

Using Contextual Information In Learners' Spoken Language Communication: An Eye-Tracking Study

Duck Geun Yoo

dyoo32@hufs.ac.kr

Hankuk University of Foreign Studies, Republic of Korea

Junkyu Lee

junkyu@hufs.ac.kr

Hankuk University of Foreign Studies, Republic of Korea

ABSTRACT

The efficiency of oral communication can vary depending on how quickly language users process not only language but also non-language information and on how that information is used. Various psycholinguistic experiments have shown that first language (L1) users incrementally form a representation by utilizing context information, and furthermore, can predict the forthcoming information. Since this topic has not been sufficiently studied using second language (L2) learners, theory and application of L2 processing are thus insufficient. This study aims to investigate, using an eye-tracking experiment, whether EFL learners can utilize visually presented contextual information during spoken language comprehension. Specifically, it eye tracked the comprehension processes of EFL learners when auditory inputs with four two-dimensional figures (e.g. Where is the large red circle?) were presented. Twenty-four learners with advanced English proficiency were asked to eye-spot figures corresponding to auditory inputs. Two conditions were manipulated in terms of figure size, color, and shape: (1) one pair of contrasting objects and; (2) two pairs of contrasting objects. It was revealed that the eye often moved to the target object before the noun (e.g. circle) was heard, adjectival information (e.g. large, red) being sufficient to restrict the domain of reference to one object in the visual display. Eye movements were thus quicker in the first condition that requires relatively little context information to distinguish the target than in the second condition that does not. The findings suggest that the subjects utilized real-time contextual information and top-down processing strategy of linguistic information in terms of Paul Grice's maxim of quantity.

Keywords: spoken language comprehension; predictive processing; context; English as a Foreign Language; eye-tracking

INTRODUCTION

In order to comprehend a single utterance, language users must interpret individual words presented within the utterance. They must assign grammatical functions to the words at fixed times and also take into account the context and situation surrounding the utterance to interpret its meaning. One of the most notable psycholinguistic controversies regarding this series of language comprehension processes has been when and how first language (L1) speakers utilize diverse information such as syntactic and pragmatic clues (Frazier, 1987; McRae, Spivey-Knowlton & Tanenhaus, 1998). Various psycholinguistic models

and theories have been proposed to answer the question of whether contextual information, which is likely to play a key role in incremental processing, is available at the initial stage of listening comprehension processes (van Gompel & Pickering, 2011, for an overview). Recent research on real-time language comprehension in the field of psycholinguistics has been gradually expanded into the domain of second language (L2). L2 research can be largely divided into two separate areas. On the one hand, at the lexical level, there is a focus on the question of how well high-level L2 learners can access lexical meaning in L2 independently from L1 (Canseco-Gonzalez et al., 2010). On the other hand, there is a growing interest in what principles/methods and information are utilized by L2 listeners for resolving ambiguity at the sentence level (Omaki & Schulz, 2011). Most of these studies, however, have been limited to examining only the effect or interference of L1 on L2 within European languages.

As there are only a few studies that have investigated the L2 comprehension of learners who use Asian languages as L1 (Witzel & Forster, 2012; Luan & Sappathy, 2011), current research results are insufficient to construct a persuasive theory regarding the L2 processing that is unique to learners in the Asian region. Furthermore, since research on cognitive processing of L2 has concentrated on the lexical or sentence level, experiment data is lacking with respect to how language learners utilize contextual information. Hence, there are insufficient resources in terms of psycholinguistic theory and experiential data that can be used to pedagogically support and cultivate learners' lexical and grammatical abilities as well as their communication abilities. In light of this, the current study aims to investigate how L2 learners in an Asian region utilize contextual information during real-time listening comprehension processes. Methodologically, the presence or absence of contextual information being used in real-time was examined by a visual world paradigm (VWP, see Literature Review for more details). By analyzing the real-time spoken language comprehension of Korean EFL listeners when visual contexts and auditory stimuli are presented for a brief interval, this study attempts to investigate whether EFL learners in an Asian region utilize both linguistic and contextual information during comprehension processes.

LITERATURE REVIEW

Sentence processing research has shown that parsing is largely incremental, i.e., language users incorporate each word into the preceding structure as it is encountered. In other words, language users do not delay sentence structure building until the end of the phrase or sentence (Marseln-Wilson, 1975). However, there is considerable controversy about when language users use different sources of information during comprehension processes. The most controversial topic is whether language users immediately use all relevant sources of information, or whether some sources of information are delayed relative to others (van Gompel & Pickering, 2011). With regard to this question, sentence processing theories can generally be divided into structural accounts and constraint-based accounts.

Structural accounts describing the real-time processing of sentences have assumed, based on “modularity” as asserted by Fodor (1983, p. 37), that individual modules existing within the cognitive system are responsible for the unique processing steps in each module and that there is encapsulation of processing where an individual

module is not affected by other modules. For instance, when processing the structure of an utterance, the module in charge of syntactic processing is not affected by the lexical information or preceding context. Similarly, Frazier (1987, p. 562) assumed the “serial processing” of information and proposed “minimal attachment” as the principle that can serve as the basis for resolving syntactic ambiguity. According to this principle, when listeners encounter ambiguous information such as sentence (1), where the prepositional phrase “with binoculars” can be attached structurally to the verb “saw” or the nominal phrase “the spy,” they prefer the interpretation method of constructing a minimal structure that takes into account only the syntactic information. In other words, the prepositional phrase is understood as modifying the verb. In the case of (1), generally the interpretation is, “The cop used binoculars to see the spy.” Various studies have confirmed that the syntactic module does not interact with lexical or pragmatic modules at the initial processing stage and is encapsulated in the utterance structure (Frazier & Rayner, 1982).

- (1) The cop saw the spy with binoculars.
- (2) The cop saw the spy with a revolver.

In contrast, constraint-based theories, which assume that there is active interplay of information at the initial comprehension process stage, have recently gained momentum (McRae, Spivey-Knowlton & Tanenhaus, 1998; Kaiser et al., 2009). For instance, McRae, Spivey-Knowlton and Tanenhaus (1998, p. 289) have asserted that when ambiguity occurs at “with” in (1) or (2) during comprehension processes, listeners activate all interpretation possibilities in parallel. The activation level of each interpretation is determined through various sources, including contextual information. Listeners finally recognize only one interpretation because information added later continues to reinforce the particular activated interpretation and thus weaken other interpretations. They argue, therefore, that listeners do not have to do the work of first attaching the prepositional phrase to the verb “saw.”

Methodologically, a precise measurement tool is required to observe in real-time which types of information are utilized in the process of comprehending utterances. In the case of written language, self-paced reading (SPR) and electroencephalography (EEG) have been widely used to measure online processing (SPR: Pan & Felser, 2011; EEG: Zhang & Boland, in press). Aside from these techniques, the visual world paradigm (VWP) has drawn researchers’ attention to understanding how visual contexts are related to the time-course of spoken language comprehension. Given that human language use frequently occurs within the nonverbal context of a scene, the paradigm of using an eye tracker, which has excellent temporal and spatial accuracy as a measuring device, can provide a significant advantage in observing patterns of understanding complex information. Just and Carpenter (1980, p. 331) hypothesized that there is “no appreciable lag between what is fixated and what is processed.” VWP is based precisely on this intimate correlation. Experimenters record and analyze in detail the patterns of participants’ fixations and saccades, which appear in response to auditory linguistic stimulations and visual nonverbal stimulations and, from this, are able to infer the listeners’ real-time processing. By presenting both visual contexts and auditory verbal stimuli at a brief interval, the VWP generally aims to find when visual context can be utilized during comprehension processing (Huettig, Rommers & Meyer, 2011). Among

the studies, the VWP used by Eberhard, Spivey-Knowlton, Sedivy and Tanenhaus (1995) has been considered a paradigmatic example and has served as the model for many subsequent studies to date. In their study, participants viewed on a computer screen figures that differed in marking, color, and shape, and were given instructions such as “Touch the starred yellow square.” The listeners then had to touch the appropriate figure as quickly and accurately as possible. Eberhard and his colleagues were seeking to examine at what point the participants could find the target figure in the course of hearing the nominal phrase “the starred yellow square” presented in the instructions. Since the figures shown to the participants had certain features in common, the image context composed of figures necessarily caused temporal ambiguity during nominal phrase processing and, accordingly, the point at which temporal ambiguity resolved also changed depending on the relationship between the figures. For instance, if one figure was the starred yellow square and another the starred red square, then the two figures had a marking feature in common and were distinguished in terms of their color feature. Hence when the participants heard the nominal phrase “the starred yellow square,” the language information up to “starred” was ambiguous because it did not provide enough information to fixate on the target figure, and it was only starting from “yellow” that the target figure could be distinguished.

According to structural accounts (Frazier, 1987) the nominal phrase “the starred yellow square” is expected to be processed based on syntactic information prior to semantic or contextual information. Within the context of VWP, therefore, structural accounts predict that the same nominal phrases will be processed in the same way regardless of visual context conditions. This assumption implies that even when a target figure can be distinguished through marking or color, the speed of finding that figure will not increase. By contrast, constraint-based accounts (McRae, Spivey-Knowlton & Tanenhaus, 1998) contend that nominal phrase processing changes depending on the contents of the objects presented on the display and that the pattern of fixations also changes accordingly. In the actual experiments, the participants, despite being given an identical nominal phrase after a word expressing a feature that could be critically distinguished from other objects on the display, fixated on the appropriate target object within an average of 75ms. For instance, after hearing only the word “starred,” and if the corresponding object could be distinguished from competing objects, the participants fixated immediately on the target object, whereas when distinction from other objects was made at “yellow,” fixation movement to the target object was slower. In other words, participants made sequences of eye movements that were closely time-locked to words in the instructions that were relevant to establishing reference. In sum, the findings show that in the processing of L1, listeners can consider (linguistic or non-linguistic) contextual information in real-time at a very early stage, and furthermore, based on contextual information, that they can continue to engage in predictive processing.

From a pragmatic standpoint as well, this experiment result showed the importance of language users’ ability to extract information implicit in utterances based on context for communication efficiency. According to Grice (1975), conversational implicature implies that what is explicitly said contains hidden/implicit information in the speech. The listener must infer such information from non-linguistic features in the conversational situation and depend on a general principle of cooperation and a series of conversation maxims during the inference process. In particular, according to the maxim

of quantity proposed by Grice (1975, pp. 45-46), since the speaker presents information keeping in mind the principle, “Do not make your contribution more informative than is required,” the listener in response to this considers the linguistic information and the related context together. Accordingly, the active consideration of context by subjects in the experiment by Eberhard et al. (1995) can be interpreted as a predictable result in that it becomes the pre-condition for performing communication. Most L2 psycholinguistic studies thus far, however, have focused on the possibility of independent access of linguistic meaning or the resolution of syntactic ambiguity. Interestingly, Clahsen and Felser (2006, p. 3) have asserted that sentence comprehension in L2 shows a critical difference from that in L1, and that “adult learners are guided by lexical–semantic cues during parsing in the same way as native speakers, but less so by syntactic information.” In other words, the syntactic component of L2 held by learners is defective, and as a result, the syntactic representations that adult L2 learners compute during comprehension are shallower and less detailed than those of native speakers. From this standpoint, it is not certain that the appearance of information utilization observed in L1 can necessarily be observed identically in L2. It is thus worthwhile to test whether L2 speakers utilize lexical-semantic and contextual information actively from the initial stage of comprehension in order to make up for the loss of syntactic ability.

THE CURRENT STUDY: WHERE IS THE LARGE RED CIRCLE?

In order to examine the role of contextual information during spoken language comprehension by EFL learners in an Asian region, this study examined the eye movements of EFL learners within the VWP, with utterances such as “Where is the large red circle?” being used as one of the questions asked. VWP allows us to study the real-time processing of referential nominal phrases directed at objects. This study chose Korean EFL learners as subjects. As discussed, the significance of this study lies in the fact that studies on how English as L2 in the Asian region is processed in real-time have so far been lacking and, above all, this study examined the correlations between English and Korean, two languages with very different grammatical structures.

RESEARCH QUESTIONS

1. When do EFL learners use contextual information during spoken language comprehension?
2. Can EFL learners utilize predictive processing based on contextual information?

PARTICIPANTS

Twenty-four students from a large University in Korea participated in the experiment. All of them majored in Foreign Language Education; their average age was 24.35 (SD = 2.3) and there were 20 females and 4 males. They were verified to be free from color blindness based on eye tests given prior to the experiment and all were EFL learners who studied English in middle and high school. To ensure the homogeneity of participants’ English abilities, selections were made from a restricted pool of college students with TOEIC scores of 800 or higher (M = 920, SD = 40.25). In addition, the participants all

scored more than 80 points out of 100 in a cloze test developed by Kobayashi (2002) ($M = 87.35$, $SD = 5.21$). Thus, the participants of this study were deemed to have an advanced level of English proficiency. The participants were not aware of the purpose of the experiment.

MATERIAL

The aim of this study was to observe the processing of a nominal phrase that performs a referential function by applying the test material of Eberhard et al. (1995) to EFL learners as subjects, which had never been previously attempted. Accordingly, two-dimensional figures were used as contextual images, while a nominal phrase with a 'definite article - adjective1 - adjective2 - noun' structure was used as the language information describing the target figure. The displays consisted of four figures that varied in terms of size (large and small), color (black, white, red, and green), and shape (circles, squares, triangles, and stars). In the previous experiment, marking was used as the first feature of the figures, but in this study, size was used as an adjective category more familiar to EFL learners. For the same reason, color feature selections were limited to focal colors. In addition, to examine how quickly EFL learners determine relevant language information in communication, two conditions of the visual display type were used: one potential contrast set (OPCS) and two potential contrast sets (TPCS). In the OPCS condition the display contained one pair of objects differing only in color as in Figure 1a.

Thus when the participants heard "Where is the large red circle?" for example, at the point of hearing 'large,' the large red circle and the large black square compete as potential target objects and temporal ambiguity was generated. However, when the word 'red' was heard, it served to confirm that ultimately the large red circle was the target object. In this condition, aside from the target object there was one competitor object with a size feature identical to the large black square of Figure 1a. The other two objects served as distractors. One of the distractors (for example, the small red circle in Figure 1a) was set to always share the color feature of the target object. Through this, the two objects could compete every time two adjectives regarding size and color in the noun phrase were being presented. In the TPCS condition the display contained two pairs of minimally contrasting objects as in Figure 1b. As in the first condition, the same questions (for example "Where is the large red circle?") containing the same noun phrases in the second condition were presented. At the onset of the first adjective expressing size, for example 'large,' the target object and competitor object, for example, the large red circle and the large black square of Figure 1b, became competing objects and generated temporal ambiguity. As in the first condition, when the second adjective, for example 'red,' was encountered, the ultimate target object could be determined.

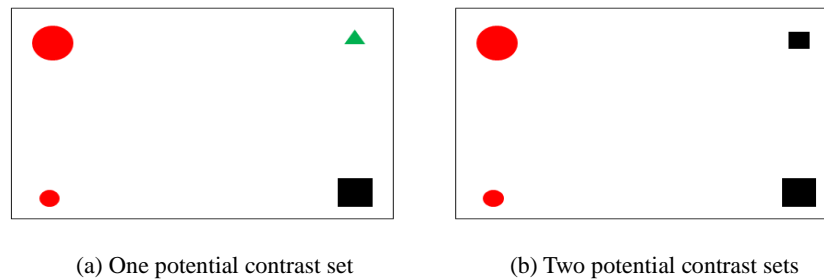


FIGURE 1. Two example display conditions serving as visual contexts for instructions

The difference between the two display conditions was as follows. In the OPCS condition the size feature of the competitor object was redundant from the perspective of Grice's (1975, p. 45) maxim of quantity ("make your contribution as informative as is required"). Compared to this, the size feature was relevant for the competitor object in the TPCS condition. For instance, the large black square in Figure 2b cannot be distinguished from the other objects just by the color feature since it shares this color feature with the small black square. In other words, while one of the two distractors in the OPCS condition had the same color features as the target object, in the TPCS condition, there were distractors with the same color features as the competitor object as well.

Therefore, if it was possible for the listener to grasp in advance the constellation of objects presented by the visual context and consider this information at the initial stage of utterance comprehension, then the processing pattern of utterance would change depending on the two display conditions. This hypothesis could be verified by specifically analyzing the movement pattern of fixations. In the first condition, on hearing the word referring to the size feature, the participants would already find the target object before the color feature was completely presented. However, in the second condition, since the ambiguity between objects could be resolved only when the size feature was presented, the fixation movement to the target object could be expected to be generated only after the word expressing the color feature was processed. In contrast, if the visual context does not have an effect on the initial comprehension stage of utterances, as asserted by structural accounts, there should be no difference in the processing of identical utterances, regardless of conditions. Hence the time of fixation movement to the target object or the proportions of fixation should not show significant differences. Particularly if the syntactic processing is viewed as preceding the processing of lexical or contextual information, the fixation proportion differences between objects would occur only after the nominal phrase structure was completed.

The two display conditions presented eight different nominal phrases to all participants. Along with these, 32 trials serving as fillers were provided. By placing the position where the ambiguity of the target object is resolved in these fillers at the first adjective of the nominal phrase or the last noun, the participants were made to experience equally the positions where ambiguity is resolved. To conduct two counterbalanced lists, we divided the 32 trials into two groups and rotated them through each of the two conditions in a Latin Square. Hence eight utterances of OPCS and eight utterances of TPCS appeared on each list, equally divided between the filler trials. The computer screen display was presented on a 22" TFT monitor at a resolution of 5,613 x 3,967 pixels. The utterances were pre-recordings of readings by a native English speaker

(American) and were edited using the GoldWave program. As seen in the table below, the utterances used in the experiment had an average duration of 2,950ms, with the first adjective onset occurring at an average of 991ms and the second adjective onset at 1,741ms.

TABLE 1. Average times for utterances used in the experiment

Utterance information	Where is the	size adjective	color adjective	noun
Utterance interval	0-991ms	991-1,741ms	1,741-2,291ms	2,291-2,950ms
Total utterance time	2,950ms			

PROCEDURE

Participants were seated at a 60cm distance from the computer screen. Eye movements were recorded with SMI RED 500, a remote eye tracker, sampling at 500 Hz. The instructions for the experiment procedure were presented to the participants through the monitor screen. Before the start of the actual experiment, a total of five trials were given as practice so that the participants could become fully familiar with the experiment method.

Each trial in the experiment comprised three stages (see Figure 2). In the first stage, the participants were presented with the screen as shown in Figure 2a, and upon hearing the instruction ‘Look at the smiley,’ they had to look at the center of the screen as instructed. When the participants’ fixations are actually positioned on the smiley at the center, a sensor detects them and presents the screen as shown in Figure 2b. In the second stage, the participants listened to the recorded utterance over the computer speaker while they viewed the corresponding visual display. A one second preview of the display preceded the presentation of each utterance, for example “Where is the large red circle?” The image information and linguistic information were presented at different points to allow the participants some time to understand the display context in advance. In the third stage, the screen as shown in Figure 2c was presented one second after utterance termination. At this point the participants had to say the number indicating the location of the appropriate target object. For example, the appropriate answer to the question “Where is the large red circle?” in the Figure 2b context is ‘one.’ⁱⁱ

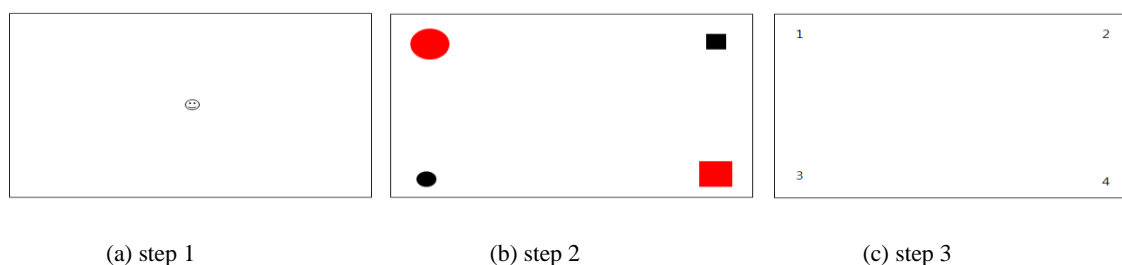
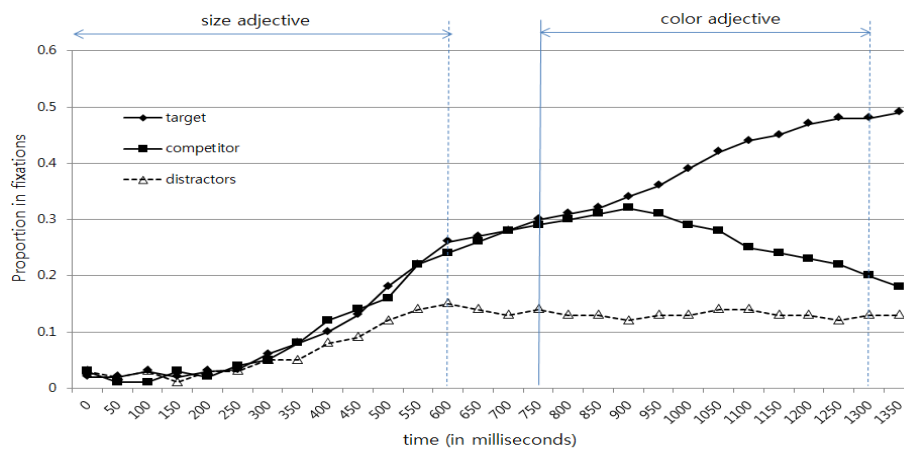


FIGURE 2. Experiment procedure

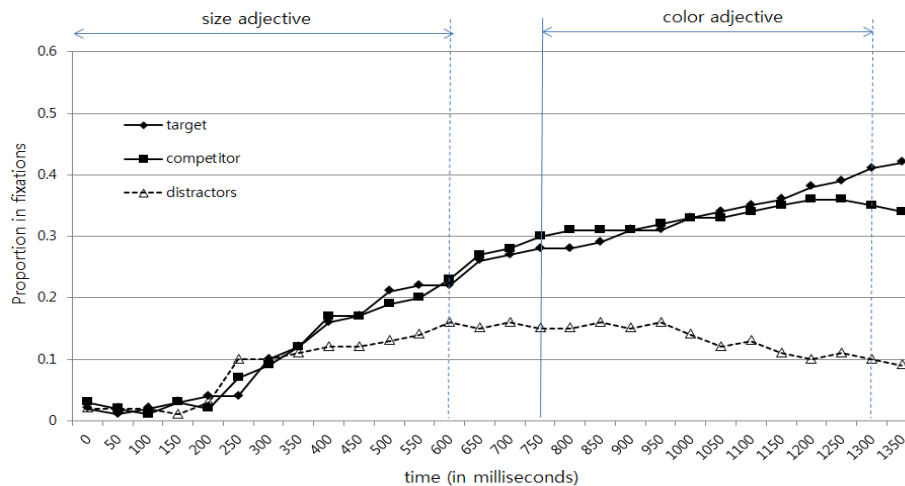
RESULTS

The data analyses in the present visual world study focus on the question of how likely the participants are able to view the target, competitor, and distractors at different times during the trials. The dependent variables are fixation proportions on the interest objects during each time window and proportions of saccades toward the objects initiated during each time window.

First, the fixation proportions on the objects for each condition in the 1,300ms interval extending from the onset of size adjective to the color adjective offset need to be examined. To do so, the proportion of fixations to each object (i.e., target object, competitor object, and distractor objects) over time (in 50ms time intervals) for each condition and each participant was calculated by adding the number of trials in which an object was fixated on during a 50ms time interval and dividing it by the total number of trials in which a fixation on any object or location was observed during this time interval. Figure 3 presents the proportions of fixations on the target and the competitor in the OPCS and TPCS conditions. In the OPCS condition, fixations on the target object and the competitor object increased with a similar slope from 200ms until around 900ms. However, starting from 950ms, while fixations on the target object continued to rise, fixations on the competitor object began to decline. In other words, the graph shows that the difference between the two objects was already visible 200ms after the presentation of the color adjective. This pattern is similar to results from other visual world experiments performed with similar themes (Tanenhaus et al., 1995; Almann & Kamide, 2009).



(a) One potential contrast set



(b) Two potential contrast sets

FIGURE 3. The proportion of fixations on the target, the competitor, and the distractors over time from the onset of the size adjective and the color adjective (in milliseconds) in the condition of one potential contrast set (a) and in the condition of two potential contrast sets (b). The dotted lines show the offset of each adjective, the first fine blue line shows the onset of color adjective, and the second vertical line indicates the head noun onset.

As shown in Figure 3b, the pattern of fixation proportions was remarkably different in the TPCS condition. Fixations on the target object and the competitor object diverged at a much later point than for the previous conditions. The competitor object's proportion of fixations continued to rise until the 1,200ms point, similarly to that of the target object, but after 1,250ms it declined in contrast to that of the target object. Put differently, in the TPCS condition, the participants showed equal interest in the target object and the competitor object during the utterance interval of two adjectives.

The first research question was concerned with the EFL learners' time-course of using contextual information during (listening) comprehension processes. In order to statistically test context effects, fixation proportions were averaged over a particular time window for each subject and item and were submitted to a one-way repeated measure ANOVA by subject (F1) or by item (F2). The time interval chosen for the analysis extended from 200ms to 1,300ms. The 200ms buffer following the first adjective onset was motivated both by the mean time required to plan and launch an eye movement, and the typical lag observed between eye movements and fine-grained phonetic detail in the speech stream (Kukona, Fang, Aicher, Chen & Magnuson, 2011). The time window extends over 1,100ms, which approximatively corresponds to the mean duration of two adjectives. The mean proportion of fixations on the target object in the OPCS condition (32.2%) was higher than the fixation proportion in the TPCS condition (26.3%), $F(1, 23) = 12.47, p < 0.01$; $F(1, 15) = 10.98, p < 0.01$. In contrast, the proportion of fixations on the competitor object was higher in the TPCS condition (26.1%) than it was in the OPCS condition (21.9%), $F(1, 23) = 4.92, p < 0.04$; $F(1, 15) = 5.01, p < 0.04$. In addition, the proportion of fixations on the distractor objects was shown to be 11.5% in the OPCS condition and 13.0% in the TPCS condition, but the difference between the two conditions was statistically not significant.

The second research question asked whether predictive processing based on contextual information is available for EFL learners. In order to establish that the effect

of predictive processing was observed from the earliest moments of lexical processing, we conducted an additional analysis over a time interval extending from 950ms to 1,250ms. As confirmed from the graph, this 300ms time interval is the segment between the times where the fixation proportions of the target object and the competitor object began to show a difference in the two conditions. Actually, prior to 950ms, the fixation proportion difference between the target object and the competitor object in the OPCS condition was not statistically significant. However, at 950ms the fixation proportion of the target object in the OPCS condition reached 36.7%, which was significantly higher than the competitor object's fixation proportion of 30.9%, $F(1, 23) = 9.98, p < 0.01$; $F(1, 15) = 12.43, p < 0.01$. In comparison, in the TPCS condition a statistically significant difference was shown between the target object's and the competitor object's fixation proportions of 39.1% and 35.1%, respectively, at 1,250ms, $F(1, 23) = 5.67, p < 0.03$; $F(1, 15) = 4.66, p < 0.05$. In the entire 300ms interval, the mean proportion of fixations on the target object was significantly higher in the OPCS condition than it was in the TPCS condition (42.1% vs. 34.5%), $F(1, 23) = 21.58, p < 0.01$; $F(1, 15) = 20.67, p < 0.01$. In contrast, the proportion of fixations on the competitor object was marginally higher in the TPCS condition than it was in the OPCS condition (33.8% vs. 26.7%), $F(1, 23) = 4.49, p < 0.05$; but $F(1, 15) = 4.45, p > 0.5$. Finally, the proportions of fixations on the distractor objects between the two conditions did not show a statistically significant difference.

Saccades are eye movements which align the fovea with objects of interest in the visual field. The selection of an object for a subsequent saccade is accompanied by a shift of focus in visual attention to the intended location (Doyle & Walker, 2002). Accordingly, together with fixation proportions, it is also meaningful to analyze the relationship between lexical processing and shifts of attention toward the target object and the competitor object. We analyzed the proportions of trials in which saccades to the target object and the competitor object were initiated. The time interval of interest extended from 950ms to 1,250ms. This 300ms time interval, as discussed above, is the interval between the times when the fixation proportions of the target object and the competitor object in each condition began to show a difference.

TABLE 2. Proportions of trials in which saccades to the target object and the competitor object were initiated.

Condition	Object	Proportion of saccades
One potential contrast set	Target	37.4%
	competitor	16.2%
Two potential contrast sets	Target	24.9%
	competitor	23.6%

As shown in Table 2, the first saccade to the target object in the OPCS was launched on 37.4% of all trials, as compared to 24.9% for the TPCS condition. First saccades to the competitor object occurred on 16.2% of trials in the OPCS condition, and 23.6% in the TPCS condition. A two-way ANOVA revealed that there was an interaction between object type (target vs. competitor) and condition (OPCS vs. TPCS), $F(1, 23) = 9.71, P < 0.01$; $F(1, 15) = 8.92, P < 0.02$. There were significantly more first saccades to the target object than to the competitor object in the OPCS condition, $F(1, 23) = 10.08, p < 0.01$; $F(1, 15) = 9.71, p < 0.01$, and the proportion of saccades to the target object was

significantly higher in the OPCS condition than it was in the TPCS condition, $F1(1, 23) = 12.32, p < 0.01$; $F2(1, 15) = 9.94, p < 0.01$. However, the first saccades to the competitor object in the TPCS condition were significantly more than to the competitor object in the OPCS condition, $F1(1, 23) = 5.36, p < 0.04$; $F2(1, 15) = 5.42, p < 0.04$, but the results showed no significant difference between the target and the competitor object in the TPCS condition.

DISCUSSION

The goal of this study was to investigate whether EFL learners are able to utilize contextual information in spoken language processing, by virtue of the visual world paradigm. In sum, when a competitor object that shares a size feature with the target object can be distinguished from other objects by only considering the color feature—that is, if the size feature is redundant—the time at which the participants began to look for the appropriate target object for the utterance information was relatively quicker than in the condition where the size feature was relevant, and the proportion of fixations and saccades was also high. As clearly shown in Figure 3, in the OPCS condition, the probability of looking at the target object at 950ms, i.e., 200ms past the color adjective onset, was already higher than the probability of looking at the other objects. By comparison, in the TPCS condition, the differences in the proportions of looking at the target object and competitor object occurred only in the time interval in which the adjective expressing the color feature was spoken. But even in this condition, the target object fixation proportion was already relatively high even before the presentation of the head noun. This finding suggests that eye movements were often initiated to the target object before the onset of the head noun when the information conveyed by the size adjective or the color adjective was sufficient to restrict the domain of reference to just one object in the visual display. Furthermore, the finding that the participants' patterns of fixations and saccades changed depending on the visual conditions shows that the constraints provided by the visual context influenced sentence processing from the earliest moments of lexical access.

To a great extent these results coincide with the patterns of fixations observed in experiments performed on L1 participants (Eberhard et al., 1995; Kunona et al., 2011). In these experiments, the participants showed different fixation patterns for identical utterance information depending on the visual context, indicating that the fixation proportions already began to show a difference in the lexical processing prior to the completion of the phrase structure. Likewise, the proportions of trials in which saccades to the target object appear differently depending on the contextual information are also similar to those in the L1 saccades analysis examples (Kamide, Altmann & Haywood, 2003). In these experiments, the proportion of saccades was analyzed to have a close correlation with the fixation proportion of the target object. Also, as seen in Figure 3, the participants had virtually no eye movements to the objects from the first adjective onset to 200ms. This 200ms buffer, similarly observed in the L1 experiments (Kukona et al., 2011), can be interpreted to be the minimum time required for planning and executing the eye movements during lexical processing (Allopenna, Magnuson & Tanenhaus, 1998). The interesting point here is that the speed of finding the appropriate object by the participants in lexical processing is not noticeably different from that of the L1

participants. In the OPCS condition, the participants began to place already relatively more fixations on the target object at 950ms, 200ms past the size adjective onset, whereas in the TPCS condition, they began to place greater fixations at 1,250ms, 50ms before the offset of the color adjective.

This is similar to the experiment results that indicate L1 participants can increase the speed or move up the point of lexical activation based on contextual information and therefore assign meaning to auditory information soon after lexical onset (van Berkum, Zwitserlood, Hagoort & Brown, 2003). Regarding the lexical processing ability of L2 learners, Kroll and Sunderman (2005, p. 113) proposed “early in acquisition there is reliance on word-to-word mappings across the two languages, but with increasing proficiency there is an increasing ability to conceptually mediate L2.” Hence when it is taken into account that the participants of the present study are high-level EFL learners, it can be conjectured that fast access to lexical meaning is fully possible. In addition, the fact that a limited number of size adjectives and color adjectives were presented to the participants can be interpreted as a cause for increasing the speed of lexical recognition.

CONCLUSION

From a new perspective, this study examined how contextual information is utilized in spoken language processing. To this end, the study performed visual world experiments using an eye tracker, with Korean EFL learners as subjects, and analyzed the interactive processing pattern of spoken language and visual context, which frequently occurs in everyday language use. The results showed that EFL learners were able to actively refer to contextual information already in the lexical processing. The findings were consistent with the predictions of constraint-based accounts that allow interactions of various types of information in the course of sentence processing (McRae, Spivey-Knowlton & Tanenhaus, 1998). Considering that the EFL participants solved the given task quickly by predicting in advance unknown information based on nonlinguistic information, the predictive processing verified for L1 (Altmann & Kamide, 2009) was also available for EFL.

Hermann (1972, p. 17) has emphasized the importance of the functional context of human communication: “In order to understand something verbal, one must consider the sender, the receiver, and the situation in which they find themselves.” In the same vein, from the perspective of psycholinguistics the EFL and ESL processing methods based on VWP certainly have value for continued research in the future. The reason for this is that VWP permits close observation of the cognitive process of language comprehension and production within a specific communication situation, and the data obtained through this can become a theoretical basis for not only simply describing the internal structure of language, but also for explaining the comprehensive communication ability of the language user. To this end subsequent studies need to present more real-life scenes for image information that functions as context. The use of realistic scenes, as Huettig, Rommers and Meyer (2011) claim, allows researchers to assess how the listeners’ perception of the scenes and their world knowledge about scenes and events affect their language understanding.

Finally, the results of this study performed on participants from a non-Indo-European language region that has thus far been the subject of minimal experiential

research in psycholinguistics, are significant in that they can serve as reference material for independently investigating the unique language understanding and language attainment methods of Asian EFL or ESL learners. However, because the participants were restricted to high-level EFL learners, this study has a clear limitation in providing a universal account. Subsequent studies therefore need to seek to examine learners of various levels and compare and analyze their language understanding methods. Furthermore, research should continue to comprehensively investigate the language processing methods of L2 learners who use different Asian languages as L1.

ACKNOWLEDGEMENT

This work was supported by a Hankuk University of Foreign Studies Research Fund given to Duck Geun Yoo and Junkyu Lee.

ENDNOTES

- i. The eyes move in a series of jumps, remaining relatively stationary between these jumps. The jumps, known as “saccades,” typically require 20-40ms. In contrast, “fixations” comprise a somewhat right-skewed normal distribution with the mean at around 200-250 ms (see Staub & Rayner, 2011).
- ii. One anonymous reviewer pointed out that verbally identifying the position of the target object could become additionally burdensome to memory. Its effect, however, can be viewed as non-significant, as in this experiment eye movements were monitored only up to the point when the instructions ended.

REFERENCES

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*. Vol 38(4), 419-439.
- Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye-movements and mental representation. *Cognition*. Vol 111(1), 55-71.
- van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*. Vol 17(3), 701-718.
- Canseco-Gonzalez, E., Brehm, L., Brick, C., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or Cárcel: Effects of age of acquisition and language proficiency on bilingual lexical access. *Language and Cognitive Processes*. Vol 25(5), 669-705.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*. Vol 27(1), 3-42.
- Doyle, M., & Walker, R. (2002). Multisensory interactions in saccade target selection: curved saccade trajectories. *Experimental Brain Research*. Vol 142(1), 116-130.

- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye Movements as a Window into Real-Time Spoken Language Comprehension in Natural Contexts. *Journal of Psycholinguistic Research*. Vol 24(6), 409-436.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*. Vol 14(2), 178-210.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.). *Attention and performance XII. The psychology of reading* (pp. 559-586). Hove, London & Hillsdale: Lawrence Erlbaum.
- van Gompel, R. P. G., & Pickering, M. J. (2011). Syntactic parsing. In M. G. Gaskell (Ed.). *The Oxford Handbook of Psycholinguistics*. England: Oxford University Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.). *Speech acts* (pp. 41-48). New York: Academic Press.
- Hermann, T. (1972). *Einführung in die Psychologie*. Bern: Huber.
- Huetig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*. Vol 137(2), 151-171.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*. Vol 87(4), 329-354.
- Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*. Vol 112(1), 55-80.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language*. Vol 49(1), 133-159.
- Kobayashi, M. (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *Modern Language Journal*. Vol 86 (4), 571-586.
- Kroll, J., & Sunderman, G. (2005). Cognitive Processes in Second Language Learners and Bilinguals: The Development of Lexical and Conceptual Representations. In C. J. Doughty, & M. H. Long (Eds.). *The Handbook of Second Language Acquisition* (pp. 104-129). Malden, MA: Blackwell.
- Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*. Vol 119(1), 23-42.
- Luan, N. L., & Sappathy, S. M. (2012). L2 Vocabulary Acquisition: The Impact of Negotiated Interaction. *GEMA Online® Journal of Language Studies*. Vol 11(2), 5-20.
- Marslen-Wilson, W. (1975). The limited compatibility of linguistic and perceptual explanation. *Proceedings of Chicago Linguistic Society 11, Papers from the Parasession on Functionalism* (pp. 409-420). Chicago: CLS.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*. Vol 38(3), 283-312.

- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*. Vol 33(4), 536-588.
- Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*. Vol 121(2), 221-236.
- Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*. Vol 268(5217), 1632-1634.
- Witzel, N. O., & Forster, K. I. (2012). How L2 words are stored: The episodic L2 hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Vol 38(6), 1608-1621.
- Zhang, Y. & Boland, J. E. (in press): The interplay between referential processing and syntactic processing. Retrieved September 28, 2012 from <http://www.personal.umich.edu/~jeboland/SyntacticDiscourseReferentialProcessing.pdf>

ABOUT THE AUTHORS

Duck Geun Yoo (Ph.D) is an Assistant Professor at the College of Occidental Languages (German) at Hankuk University of Foreign Studies. He received an award for the best doctoral dissertation from Westfälisch-L. Universitätsverband & Berstelsmann in German. His research interests include cognitive language processing, SLA, syntax, and pragmatics.

Junkyu Lee (Ph.D) is an Assistant Professor at the Graduate School of Education (English Education) at Hankuk University of Foreign Studies. His articles can be found in major L2 journals including *Studies in Second Language Acquisition*, *Modern Language Journal*, and *International Journal of Bilingual Education and Bilingualism*. His research interests include cognition and SLA, instructed SLA, and L2 research methodology.