

The Role of the Verb in Grammatical Function Assignment in English and Korean

Heeju Hwang

University of Southern California and Stanford University

Elsi Kaiser

University of Southern California

One of the central questions in speech production is how speakers decide which entity to assign to which grammatical function. According to the lexical hypothesis (e.g., Bock & Levelt, 1994), verbs play a key role in this process (e.g., “send” and “receive” result in different entities being assigned to the subject position). In contrast, according to the structural hypothesis (e.g., Bock, Irwin, & Davidson, 2004), grammatical functions can be assigned based on a speaker’s conceptual representation of an event, even before a particular verb is chosen. In order to examine the role of the verb in grammatical function assignment, we investigated whether English and Korean speakers exhibit semantic interference effects for verbs during a scene description task. We also analyzed speakers’ eye movements during production. We found that English speakers exhibited verb interference effects and also fixated the action/verb region *before* the subject region. In contrast, Korean speakers did not show any verb interference effects and did not fixate the action/verb region before the subject region. Rather, in Korean, looks to the action/verb region sharply increased following looks to the object region. The findings provide evidence for the lexical hypothesis for English and are compatible with the structural hypothesis for Korean. We suggest that whether the verb is retrieved before speech onset depends on the role that the verb plays in grammatical function assignment or structural choice in a particular language.

Keywords: verb, grammatical function assignment, picture–word interference paradigm, visual world eye-tracking, English/Korean

In order to transform a prelinguistic message into linguistic form, speakers need to select necessary lexical items and to assign them to the appropriate grammatical functions, such as subject and object. One of the central issues in production research is how speakers decide which entity to assign to which grammatical function (Blumenthal, 1970; Bock & Ferreira, *in press*; Bock & Levelt, 1994; Pickering & Ferreira, 2008). There are two major hypotheses concerning grammatical function assignment, namely, the lexical hypothesis and the structural hypothesis.

The *lexical hypothesis* suggests that grammatical functions are contingent on the verb’s argument structure (e.g., Bock & Levelt, 1994; deSmedt, 1996; Ferreira, 2000; see also Grimshaw, 1990). For example, the verb *send* requires the recipient role to be assigned to the indirect object as in “Mary sent John a letter,” whereas the verb *receive* assigns the recipient role to the subject as in “John received a letter from Mary.” Similarly, the lexical properties of verbs also constrain structural options. The verbs *donate* and *give* have similar meanings, but *donate* only allows a prepositional object structure (e.g., *The man donated some money*

to the charity/**The man donated the charity some money*), whereas *give* allows both a prepositional object and a double object structure (e.g., *The man gave some money to the charity*/*The man gave the charity some money*). Thus, according to this view, speakers must select a verb lemma before they start to speak, since the verb dictates the syntactic plan of the sentence.

In contrast, according to the *structural hypothesis*, speakers’ choices about which entity to realize in which grammatical function is not dictated by a preselected verb lemma, but rather by speakers’ conceptual representation of the event or the message (e.g., Bock, Irwin, & Davidson, 2004; Fisher, Gleitman, & Gleitman, 1991; Lashley, 1951; Wundt, 1900/1970; see also Griffin & Bock, 2000, for related work). According to this view, grammatical function assignment depends on how speakers encode the relationship between different scene entities. For example, if speakers interpret the recipient John as the “figure” and the agent Mary as “background,” they are likely to produce a sentence where the prominent entity John is realized as the subject (e.g., *John received a letter from Mary*). Thus, speakers’ holistic figure-ground conception plays a key role in determining sentence structure, rather than a single element of an event (see Bock et al. 2004; Bock & Ferreira, *in press*, for more details). Crucially, under this account, the selection of a verb lemma is *not* necessary for grammatical function assignment, as the subject can be chosen based on the relationship among the event elements.

In sum, the lexical hypothesis suggests that the verb lemma needs to be selected before speech onset, whereas according to the structural hypothesis, selection of the verb lemma before speech onset is not necessary. Broadly speaking, the claim of the lexical

This article was published Online First June 2, 2014.

Heeju Hwang, Department of Linguistics, University of Southern California, and Department of Linguistics, Stanford University; Elsi Kaiser, Department of Linguistics, University of Southern California.

Correspondence concerning this article should be addressed to Heeju Hwang, Department of Linguistics, 450 Serra Mall, Margaret Jacks Hall, Stanford University, Stanford, CA 94305. E-mail: heeju.hwang@gmail.com

hypothesis that grammatical functions are contingent on the verb's argument structure, and the claim of the structural hypothesis that speakers' choices about which entity to realize in which grammatical function are influenced by speakers' conceptual representation of the event or the message, are not mutually exclusive (Bock & Ferreira, *in press*). What is at issue here is the *relative contributions* of an individual lexical item (the verb) and relational information to early sentence formulation. The present research aims to explore this issue for two typologically different languages, English and Korean. Using a picture–word interference paradigm coupled with visual-world eye tracking, we tested whether speakers of these two languages retrieve verb lemmas before speech onset when producing active transitive sentences.

Previous Work on Verb Processing Before Speech Onset

Previous research investigating whether verb lemmas are retrieved before speech onset has led to mixed results. Compatible with the lexical hypothesis, Lindsley (1975) suggests that English speakers may process some aspects of verbs before speech onset. In his study, participants saw transitive events (e.g., a man greeting a woman) and were asked to produce (a) only the subject (e.g., *the man*), (b) the subject and the verb (e.g., *the man greets*), or (c) the full subject-verb-object sentence (e.g., *the man greets the woman*). What is relevant for our aims here is Lindsley's finding that speakers took longer to initiate the subject-verb utterances (e.g., *the man greets*) than the subject-only utterances (e.g., *the man*). Lindsley (1975) suggested that this is because speakers processed the verb before initiating a subject-verb utterance, in line with the lexical hypothesis. Kempen and Huijbers (1983) report similar findings in Dutch, employing a task similar to Lindsley's (1975).

Further evidence supporting the lexical hypothesis comes from recent work on Japanese by Iwasaki, Vinson, Vigliocco, Watanabe, and Arciuri (2008). Iwasaki et al. used a picture–word interference paradigm, where speakers were asked to name the depicted actions/verbs either in the citation form (e.g., *naku* “cry”) or in the progressive form (e.g., *naku + te iru* “is crying”), while ignoring auditory distractor words (e.g., *hohoemu* “smile”). The distractor words were either semantically related to the target verbs (e.g., “smile”) or unrelated to them (e.g., “crowd”). Iwasaki et al.'s design builds on the well-known *semantic interference effect*. In picture–word interference experiments, semantically related distractor verbs interfere with the selection of the target verb lemmas and consequently delay the utterance latencies for naming target verbs (e.g., Roelofs, 1993). Thus, semantic interference effects for verbs can be used to test whether speakers select verb lemmas before speech onset. Iwasaki et al. found that semantically related verb distractors delayed utterance onset latencies compared to unrelated distractor words, providing evidence for the retrieval of verb lemmas before speech onset in Japanese.

Using eye-tracking technology, Meyer and Dobel (2003) further supported the possibility that speakers retrieve the verb lemma before speech onset. When Meyer and Dobel monitored eye movements of Dutch speakers describing ditransitive events (e.g., *The cowboy gives the hat to the clown/The cowboy gives the clown the hat*), they found that the speakers usually fixated the agent (e.g., *cowboy*) and the object (e.g., *hat*) before speech onset. The speak-

ers rarely fixated the recipient (e.g., *clown*) before speech onset. Given the previous findings of eye-tracking production experiments that speakers fixated scene entities one after the other in the order of mention (i.e., fixations on the subject precede fixations on the object; Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000; Van der Meulen, 2001; Van der Meulen, Meyer, & Levelt, 2001, among others), the initial fixations to the object are rather surprising—as the object is not the first element in a canonical ditransitive sentence. Also, as Meyer and Dobel noted, it also would not be the first element to be fixated upon even if the speakers inspected the event elements from left (the location of the agent) to right (the location of the recipient) on the display. Meyer and Dobel suggested that the speakers fixated the object to encode the action and select the verb as the area around the object depicted the action (e.g., *giving*). Consistent with this possibility, they found that speakers fixated the object more often when asked to produce a sentence to describe the event than when asked to only name the event participants (e.g., *a cowboy, a hat, a clown*).

There is, however, evidence suggesting that verb lemmas are not retrieved until later, consistent with the structural hypothesis. For example, using the picture–word interference paradigm, Schriefers, Teruel, and Meinshausen (1998) provided evidence from German suggesting that verb lemmas are not obligatorily retrieved before speech onset. In German main clauses, the verb occurs immediately after the first constituent, irrespective of whether that first constituent is the subject or not. In subordinate clauses, however, the verb occurs in the clause-final position, following both the subject and the object. In a series of picture–word interference experiments, Schriefers et al. found that semantically related verb distractors delayed utterance onset latencies only for main clauses (where the verbs occurred in the second position), but crucially not for subordinate clauses (where verbs appeared in the utterance-final position). More specifically, before a picture was presented, participants heard two types of sentence beginnings (i.e., lead-in fragments). One type of fragment (e.g., *auf dem nächsten Bild sieht man wie . . .* “on the next picture one sees how”) required them to produce a verb-final subordinate clause (e.g., Subject-Object-Verb) upon the presentation of a picture, while the other type (e.g., *und auf dem nächsten Bild . . .* “and on the next picture”) required a verb-initial main clause (e.g., Verb-Subject-Object). Schriefers et al. found verb interference effects only when the lead-in fragments forced participants to produce the verb-initial main clause. This suggests that (a) German speakers can assign grammatical functions without retrieving a verb, and (b) whether German speakers retrieve verb lemmas before speech onset may depend on word order, with verb lemmas being retrieved when the verb occurs in the second position (immediately after the first constituent) but not when the verb occurs in clause-final position.

Griffin and Bock (2000) provided further support for the structural hypothesis. They presented participants with pictured events and compared eye-movement patterns when participants performed a linguistic task (picture description) or a nonlinguistic task (patient detection). In the patient detection task, where participants were asked to locate the patient entity, they fixated the patient entity approximately 300 ms after the image appeared on the screen. Because patient detection requires an understanding of relations among scene entities (e.g., who did what to whom), Griffin and Bock suggested that during the initial viewing period,

participants extract the event structure. It is generally suggested that a brief glimpse (500 ms or less) of an event is enough to grasp the general nature of the event (Bock & Ferreira, *in press*; Schyns & Oliva, 1994). Critically, the initial viewing period was also observed in the linguistic picture description task; speakers fixated the subject entity at about 300 ms after picture onset. Thus, Griffin and Bock suggested that speakers identify the causal or relational structure of an event (e.g., who did what to whom) during the initial viewing period and then direct their gaze to the scene entity that is established as the suitable starting point based on the relational scheme (for further evidence consistent with the structural account for English, see Bock, Irwin, Davidson, & Levelt, 2003; Bock et al., 2004; Kuchinsky, Bock, & Irwin, 2011).

In sum, evidence concerning retrieval of verb lemmas is mixed. Whereas Lindsley (1975) (see also Kempen & Huijbers, 1983), Iwasaki et al. (2008) and Meyer and Dobel (2003) suggested that some planning of verbs is necessary before initiating utterances, the findings of Schriefers et al. (1998) and Griffin and Bock (2000) suggested that verb selection is not obligatory before utterance onset. However, it is worth noting that in all of the production studies (Iwasaki et al., 2008; Kempen & Huijbers, 1983; Lindsley, 1975; Schriefers et al., 1998), the experimental designs were such that speakers could potentially infer/determine the grammatical functions of the nouns based on *factors other than the selection of verb lemmas*. In Lindsley (1975) (see also Kempen & Huijbers, 1983) the subject entities were always located on the left side of the picture, and the object entities on the right side of the picture. In Schriefers et al. (1998), the subjects were always animate (human; e.g., man), whereas the objects were inanimate (bucket). Because these studies did not include any filler trials, it is possible that speakers simply used the location or animacy of the scene characters to determine their grammatical role. Iwasaki et al. (2008) included filler trials, but (a) the subjects were given (known) to the speakers beforehand (before a picture appeared), and (b) speakers were simply asked to complete their utterances by naming the action upon the presentation of a picture. Thus, speakers did not have to assign the subject function via the selection of a verb. As a result, speakers' retrieval of verb lemma in these studies does not necessarily reflect the role of the verb in grammatical function assignment.

Evidence from eye tracking studies is not conclusive, either. Meyer and Dobel (2003) suggested that speakers initially fixated the object to encode the action and select the verb as the area around the object depicted the action. However, the object appeared in the middle of the screen, directly below the fixation point. Thus, it is possible that speakers fixated the object because it was a good vantage point to gain an overview of the scene, as Meyer and Dobel noted, or because it was simply close to the fixation point, but crucially not because it was relevant to the selection of the verb. In Griffin and Bock (2000), it is hard to access whether speakers selected a verb before speech onset, since the eye movements were analyzed in terms of the subject and the object without including the verb or the action, as in most eye movement production studies (e.g., Gleitman et al. 2007; Van der Meulen, 2001; Van der Meulen et al., 2001, among others).

Because previous studies do not provide clear empirical support for the lexical hypothesis or against the structural hypothesis due to their experimental designs or analyses, the question of whether

the selection of a verb lemma is necessary for grammatical function assignment is still open.

The Present Study

The present study aims to examine the role of verb in grammatical function assignment in two typologically different languages, English and Korean, in order to explore the validity of the lexical hypothesis and the structural hypothesis for these languages. In order to assess the time course of lemma access, we used a picture–word interference paradigm coupled with visual-world eye tracking. Participants were asked to describe a pictured event, while ignoring an auditorily presented distractor word. The distractor word was (a) semantically related to the verb, (b) semantically related to the object, or (c) not related to either the verb or the object. As participants carried out the production task, we recorded their speech and eye-movements. When analyzing participants' eye-movements, we compared fixations to the subject, the object and the action regions of each image. Action regions are defined as areas where the two scene characters make contact/interact with each other—i.e., the areas that provide crucial information about what action is being depicted in the image. For example, for the item shown in Figure 1, the action region is the circled area. (The circle was not part of the image shown to participants.) The same images and the same critical regions were used for both English and Korean.

Our design differs from existing studies in certain key ways, in order to avoid some of the challenges faced by earlier work. For example, our design minimizes the possibility of speakers' assigning grammatical functions on the basis of cues, such as location or animacy of scene entities, other than verb selection. To make sure that speakers cannot use such cues to "bypass" the function assignment processes, we counterbalanced the left/right locations of subjects and objects, and only used images depicting actions with both an animate subject and an animate object. Filler trials were also included to prevent speakers from associating a particular character with a particular action or employing other strategies to assign grammatical functions.

In addition to analyzing speakers' speech latencies in the picture description task, we also analyzed their eye-movement patterns. In

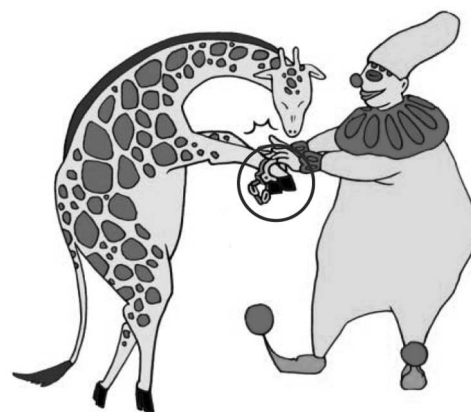


Figure 1. A sample target image (the circle was not part of the image shown to participants and is included to provide an example of the action regions).

particular, to access information about the process of verb lemma retrieval, we categorized fixations into the subject, the object and the verb/action regions (as explained below), instead of just the subject and the object as was done in most earlier eye-tracking studies (e.g., Gleitman et al., 2007; Van der Meulen, 2001; Van der Meulen et al., 2001, among others). We also used an initial fixation point at the start of each trial that is located neutrally among the event components in order to avoid confounding initial fixations with fixations on the verb or action region, as was the case in Meyer and Dobel (2003).

Predictions for Speech Latencies

Semantic interference effects in the picture–word interference paradigm can indicate whether speakers retrieve verb lemmas before they initiate an utterance. If *grammatical function assignment is contingent upon verb selection* in both English and Korean—as suggested by the lexical hypothesis—then speakers must retrieve verb lemmas before speech onset. In this case, we expect to find semantic interference effects for verbs in both English and Korean, despite word order differences between the two languages (i.e., regardless of whether the verb occurs early in an utterance as in English or late as in Korean). In other words, verb-related distractors should slow down speech latencies in both English and Korean.

In contrast, if *verb lemmas are not crucial for grammatical function assignment* and thus are not retrieved before speech onset—in line with the structural hypothesis—then we do not expect that English and Korean speakers will exhibit semantic interference effects for verbs. In other words, verb-related distractors should not slow down speech latencies. Yet, even if the retrieval of a verb lemma is not necessary for grammatical function assignment, we might observe verb interference effects in English due to the relatively early occurrence of a verb in a sentence (SVO). If the verb interference effect in English were caused by the relatively early occurrence of the verb in a sentence, then (by parity of argumentation) we might expect that Korean speakers would show object interference effects, given that the object immediately follows the subject in Korean (SOV).

Predictions for Eye-Movements

To supplement the speech onset latency data from the picture–word interference paradigm, we used visual-world eye tracking.

Eye tracking provides provide fine-grained temporal information about when people are attending to which aspects of the event. Because speakers are likely to fixate the scene element (e.g., subject, object) to which they are currently attending (e.g., Bock et al., 2004; Brown-Schmidt & Konopka, 2008; Rayner, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), the timing of successive fixations on the scene elements during the preparation of a sentence allows us to make inferences about the time course of the cognitive preparation of retrieving and arranging lexical items (Griffin & Bock, 2000). That is, the successive fixations can be used to tap into information about the order of word selection.

If English and Korean speakers' choice of the subject (or more broadly, the process of grammatical function assignment) requires that the verb already be selected—as suggested by the lexical hypothesis—then we expect to see both English and Korean speakers initially fixating the region of a picture that is informative or important for the selection of a verb (e.g., the part of the image that depicts the interaction of the two characters, see Figure 2). Indeed, this is what Meyer and Dobel (2003) suggested. Critically, given that the fixations indicate word selection (e.g., Griffin & Bock, 2000), we expect fixations on the verb-related region to emerge *before* fixations on the subject, assuming that the selection of the subject depends on (and thus follows) the selection of the verb.

In contrast, if grammatical functions can be assigned on the basis of a speaker's conceptual representation of an event without accessing a specific verb lemma—as suggested by the structural hypothesis—then English and Korean speakers are *not* expected to fixate the verb-related region of the image before fixating the subject. As Griffin and Bock (2000) suggested, we expect that English and Korean speakers should first fixate the subject entity after the initial event conceptualization period (500 ms or less)—as the subject can be established based on speakers' construal of an event without selecting the verb. If the retrieval of the verb, rather, depends on its position in an utterance (assuming that grammatical functional assignment depends on conceptual representations), then we expect that speakers should fixate the event elements in the order of mention similar to the previous findings (e.g., Gleitman et al., 2007; Griffin & Bock, 2000; Van der Meulen, 2001; Van der Meulen et al., 2001, among others). That is, (a) English speakers will fixate the action region *after* fixating the subject character, and (b) Korean speakers will not fixate the action region until after fixating the subject character and the object character.

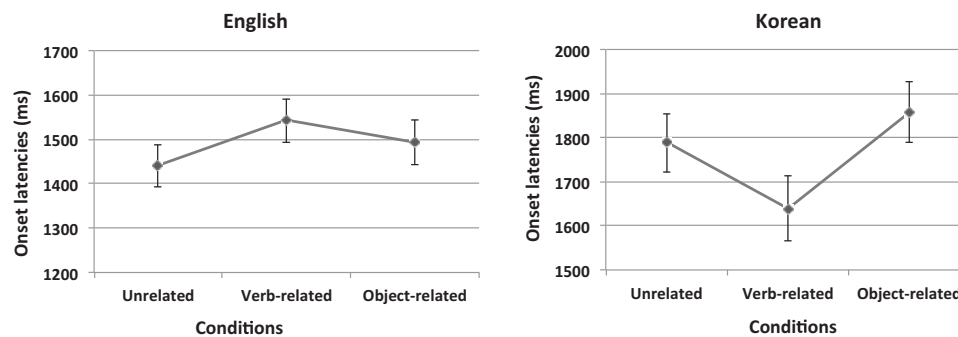


Figure 2. The onset latencies of utterances in the unrelated, verb-related, and object-related distractor conditions in English and Korean. Error bars indicate $\pm 1 SE$.

Experiment

Method

Participants. Eighteen native speakers of English and 16 native speakers of Korean from the University of Southern California community participated in the experiment in exchange for \$10 per hour.

Materials and design.

Character norming study. In order to select the characters for the target trials, a norming study was conducted on a separate group of 16 English speakers and 14 Korean speakers. Participants saw images of 20 individual characters and were asked to name the character they saw (e.g., giraffe, clown, pirate). Based on naming latencies and error rates, eight characters that had similar naming latencies and low error rates in both languages were selected. In English, the mean naming latency of the selected characters was 1,124 ms ($SE = 103$ ms), and the error rate was less than 2%. In Korean, the mean naming latency of the selected characters was 1,386 ms ($SE = 80$ ms), and the error rate was 4%.

Verb norming study. A second phase of the norming study was used to ensure that the interfering verbs to be used in the main picture–word interference task were indeed semantically close enough to the target verbs to create interference effects. We conducted a norming study specifically for the verbs, because retrieval of verbs has not been studied as much as retrieval of nouns in the picture–word interference paradigm (Iwasaki et al., 2008). In the norming study, participants were asked to name depicted actions, while they heard semantically related or unrelated distractor verbs. Following Roelofs (1993), semantically related verb distractors were chosen from cohyponyms (e.g., if the depicted action is “crying,” the related distractor word could be “laugh”), which share the same hypernym as the target verbs (e.g., emotions). We found that out of the 10 pairings of depicted verbs and semantically related distractor verbs that we tested, only four showed reliable interference effects in both languages: The average interference effect for these four pairs was 99 ms in English and 170 ms in Korean. Thus, these four pairs were selected for the main picture–word interference task. The average interference effect for our selected items is greater than the 74 ms reported in Schriefers et al. (1998).

Experimental stimuli. On the basis of the norming studies, eight target images were constructed using the eight selected characters and the four selected actions (see Appendix). Each action was used twice, but the same characters were never involved in the same action more than once. Over the course of the experiment each character occurred once as an agent (the doer of the action or the subject) and once as a patient (the acted-upon entity of the action or the object). The two characters for each picture were chosen so that they belonged to different semantic categories—animal and human. We paired humans with animals in order to rule out any chance of the object-related distractor words (e.g., *deer*) accidentally priming the subject characters. Humans and animals occurred as the subject and the object an equal number of times. The location of the agent and the patient (left or right) were counterbalanced.

To ensure that the results from the English and Korean studies are maximally comparable, the frequency and word length of the target names were matched as closely as possible both across and

within languages. In English, the target words had a mean frequency of 8.75 per million (CELEX database; Baayen, Piepenbrock, & Gulikers, 1995). The mean number of syllables was 1.58. In Korean, the target words had a mean frequency of 10.08 per million (mean frequency calculated based on Kang & Kim, 2009). The mean number of syllables was 2.66. Although the frequencies for the words in both English and Korean were relatively low, the characters were easily identifiable, as shown by the low error rates in the norming study.

The target pictures were presented with one of three types of auditory distractor words: (a) *a verb-related distractor* (a verb that is semantically related to the action depicted in the picture; e.g., *sip* as a distractor for the target verb *lick*), (b) *an object-related distractor* (a noun that is semantically related to the patient depicted in the picture; e.g., *deer* as a distractor for the patient *giraffe*), or (c) *an unrelated distractor* (an adjective that is semantically unrelated to the event in the picture; e.g., *vain* as an unrelated distractor for *lick* and *giraffe*). The noun distractors were semantically related to the target objects and of the same semantic category as the target objects (e.g., *deer* for the picture of a giraffe and *bandit* for the picture of a pirate; e.g., Glaser & Glaser, 1989; La Heij, Dirks, & Kramer, 1990; Lupker, 1979). The distractors were selected using the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998). Following Roelofs (1993), the verb distractors were cohyponyms of the target verbs (e.g., *sip* for the licking). Unrelated distractors were adjectives that had no connection to the depicted event (e.g., *vain* for the licking event). We chose adjectives as unrelated distractors because they provide a uniform baseline for potential verb and object interference effects by avoiding any possible syntactic interference effects. This is because earlier work found that distractors in the same grammatical class as the target word (e.g., nouns and nouns, verbs and verbs) could cause interference effects on naming (e.g., Pechmann, Garrett, & Zerbst, 2004, for English; Pechmann & Zerbst, 2002, and Schriefers et al., 1998, for German; Vigliocco, Vinson, & Siri, 2005, for Italian). The use of adjective distractors also minimizes the repetition of the target images in a within-subject design and thus helps to prevent speakers from associating a particular character with a particular action. In English, the distractor words had a mean frequency of 10.33 per million and the mean number of syllables was 1.70. In Korean, the target words had a mean frequency of 8.04 per million and the mean number of syllables was 2.95.

The three conditions (baseline/unrelated condition, verb-related condition, and object-related condition) were tested within subject; Each participant saw each of the eight target pictures three times over the course of the experiment: once in the unrelated condition, once in the verb-related condition, and once in the object-related condition. As in previous studies (e.g., Iwasaki et al., 2008; Schriefers et al., 1998), the repetition of the target images within subject allows us to evaluate the effects of the different distractor types within subject.

In addition to the eight target images, 12 filler images were constructed using 12 other characters and six other actions from the norming study. The filler images were paired with words that were *not* semantically related to the objects or the verbs (e.g., *banana* for *nun*, *donate* for *peck*).

The 24 critical items (created from the pairings of eight pictures and three distractors of different types) and 36 filler items (created

from the pairings of 12 pictures and three distractors of different types) were assigned to three blocks, adding up to a total of 60 trials (20 trials per block). Each target image and filler image occurred only once in each block. The order of items within each block was pseudorandomized so that (a) a target image did not appear as the first item in a block, (b) there were no more than two consecutive target images, and (c) target images that shared a character or an action were separated by at least three filler items.

The ordering of the blocks was rotated across participants. After each block, participants performed a word recognition task that probed their recall of the auditorily presented distractor words: Participants heard words and had to decide whether these words had occurred in the immediately preceding block. This was to make sure that participants paid attention to the distractor words. Before the main experiment, participants were presented with five practice trials and two word recognition trials.

Procedure. Following Schriefers et al. (1998), prior to the experiment, participants were given a booklet that included the instructions and the target and the filler pictures used in the main experiment. The pictures were not identified as targets or fillers and were presented in a randomized order. The names of each character and action were printed beside each picture, as in Schriefers et al. (1998). (For a similar approach, see also Dhooge & Hartsuiker, 2012; Schriefers, Meyer, & Levelt, 2002; Maess, Friederici, Damian, Meyer, & Levelt, 2002; Meyer & Damian, 2007, among others.) The nouns and the verbs were presented in bare forms without articles or inflections. Participants were told that, in the main experiment, they should produce active-voice descriptions of an event using those words shown in the booklet.

As soon as participants indicated that they had understood the instructions and familiarized themselves with the words, the experimenter asked them to name some of the characters and actions in target pictures to ensure that they had familiarized themselves with the characters and actions. After that, participants were seated in front of a 21-in. (53.34-cm) computer screen and presented with five practice items before proceeding to the main experiment.

Participants went through a 9-point calibration procedure. Then, to start each trial, participants fixated a point in the center of the computer screen and validated the fixation by pressing a controller button. After the validation procedure, participants were presented with a crosshair for 1,000 ms, which was neutrally located between the characters. They were asked to look at the cross and to press a button on a game controller. Immediately after the button press, the picture and the auditory distractor were presented simultaneously on each trial. Participants described the picture while viewing it and their spoken descriptions were recorded. Participants' eye movements were also recorded during the description. An SR Research EyeLink II head-mounted eye-tracker, sampling at 500 Hz, was used to collect and store eye-tracking data. At the end of each description, participants pressed a controller button to proceed to the next trial. The entire experiment lasted approximately 30–40 min.

Analyses. Participants' speech was transcribed for analysis. Trials in which participants did not use the expected character names or verbs were categorized as errors and removed from the analyses. For the remaining trials, onset latencies were manually determined using the phonetic software package *Praat* (Boersma & Weenink, 1992–2011). Following Griffin (2001), the onset of the subject noun (e.g., *pirate* in "the pirate") was used to mark the

beginning of sentences. In addition to measuring raw onset latencies, we also computed modified onset latencies by subtracting the length of auditory distractors from the utterance onsets of the recorded waveform files. This was done to control for any potential effects due to differences in the length of the distractor words: Each distractor word was, by necessity different in duration (within and between languages), since a range of distractor words was used. Thus, one might have concerns about longer distractors leading to longer onset latencies. To prevent the duration of distractors from affecting onset latencies, the length of distractors was subtracted from the utterance onset times in order to obtain the modified onset latencies. Unless specified, the onset latencies reported in the results refer to these modified onset latencies (the same basic patterns are also obtained if the raw onset latencies are used). Using the MAD-median rule (Wilcox, 2012), trials with outlier latencies were also removed. In sum, about 4% of trials were removed in English (13 out of 324 trials) and in Korean (12 out of 288 trials) for one of the reasons.

Results

In this section, we first report analyses of speech onset latencies in English and Korean and then turn to the eye movement analyses. For all results reported here, the data were analyzed with linear mixed-effects models (Baayen, 2008; Bates & Sarkar, 2007). These models are well-suited for analyzing categorical data as in our study and are better able to deal with unbalanced data sets than analyses of variance (ANOVAs; see Jaeger, 2008). Our analyses closely follow the approach argued for by Baayen, Davidson, and Bates (2008). We ran separate mixed-effects models for English and Korean, with distractor condition as a fixed effect and participant and item as random effects. We also ran a model over both languages including main effects of distractor condition and language as well as the interaction between the two. For each model, stepwise model reduction was performed in order to determine the random effect structure warranted by the data (see Barr, Levy, Scheepers, & Tily, 2013, and Jaeger, Graff, Croft, & Pontillo, 2011, for discussions of how to choose the random effect structure of mixed models). At each step, the least significant control predictor was removed from the model as long as $p > .1$. This procedure was repeated until no control predictor remained in the model or no control predictor that still was in the model was less significant than $p = .1$. Factor labels were transformed into numerical values, and centered prior to analysis, so as to have a mean of 0 and a range of 1. This procedure minimizes collinearity between variables (Baayen, 2008). For each result, we report the coefficient for each independent variable and its level of significance. Coefficients in mixed logit models are given in log-odds.

Speech onset latencies. Figure 2 plots the mean utterance onset latencies in the unrelated, verb-related and object-related distractor conditions in English and Korean. In English, the verb-related distractors significantly delayed utterance onset compared to unrelated distractors by 102 ms (1,543 ms vs. 1,441 ms; $\beta = 100.73$, $t = 1.992$, $p < .05$, $SE = 50.57$). In Korean, however, the verb-related distractors did *not* significantly delay utterance onset when compared to unrelated distractors (1,639 ms vs. 1,788 ms; $\beta = -140.02$, $t = -1.693$, $p = .09$, $SE = 82.69$). Rather, Korean speakers tended to initiate utterances faster in the verb-related distractor condition. In fact, when we compared utterance onset

latencies in the unrelated and the verb-related distractor conditions between English and Korean, there was a significant interaction effect of distractor condition and language ($\beta = -243.85$, $t = -2.545$, $p < .05$, $SE = 95.81$). The significant verb interference effect in English but the lack thereof in Korean suggests that only English speakers retrieved verb lemmas before speech onset. Object-related distractors did not delay utterance onset compared to unrelated distractors in either English or Korean (English: 1,441 ms vs. 1,493 ms; Korean: 1,788 ms vs. 1,858 ms; $ps > .1$), suggesting that neither English nor Korean speakers accessed object lemmas before they started to speak.

Eye movements. The eye-movement data provide another means of assessing the timing of lemma retrieval. In this section, we present the eye-movement patterns collapsed across the three conditions (unrelated, verb-related and object-related distractors).¹ This is because detailed analyses of eye-movement patterns indicate that, in both English and Korean, the timing of lemma retrieval is largely independent of condition—but differs in interesting ways across languages.

Before discussing the timing of verb lemma retrieval, we first checked whether our data matches what is currently known about the eye-voice span, i.e., the temporal relation between looking at a character before speech onset and naming it. Previous eye-tracking production experiments found strong effects of word order on eye fixations at utterance onset (e.g., Gleitman et al., 2007; Griffin & Bock, 2000; Van der Meulen, 2001; Van der Meulen et al., 2001 among others); speakers fixate a scene entity just prior to mentioning it. If our participants are behaving like those in other studies, we predict that both English and Korean speakers should fixate scene entities one after the other in the order of mention, i.e., the subject entity first and then the object entity. Indeed, this can be seen in Figure 3, which plots the mean proportion of looks to the subject, the object, and the action regions relative to utterance onset in English and Korean. As expected, participants fixated the subject region immediately before starting to speak and then fixated the object region. Additionally, in Korean we see that speakers fixated the action region after fixating on the object region at about 800 ms after speech onset. In English, although looks on the action increase following looks to the subject, they do not exceed looks to the object.

Let us now consider the timing of verb lemma retrieval, and what we might expect the eye-movements to look like given the speech latency results presented above. In English, the speech onset latencies revealed a verb interference effect. This suggests that speakers accessed verb lemmas before starting to articulate their utterances. An important question, then, is whether English speakers accessed verb lemmas *before* or *after* subject lemmas. The timing of the verb lemma retrieval is crucial in determining the role of the verb in grammatical function assignment: If English speakers determine the subject on the basis of the verb, they should fixate the action region *before* the subject region. If English speakers retrieve the verb lemma due to its linear position in the sentence, they should fixate the verb/action region *after* the subject region.

For Korean speakers, the analyses of speech onset latencies revealed no significant verb interference effect. This suggests that in Korean, the verb lemma was not retrieved prior to speech onset. In light of this finding, we expect that Korean participants' eye-movements should differ strikingly from English speakers' eye-

movements. In particular, if Korean speakers do not access verb lemmas before speech onset, they are *not* expected to preferentially fixate the action region *before* fixating the subject region, in contrast to English speakers.

To investigate the timing of lemma retrieval, we plotted the proportion of looks to the subject, the object, and the action/verb regions, relative to the time that the picture appeared on the screen (which we refer to as “picture onset”). The fixation patterns in English and Korean are shown in Figure 4.

Given that 200 ms is generally assumed to be necessary for programming and initiating an eye movement (e.g., Altmann & Kamide, 2004; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Huettig & Altmann, 2005; Matin, Shao, & Boff, 1993; Salverda, Kleinschmidt, & Tanenhaus, 2014; Saslow, 1967), we report and analyze our data in 200-ms time-windows starting at 200 ms after picture onset.

During the 200- to 400-ms time-window, neither English nor Korean speakers showed a selective preference for any one region. The mean proportions of looks to the subject (S), the object (O), and the action regions (V) were 0.24, 0.22, and 0.25, respectively, in English and 0.27, 0.27, and 0.21 in Korean (English: S vs. V: $\beta = -0.01044$, $t = -0.241$, $p = .80$, $SE = 0.04328$; S vs. O: $\beta = 0.01743$, $t = 0.409$, $p = .68$, $SE = 0.04266$; V vs. O: $\beta = 0.02787$, $t = 0.653$, $p = .51$, $SE = 0.04266$; Korean: S vs. V: $\beta = 0.06006$, $t = 1.408$, $p = .16$, $SE = 0.04266$; S vs. O: $\beta = -0.0006$, $t = -0.014$, $p = .98$, $SE = 0.04598$; V vs. O: $\beta = -0.0607$, $t = -1.320$, $p = .188$, $SE = 0.0459$).

During the 400- to 600-ms time-window, however, English speakers fixated the *action region* significantly more than the subject region and the object region. As can be seen in Figure 4, there is a sharp increase in looks to the action region, relative to the subject region and the object region. This is confirmed by statistical analyses. The mean proportions of looks to the subject, the object, and the action regions were 0.25, 0.26, and 0.41, respectively (S vs. V: $\beta = -0.15597$, $t = -3.296$, $p < .01$, $SE = 0.04732$; S vs. O: $\beta = -0.0064$, $t = -0.135$, $p = .89$, $SE = 0.0475$; V vs. O: $\beta = 0.1495$, $t = 3.146$, $p < .01$, $SE = 0.0475$). We see that fixations on the subject region surpass the fixations on the object region and the action region at around 600 ms after image display.

Whereas English speakers showed a striking preference to look at the action region, Korean speakers showed *no significant preference for the action region* over the subject region or the object region during the 400- to 600-ms time-window. The mean proportions of looks to the subject, object, and action regions in Korean were 0.30, 0.35, and 0.27, respectively (S vs. V: $\beta = 0.02745$, $t = 0.616$, $p = .53$, $SE = 0.04457$; S vs. O: $\beta = -0.04692$, $t = -0.984$, $p = .32$, $SE = 0.0477$; V vs. O: $\beta = -0.07437$, $t = -1.559$, $p = .12$, $SE = 0.0477$). In fact, when we analyzed the fixations on the action region during the 400- to 600-ms time-window with language as a factor, they differed significantly between English and Korean ($\beta = 0.13848$, $t = 2.202$, $p < .5$, $SE = 0.06288$). Instead, in Figure

¹ Target pictures that depicted the chasing event were not included in the eye movement analyses. This is because there were no areas where two scene characters were in physical contact with each other, and thus, action regions (for the purposes for eye-tracking data analysis) could not be defined for those items.

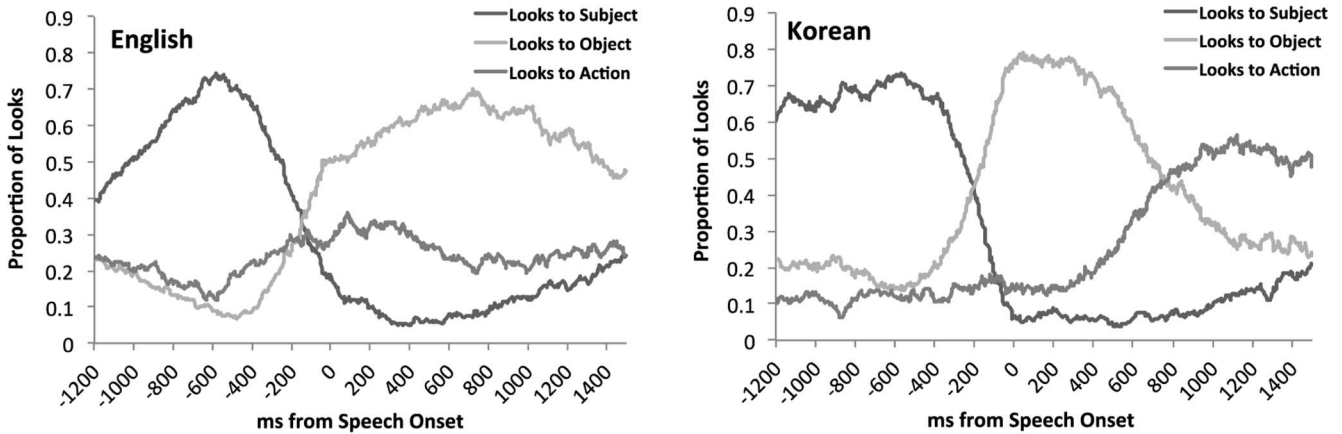


Figure 3. Proportion of looks to subject, object, and action regions in critical items relative to speech onset in English and Korean (collapsing all three distractor conditions within each language).

3, we see that in Korean, looks to the action region sharply increase after looks to the subject region and the object region. The significant fixations on the action following the fixations on the object—combined with the absence of initial fixations on the action—strongly suggest that Korean speakers retrieve a verb lemma only after speech onset. When we look beyond the 400–600 ms time window, we see that the looks to the subject region start to diverge from looks to the object region and the action region at around 600 ms after picture onset. This is at the same time window where we found increased looks to the subject region for English.

The findings for English that (a) fixations on the action region emerged rapidly during the 400–600 ms of image display and that (b) these fixations emerge *before but not after* fixations to the subject region suggest that English speakers retrieved verb lemmas immediately after inspecting the depicted scene and constructed sentences around the verbs. This is in line with the lexical hypothesis. In striking contrast to what we found for English, we found no indication of Korean speakers preferentially fixating the action

region before fixating the subject region. In fact, looks to the action region emerged only after looks to the object region. This points toward verb-independent function assignment in Korean, which is compatible with the structural hypothesis.

Discussion

The present study set out to investigate whether the selection of a verb lemma is essential for the process of grammatical function assignment in English and Korean, using the picture–word interference paradigm coupled with visual-world eye-tracking. If verbs play a central role in grammatical function assignment as suggested by the *lexical hypothesis* (e.g., Bock & Levelt, 1994), then English and Korean speakers should exhibit semantic interference effects for verb-related distractor words—i.e., verb-related distractors should delay the selection of the target verbs, thus prolonging utterance onset latencies. In terms of eye movements, we expect that before fixating the

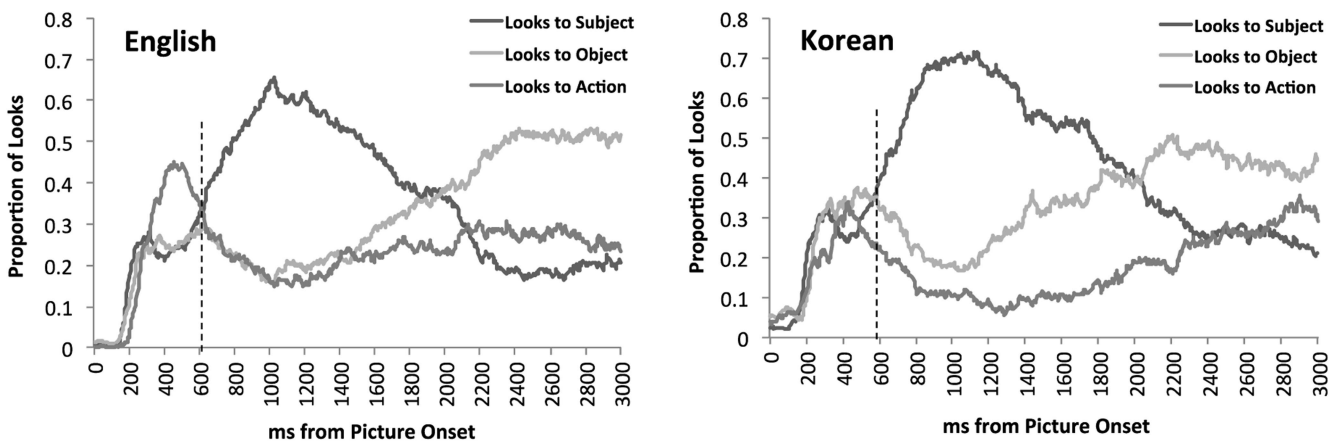


Figure 4. Proportion of looks to subject, object, and action regions in critical items relative to picture onset in English and Korean (collapsing all three distractor conditions within each language). The dotted lines indicate when looks to the subject diverge from looks to the object and the verb.

subject region, speakers should fixate the action region in order to select a verb lemma.

However, if English and Korean speakers can select the subject of a sentence without selecting a verb lemma as suggested by the *structural hypothesis* (e.g., Griffin & Bock, 2000), then we expect that verb-related distractors should not delay utterance onsets. In terms of eye movements, both English and Korean speakers are expected to fixate the subject region *before* fixating the action region or the object region, as the subject is the first element in a sentence.

Yet, even if a verb does not play a key role in grammatical function assignment, it is possible that verb interference effects could arise in English due to its early occurrence in a sentence. That is, if the retrieval of a verb lemma is instead sensitive to word order, then we expect that only English speakers would show interference effects for the verb-related distractors due to the SVO word order of English. If anything, Korean speakers might show interference effects for the object-related distractors, since the object immediately follows the subject in Korean (which has SOV word order). Crucially in terms of eye movements, if the retrieval of a verb lemma is instead sensitive to word order, we would expect English and Korean speakers should fixate the subject region first. Thus, eye movements help us to tease apart whether the verb retrieval is dependent on word order or its role in grammatical function assignment.

Our findings for utterance onset latencies show that in English, verb-related distractors significantly delayed utterance onset. This suggests that English speakers retrieved verb lemmas before speech onset. The eye movement analyses further showed that in general, English speakers looked at the action region early on (400–600 ms after image appears), *before and not after* fixating the subject region (600 ms after image appears). Taken together, these findings suggest that English speakers determined the subject *after* selecting the verb lemma. That is, English speakers build sentences around the verbs, as suggested by the *lexical hypothesis*. This fits with existing work by Lindsley (1975) (see also Kempen & Huijbers, 1983); Iwasaki et al. (2008); and Meyer and Dobel (2003).

However, it is important to note that the lexical and the structural hypotheses are not mutually exclusive. Indeed, during the 200–400 ms time-window English speakers did not preferentially fixate any event element. This, according to Griffin and Bock (2000), indicates that English speakers were initially processing the relational information between scene entities. Yet, the subsequent fixations on the verb region *before* fixations on the subject suggest that the verb is crucial in the sentence formulation processes in English. For example, the verbs *donate* and *give* assign the agent/source to the subject, but *donate* requires the theme to precede the goal unlike *give*.

In contrast to the clear verb interference effects in English, verb-related distractors did not delay utterance onsets in Korean. Korean speakers also did not significantly fixate the action region before fixating the subject region. Rather, they fixated the action region *after* fixating the object region. The absence of verb interference effect and the failure to find an effect of initial fixation on the action region in Korean suggest that Korean speakers did not access verb lemmas before deciding the subjects, which is incompatible with the lexical hypothesis. In fact, we suggest that these results point toward verb-independent function assignment in Ko-

rean. More specifically, the finding that Korean speakers did not show a systematic preference to fixate any one region before the emergence of a preference for the subject region suggests that sentence formulation in Korean may involve an initial time period of scene investigation, which shapes the subsequent determination of the subject and the object (see Griffin & Bock 2000, for related discussion). Our finding that participants showed a significant pattern of fixations on the action region only after having fixated the object further suggest that Korean speakers retrieve a verb lemma after speech onset. This situation is compatible with the *structural hypothesis*.

Note that our findings in Korean are different from the findings of Iwasaki et al. (2008) regarding Japanese, a language that is typologically similar to Korean. We suspect that the different findings might stem from different experimental designs between ours and Iwasaki et al.'s. As noted in the introduction, in Iwasaki et al. (2008), participants were presented with the subject word beforehand and their task was to complete a sentence by naming the action. Because participants had to simply name the action upon the presentation of a picture (and did not have to establish the subject since it was already provided as part of a “sentence fragment to be completed”), we think that verb interference effects are not surprising, but in fact expected. This is because in this task, the only and the first word that participants had to retrieve was the verb. In our experiments, however, the participants' task was not to name the action but to formulate a complete utterance on their own. That is, participants did not have to retrieve a verb lemma immediately upon the presentation of a picture. In our design, the retrieval of a verb lemma is part of the sentence formulation process. Our results suggest that Korean speakers could plan their utterance without necessarily retrieving a verb lemma.

Despite the semantic interference effects, note that Iwasaki et al. (2008) did not find any effect of the distractors' grammatical class in Japanese. This contrasts with the results from other languages. For example, Pechmann and Zerbst (2002) and Pechmann et al. (2004) found that distractors sharing the grammatical class of the targets had interfering effects on naming latencies in German and English, respectively. The effect of grammatical class is attributed to the retrieval of syntactic information (e.g., grammatical gender of the noun). Iwasaki et al. found that both noun and verb distractors that were semantically similar to the target verbs caused semantic interference, but the grammatical class of distractors did not have any effect on naming latencies in Japanese. Together with our finding, this provides further support for the reduced role of the verb in head-final languages like Japanese and Korean.

The fact that Korean speakers (a) fixated the subject during the same time window as English speakers (pointing toward a similar time course of sentence formulation in both languages) but, quite strikingly, (b) did *not* fixate the action region suggests that in both languages eye-movements provide a window into speech planning processes and that the differences we observed point to a real distinction between the utterance formulation process in English and Korean. Additionally, the fact that Korean speakers neither showed significant object interference effects nor fixated the object region before fixating the subject region suggests that the verb retrieval in English cannot be simply attributed to its early position in the sentence, further

highlighting the role of the verb in grammatical function assignment in English.

In sum, utterance latencies and fixation patterns support the lexical hypothesis for English, and the structural hypothesis for Korean. Thus, this research provides the first evidence that—when dealing with the exact same pictures, with the exact same methodology—speakers of different languages engage in language production in different ways at least for transitive sentences. More specifically, our results suggest that the process of grammatical function assignment at least in transitive sentences works in different ways in different languages. The implication is that from a cross-linguistic perspective, neither the lexical hypothesis nor the structural hypothesis is a universal model of grammatical function assignment.

Exploring the reasons for cross-linguistic differences. In light of the divergent results we observed for grammatical encoding in English and in Korean, in this section we discuss how the differences between Korean and English may follow naturally from the grammatical properties of these two typologically different languages. Broadly speaking, we suggest that the different mechanisms of grammatical function assignment in English and Korean are related to the relative contribution that verbs make to sentence production in these languages. Let us consider this in some more detail.

English is a head-initial SVO language. It has fairly strict word order and defines the grammatical functions of arguments in terms of their relative order to the verb. For example, in a canonical active transitive sentence, the subject precedes the verb and the object follows the verb. When English speakers are given a sentence with two nouns and a verb and asked to identify the subject of the sentence, they predominantly chose the preverbal noun as the subject (e.g., Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; MacWhinney, Bates, & Kliegl, 1984; Townsend & Bever, 2001). Thus, verbs play a powerful role in defining grammatical functions in English. In addition, verbs also play an important role in constraining structural options in English. For example, the verb *donate* allows a prepositional object structure (e.g., *The man donated some money to the charity*), whereas the synonymous verb *give* allows both a prepositional object and a double object structure (e.g., *The man gave some money to the charity/The man gave the charity some money*). If verbs provide crucial information to grammatical functions and constrain syntactic structures, it is only natural for verbs to play a pivotal role in English production, and for delays in verb retrieval to lead to disfluency (see Iwasaki, 2011, for related discussion).

The key role of the verb in English function assignment is also evidenced by comprehension research. Garden path sentences in particular clearly demonstrate that English speakers' parsing process is greatly constrained by the subcategorization frames of verbs (e.g., Britt, 1994; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Snedeker & Trueswell, 2004). For example, in a sentence like *put the frog on the napkin into the box*, English speakers tend to initially misinterpret "on the napkin" as a Goal for the verb *put* because it has a strong subcategorization bias in favor of a goal thematic role (Tanenhaus et al., 1995; Trueswell, Sekerina, Hill, & Logrip, 1999).

Unlike English, *Korean* is a head-final SOV language. In Korean, grammatical functions are indicated by case markers

and not by word order. For example, in an active transitive sentence, the subject is marked with the nominative marker, and the direct object is marked with the accusative marker. Because case markers provide information about the grammatical functions of arguments, word order relative to the verb does not indicate grammatical functions in Korean. For example, Sentences 1a and 1b represent a chasing event where a dog is the agent and a cat is the patient. In 1a, the agent *dog* is the sentence-initial noun (SOV order), whereas in 1b the patient *cat* is the sentence-initial noun (OSV order). Yet in both sentences, *dog* is the subject and *cat* is the object because *dog* is marked with nominative case and *cat* with accusative case.

- Kay-ka koyangi-lul cchochnunta. [Active: Agent-Patient-Verb]
 1a. dog-NOM cat-ACC chase
 "A dog is chasing a cat."
 Koyangi-lul kay-ka cchochnunta. [Active: Patient-Agent-Verb]
 1b. cat-ACC dog-NOM chase
 "A dog is chasing a cat."

As grammatical functions of arguments are identified by case markers and not by their order relative to the verb, the role of the verb is much less reduced in Korean than in English. In fact, the reduced role of the verb in defining grammatical functions in Korean is well demonstrated in a sentence such as Sentence 2.

- John-i cha-lul kochoyesta.
 2. John-NOM car-ACC repair
 "John had his car repaired."

The sentence is about John, but crucially John is not the agent of repairing in Sentence 2, suggesting that John is not selected by the verb *repair*. (John is a topic; see Li & Thompson, 1976, for detailed discussion.) This is in striking contrast to its literal equivalent in English "John repaired his car," where John is the agent of repairing. Constructions like Sentence 2 then show that the verb exerts much less influence on grammatical function assignment or sentence structure in Korean compared to English.

The limited role of the verb in grammatical function assignment receives further support from comprehension research. A recent study by Choi and Trueswell (2010) showed that Korean speakers could assign grammatical functions independently of verbs during comprehension. For example, Sentence 3 is temporarily ambiguous because the case particle *-ey* is ambiguous between the genitive and the locative marker.

- Naypkhin-ey kaykwuli-lul nohuseyyo/cipuseyyo.
 3. Napkin-ey frog-ACC put/pick up
 "Put/Pick up the frog on the napkin."
 (from Choi & Trueswell, 2010)

The temporal ambiguity is resolved either in favor of the modifier interpretation upon hearing *pick up* ("pick up the frog on the napkin") or the goal interpretation upon hearing *put* ("put the frog on the napkin"). Choi and Trueswell (2010), however, showed that when encountering *-ey*, Korean speakers did not wait until the verb became available but often interpreted it as the locative marker, because *-ey* is used much more frequently

as the locative case (see Sells, 2008, for syntactic evidence of verb-independent case assignment).

In related work on Japanese comprehension, Kamide, Altmann, and Haywood (2003) provided evidence that function assignment during the parsing of a head-final language can be guided by speakers' understanding of events. In their study, Japanese speakers looked at pictured events while listening to sentences that described what might happen in the picture. Kamide et al. found that Japanese speakers were able to anticipate a forthcoming referent in the picture on the basis of the meanings and grammatical roles of the previously occurred arguments in the pictured event. These findings suggest that speakers of a verb-final language do not depend on verbs to assign grammatical functions but, rather, make use of the relationships among the arguments consistent with a structural hypothesis.

The verb-final Korean structure also does not fare well with the lexical hypothesis, considering the incremental nature of speech production (i.e., speakers do not necessarily plan a complete sentence before starting to speak; e.g., Brown-Schmidt & Konopka, 2008; Griffin, 2001). Under the lexical hypothesis, the verb needs to be selected first—thus, Korean speakers should first access the *sentence-final verb* in order to be able to assign grammatical functions (see also Kamide et al., 2003, on Japanese comprehension). But this idea clashes with the findings highlighting the extreme incrementality of language production. The claim that Korean speakers should first access the sentence-final verb also violates principles of efficient processing (Hawkins, 1994, 2004, 2007)—as it would mean that Korean speakers have to keep the verb active in memory throughout the production of the sentence. Crucially, however, under the structural account, such difficulties do not arise: If Korean speakers assign grammatical functions under conceptual guidance (i.e., according to their conceptual representation of an event, e.g., figure vs. ground), they can initiate their utterances prior to selecting the sentence-final verb, and thus can produce utterances incrementally. In fact, if case marks the relationship between sentence elements (e.g., Haspelmath, 2006), it could be assigned based upon speakers' coding of an event that includes information about functional or hierarchical dependencies between referents and their ordering (in much the same way as the listeners could construct the relation between event elements based upon case-marking in Choi & Trueswell, 2010; Kamide et al., 2003). In sum, verb-independent grammatical function assignment in Korean makes sense if we consider the grammatical properties of Korean as well as the incremental sentence processing mechanism.

As a whole, our data—combined with grammatical properties of English and Korean—suggest that different languages may adopt different grammatical function assignment mechanisms to allow production to proceed in a smooth and efficient way. Keeping this in mind, let's think back to German, where verbs occur in the second position in main clauses but are clause-final in subordinate clauses. In their work on German, Schriefers et al. (1998) found verb interference effects in main clauses (where a verb was the first sentence element that participants had to produce) but not in subordinate clauses (where a verb was the last element to be produced). We suspect that this asymmetry may be due to speakers' use of a strategy

(see Ferreira & Swets, 2002), rather than the different role of the verb in a main and a subordinate clause in German—as we do not believe that the role of the verb changes depending on a clause type.² Recall that in Schriefers et al. (1998), speakers could determine the subject without necessarily depending on the selection of the verb. The subject was always animate (e.g., man), whereas the objects were inanimate (e.g., bucket). Thus, when forced to produce the verb first immediately after a lead-in fragment in a main clause, German speakers might have simply retrieved the verb due to their early position. When the verb appears in the clause-final position in a subordinate clause, however, they might have used the strategy of assigning the animate entity to the subject without retrieving the verb. However, as German uses both word order and case marking for function assignment, and the position of the verb in matrix clauses and embedded clauses is different, future research might take advantage of this variability in order to better understand how the retrieval of the verb lemma depends on the role of the verb, the canonical word order, and the flexibility of word order in a language.

Conclusions. The two picture–word interference experiments that we conducted on English and Korean showed that verb distractor words delayed speech onsets in English but not in Korean. We also found that English speakers looked early on at the action region before looking at the subject region. In contrast, Korean speakers did not show clear preferential fixation on one region before fixating the subject region. As a whole, our results suggest that English speakers accessed verb lemmas before subject lemmas, indicating that English speakers built sentences around the verbs at least for transitive sentences. In contrast, our findings for Korean suggest that Korean speakers did not necessarily retrieve verb lemmas before speech onset, which in turn indicates that verb lemmas do not play a central role in Korean production.

The asymmetry between English and Korean suggests that languages may differ in their function assignment mechanisms and that neither the lexical hypothesis nor the structural hypothesis is a universally valid model of grammatical function assignment. We suggest that the different function assignment mechanisms used in English and Korean can be connected to the grammatical differences between these two languages.

To conclude, the present study provides first evidence that the role of verb in function assignment is not universal (at least for transitive sentences) and suggests that different languages can adopt different grammatical function assignment mechanisms. More generally, this study contributes to our understanding of the broad question of whether and how the specific grammatical properties of a language interact with the architecture of real-time language production.

² We believe what might determine the retrieval of a verb before speech onset is the role of the verb in grammatical function assignment or structural choice in that language. English allows object-initial sentences such as preposing (e.g., *carrots, I like*; see Birner & Ward, 1998; Ward, 1988). Crucially, in these sentences the verb still constrains grammatical function assignment or sentence structure regardless of its position in a sentence. For example, the verb *donate* allows “To the church, I donated toys,” but not “The church, I donated toys.” Thus, we suggest that regardless of word order, the role of a verb may remain invariable in a language.

References

- Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York, NY: Psychology Press.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511801686
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. doi:10.1016/j.jml.2007.12.005
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Retrieved from <http://www.celelex.mpi.nl>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using Eigen and Eigen++ (R package version 0.99875-6) [Computer software].
- Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence comprehension: A cross-linguistic study. *Cognition*, 11, 245–299. doi:10.1016/0010-0277(82)90017-8
- Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam, the Netherlands: Benjamins.
- Blumenthal, A. L. (1970). *Language and psychology: Historical aspects of psycholinguistics*. New York, NY: Wiley.
- Bock, J. K., & Ferreira, V. S. (in press). Syntactically speaking. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production*. Oxford, England: Oxford University Press.
- Bock, J. K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The integration of language, vision, and action: Eye movements and the visual world* (pp. 249–278). New York, NY: Psychology Press.
- Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, 48, 653–685.
- Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.
- Boersma, P., & Weenink, D. (1992–2011). *Praat: Doing phonetics by computer*. Retrieved from <http://www.praat.org/>
- Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33, 251–283. doi:10.1006/jmla.1994.1013
- Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, 109, 274–280. doi:10.1016/j.cognition.2008.07.011
- Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden-paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106, 41–61. doi:10.1016/j.jecp.2010.01.003
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.
- deSmedt, K. J. M. J. (1996). Computational models of incremental grammatical encoding. In T. Dijkstra & K. deSmedt (Eds.), *Computational psycholinguistics* (pp. 279–307). London, England: Taylor & Francis.
- Dhooge, E., & Hartsuiker, R. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language*, 66, 163–176. doi:10.1016/j.jml.2011.08.004
- Ferreira, F. (2000). Syntax in language production: An approach using tree-adjointing grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Hove, England: Psychology Press.
- Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57–84. doi:10.1006/jmla.2001.2797
- Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic context of subcategorization frames. *Cognitive Psychology*, 23, 331–392. doi:10.1016/0010-0285(91)90013-E
- Garney, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93. doi:10.1006/jmla.1997.2512
- Glaser, W. R., & Glaser, M. O. (1989). Context effects on Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13–42. doi:10.1037/0096-3445.118.1.13
- Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544–569. doi:10.1016/j.jml.2007.01.007
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82, B1–B14. doi:10.1016/S0010-0277(01)00138-X
- Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279. doi:10.1111/1467-9280.00255
- Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
- Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, 42, 25–70. doi:10.1017/S0022226705003683
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, England: Cambridge University Press.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammar*. Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780199252695.001.0001
- Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. *New Ideas in Psychology*, 25, 87–107. doi:10.1016/j.newideapsych.2007.02.003
- Huetig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96, B23–B32. doi:10.1016/j.cognition.2004.10.003
- Iwasaki, N. (2011). Incremental sentence production: Observations from elicited speech errors in Japanese. In Y. Hiroko, Y. Hirose, & J. Packard (Eds.), *Studies in theoretical psycholinguistics: Vol. 38. Processing and producing head-final structures* (pp. 131–151). New York, NY: Springer Science.
- Iwasaki, N., Vinson, D. P., Vigliocco, G., Watanabe, M., & Arciuri, J. (2008). Naming action in Japanese: Effects of semantic similarity and grammatical class. *Language and Cognitive Processes*, 23, 889–930. doi:10.1080/01690960801916196
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. doi:10.1016/j.jml.2007.11.007
- Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, 15, 281–319. doi:10.1515/lity.2011.021
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156. doi:10.1016/S0749-596X(03)00023-8
- Kang, B., & Kim, H. (2009). *Korean word frequencies*. Seoul, Korea: Hankukmunhwa Press.
- Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14, 185–209. doi:10.1016/0010-0277(83)90029-X
- Kuchinsky, S. E., Bock, J. K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of*

- Experimental Psychology: Learning, Memory, and Cognition*, 37, 748–756. doi:10.1037/a0022637
- La Heij, W., Driks, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, 81, 511–525. doi:10.1111/j.2044-8295.1990.tb02376.x
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanism in behavior* (pp. 112–136). New York, NY: Wiley.
- Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 458–489). New York, NY: Academic Press.
- Lindsay, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, 7, 1–19. doi:10.1016/0010-0285(75)90002-X
- Lupker, S. J. (1979). The semantic nature of response competition in the picture–word interference task. *Memory & Cognition*, 7, 485–495. doi:10.3758/BF03198265
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning & Verbal Behavior*, 23, 127–150. doi:10.1016/S0022-5371(84)90093-8
- Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, 14, 455–462. doi:10.1162/089892902317361967
- Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics*, 53, 372–380. doi:10.3758/BF03206780
- Meyer, A., & Damian, M. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, 35, 494–503. doi:10.3758/BF03193289
- Meyer, A. S., & Dobel, C. (2003). Application of eye tracking in speech production research. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 253–272). Amsterdam, the Netherlands: Elsevier. doi:10.1016/B978-044451020-4/50015-3
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from <http://www.usf.edu/FreeAssociation/>
- Pechmann, T., Garrett, M., & Zerbst, D. (2004). Observation: The time course of recovery for grammatical category information during lexical processing for syntactic construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 723–728. doi:10.1037/0278-7393.30.3.723
- Pechmann, T., & Zerbst, D. (2002). The activation of word class information during speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 233–243. doi:10.1037/0278-7393.28.1.233
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134, 427–459. doi:10.1037/0033-2909.134.3.427
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. doi:10.1037/0033-2909.124.3.372
- Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, 47, 59–87. doi:10.1016/0010-0277(93)90062-Z
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71, 145–163. doi:10.1016/j.jml.2013.11.002
- Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, 57, 1030–1033.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (2002). Exploring the time course of lexical access in language production: Picture word interference studies. In G. Altmann (Ed.), *Psycholinguistics: Critical concepts in psychology* (Vol. 5, pp. 168–191). London, England: Routledge.
- Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language*, 39, 609–632. doi:10.1006/jmla.1998.2578
- Schyns, P., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5, 195–200. doi:10.1111/j.1467-9280.1994.tb00500.x
- Sells, P. (2008). *Oblique case marking on core arguments in Korean and Japanese: Workshop on non-canonical perspectives on case*. Konstanz, Germany: University of Konstanz.
- Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49, 238–299. doi:10.1016/j.cogpsych.2004.03.001
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. doi:10.1126/science.7777863
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89–134. doi:10.1016/S0010-0277(99)00032-3
- Van der Meulen, F. F. (2001). *Moving eyes and naming objects: MPI series in psycholinguistics*. Nijmegen, the Netherlands: UB Nijmegen.
- Van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. M. (2001). Eye movements during the production of nouns and pronouns. *Memory & Cognition*, 29, 512–521. doi:10.3758/BF03196402
- Vigliocco, G., Vinson, D. P., & Siri, S. (2005). Semantic similarity and grammatical class in naming actions. *Cognition*, 94, B91–B100. doi:10.1016/j.cognition.2004.06.004
- Ward, G. (1988). *The semantics and pragmatics of preposing*. New York, NY: Garland.
- Wilcox, R. (2012). *Modern statistics for the social and behavioral sciences: A practical introduction*. New York, NY: CRC Press.
- Wundt, W. (1970). The psychology of the sentence. In A. L. Blumenthal (Ed. and Trans.), *Language and psychology: Historical aspects of psycholinguistics* (pp. 20–31). New York, NY: Wiley. (Original work published 1900)

(Appendix follows)

Appendix

Experimental Items

Target	Verb-related distractor	Object-related distractor	Unrelated distractor
A pirate tickles a penguin.	fiddle	seal	sparse
A giraffe licks a pirate.	sip	bandit	vain
A penguin chases a clown.	stalk	magician	uneven
A clown arrests a giraffe.	shackle	deer	faulty
A chef tickles a turtle.	fiddle	toad	petty
A turtle licks an alien.	sip	monster	dense
A zebra chases a chef.	stalk	baker	creamy
An alien arrests a zebra.	shackle	donkey	crude
Filler items			
A witch buries a boar.			
A nun hugs a duck.			
A knight pats a squirrel.			
A duck pecks a nun.			
A boar rams a vampire.			
A fox scratches a witch.			
A dwarf buries a fox.			
A knight hugs a deer.			
A mermaid pats a peacock.			
A peacock pecks a dwarf.			
A deer rams a vampire.			
A squirrel scratches a mermaid.			

Received July 24, 2013
Revision received February 4, 2014
Accepted March 12, 2014 ■