CrossMark

Bilingualism: Language and Cognition 21 (2), 2018, 375–383 © Cambridge University Press 2017 doi:10.1017/S1366728917000256

Automatic semantic integration during L2 sentential reading*

SUNYOUNG AHN
Harvard University
NAN JIANG
University of Maryland

(Received: March 2, 2016; final revision received: April 5, 2017; accepted: April 25, 2017; first published online 22 May 2017)

Research has shown that L1 speakers can routinely generate mental imagery corresponding to sentence meaning in reading comprehension. This may reflect an efficient process of semantic integration in which information from the input combines with an individual's linguistic and real world knowledge to form a semantic representation of a sentence. Semantic integration in L2 reading has received limited attention. The present study attempted to examine the activation of mental imagery in reading among L2 speakers and thus to assess the efficiency of semantic integration during L2 reading. L2 Korean learners were tested on a sentence-based picture recognition task in which they were asked to respond to a picture immediately after reading a sentence related to the picture. Results showed that L2 Korean learners resembled native Korean speakers in their ability to generate mental imageries that reflected subtle semantic differences in sentence input.

Keywords: semantic integration, embodied cognition, mental imagery, L2 reading, L2 Korean learners

Introduction

Sentence comprehension is believed to include three processes: word recognition, parsing, and semantic integration (Gernsbacher, 1990; Pickering, Clifton & Crocker, 2000). To understand the meaning of a sentence, a language user must first recognize its component words so that their lexical meanings and syntactic properties become available. In the process of parsing, a user builds a syntactic structure in terms of how words are syntactically related. Based on the information obtained from the first two processes, individuals can then build a semantic representation of the sentence by relating the information to their linguistic, contextual, and real-world knowledge. The present study was pursued to probe the efficiency of sentence comprehension among second language (L2) learners as compared to native speakers at the level of semantic integration.

Abundant research has shown that semantic integration is highly efficient in first language (L1) processing. During sentence comprehension L1 speakers can easily sense semantically unacceptable words (Ardal, Donald, Meuter, Muldrew & Luce, 1990), actively predict upcoming words (Wicha, Moreno & Kutas, 2003), and routinely generate mental simulation according to the sentence context in the course of completing the semantic representation.

This last finding, which is particularly relevant to the present study, was demonstrated in several studies published by Zwaan and colleagues (Pecher, Van Dantzig, Zwaan & Zeelenberg, 2009; Stanfield & Zwaan, 2001; Zwaan, Stanfield & Yaxley, 2002). In these studies, participants were asked to read sentences for comprehension. Reading of each sentence was followed by the display of a picture. After reviewing the picture, the participants had to decide whether the displayed object was mentioned in the sentence. For items that required a positive response (i.e., objects that were mentioned in the sentences), the authors manipulated the relationship between the sentence and the picture such that the object implied in a sentence was either congruent or incongruent with the shape or orientation of the object in a picture. For example, a picture of a nail in its horizontal orientation may follow two contrasting sentences, such as those seen in Example 1. Sentence 1a implies a horizontal orientation (compatible) and 1b a vertical orientation (incompatible).

1a. *John hammered the nail into the wall.*

1b. *John hammered the nail into the floor.*

The picture and the first sentence were compatible in the sense that a nail hammered into the wall is likely to appear in a horizontal orientation rather than a vertical one. A picture of a nail in its vertical orientation would be compatible with the second sentence. Reaction

* This study was supported, in whole or in part, by funding from the Ann G. Wylie Dissertation Fellowship and the Second Language Acquisition program at the University of Maryland. Any findings and conclusions expressed in this material are our own and do not necessarily reflect the views of any of our institutions. We thank Dr. Sunhwa Jeong, previously at Chung–Ang University; Dr. Young–Joo Kim, Kyung Hee University; Dr. Mark Peterson, Brigham Young University; and Dr. Youngkyu Kim, Ewha Womans University for crucial help with data collection. Finally, we thank two anonymous reviewers for support and comments on the earlier version of this manuscript.

Address for correspondence:
Sunyoung Ahn, Department of East Asian Languages and Civilizations, Harvard University, 5 Bryant Street, Cambridge, MA
sunyoungahn@fas.harvard.edu

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

time data in these studies showed that the participants were faster in producing a positive response when the object was displayed in a way consistent with what was implied in the sentence, even though the participants were not asked to pay any attention to object orientation or shape. These results suggested that reading a sentence automatically resulted in the generation of a mental image that was consistent with the meaning of the sentence. These findings have been taken as evidence for theories of embodied cognition (for a review of embodied cognition, see Pecher & Zwaan, 2005).

Semantic integration in L2 sentence processing has received only limited attention. One line of evidence for efficient semantic integration in L2 reading comes from research that employed electrophysiological measures. These studies found that participants showed a native-like sensitivity to semantic anomaly in L2 sentence processing. Ardal and colleagues (1990) were among the first to demonstrate that both native and non-native speakers showed the N400 effect in responding to semantically anomalous sentences, even though the effect was both delayed and smaller when sentences were in participants' L2 rather than their L1. This native-like sensitivity to semantic anomaly in L2 sentence processing has been replicated in many subsequent studies (e.g., Hahne, 2001; Hahne & Friederici, 2001; Kutas & Kluender, 1994; Ojima, Nakata & Kakigi, 2005; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre & Molinaro, 2006; Weber-Fox & Neville, 1996).

Additional evidence comes from results reported by Hu and Jiang (2011). They asked L2 speakers to perform a lexical decision task on targets that appeared visually on a computer monitor while listening to sentences for comprehension. As shown in Example 2, the target words (underlined in sample sentences) were either semantically congruent or incongruent with the sentence context. The L2 speakers tested in the study showed a reliable sentence context effect in that they responded to the targets incongruent to the sentence more slowly than they did to congruent targets, just as native English speakers did. These results showed that, while listening to sentences, L2 speakers are able to quickly build a semantic representation of the incoming input, which helps them to anticipate the upcoming words.

2a. *The girl mailed the letter without a stamp.*

2b. *The girl mailed the letter without a stone.*

However, there is also evidence suggesting that L2 speakers were less capable of efficient semantic integration – specifically in anticipating upcoming input – in sentence processing. For example, in using sentences such as those in Example 3 below, Martin, Thierry, Kuipers, Boutonnet, Foucart and Costa (2013) tested L1 English speakers and late L1 Spanish–L2 English speakers in a sentence reading task. Semantic integration was assessed by examining the ERP (event-related potential) component N400 a) for the more expected and less expected nouns (*chair* vs. *armchair*) and b) for the articles preceding the nouns (*a* being compatible with the expected noun chair but not *an*). The latter was more concerned with assessing the participants' ability to anticipate upcoming input. In line with the findings from previous studies, both L1 and L2 speakers showed a similar N400 for the less predictable nouns. However, unlike native speakers, L2 speakers did not show a difference between the two article conditions. There was no increase in N400 amplitude for the article incompatible (e.g., *an*) to the more predictable ending (e.g., *chair*), in comparison to the compatible article (e.g., *a*). This latter finding was interpreted "as evidence that L2 comprehenders do not actively predict upcoming words during sentence comprehension to the same extent as L1 comprehenders" (p. 574). However, subsequent studies showed that L2 speakers can anticipate incoming information as efficiently as L1 speakers under some conditions (e.g., Foucart, Romero-Rivas, Gort & Costa, 2016; Foucart, Martin, Moreno & Costa, 2014).

3a. *He was very tired so he sat on a chair.*

3b. *He was very tired so he sat on an armchair.*

The present study

In light of the conflicting findings and limited research on this topic, the present study aimed to examine semantic integration in L2 processing by considering the activation of mental imagery in L2 sentence reading. To do so, a sentence-based picture recognition task was designed for L2 Korean speakers. As in the original study (Stanfield & Zwaan, 2001), participants read each sentence and were then asked to decide whether the object presented in the picture was mentioned in the sentence. They responded using a key press and were instructed to press *yes* if the object was mentioned, and *no* if it was not. Reaction time (RT) was measured from the time an image appeared to the time participants pressed the response key.

The test materials included both items that required *yes* responses and an equal number of items that required *no* responses. Among the former, a manipulation was made such that an object was displayed in an orientation or shape that was consistent or inconsistent with what was implied by the sentence. Based on the assumption that mental imagery is automatically generated in sentence processing (Barsalou, 1999) and on findings from previous research (Coppens, Gootjes & Zwaan, 2012; Stanfield & Zwaan, 2001; Taylor & Zwaan, 2009; Zwaan, Stanfield & Yaxley, 2002; Zwaan & Yaxley, 2003), we expected native Korean speakers to show a faster response time when responding to congruent pictures than they did

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

responding to incongruent pictures. More importantly, we wanted to determine if L2 speakers would show a similar congruency effect. If they did, we would take the finding as evidence for efficient semantic integration in L2 sentence processing. Otherwise, we would conclude that L2 speakers were less capable of efficient semantic integration.

Method

Participants

We recruited 60 L2 Korean speakers to participate in the experiment. The majority of participants used either Chinese ($n = 32$) or English ($n = 23$) as their native and dominant language. The remainder had French ($n = 1$), German ($n = 1$), Japanese ($n = 1$), Mongolian ($n = 1$), and Thai ($n = 1$) as their L1s. All had advanced proficiency in Korean, when measured using the Korean C–test (modified from Lee-Ellis, 2009). The test consisted of four passages containing 25 incomplete words in each passage with the second half of letters missing. Participants were asked to fill in the blanks. Raw test scores were converted into percentage accuracy scores. These ranged from 60% to 94% with an average of 79%.[1]

Furthermore, all L2 Korean-speaking participants were late L2 learners who began learning Korean at age 18 or older. All were enrolled in institutions of higher education and were taking advanced Korean language courses for academic purposes (e.g., Korean literature, Korean media, and Teaching Korean as a Second Language) at the time of testing. All had at least some experience of staying in Korea, ranging from one month to six years of residency. Their age at testing was 24.9 on average. Table 1 summarizes the participants' language backgrounds.

The control group was comprised of 36 L1 Korean speakers whose native and dominant language is Korean, and who live in either the United States or Korea. Their age at testing was 30.2 on average. [2] All participants were paid USD 10 per hour.

[1] L2 Korean (non–heritage) speakers, assessed as advanced L2 Korean learners by their institutions or in self–reports, were initially recruited in the United States and Korea. After recruitment, their Korean proficiency was measured by a Korean C-Test adapted from Lee-Ellis (2009) to confirm that their Korean proficiency levels were advanced enough to participate in the present study. As a result, 60 of 94 recruits were chosen to participate in the present study.

[2] As one reviewer suggested, given the participants' language backgrounds, we will be interested to see if differences emerge within a group. Differences could appear in such groups as residency (native Korean speakers living in Korea and native Korean speakers living in the United States), and L1 effects and L2 proficiency effects among L2 speakers. The L2 proficiency effect appeared significant among L2 English speakers in Ahn, Jiang and Osthus (2012). Other factors will be dealt with in our follow-up study. We thank the reviewer for this comment.

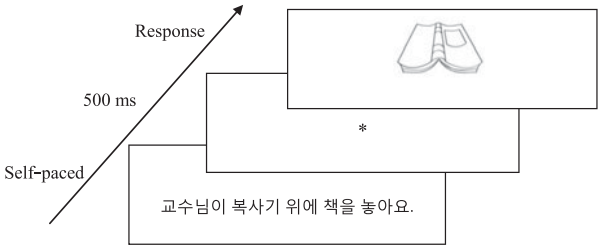


Figure 1. The stimulus sequence within a trial of the sentence-based picture recognition task.

Materials and design

The sentence-based picture recognition task consisted of 192 items. Half (96 items) included a picture of an object mentioned in the sentence and thus required a *yes* response. The other half included items requiring *no* response. Of *yes*-response items, half (48 items) were critical items, and the other half were fillers. For the critical items, two versions of the sentences were created. They were matched to the same picture, such that the picture was congruent with one version of the sentence (i.e., with object orientation or shape) and incongruent with the other version. Example 4 illustrates this manipulation. The two sentences were paired with the same picture, showing an opened book facing down (see Figure 1). Thus, the first sentence (4a) and the picture presented a congruent item (i.e., a match condition). That is, the picture showed an object that was not only mentioned in the sentence, but also congruent in shape to the meaning implied by the sentence. The second version of the sentence (4b) and the picture formed an incongruent item (i.e., a mismatch condition): the object in the picture was mentioned in the sentence but was not in the shape implied by the sentence (i.e., a closed book).

4a. 교수님이 복사기 위에 책을 놓아요 .
(*The professor placed the book on the copy machine.*)

4b. 교수님이 가방 안에 책을 넣어요 .
(*The professor put the book into the backpack.*)

An initial list of sentences was created by the first author, including five items from Zwaan's studies (e.g., Stanfield & Zwaan, 2001). The sentence stimuli were matched in terms of length and position of a target object noun in a sentence across conditions. A professional illustrator drew images as black-and-white line drawings based on information reported in previous studies as sources for typicality, complexity, and familiarity (Snodgrass & Vanderwart, 1980; Zwaan et al., 2002). Each picture was scaled to fit into a 3-inch square.

The critical stimuli were presented to five L1 Korean speakers who were Korean linguists as well as Korean language instructors with a teaching experience of

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

Table 1. *Summary of late L2 Korean speakers' language backgrounds.*

	Korean C-test score (%)	Age of L2 Acquisition	Length of Residence in Korea (m)	Age at Test
Mean	79	19.6	22.3	24.9
SD	15.4	3.6	16.4	2.3
Minimum	60	18	1	19
Maximum	94	30	72	33

10 years on average. They were asked to check (a) whether the proficiency level of the sentences was appropriate for the participants, and (b) whether the congruency status of the sentences in relation to the object pictures was sufficient. For assessing the latter, the five informants were presented with two versions of the same sentence, along with the pictures. The informants had to decide whether the shape or orientation of the object was implied (a) by one version of the sentence, (b) by both versions of the sentence, or (c) by neither version of the sentence. From the initial list of 60 items, 48 items were chosen on the basis that all five informants reached the same conclusion that only one version, the same version, was congruent with the picture.

In addition to the 48 critical items, 48 fillers requiring a *yes* response were created to direct participants' attention away from the manipulation of object shape and orientation in the critical items. The test materials also included 96 items that required a *no* response. Examples 5 and 6 illustrate such filler items; Sentence 5 was followed by an image of a watch, and sentence 6 was followed by an image of a ring.

5. 친구가 이 시계를 좋아해요.
(*My friend likes this watch.*)

6. 남자가 식탁 위에 꽃을 놓아요.
(*A man puts the flowers on the table.)*

Using these items we constructed two presentation lists of 192 items. The 48 critical items were counterbalanced across the two lists such that the same picture appeared only once on any list. It was matched with a congruent sentence on one list and with an incongruent sentence on the other list. The two lists had the same filler items.

As part of the material development, a group of 34 L1 Korean speakers was tested in a pilot study to ensure that a reliable sentence-picture congruency effect could be obtained among native Korean speakers. As expected, a reliable congruency effect was found ($r > .78$).

Procedure

The participants were tested individually and assigned to one of the two presentation lists randomly. Each trial consisted of several steps. First, a sentence was presented in the center of the computer screen. When the participants finished reading the sentence, they pressed the right shift key on a keyboard. A fixation mark (*) then appeared in the center of the screen for 500 ms, followed by a picture of an object. The participants' task was to decide whether the object depicted in the picture was mentioned in the sentence. They responded by pressing the right-shift key for *yes* and the left-shift key for *no*. The next trial (*sentence* → * → *image*) began automatically (Figure 1) after a response was provided.

The task consisted of two phases: a practice phase and a test phase. The participants completed 12 practice trials. When they reached 80% accuracy or higher, they could start a test phase. Among the 192 trials, 48 trials were followed by a comprehension question after participants provided a response to a picture. For example, for the sentence그는 연필을 컵에 넣어요 *(He puts the pencil in the cup.)*, the comprehension question was 그가 연필을 갖고 있어요*?* (*Does he have a pencil?*). The purpose of including the comprehension questions was to ensure that the participants were paying attention to sentence semantics, since understanding sentence meaning is essential to successfully examining sentence-picture congruency effects (Jiang, 2012). Data collection was done with DMDX (Forster & Forster, 2003). The experiment took about 50 minutes for L2 Korean speakers to complete. Participants took a post-test vocabulary test to assess whether they were familiar with the words used in the stimuli.

Results

The mixed effects models (MEMs) were used to analyze participants' RTs on the sentence-based picture recognition task.[3] In this model, the independent variable of primary interest was a condition with two levels (match and mismatch) set up as a fixed effect. Random effects were imposed on both participants and items to control for the variation across them. We imposed these because participants and items were randomly selected from a

[3] The use of MEMs has advantages over analysis of variance (ANOVA) to help us better understand this type of data, in which participants are measured on a series of items and vice versa (Baayen, 2008; Cunnings, 2012).

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

Table 2. *Error rates (in mean) of participants and items (standard deviations in parentheses).*

	Participant	Item
L1 Speakers	1.7 (1.8)	1.7 (4.4)
L2 Speakers	5.7 (4.3)	5.3 (6.0)

Table 3. *L1 and L2 Korean speakers' RTs (in mean) in the two conditions (standard deviations in parentheses).*

Group	L1 Speakers		L2 Speakers	
Condition	match	mismatch	match	mismatch
Mean	724	753	837	881
(*SD*)	(278)	(290)	(397)	(441)
Difference*	29**		44**	

Notes: * Difference between match and mismatch conditions.
** Difference significant at a level of $\alpha = .05$.

larger population, and random sampling can presumably affect the results.

We analyzed 4608 data points that 96 participants (36 L1 and 60 L2 Korean speakers) produced by responding to 48 critical items in the sentence-based picture recognition task. This data set was inspected in terms of error rates per participant as well as item, vocabulary knowledge (a post-test vocabulary check), incorrect responses, and outliers. First, error rates were examined for each participant and item. If a participant had an error rate greater than 20%, he or she was excluded from analyses. If an item had an error rate greater than 40%, it was excluded from analyses. None of the participants or items used in this study was rejected based on this criterion (see Table 2).

However, two items were rated as very difficult by L2 Korean speakers in the post-test vocabulary check (i.e., #14 and #44, by 5 and 8 participants, respectively). Thus, these two items were removed from further analyses. Furthermore, 286 (6%) reaction times were deleted due to incorrect decision responses, and 346 (7%) reaction times fell outside the range of acceptable latencies (mean +/− 2 *SD*). The data-trimming procedure accounted for a total data loss of 16%. Table 3 summarizes the means and standard deviations (*SD*) of the remaining reaction times in each condition and group for the experiment.

All data analyses were conducted with R, an open-source statistical package R (R core team, 2012). The lme4 package (Bates, 2005) was used to run MEMs.

Sentence-picture congruency effect in L2 sentence reading

L2 speakers' RTs were analyzed and compared with L1 speakers' to determine the mental imagery generation in L2 sentential reading. L1 speakers' RTs were 29 ms slower in the mismatch condition (753 ms, *SD* 290) than in the match condition (724 ms, *SD* 278), and L2 speakers' RTs were 44 ms slower in the mismatch condition (881 ms, *SD* 441) than in the match condition (837 ms, *SD* 397), as shown in Table 3. To test the significance of condition differences and of group difference, the *lmer()* function in R was used to construct MEMs. The command used in R to run the model was as follows:

> Model 1 = lmer (logged RTs ~ condition * group + (1 | subject) + (1 | item), data = SPT)

As shown in Model 1, we used log transformed RTs (i.e., logged RTs in the formula) as a dependent variable to better follow the statistical assumption of the model that the error term follows the normal distribution for a dependent variable. The independent variables of main interest were condition and the interaction effect between the condition variable and the group variable, set up as fixed effects (i.e., condition * group in the formula). The purpose of setting up these two factors as fixed effects was to see whether participants' picture recognition was significantly different according to the condition, and whether there was any meaningful difference in the degree of condition effects between the groups. Also, participants and items effects were controlled as random effects. The formula (1 | subject) + (1 | item) specifies crossed random effects for participants and items.

The results of Model 1 revealed a SIGNIFICANT condition effect on picture recognition in sentential reading OR sentence-picture congruency effect among the all participants ($|t| \geq 2.0$ at $\alpha = .05$).[4]

We ran another model in which by-subject random slope for condition was included. Given that participants were repeatedly measured on match and mismatch items, participants may vary in sensitivity to the manipulation (e.g., some participants may respond to the mismatch items much more slowly than to the match items, while other participants' response times across two conditions may show minimal difference.) Thus, the second model was run using the following formula:

> Model 2 = lmer (logged RTs ~ condition * group + (1 + condition | subject) + (1 | item), data = SPT)

The results of the second model showed the significant condition effect as well. That is, even when allowing participants to vary in sensitivity to the manipulation (condition), all participants responded significantly more slowly in mismatch condition than in match condition

[4] The significance in the MEMs using R can be determined based on a *t* value instead of a *p* value; if the absolute *t* value is equal to or larger than 2.0, the null hypothesis is rejected and the conclusion is that the parameter is significantly different from zero. For more on the rationale for using *t*, see Cunnings (2012, pp. 377-378).

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

(0.04184), and the difference was statistically significant ($|t| \geq 2.0$ at $\alpha = .05$).

In addition, we ran a third model that included by-item, as well as by-subject, random slopes for condition. Given that items were repeatedly measured on match and mismatch conditions within minimal pairs of sentences (i.e., a sentence in match condition can be measured in mismatch condition with a minimal pair changed), items may vary in sensitivity toward the manipulation (e.g., some item pairs may trigger much more delayed responses in mismatch condition than in match condition while other item pairs across two conditions show minimal difference.) Thus, the third model was run using the following formula:

> Model 3 = lmer (logged RTs ~ condition* group + (1+condition | subject) + (1 + condition | item), data = SPT)

The result of the third model showed the significant condition effect as well. That is, even when allowing different items, in addition to different participants, to vary in sensitivity to the manipulation (condition), all participants responded more slowly in a mismatch condition than in a match condition (0.04371), and the difference was statistically significant ($|t| \geq 2.0$ at $\alpha = .05$), as shown in Table 4.

To test whether one model provides a significantly improved fit to the data than the others, likelihood ratio tests were conducted with the *anova ()* function in R. As shown in Table 5, the result, expressed as a chi-squared statistic, demonstrates that the third model provides a significantly better fit to the data than the others ($p < .001$). This result indicates that the inclusion of both by-subject and by-item random slopes have improved model fit, and thus need to be included in the model for this data set.

Second, the result in Table 4 showed that the difference in sentence-picture congruency effect between L1 and L2 groups was NOT SIGNIFICANT. Specifically, L2 speakers (listed as *group_L2* in Table 4) were significantly slower to respond than L1 speakers overall (0.11186, $t = 2.60$). However, their sentence-picture congruency effect (0.01001) was NOT SIGNIFICANTLY different from that of L1 speakers ($|t| < 2.0$ at $\alpha = .05$). The non-significant group difference in condition effect (as shown as indicated by *cond mismatch: group_L2* in Table 4) indicates that the level of automatic semantic integration of L2 speakers during L2 sentence reading was not significantly different from that of L1 speakers. [5]

Table 4. *The sentence-picture congruency effect.*

	Fixed effects		
	Estimate Std	Error	*t* value
(Intercept)	6.53723	0.03985	164.05
cond mismatch	**0.04371**	**0.02176**	**2.01**
group_L2	0.11186	0.04306	2.60
cond mismatch: group_L2	**0.01001**	**0.02276**	**.44**

In conclusion, the best fitting-model confirmed that L2 speakers' picture recognition after L2 sentence reading was significantly slower when the object was portrayed in a direction or shape incongruent with implied sentence meaning. Picture recognition was faster when the object was congruent with implied meaning. Also, the degree of the delay in participant responses was similar to that of L1 speakers.

Discussion

The present study was conducted to address the extent to which late L2 speakers can be as efficient as L1 speakers in semantic integration during sentence reading. Semantic integration was observed by examining whether L2 speakers automatically generated mental imageries during reading, a process which we assessed by measuring a sentence-picture congruency effect. A significant sentence-picture congruency effect was found in both L1 and L2 speakers in the sentence-based picture recognition task. When the displayed image did not match the image implied by the sentence meaning in either shape or orientation, they showed a delay in responding to the pictures. This finding suggested that L2 speakers were similar to L1 speakers in their ability to quickly generate a semantic representation of a sentence while reading; this representation included a mental image of an object in its particular shape or orientation as implied by the sentence meaning.

This finding is consistent with those of several previous studies, which demonstrated native-like sentence processing performance among adult L2 speakers. For example, in Hu and Jiang (2011), Chinese ESL speakers listening to English sentences showed a delay in responding to

[5] We agree with one reviewer's comment that the difference in overall RTs between groups still shows a group difference, although this difference may not say anything about the efficiency of L2 semantic integration. Such a phenomenon reflects the nature of L2 processing. That is, although participants in the present study's L2 group are advanced L2 Korean speakers, they are likely to be less fluent in Korean than the L1 group, and to require more time to perform the sentence-based picture recognition task than L1 speakers do. To determine whether reading time had any effects on our study results, we ran MEMs including reading time as a variable. The results regarding our main interest stayed the same. Specifically, the significance of the sentence-picture congruency effect remained the same, and there was still no interaction effect between the group variable and the sentence-picture congruency effect. Therefore, we would like to keep our current model without adding a reading time variable.

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

Table 5. *Model comparison for the sentence-picture congruency effect.*

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model 1	7	3191.8	3235.7	−1588.9	3177.8			
Model 2	9	3195.7	3252.1	−1588.9	3177.7	0.0875	2	0.9572
Model 3	11	3179.5	3248.4	−1578.7	3157.5	20.2329	2	4.041e-05***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

an incongruent (as compared to a congruent) ending. In Hahne and Friederici's study (2001), late L1 Japanese–L2 German speakers were compared with L1 German speakers on N400 latency as they listened to semantically correct and incorrect sentences in German. Researchers in this study found that the patterns of N400 produced by semantic anomalies were not significantly different between L1 and L2 speaker groups, suggesting that even late L2 speakers who start learning L2 after puberty can automatically activate and integrate semantic information during L2 sentence comprehension.

In light of the findings of this present study, one may ask whether the lack of native-like sensitivity that researchers such as Martin et al. (2013) found truly reflected L2 speakers' limitation in semantic anticipation. Instead, it may have resulted from the use of the particular structure used to index anticipation, that is, the difference between the two different forms of the indefinite article of English, *a* and *an*, as pointed out by Foucart and colleagues (2014). Indeed, when a similar paradigm was used in the latter study, which tested Spanish monolinguals, French–Spanish late bilinguals, and Spanish–Catalan early bilinguals on Spanish sentences such as those in Example 7, late L2 speakers showed a native-like ERP pattern.

7a. *El pirata tenía el mapa secreto, pero nunca encontró el tesoro [masc] que buscaba*. [Expected]
(The pirate had the secret map, but he never found the treasure he was looking for.)

7b. *El pirata tenía el mapa secreto, pero nunca encontró la gruta [fem] que buscaba.* [Unexpected]
(The pirate had the secret map, but he never found the cave he was looking for.)

To conclude, by examining the activation of mental imagery in L2 reading, the present study provided converging evidence for a native-like semantic integration process among non-native speakers in sentence processing, which was already documented in several previous studies that examined semantic integration in terms of anticipation and semantic congruency (Hahne & Friederici, 2001; Hu & Jiang, 2011; Weber-Fox & Neville, 1996). Based on these findings, one may conclude that adult L2 speakers are able to develop an efficient process of semantic integration in sentence processing.

However, we are just beginning to understand semantic integration in L2 processing. Against the background of an overall similarity between L1 and L2 speakers observed in these studies, some differences were already apparent. For example, in several studies that compared L2 and L1 speakers with ERP measures, the N400 component – although present in L2 speakers' performance – was found to be delayed and smaller in magnitude (e.g., Ardal et al., 1990). Such differences were sometimes dismissed as being quantitative rather than qualitative in nature. However, whether such differences affect L2 reading in some fundamental ways is yet to be determined. In Hu and Jiang (2011), L2 speakers showed a native-like facilitative effect for congruent endings, but they differed from native speakers in that they were unable to produce an inhibitory effect when incongruent endings were compared to neutral endings. Furthermore, it is likely that the native-like performance in semantic integration observed so far may be due in part to the methodological approaches taken. Some qualitative differences may emerge when a more sensitive measure or a more powerful paradigm is adopted.

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

Appendix. *Samples of the picture and sentence stimuli used in the experiment. English translations in parentheses.*

Image	Match sentences	Mismatch sentences
	아이가 나무에 있는 새를 봐요. (A child sees a bird in the tree.)	아이가 하늘에 있는 새를 봐요. (A child sees a bird in the sky.)
	그는 필통에 연필을 넣어요. (He puts a pencil in a pencil box.)	그는 컵에 연필을 넣어요. (He puts a pencil in a cup.)
	그녀가 접시에 있는 계란을 봐요. (She sees an egg on a plate.)	그녀가 상자에 있는 계란을 봐요. (She sees an egg in a box.)
	기린이 땅 위의 풀을 먹어요. (A giraffe eats grass on the ground.)	기린이 나무 위의 잎을 먹어요. (A giraffe eats leaves on the tree.)
	개구리가 계단을 올라가요. (A frog jumps up the stairs.)	개구리가 계단에서 쉬어요. (A frog rests on the stairs.)

References

Ahn, S., Jiang, N., & Osthus, P. (2012). *Mental imagery in second language processing*. Presented at the Second Language Research Forum 2012, Pittsburgh, PA.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language, 39,* 187–205.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. New York: Cambridge University Press.

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

Bates, D. M. (2005). Fitting linear mixed models in R. *R News, 5,* 27–30.

Coppens, L. C., Gootjes, L., & Zwaan, R. A. (2012). Incidental picture exposure affects later reading: evidence from the N400. *Brain and Language, 122,* 64–69.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28,* 369–382.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124.

Foucart, A., Romero-Rivas, C., Gort, L. B., & Costa, A. (2016). Discourse comprehension in L2: Making sense of what is not explicitly said. *Brain and Language, 163,* 32–41.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1461–1469.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30,* 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4,* 123–141.

Hu, G., & Jiang, N. (2011). Semantic integration in listening comprehension in a second language: Evidence from cross-modal priming. In P. Trofimovich & K. McDonough (eds.), *Applying priming methods to L2 learning, teaching and research*, pp. 199–218. Philadelphia: John Benjamins Publishing Company.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Kutas, M., & Kluender, R. (1994). What is who violating? A reconsideration of linguistic violations in light of event-related brain potentials. In H. J. Heinze, T. F Munte, & G. R. Mangun (eds.), *Cognitive electrophysiology*, pp. 183–210. Boston: Birkhauser.

Lee-Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing, 26,* 245–274.

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language, 69,* 574–588.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17,* 1212–1228.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56,* 199–230.

Pecher, D., & Zwaan, R. A. (2005). *Grounding of cognition: The role of perception and action in memory, language, and thinking*. New York: Cambridge University Press.

Pecher, D., Van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology, 62,* 1108–1114.

Pickering, M. J., Clifton, C., & Crocker, M. W. (2000). Architectures and mechanisms in sentence comprehension. In M. W. Crocker, M. J. Pickering and C. Clifton (eds.), *Architectures and mechanisms in language processing*, pp. 1–28. New York: Cambridge University Press.

R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press

for Statistical Computing (Version 2.15.3) [Software]. Retrieved from http://www.R-project.org/

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6,* 174–215.

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

Taylor, L. J., & Zwaan, R. A. (2009). Action in cognition: The case of language. *Language & Cognition, 1,* 45–58.

Weberâ–Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8,* 231–256.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex, 39,* 483–508.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic–relatedness judgments. *Psychonomic Bulletin & Review, 10,* 954–958.

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

https://doi.org/10.1017/S1366728917000256 Published online by Cambridge University Press