Visual Scene Description and Recall: On Differences between Korean and English Speakers*

Mun-Hong Choe
(Chonnam National University)

Choe, Munhong. (2015). Visual Scene Description and Recall: On Differences between Korean and English Speakers. ***Language Research*****, 51.1, 197-223.**

Scientific research into the relationship between language and thought has profound implications for the understanding of second language learners and their learning process. The present study focuses on the connections between verbalization patterns and perceptual orientations. Adult monolingual speakers of Korean and English were asked to describe a set of dynamic scenes at two different presentation durations, 1.5 and 3 seconds, and then recall the focal figures and backgrounds of the depicted situations. Participants' utterances were counted for comparison with their performance on the subsequent recall test. The study found that Korean speakers (KSs) mentioned more specific subjects at longer durations whereas English speakers' (ESs) choice of subject referents was not significantly influenced by time. Further, in contrast to previous research, ESs produced more background details than KSs and were inclined to add more descriptions about figures at longer durations while dismissing a commensurate amount of background information. In the recall test, KSs remembered background details more accurately than ESs. This asymmetry in linguistic encoding and recall in function of presentation durations suggests that the effect of a particular language on the language speakers' attentional allocation and information storing is far more intricate than was reported in earlier works.

Keywords: attention, linguistic relativity, memory, picture description, thinking for speaking

1. Introduction

The scope of inquiry in language teaching and learning continues to diversify with varying interests, methods, theoretical and empirical pursuits

* This research was supported by the 2012 Chonnam National University research grant.

allying themselves to enhance their scientific base and utility. Nevertheless, the fundamental question of whether thought is conditioned by language has long been of recurring interest. Debates on the interaction between language and thought have recently been revitalized in applied linguistics (e.g., Cadierno, 2008; Casasanto, 2008; Odlin, 2008; Pae, 2012). There is renewed focus on the rationale that cross-language studies concerning linguistic relativity are essential to understand not only linguistic features affecting cognitive domains but also first language (L1) influences on second language (L2) development.

It is known that a major difficulty in learning an L2 arises when learners project concepts from their L1 into L2 processing. If L1 formulation of events for verbal encoding is markedly different from that of an L2, L2 learning requires not just the learning of its vocabulary but the learning of alternative ways of information structuring. Language-specific information structures may direct the speaker's attention to specific aspects of objects and events which must be encoded for reliable communication in the language community. It is thus reasonable that L2 researchers have a growing interest in L1 influence on the speaker's selective attention and the influence of the forestalled L1 cognitive patterns on L2 acquisition (e.g., Cadierno, 2010; Ellis, 2008; Han & Cadierno, 2010; Odlin, 2008; Tajima & Duffield, 2012). Most writers in the field concede that one's experiences are filtered through language into verbalized events, so any utterance is a selective schematization of a concept, which is dependent on the grammaticized meanings of the speaker's language (Slobin, 1996, 2003).

Given that people are born with a common set of perceptual and cognitive tools while the languages they learn differ in detail with respect to grammar and discourse structure, it is natural to ask whether crosslinguistic differences affect how people perceive the world. Speakers of typologically dissimilar languages are required to verbalize different aspects of reality when constructing linguistic messages. For example, languages differ in how to label colors and spatial relations and this may affect the speakers' categorization, similarity judgments, and memory of a referent's physical properties (Bowerman, 1996; Kay & Regier, 2006; Levinson, 2003; Li

& Gleitman, 2002; Lucy, 1992; Regier & Kay, 2009). The opposing view, often referred to as the linguistic universality hypothesis, posits that beneath surface differences in how people express the world, there is a universal repertoire of thought among speakers of all languages (Fodor, 1983; Pinker, 1994). The strong universality perspective predicts that even though languages differ in their semantic distinctions, speakers retain universal perceptual orientations. For example, colors are categorized differently across languages, but they can be perceived by all humans in a universal spectrum (Kay & Kempton, 1984; Roberson, Hanley, & Pak, 2009; but see also Winawer, Witthoft, Frank, Wu, Wade, & Boroditsky, 2007). It seems no longer tenable to contend that language entirely determines cognition. However, it is still a matter of debate whether and how far language affects it.

In this study, we investigate the link between language and perception by testing whether speakers of different languages show characteristic differences in their description and retention of visual scenes. Previous research has shown that East Asians attend more to background elements and Westerners attend more to focal objects. For example, when Americans and Japanese process visual scenes, the American descriptions focus on the salient focal objects and the Japanese descriptions focus more on the background context (Masuda & Nisbett, 2001). When they were asked to compare two consecutively presented images in a change detection task, Americans detected more changes in focal objects while Japanese speakers detected more changes occurring in the backgrounds (Masuda & Nisbett, 2006). Studies using eye trackers have also demonstrated that East Asians are inclined to shift quickly away from focal figures and toward backgrounds while Americans' fixation on focal figures is relatively sooner and longer than that of East Asians (Chua, Boland, & Nisbett, 2005). These cognitive differences may reflect broad-range cultural differences in Asian and Western ways of thinking (Boduroglu, Shah, & Nisbett, 2009; Masuda & Nisbett, 2001). More recently, however, in a study involving Chinese and Japanese speakers, Tajima and Duffield (2012) found that at least a significant part of the explanation for the East-West differences in scene description and recall are attributable to

grammatical structure rather than to pan-Asian cultural factors.

It has been widely attested that crosslinguistic variations in lexicalization patterns affect the speakers' cognitive processes (Choi & Bowerman, 1991; Oh, 2003; Slobin, 2003; Talmy, 1985, 2000). But the question of whether those in sentence structure do so has remained a matter of speculation to a large extent. Moreover, empirical studies involving Koreans are few (e.g., McDonough, Choi, & Mandler, 2003; Munnich, Landau, & Dosher, 2001; Roberson, Pak, & Hanley, 2008). Only a handful of studies have addressed the linguistic relativity hypothesis directly, and most of them have dealt with a small set of lexical items, hence failing to provide direct evidence regarding cognitive relativity (Choe, 2012; Lucy, 1996). The purpose of this study is two-fold. First, we investigate Korean L1 speakers' visual scene processing which has not been reported in the relevant literature. Using a picture description and recall paradigm, we draw a comparison between Korean and English speakers in terms of the number of figure and background elements included in their time-constrained online utterances and the extent to which it correlates with their subsequent recall of such elements. Second, we observe perceptual process as well as outcome by varying the length of image display time so as to see what information is given priority in linguistic encoding when scenes are presented very briefly (1.5 sec) and what details are added when they are presented for a longer time (3 sec). In so doing, we try to identify more exact points of distinction between the two groups' message formulation in relation to their usual manner of perception.

In short, the present study explores whether crosslinguistic variations in grammar have an effect on the speakers' perceptual process and retention of visual stimuli that can be selectively verbalized by the language spoken. This is addressed as an attempt to identify a putative source of difficulty in L2 acquisition caused by the learner's cognitive patterns habituated over the course of L1 acquisition.

2. Background of the Study

2.1. Linguistic Relativity in the Context of L2 Acquisition

The Whorfian hypothesis, which posits that the language we speak influences the way we perceive, understand, and interpret the world, has been a topic of intense interest for well over a century. Some scholars argue that language affects nonlinguistic cognition (Levinson, 1996, 1997; Lucy & Gaskins, 2001; McDonough, Choi, & Mandler, 2003; Pederson et al., 1998; Yoshida & Smith, 2005) as well as linguistic processes (Gennari, Sloman, Malt, & Fitch, 2002; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). When English speakers are taught a name for a novel thing, for example, they typically infer the word refers to a kind of object and extend it to other objects with the same shape. In contrast, Japanese speakers more often interpret the word as referring to the thing's substance (Athanasopoulos, 2006; Imai & Gentner, 1997). Evidence against the strong form of linguistic determinism is also found in numerous studies (Hunt & Agnoli, 1991; Li & Gleitman, 2002; Mazuka & Friedman, 2000; Munnich, Landau, & Dosher, 2001). For instance, when presented with novel objects varying in shape, substance, and complexity, speakers of Chinese, Japanese, and English behave alike in rating whether or not they are objects (Li, Dunham, & Carey, 2009). Moreover, Chinese-English bilinguals tend to extend words by shape more often when tested in English than when tested in Chinese, indicating that it is linguistic cues rather than learned conceptual distinctions that make speakers more likely to extend novel words to objects (Barner, Inagaki, & Li, 2009). In the study of Coventry, Valdés, and Guijarro-Fuentes (2010) involving Spanish and English speakers, the researchers predicted that since Spanish has one lexeme (*en*) for two spatial prepositions in English (*in* and *on*), their sensitivity to spatial relations such as containment and support might differ. It turned out however that there was no significant difference between them. These discrepancies in earlier findings imply that a rigorous assessment of the extent to which differences in linguistic features across languages yield differences in the speakers' perceptions of the world is still required.

Speech begins with message construction – an interface linking thought and language where concepts or intentions are represented in (allegedly) primitive form (Vigliocco, Vinson, Lewis, & Garrett, 2004). These concepts are converted into lexemes with semantic and syntactic features, which are merged by syntactic operations. Then, vocabulary items, the phonological exponents of the lexemes, are inserted and produced by virtue of articulators. In this view, language processing consists of modules of grammar and interfaces with conceptual-intentional and articulatory-perceptual systems (Hauser, Chomsky, & Fitch, 2002). Though languages share in common a greater part of concepts, each language has its own ways of combining them into a sentence. In the process of sentence construction, the speaker tries to make optimal use of memory by way of being moderately incremental (Ferreira & Swets, 2002; Kempen & Hoenkamp, 1987; Levelt, 1989). This incremental pattern is evident when speakers reduce processing burden by uttering sentences gradually with intermediate linkers such as pauses, fillers, syllabic lengthening, etc., especially when they spell out a composite of multiple phrases or clauses. Therefore, it can be argued that language-cognition interfaces are customized in accordance with the structural properties of a given language; that is, grammar plays a central role in shaping our cognitive processes from initial message planning to phonological output, at least at the moment of speaking. This thesis is embodied recently in Slobin's (1996, 2003) thinking-for-speaking hypothesis.

2.2. Grammatical Contrasts between Korean and English

Grammatical relations expressed in a sentence can be classified into two kinds: principal relations that necessarily appear in a well-formed sentence and adjunctive relations that add or modify meanings irrespective of syntactic well-formedness. For example, the relation of a predicate and its subject (or vice versa) must be expressed in a grammatical sentence whereas the relation of a predicator and an adverbial phrase in its projection is syntactically adjunctive. Speakers typically utter a grammatically coherent segment of discourse with a regular pattern of phonological boun-

dary, known as intonation units (or thought groups). A complete intonation contour is almost always accompanied by a grammatical completion (Ford & Thompson, 1996). In English, most overt functional morphemes for agreement and case have disappeared, and instead grammatical relations are normally denoted by word order. On the other hand, Korean employs a number of functional morphemes to express grammatical relations, and thus allows relatively free word order. English is an S-VO-A(djuncts) language in which verbs are spelled out before their complements and/or adjunctive phrases, while the basic word order of Korean is S-A-OV where verbs appear after their complements or other contextual elements. If this grammatical parameter has a bearing on the speakers' sentence production process, viz., if word order affects the order of planning a sentence online, we may hypothesize that ESs are inclined to focus first on the main figure and its characteristic state or action and then move toward background details. So, there is no intermediary delay in the selection of a subject and a predicator. This (hypothetical) language-biased perceptual orientation can be seen as field-independent. On the other hand, KSs are inclined to consider the figure and background together before they opt for a predicator. In other words, the Korean language requires its speakers to place background elements ahead of a predicator, and so they are predisposed to attend to contextual information and take it into account in the selection of a predicator. In this sense, KSs are field-dependent in the process of sentence formulation.

Another relevant point of contrast between English and Korean is known as the null-subject parameter in the generative tradition (Huang, 1989; Jaeggli & Safir, 1989). While English requires a subject to be expressed in a tensed clause, Korean can employ a null-subject (i.e., an unpronounced understood subject) when the subject referent has already been topicalized in the preceding context or it is informationally marginal and thus should rather be backgrounded instead of being highlighted in sentence initial position. One of the widely held postulates in theoretical linguistics is that verbs in English are represented in the lexicon with their argument structures and subcategorization frames. Since the arguments of a verb are regularly realized in the sentence, it is likely that

ESs encode a verb and its arguments in the first place, thereby the core structural template of a sentence is determined. This characteristic of English in sentence structure appears to be associated with the fact that the language lacks functional morphemes to mark grammatical relations. In contrast, Korean speakers settle on a topic at first, constructing a message by way of electing a topic accessible in the given discourse context (Li & Thompson, 1976; Rutherford, 1983; cited in Choe, 2012). In other words, the process is not of configuring a verb and its core participants, but of picking up an information unit to be predicated of.

Furthermore, the difference in question also manifests itself when speakers refer to an identical referent in the subsequent discourse. ESs ordinarily replace it with a pronoun specified for person, number, and gender, and so they have to pay attention to the grammatical features in order to employ an appropriate anaphor afterwards. On the other hand, KSs often leave out the referent if it has been topicalized in the preceding discourse, or instead use a demonstrative determiner plus a noun classifier. They rarely use gender-specified pronouns in spoken communication. Since Korean has no regular subject-verb agreement other than honorifics, the referent of a null-subject is inferred from the context. The form in which one receives and produces information may influence how that information is understood, stored, and accessed later. This leads to the hypothesis that ESs will pay more attention to a human referent's gender than KSs will, especially when the referent is newly introduced. Thus, in a picture description and recall task, ESs would more likely retain the focal figure's gender information than KSs would. This prediction may be fundamentally flawed because Korean has a rich set of gender- and age-specified noun classifiers, such as *namca* (man), *yeca* (woman), *sonyen* (boy), *sonye* (girl), *ay/ai* (person junior to the speaker), *pwun* (person senior to the speaker), *noin* (old person), etc. However, the choice of a specific classifier in a given context is not as grammaticalized as the English pronoun system. It is also noteworthy that Korean makes use of age-specified anaphoric referents and predicators that are not found in English. Then, will it be the case that KSs pay more attention to a human referent's age than ESs? These hypotheses are at least testable.

More importantly, in order for the reasoning to be any more than speculation, it is essential to make a direct comparison between the speakers' actual utterances and perceptual predispositions, which can help provide empirical justifications for the hypotheses.

Languages do not differ from one another only in terms of grammatical features realized in referential expressions but in terms of semantic components incorporated in motion verbs. According to Talmy (1985, 1991, 2000), English and most Indo-European languages form one group of languages, namely, satellite-framed languages (S-languages), in which motion and manner are encoded together within a verb, with the path of the motion encoded by a separate particle (i.e., satellite). On the other hand, Korean, Japanese, and Semitic languages form another group, verb-framed languages (V-languages), in which motion and path are encoded together within a verb, and manner information is encoded by separate adverbials. In S-languages where path information can be accumulated with multiple particles or prepositional phrases, the speakers' attentional focus is on actors as path satellites indicate changes in their location. In V-languages, however, speakers cannot mount paths in such a compact way, and so instead of tracking actors, they must attend to the scenes holistically in order to locate actors in relation to the context. These give rise to the hypothesis that speakers of S-languages are inclined to specify the details of trajectories and leave the settings to be inferred whereas speakers of V-languages tend to describe the static aspects of scene settings, leaving trajectories to be inferred (Slobin, 1996, 2003). It has indeed been attested that speakers of V-languages show a marked difference from those of S-languages in describing the path and manner of motion (Cadierno, 2010; Stam, 2010). It is thus predicted that ESs will focus more on figures and less on background settings than KSs, especially when describing dynamic scenes.

Drawing on the foregoing general discussion, the present study addresses the following three research questions: (a) Do Korean and English native speakers produce an equivalent amount of figure and background information when they describe briefly presented scenes? (b) If they are allowed to watch the scenes for a longer time, what information will

be added in their descriptions? (c) What kinds of information do they recall more accurately after the time of utterance?

3. Method

3.1. Participants

Thirty-nine paid participants were recruited from a university in Korea. Twenty were Korean college students whose ages ranged from 20 to 24 (9 male, 11 female). They had been learning English as a foreign language for over six years in schools. Their TOEIC (the Test of English for International Communication) scores were lower than 500. No one had the experience of studying abroad. The other nineteen were English native speakers from the U.K., U.S.A., Canada, and New Zealand, who were teaching English conversation skills in the university. They had lived in Korea for 1.3 to 5 years, whose ages ranged from 25 to 36 with an average of 29 (12 male, 7 female). No one was proficient in L2 Korean.

3.2. Instruments

Twenty-four uncolored drawings created by a professional painter and eight images from the Internet sources were used. They all depicted scenes in which a person was carrying out a common activity such as hiking, washing a car, picking up an object on the street, etc. The size of the person occupied approximately one third of the whole image presented on a white background. See Figure 1 for sample items.

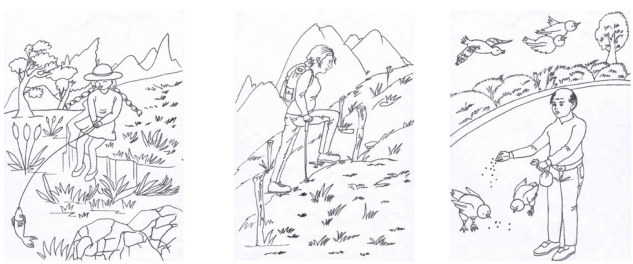

Figure 1. Sample Picture Items.

Each picture contained one person in a particular place. The person's age and gender and the place where s/he acts could be clearly recognized as assured by a preliminary test with ten English and ten Korean speakers who unanimously agreed on the judgment of the three features. The scenes were presented sequentially, each appearing on the screen in three different duration conditions: 1.5, 2, and 3 seconds. Among the total of thirty-two pictures, eighteen were critical items presented to participants for either 1.5 or 3 seconds. The rest were presented for 2 seconds serving as fillers that kept participants from making a prediction on how long a picture would remain on the screen. In order to determine what information in the pictures participants attended to, a five-point Likert-scale questionnaire was distributed five minutes after stimuli presentation. This questionnaire consisted of eighteen true or false statements regarding three units of information: age and gender of the main figure and the place where the event occurs. For each type of information, six descriptions were allotted equally, among which three were true and three were false. Participants were asked to choose only one option that they thought fit the best. Consider Table 1 for sample test items.

Table 1. Sample Memory Test Items

	Definitely wrong	Probably wrong	Don't remember	Probably correct	Definitely correct
It's a middle-aged person who is picking up a book on the street.					
It's a female who is standing at the crossroads.					
It's in the playground that a person is reading a book.					

3.3. Procedure

Participants were told that they were taking part in an experimental study of cross-cultural differences in visual scene description. They were instructed to watch pictures on a computer screen and describe each orally in one complete sentence. Before the main experiment, three practice items were given. As mentioned before, critical items were presented for

either 1.5 or 3 seconds at random while filler items distributed irregularly in between them were presented for 2 seconds. As soon as a picture disappeared, there was a five-second interval within which the speaker had to finish uttering a sentence. Their oral responses were recorded and then transcribed later. One person each from the two groups was not recorded due to malfunction of the recorder.

Five minutes after the picture description task, a memory test was administered to each participant who was not given a notice in advance. Participants were asked to read the written direction on the test sheet and then choose the best answer for each question among the given fives as seen in Table 1 above. Instead of using a dichotomous true/false scale, the questionnaire was designed to elicit the respondents' scaled degree of certainty. This method was thought to be more effective not only in terms of the prevention of chance or careless performance but also in terms of the estimation of one's memory on a finer scale, thereby permitting the application of more powerful statistical tests.

3.4. Data Coding

Sentences produced by each participant were analyzed using a coding scheme with four information categories: figure, event, background, and others. Here the term "figure" is used to refer to the actor in each picture, usually encoded as the grammatical subject of the sentence. It was classified into five subtypes: subjects without any specification (S[Ø]), subjects with a specification for age (S[age]), subjects with a specification for gender (S[gender]), subjects with a specification for both age and gender (S[age][gender]), and other subject-modifiers (S-modifiers) such as adjectives and relative clauses. As we will see in the next section, variability in the description of a figure was large enough to be analyzed and interpreted in a meaningful way. Background refers to the contextual setting in which a depicted event or situation takes place, most often realized by locative pre-/post-positional phrases. Event means a dynamic verb accompanied by its semantic and syntactic complement. Recall that every experimental picture portrayed an actor who was performing a certain

action. This was because if the figure had been stationary, speakers might have focused predominantly on background elements. The category "others" include such information as cause, reason, and purpose, often expressed in the form of adverbials. Since they stemmed from the speaker's subjective interpretation, not directly from the pictures, they were excluded from the counting.[1] Two researchers coded the data independently, and the intercoder agreement was over 98%.

A participant's response to each question in the memory task was scored from 0 to 4. In case the given statement was true, those who marked 'definitely wrong' gained 0 points, 'probably wrong,' 1 point, 'don't know,' 2 points, 'probably correct,' 3 points, and 'definitely correct,' 4 points. Likewise, when the given statement was false, those who marked 'definitely wrong' gained 4 points, 'probably wrong,' 3 points, 'don't know,' 2 points, 'probably correct,' 1 point, and 'definitely correct' 0 points. All the responses from the thirty-nine participants were collected successfully. Each participant's total sum of scores was used as data for statistical analysis.

4. Results

This section reports the results from the two experimental tasks. It begins with a frequency analysis of figure and background information produced by Korean and English speakers, which is followed by an analysis of how accurately they recall that information.

4.1. The Analysis of Utterances

Table 2 provides an overview of the mean frequencies of subject and background elements produced by Korean and English speakers. Since there were eighteen experimental items in total, if a speaker produced

1) As a matter of fact, both groups of speakers similarly produced a lot of nonfactual (i.e., inferential) details. One possible reason might be that the instruction did not expressly demand them to speak only what could be referred to in a given picture. We leave this methodological issue in the picture description paradigm for further investigation in the future.

one subject and one background in each sentence, then the total number of subjects and backgrounds produced would be eighteen.

Table 2. Descriptive Statistics of Participants' Utterance Data

	Group	*N*	*M*	*SD*
S[Ø]	Korean	19	.632	.831
	English	18	1.722	1.708
	Total	37	1.162	1.424
S[age]	Korean	19	.368	.831
	English	18	.500	1.200
	Total	37	.432	1.015
S[gender]	Korean	19	9.526	3.565
	English	18	11.278	4.336
	Total	37	10.378	4.002
S[age][gender]	Korean	19	6.053	3.027
	English	18	7.167	2.813
	Total	37	6.595	2.939
S-modifiers	Korean	19	2.211	2.679
	English	18	4.167	3.111
	Total	37	3.163	3.023
Background	Korean	19	7.368	3.148
	English	18	13.500	3.730
	Total	37	10.351	4.602

The mean frequencies of S[Ø] (e.g., a person, someone, etc.) were 0.632 and 1.722 for KSs and ESs, respectively, and those of S[age] (e.g., a child, an elderly person, etc.) were less than one in both groups. These suggest that the two categories are rarely used in ordinary conversation and cannot be subjected to further analysis. Both groups used S[gender] most frequently, followed by S[age][gender] and S-modifiers. It was observed that ESs produced more tokens of each type than KSs did, indicating that the former provided more detailed figure information than the latter. This is partially compatible with the findings of Chua et al. (2005) and Masuda and Nisbett (2001) in which Americans focused on focal objects longer than East Asians. Nonetheless, the largest between-group difference

was found in the production of background information: Korean = 7.368, English = 13.500. ESs produced a lot more background details than Korean speakers.

In order to determine if these observed differences are statistically significant, a multivariate analysis of variance was conducted. The results indicate that there is a significant difference between the two groups on the combined dependent variables: Wilks' Lambda = .437, F (6, 30) = 6.452, p = .000, partial eta squared = .563. The post-hoc one-way group comparisons with respect to each dependent variable are presented in Table 3 with the alpha level adjusted to .012 for multiple comparisons.

Table 3. One-way Tests of Between-group Differences in Utterance Tokens

Source	Variables	*SS*	*df*	*F*	*p*
Group	S[gender]	28.355	1	1.810	.187
	S[age][gender]	11.472	1	1.341	.255
	S-modifier	35.369	1	4.216	.048
	Background	347.511	1	29.314	.000

A statistically significant difference is found only in the background category, but not in the other categories. Both groups behaved alike when it came to subject details. However, ESs included significantly more background details in their sentences than KSs did, and as its effect size (η^2 = .456) suggests, the between-group difference found in background production accounts for a large portion of the total variance. This result is at odds with Tajima and Duffield (2012) where Japanese speakers mentioned more peripheral (background) items than ESs in a picture description task. We will return to this discrepancy shortly.

We now turn to look at how each group performed at two different presentation durations, 1.5 and 3 seconds of scene display. If the types and tokens of information produced in the two conditions differ markedly, it is assumed to reflect the speaker's predisposed perceptual priority and movement. As aforementioned, S[Ø] and S[age] were excluded from further consideration because the observed cases were too few to be treated in greater detail. Table 4 shows KSs' utterance frequencies together with

their *z* values. Since the data did not meet the assumptions of parametric tests, the Wilcoxon Signed Rank Test was conducted.

Table 4. Wilcoxon Signed Ranks Test for Korean Speakers' Utterances by Time

Information	*N*	1.5 sec		3 sec		*z*	*p*
		M	*SD*	*M*	*SD*		
S[gender]	19	5.474	1.925	4.053	2.041	-2.854	.004
S[age][gender]	19	1.789	1.512	4.263	1.881	-3.644	.000
S-modifiers	19	.368	.831	1.842	2.363	-2.407	.016
Background	19	4.263	2.257	3.105	1.560	-1.904	.057

It turned out that KSs were significantly affected by time in their description of subjects but not of backgrounds. They produced considerably more S[gender] at short durations and more S[age][gender] and S-modifiers at longer durations, but their production of background elements in the two conditions did not differ significantly. These findings suggest that KSs first attended to focal figures for a time sufficient to identify their gender and then moved toward backgrounds. When given more time, they moved back to figures and added more details about them.

As shown in Table 5 below, ESs' referential expressions for subject referents were not significantly influenced by the length of time for which they were presented with the scene.

Table 5. Wilcoxon Signed Ranks Test for English Speakers' Utterance Tokens by Time

Information	*N*	1.5 sec		3 sec		*z*	*p*
		M	*SD*	*M*	*SD*		
S[gender]	18	5.444	2.770	5.833	1.886	-.735	.462
S[age][gender]	18	3.444	1.688	3.722	1.320	-.965	.334
S-modifiers	18	1.667	1.495	2.500	1.948	-2.101	.036
Background	18	7.500	2.065	6.000	2.223	-2.370	.018

This implies that they attended to subject referents for a stable time and

chose either S[gender] or S[age][gender] depending not on time but on some other factors. It was also observed that ESs produced significantly more background information in the 1.5-second condition and more S-modifiers in the 3-second condition. The increased number of S-modifiers tells us that when given more time, both Korean and English speakers turned back toward focal figures and described more details about them. In contrast, the significant decrease in the number of backgrounds indicates that ESs were more likely to discard the previously processed background information as they focused back on figures.

4.2. The Recall Test

The mean scores of the two groups on the recall test were shown in Table 6. Since each information type contained six items which were divided in half for either time condition and scored on the Likert scale of 0 to 4, the maximum value for each cell is 3×4 = 12. One English speaker did not want to participate in the recall task, so the total number of English participants was 18.

Table 6. Descriptive Statistics of the Recall Test

	Group	*N*	1.5 sec		3 sec	
			M	*SD*	*M*	*SD*
Gender	Korean	20	8.650	1.954	8.900	1.889
	English	18	8.500	2.149	8.889	2.193
	Total	38	8.579	2.022	8.895	2.011
Age	Korean	20	9.100	1.714	9.650	1.461
	English	18	7.722	2.347	8.056	1.893
	Total	38	8.447	2.127	8.895	1.842
Background	Korean	20	8.600	2.037	8.050	2.481
	English	18	5.944	3.058	5.222	2.487
	Total	38	7.342	2.869	6.711	2.837

The two groups did not show any substantial difference in recalling the gender of figures. Regarding age and background, however, KSs outperformed ESs to a considerable degree. In particular, KSs remembered

background elements more accurately than ESs. The effect of time appears not to be salient. Both groups exhibited a consistent pattern across all the three variables. There was a slight tendency for them to better recall age and gender in the long-time condition. On the contrary, they recalled backgrounds better when given a shorter time. A mixed between-within subjects ANOVA was conducted to assess the effect of group, time, and their interaction on participants' scores on the recall test. The results are summarized in Table 7 below.

Table 7. Multivariate Tests of Between- and Within-Group Differences

Effects		Wilks' λ	F (3, 34)	p	η_p^2
Between					
	Intercept	.025	442.962	.000*	.975
	Group	.684	5.230	.004*	.316
Within					
	Time	.629	6.696	.001*	.371
	Group × Time	.988	.135	.938	.012

There was no significant interaction between group and time, as shown by Wilks' Lambda = .988, F (3, 34) = .135, p = .938. This means that the two groups' scores were not differentially impacted by time. The main effect of group on the aggregated dependent variables and that of time were both significant, suggesting that overall KSs performed significantly better than ESs on the recall test and that both groups behaved differently under the two time conditions. In order to look at where exactly the differences lie, post-hoc comparisons were conducted as follows.

Table 8. One-Way Tests of Group and Time Effects

Source	Variables	SS	df	F	p	η^2
Group	Gender	.123	1	.017	.897	.000
	Age	41.846	1	6.742	.014	.158
	Background	142.422	1	12.221	.001	.253
Time	Gender	1.933	1	1.740	.196	.046
	Age	3.696	1	4.843	.034	.119
	Background	7.667	1	6.852	.013	.160

The largest between-group difference was found in background information, which accounts for about 25% of the total variance. The effect of time on participants' recall scores was not significant when measured with respect to each individual variable.

5. Discussion

The finding that KSs were able to remember background information significantly more accurately than ESs gives rise to the question of why ESs who produced more background elements in their utterance failed to recall them in the subsequent memory test. One possible explanation is hinted at by the two groups' non-uniform behavior under the short and long time conditions. We saw that KSs tended to move away from figures and toward backgrounds as soon as they identified their gender. When given a longer time, they produced more age-specified subjects and other modifiers with no significant decrease in background information. This implies that KSs focus on backgrounds sooner than ESs and their attentional allocation to subject referents are distributed (with the backgrounds of the scenes placed in between) to a greater extent than shown among ESs. It seems that this relatively more distributed perceptual pattern allows KSs to retain background information longer and recall it successfully later.

ESs' perceptual patterns were quite different from KSs'. The manipulation of time had little effect on ESs' production of S[gender] or S[age][gender]. This suggests that their use of either referential expression has more to do with intervening variables such as visual saliency and informational significance than with the time required to observe the background context. As being less affected by backgrounds, their attention to subject referents seems to be of concentration rather than of distribution, at least for a time until they can detect the inherent subject features to be linguistically encoded within a referential expression. It is also worth noting that overall ESs mentioned more subject details than KSs, which lends support to the claim that the former tend to pay more attention

to focal figures than the latter. Even if it might be the case that ESs moved toward backgrounds after opting for either S[gender] or S[age][gender] unlike KSs whose attention was more quickly oriented toward backgrounds, we have to account for the fact that ESs managed to include background elements within one-and-a-half seconds, and even more so than KSs.

There is reason to suppose that this might be due to the interplay between the recency effect – people find it easier to recall the most recently processed items than the previous ones – and, as it were, the maxim of quantity – people are predisposed to cooperate in conversation by providing a normative amount of information. Recall that ESs produced significantly less background information when they watched the scenes for a longer time. They instead focused back on figures and added more details about them. A similar backward zoom-in pattern was observed for KSs, but a striking difference between the two was found in the retention of background information; that is, KSs did not show a significant loss of backgrounds at longer presentation durations whereas ESs did. Drawing on these, it can be inferred that ESs produced more background elements at short presentation durations because they were processed most recently in their working memory, and hence the most readily available resources from which they could construct their sentences in a more informative way, as a kind of compensation strategy for the lack of figure information.

The recency effect was significantly lessened when ESs were allowed to watch the scenes longer and so had a chance to focus back on figures. This indicates that they prefer to describe subjects in more detail at the expense of backgrounds. However, KSs were less likely to lose background information even when they added more subject details at longer durations. Provided that the two groups paid an equivalent amount and length of attention to the scenes as a whole, it can be said that ESs attended to figures relatively longer than KSs and, on the contrary, KSs attended to backgrounds relatively longer than ESs. All in all, this asymmetry between utterance tokens and recall accuracy at two different presentation durations underscores that the effect of a language on the language speakers' cognitive processes is far more complex than was discussed in earlier

studies.

Finally, one may argue that the between-group differences found in the recall task might have arisen because of cultural differences in expressing one's attitude towards certainty. This may take the form of saying, for example, KSs have a strong tendency to give definite answers (i.e., 'definitely right' or 'definitely wrong') whereas ESs prefer to choose answers with less certainty (i.e., 'probably right' or 'probably wrong') even though they are actually certain that their answers are correct. As seen in the results, this line of argument would be invalid on the grounds that the mean scores of the two groups were significantly different only in background information, but not in the other categories. This also means that the use of scaled options in the assessment of one's recall accuracy can be potentially more accurate than simple judgment tasks.

6. Conclusion

Although this study has successfully provided evidence for the claim that there are considerable differences in ways Korean and English speakers describe and recall visual scenes, particularly in terms of the production and retention of figure and background information, many issues remain to be addressed. Among them are whether the current findings will be consistent with larger and more varied samples of the populations, to what extent and in what domains crosslinguistic differences in perceptual processes may affect L2 development, and ultimately how such findings can be effectively incorporated into L2 teaching and learning practices. Moreover, it is still speculative that the observed differences are indeed attributable to the linguistic factors presumed here. It was argued, for instance, that ESs would tend to focus on the gender of a subject referent because the English grammar predisposes its speakers to encode it in a referential expression and employ a gender-specified anaphoric pronoun in the following discourse, and hence they would have an advantage over null-subject language speakers in recalling the subject feature later. Unfortunately, this does not point to the conclusion that cognitive differ-

ences across language speakers, if any, are necessarily caused by the grammar in question. Empirical evidence is not sufficient to claim that the phenomenon is associated not so much with anything else as with the particular linguistic feature. Therefore, it is more appropriate to see it suggestive rather than confirmatory.

Despite these limitations, the findings of the present study lend strong support to the interrelationship between language and perception. The role of unconscious processes has been of intense debate in the field of L2 acquisition, but it has gained wide support out of intuition rather than on the basis of empirical evidence. Granted that each language is a selective fabric and mode of human experience, L2 learners will be able to use the target language only when they view the world from that particular perspective and construct messages as such. In other words, learning an additional language for communicative purposes cannot but involve the process of attaining an alternative standpoint of the world. Then, what kinds of knowing should L2 teachers and materials cater to learners so as to facilitate them to actually put the learned knowledge into practice? What the study of linguistic relativity suggests is that learners can benefit from engaging in activities that enable them to internalize another way of cognitive and perceptual dealings with the world. Lastly, there has been virtually no attempt to track down the time course of linguistic relativity effects (cf. Klemfuss, Prinzmetal, & Ivry, 2012). If language influences a particular cognitive domain, does it do so upon acquisition of a new language, or how can L2 sentence structure gradually be proceduralized as something more than a learned code? It is this issue that L2 studies in the framework of the linguistic relativity hypothesis have to face in the immediate future.

References

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition* 9, 89-96.

Barner, D., Inagaki, S., and Li, P. (2009). Language, thought, and real nouns.

Cognition 111, 329-344.

Boduroglu, A., Shah, P., and Nisbett, R. E. (2009). Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology* 40.3, 349-360.

Bowerman, M. (1996). Learning how to structure space for languages: A cross-linguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 385-436). Cambridge, MA: MIT Press.

Cadierno, T (2010). Motion in Danish as a second language: Does the learner's L1 make a difference. In Z. H. Han & T. Cadierno (Eds.), *Linguistic relativity in second language acquisition: Thinking for speaking* (pp. 1-33). Tonawanda, NY: Multilingual Matters.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning* 43, 1-22.

Choe, M-H. (2012). Does linguistic variation in null-subject licensing lead to the speaker's perceptual difference? A pilot study. *Studies in English Education* 17.1, 1-26.

Choi, S. and Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* 41, 83-121.

Chua, H. F., Boland, J. E., and Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences* 102, 12629-12633.

Coventry, K. R., Valdés, B., and Guijarro-Fuentes, P. (2010). Thinking for speaking and immediate memory for spatial relations. In Z. H. Han & T. Cadierno (Eds.), *Linguistic relativity in second language acquisition: Thinking for speaking* (pp. 84-101). Tonawanda, NY: Multilingual Matters.

Ellis, N. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372-405). NY: Routledge.

Ferreira, F. and 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* 46.1, 57-84.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Ford, C. E. and Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar*

(pp. 134-184). NY: Cambridge University Press.

Gennari, S. P., Sloman, S., Malt, B., and Fitch, T. (2002). Motion events in language and cognition. *Cognition* 83, 49-79.

Han, Z. H. and Cadierno, T. (2010). *Linguistic relativity in second language acquisition: Thinking for speaking*. Tonawanda, NY: Multilingual Matters.

Hauser, M. D., Chomsky, N., and Fitch W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569-1579.

Huang, C.-T. J. (1989). Pro-drop in Chinese: A generalized Control theory. In O. Jaeggli & K. J. Safir (Eds.), The null subject parameter (pp. 185-214). Dordrecht: Kluwer.

Hunt, E. and Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review* 98, 377-389.

Imai, M. and Gentner, D. (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition* 62, 169-200.

Jaeggli, O. and Safir, K. J. (1989). The null subject parameter and parametric theory. In O. Jaeggli & K. J. Safir (Eds.), *The null subject parameter* (pp. 1-44). Dordrecht: Kluwer.

Kay, P. and Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist* 86.1, 65-79.

Kay, P. and Regier, T. (2006). Language, thought and color: Recent developments. *Trends in Cognitive Science* 10, 51-54.

Kempen, G. and Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science* 11, 201-258.

Klemfuss, N., Prinzmetal, W., and Ivry, R. B. (2012). How does language change perception: a cautionary note. *Frontiers in psychology* 3.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. Oederson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109-169). Cambridge, MA: MIT Press.

Levinson, S. C. (1997). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology* 7, 98-131.

Levinson, S. C. (2003). *Space in language and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Li, C. N. and Thompson, S. A. (1976). Subject and topic: A new typology of languages. In C. N. Li (Ed.), *Subject and topic* (pp. 457-490). NY: Academic Press.

Li, P., Dunham, Y., and Carey, S. (2009). Of substance: The nature of language effects on construal. *Cognitive Psychology* 58, 487-524.

Li, P. and Gleitman, L. (2002). Turning the tables: Spatial language and spatial cognition. *Cognition* 83, 265-294.

Lucy, J. and Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257-283). Cambridge: Cambridge University Press.

Lucy, J. A. (1996). The scope of linguistic relativity: An analysis and review of empirical research. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 37-69). Cambridge: Cambridge University Press.

Masuda, T. and Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology* 81.5, 922-934.

Masuda, T. and Nisbett, R. E. (2006). Culture and change blindness. *Cognitive Science* 30, 1-19.

Mazuka, R. and Friedman, R. (2000). Linguistic relativity in Japanese and English: Is language the primary determinant in object classification? *Journal of East Asian Linguistics* 9, 325-351.

McDonough, L., Choi, S., and Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology* 46, 229-259.

Munnich, E., Landau, B., and Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition* 81, 171-207.

Odlin, T. (2008). Conceptual transfer and meaning expression. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 306-340). NY: Routledge.

Oh, K-j. (2003). *Language, cognition, and development: Motion events in English and Korean*. Unpublished doctoral dissertation, University of California, Berkeley.

Pae, H. K. (2012). Linguistic relativity revisited: The interaction between L1 and L2 in thinking, learning, and production. *Psychology* 3.1, 49-56.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., and Senft, G. (1998). Semantic typology and spatial conceptualization. *Language* 74, 557-589.

Pinker, S. (1994). *The language instinct*. New York, NY: William Morrow and Company.

Regier, T. and Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences* 13, 439-446.

Roberson, D., Hanley, J., and Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition* 112, 482-487.

Roberson, D., Pak, H., and Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition* 107.2, 752-762.

Rutherford, W. E. (1983). Language typology and language transfer. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 358-470). Rowley, MA: Newbury House.

Slobin, D. I. (1996). From thought and language to thinking for speaking. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Measow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157-192). Cambridge, MA: MIT Press.

Stam, G. A. (2010) Can an L2 Speaker's patterns of thinking for speaking change? In Z. H. Han & T. Cadierno (Eds.), *Linguistic relativity in second language acquisition: Thinking for speaking* (pp. 59-83). Tonawanda, NY: Multilingual Matters.

Tajima, Y. and Duffield, N. (2012). Linguistic versus cultural relativity: On Japanese-Chinese differences in picture description and recall. *Cognitive Linguistics* 23.4, 675-709.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon* (pp. 36-149). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics: Typology and process in concept structuring*. Cambridge, MA: The MIT Press.

Vigliocco, G., Vinson, D., Lewis, W., and Garrett, M. F. (2004). Representing the meaning of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology* 48, 422-488.

Vigliocco, G., Vinson, D., Paganelli, F., and Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology* 134, 501-520.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences* 104.19, 7780-7785.

Yoshida, H. and Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science* 16, 90-95.

Mun-Hong Choe
Department of English Education
Chonnam National University
77 Yongbong-ro, Buk-gu, Gwangju, 500-757, Korea
E-mail: munhong@jnu.ac.kr

Received: March 5, 2015
Revised version received: April 2, 2015
Accepted: April 7, 2015