Hillsdale, NJ: Erlbaum.

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

Kamide, Y., G. T. M. Altmann & S. L. Haywood. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49. 133–156.

Kamide, Y., C. Scheepers & G. T. M. Altmann. 2003. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research* 32(1). 37–55.

Kaup, B., J. Lüdtke & R. A. Zwaan. 2006. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics* 38. 1033–1050.

Knoeferle, P., M. W. Crocker, C. Scheepers & M. J. Pickering. 2005. The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye movements in depicted events. *Cognition* 95. 95–127.

Levy, R. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.

MacWhinney, B. 1999. The emergence of language from embodiment. In B. MacWhinney (ed.), *Emergence of language*, 23–256. Hillsdale, NJ: Erlbaum.

MacWhinney, B. 2005. The emergence of linguistic form in time. *Connection Science* 17. 191–211.

Mazuka, R. & K. Itoh. 1995. Can Japanese speakers be lead down the garden path? In R. Mazuka & N. Nagai (eds.), *Japanese sentence processing*, 295–329. Hillsdale, NJ: Lawrence Erlbaum.

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

Miyagawa, S. 1997. Against optional scrambling. *Linguistic Inquiry* 28. 1–25.

Miyamoto, E. T. & S. Takahashi. 2002. Sources of difficulty in the processing of scrambling in Japanese. In M. Nakayama (ed.), *Sentence processing in East Asian languages*, 167–188. Stanford, CA: CSLI.

Nakatani, K. & E. Gibson. 2010. An on-line study of Japanese nesting complexity. *Cognitive Science* 34. 94–112.

Sedivy, J. C., M. K. Tanenhaus, C. G. Chambers & G. N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71. 109–147.

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

Tabossi, P. 1988. Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language* 27. 324–340.

Talmy, L. 1991. Path to realization: A typology of event conflation. *Berkeley Linguistics Society* 17. 480–519.

Taylor, L. J. & R. A. Zwaan. 2008. Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology* 61. 896–904.

Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard & J. C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–1634.

Traxler, M. J., M. D. Bybee & M. J. Pickering. 1997. Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology* 50A(3). 481–497.

Tucker, M. & R. Ellis. 1998. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* 24. 830–846.

Tucker, M. & R. Ellis. 2004. Action priming by briefly presented objects. *Acta Psychologica* 116. 185–203.

Vasishth, S. & R. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82. 767–794.

Yamashita, H. 1997. The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research* 26. 163–188.

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

Zeelenberg, R., D. Pecher, R. M. Shiffrin & J. G. W. Raaijmakers. 2003. Semantic context effects and priming in word association. *Psychonomic Bulletin & Review* 10. 653–660.

Zwaan, R. A. 2004. The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (ed.), *The psychology of learning and motivation*, 35–62. New York: Academic Press.

Zwaan, R. A. & G. A. Radvansky. 1998. Situation models in language and memory. *Psychological Bulletin* 123. 162–185.

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