The role of non-syntactic factors in the production and processing of English relative clauses by Korean EFL learners: A L2 corpus-based approach[*]

Keun Young Shin

(Chonnam National University)

Shin, Keun Young. 2024. The role of non-syntactic factors in the production and processing of English relative clauses by Korean EFL learners: A L2 corpus-based approach. *Linguistic Research* 41(Special Edition): 25-50. There have been a growing number of corpus studies exploring the role of linguistic experience or input in sentence processing. Previous studies have used frequencies in first language (L1) corpora to represent the input of second language (L2) learners, but it has been claimed that L2 output as well as L1 output contribute to L2 acquisition and processing (Ellis 2002; Ellis and Collins 2009; Ellis and Wulff 2020). By analyzing an L2 corpus produced by Korean adult learners of English, this paper aims to establish more accurate input for L2 learners and investigate why Korean adult learners of English are sensitive to noun animacy and topicality in processing English relative clauses, as native speakers of English are. The results of the L2 corpus analysis reveal that object and subject relative clauses display different distributional patterns with respect to head noun animacy, embedded noun animacy, and the topicality of the embedded NP due to their strong tendency to have unmarked subjects and objects in terms of animacy and topicality. This paper proposes that the experience of subjecthood and objecthood, which are crosslinguistically associated with noun animacy and topicality, plays a role in producing and processing relative clauses in L2 which is syntactically different from the native language of L2 speakers. In addition, I discuss that some processing differences between L1 and L2 can arise due to L2 learners' experience with their native language. **(Chonnam National University)**

Keywords relative clause, L2 learner corpus, production, processing, subjecthood

1. Introduction

There have been a growing number of corpus studies regarding relative clauses (RCs) seeking to identify the kinds of information that influence the incremental processing of RCs. The general idea behind this research is that input frequencies play a role in language processing. The frequency-based or experience-based approach has been supported by previous corpus studies: object-extracted relative clauses (ORCs) and subject-extracted relative clauses (SRCs) show very different distributions with respect to the factors claimed to modulate the processing difficulties of ORCs compared to SRCs,[1] that is, the animacy of the head noun (Roland et al. 2007; Gennari and MacDonald 2008, 2009) and the discourse form or topicality of the embedded argument (Reali and Christiansen 2007; Roland et al. 2012).

Some experimental studies (Baek 2012; Kim 2016; Choe and Deen 2020) have shown that Korean adult learners of English as a foreign language (EFL) process English ORCs more easily when they have inanimate head nouns or embedded pronouns, as native speakers of English do. One may argue that the processing similarities between the first language (L1) and the second language (L2) stems from exposure to the same L1 input – the language produced by native speakers of English (Kim 2016). However, as Ortega (2016) and Monteiro et al. (2020) point out, it is problematic to use L1 corpora to represent L2 input when one considers the circumstances under which L2 learners are exposed to the target language. L2 learners generally receive input from other L2 learners in their home countries. Indeed, recent L2 acquisition models have suggested that not only L1 output frequencies but also L2 output frequencies contribute to L2 learning and processing (Ellis 2002; Ellis and Collins 2009; Ellis and Wulff 2020). To date, however, few or no L2 corpus studies have been conducted to compare the distributional patterns of ORCs and SRCs with respect to non-syntactic factors.[2] By analyzing an L2 English corpus produced by Korean EFL learners, this paper investigates whether an experience-based approach, which is in line with usage-based approaches in L2 acquisition studies, can explain why Korean EFL learners are sensitive to noun animacy and topicality in processing

---

1   In this paper, SRCs, which are compared with ORCs, refer to SRCs with a transitive verb taking an NP argument.

2   Most L2 corpus studies (e.g., Sung 2014; Lee and Shin 2017; Kang and Oh 2022) have focused on the question of whether Korean EFL learner's RC acquisition can be explained by Keenan and Comrie's (1977) noun phrase accessibility hierarchy hypothesis.

English RCs.

Language comprehension is a complex process in which multiple factors of syntax, semantics, and pragmatics interact with one another (Hsiao and MacDonald 2013; Kim 2016). It has been claimed that the effects of head noun animacy and embedded pronouns on RC processing vary depending on embedded noun animacy (Baek 2012; Roland et al. 2012; Hsiao and MacDonald 2013). One recent study of an L1 English corpus (Shin 2020) argued that ORCs and SRCs showed different distributional patterns with respect to embedded noun animacy as well as head noun animacy and the topicality of the embedded NP. In other words, both ORCs and SRCs tend to have prototypical subjects and objects in terms of noun animacy and topicality, despite their different word orders. The results derived from this corpus study put a new perspective on RC processing by suggesting that the integrated effects of noun animacy and topicality should be considered in accounting for RC processing.

Given that noun animacy and topicality are crosslinguistically associated with subjects and objects (Silverstein 1976; Dixon 1979, 1994; Comrie 1989; Aissen 1999, 2003), this corpus study is conducted to explore whether the non-syntactic features of subjecthood and objecthood can also explain the distributional patterns of ORCs and SRCs produced by L2 learners whose native language differs syntactically from the target language in terms of RC structures. I examine whether the processing similarities between Korean EFL learners and native English speakers are attributable to the crosslinguistic experience of subjecthood and objecthood. In addition, I discuss the role of L1-specific experience in processing RCs in L2, which can provide an account of some slight processing differences between L1 and L2.

2. Previous corpus and processing studies of RCs in L1 and L2

2.1 L1 corpus and processing studies

There has been a longstanding assumption that ORCs are more difficult to process than SRCs due to their syntactic differences. Compare (1a) and (1b). In ORC (1a), both the head noun and the embedded noun occur before the RC verb. Considering an online incremental sentence processing mechanism, the embedded noun as well as the head noun in (1a) are placed in working memory until the parser encounters

the RC verb that determines the grammatical and thematic roles of the nouns (Gibson 1998, 2000; Warren and Gibson 2002; Grodner and Gibson 2005). The embedded noun does not cause a memory burden during the processing of the SRC in (1b), as it follows the verb immediately.

(1) a. [The reporter that the senator attacked] admitted the error.
    b. [The reporter that attacked the senator] admitted the error.

However, this syntax-based account has been challenged. No processing differences between ORCs and SRCs have been found when head nouns are inanimate (Traxler et al. 2002, 2005; Baek 2012). It has also been reported that ORCs are even read faster than SRCs when embedded noun phrases (NPs) are pronouns or topical NPs referring to discourse-old referents (Reali and Christiansen 2007; Roland et al. 2012). Such non-syntactic effects on RC processing have been claimed to result from language users' linguistic experience or input frequencies (Reali and Christiansen 2007; Roland et al. 2007, 2012; Gennari and MacDonald 2008, 2009). That is, ORCs with inanimate head nouns or embedded pronouns are processed more easily because comprehenders encounter such ORCs frequently in the input. Several L1 corpus studies have supported this experience-based account by demonstrating that ORCs occur more frequently than SRCs when head nouns are inanimate (Roland et al. 2007; Gennari and MacDonald 2008, 2009) or when embedded NPs are personal pronouns (Reali and Christiansen 2007; Roland et al. 2012).

2.2 L2 English relative clause processing by L1 Korean speakers

Just like native speakers of English, Korean adult EFL learners are sensitive to semantic and discourse information when processing English RCs (Baek 2012; Kim 2016; Choe and Deen 2020). Korean EFL learners process SRCs more readily than ORCs when RCs modify animate head nouns, as in (1), but not when they have inanimate head nouns (Baek 2012). Korean EFL learners process ORCs with pronouns more easily than those with descriptive NPs headed by common nouns (Kim 2016; Choe and Deen 2020).

Kim (2016) claimed that the pronominal effect on ORC processing in L2 English

cannot be due to the transfer of L1 Korean. In Korean, which is a head-final language unlike English, an RC verb follows an embedded argument NP in both ORCs and SRCs, as exemplified in (2). The case-markers attached to the embedded arguments overtly indicate their grammatical roles.

(2) a. haksayng-i   sa-n   chayk-i   salaci-ess-ta.
       student-nom   buy-rel   book-nom   disappear-pst-decl
       'The book that the student bought disappeared.'
    b. chayk-ul   sa-n   haksayng-i   salaci-ess-ta.
       book-acc   buy-rel   student-nom   disappear-pst-decl
       'The student that bought the book disappeared.'

Kim (2016) suggested that the processing similarities between L1 and L2 result from L2 learners being exposed to the target language produced by native English speakers. When Korean students learn English in their home country, however, they are frequently exposed to English produced by other L2 learners such as their English teachers and classmates. Indeed, it is claimed that L2 output as well as L1 output contribute to L2 acquisition and processing (Ellis 2002; Ellis and Collins 2009; Ellis and Wulff 2020). According to Ortega (2016) and Monteiro et al. (2020), both L1 and L2 corpora should be considered when establishing L2 input frequencies. Nevertheless, no L2 corpus studies have examined the distributional patterns of ORCs and SRCs in terms of noun animacy and topicality, which are claimed to affect RC processing in L2.

Furthermore, according to the production-distribution-comprehension (PDC) model (MacDonald 1999, 2013; Gennari and MacDonald 2009), sentence comprehension processes are similar to sentence production processes, which in turn affect the distribution of linguistic forms in corpora. Language users utter forms that are biased in their attempts to mitigate the cognitively demanding difficulty of production. Mechanisms involved in a production process influence the distributions of utterance forms in language, which in turn impact the incremental process of comprehension. Gennari and MacDonald (2009) examined ORCs and passive SRCs with verbal predicates with the argument structure <Theme, Experiencer> (e.g., *please*) in the *British National Corpus* (BNC). They found that ORCs rarely occurred when the Theme was inanimate and the Experiencer was animate, but ORCs occurred as

frequently as passive SRCs when both arguments were animate. Their experimental studies further showed that the distributional patterns of ORCs and passive SRCs mirrored the production processes of RCs and matched their comprehension processes. Considering such a close relationship among distribution, production, and comprehension, a corpus study of L2 English can therefore provide insight into L2 processing by revealing what kinds of information play a role in the process of L2 production.

2.3 Subjecthood effects: Non-linear interactions of multiple factors

One limitation of most prior corpus studies is that they have focused on the effect of one single factor on RC distribution. Little consideration has been given to possible interactions among multiple factors. Nevertheless, it has been claimed that embedded noun animacy modulates the effect of head noun animacy (Traxler et al. 2002, 2005; Baek 2012) or that of topicality (Reali and Christiansen 2007; Roland et al. 2012) on RC processing. Shin (2020) argued that these interactions are not accidental by showing that ORCs and SRCs in L1 English corpora exhibit very different distributional patterns with respect to head noun animacy, embedded noun animacy, and the form of the embedded NP that reflects its topicality or givenness (Gundel et al. 1993; Mulkern 1996).[3] She contended that the integrated effects of the three non-syntactic factors can be captured by the non-syntactic notion of subjecthood or objecthood, which expresses the semantic and pragmatic (un)markedness of the grammatical relations (Shin 2020).

Numerous theoretical and empirical studies in functional and typological linguistics (Silverstein 1976; Dixon 1979, 1994; Comrie 1989) have claimed that subjects and objects are crosslinguistically associated with non-syntactic features such as animacy, humanness, definiteness, and specificity. According to Comrie (1989), Bossong (1991), and Aissen (1999, 2003), these features can be analyzed into two dimensions of prominence, that is, animacy and topicality, as illustrated in (3).

---

3   This study was inspired by Mak et al. (2002, 2006), who insightfully recognized embedded noun animacy as another key factor that explains the processing and distributional differences between ORCs and SRCs in Dutch.

(3) a. Animacy Scale: Human > Animate > Inanimate
   b. Topicality Scale: Pronoun > Proper noun > Definite NP > Indefinite NP

Prototypical, unmarked subjects are high in prominence, whereas typical objects are low in prominence. In other words, NPs high on the animacy and topicality hierarchies are unmarked for subjects but marked for objects, whereas NPs low on these hierarchies are marked for subjects but unmarked for objects. In Shin's (2020) L1 corpus data, ORCs and SRCs had a strong tendency to contain prototypical subjects (i.e., animate pronouns/proper nouns) and prototypical objects (i.e., inanimate definite/indefinite NPs headed by common nouns), respectively. ORCs and SRCs showed very different distributions when head nouns and embedded nouns differed in terms of animacy because both types of RCs tend to have animate subjects and inanimate objects. Thus, the distributional differences between ORCs and SRCs found in Shin's L1 data are due to their tendency to have prototypical semantic and discourse structures.

The subjecthood account of the RC distribution added a new perspective to the prior literature on RC processing. That is, head noun animacy, embedded noun animacy, and topicality can interact in the incremental processing of RCs (Shin 2019, 2020). As discussed in detail in Section 5, the subjecthood account can provide a better understanding of the well-known processing difficulties of RCs found in previous experimental studies.

Therefore, this paper examines the distributional patterns of ORCs and SRCs with respect to the non-syntactic features of subjecthood in an L2 corpus of Korean EFL learners. The aim is to investigate whether Korean EFL learners are sensitive to noun animacy and topicality when processing English RCs due to their experience of subjecthood in the input. Therefore, the main research questions are as follows.

1) How do the distributional patterns of ORCs and SRCs differ in terms of head noun animacy, topicality, and embedded noun animacy in a corpus of Korean learners of English? In other words, do RCs frequently have unmarked semantic and discourse structures in an L2 English corpus as they do in L1 English corpora?

2) Do the distributional patterns of RCs mirror their processing difficulties in L2? Can the effects of noun animacy and topicality on L2 RC processing

be explained by L2 learners' linguistic experience?

3. Data extraction and coding

A corpus study was conducted using the *Yonsei English Learner Corpus* (YELC), which is available from Yonsei University's English Informatics Laboratory upon request. The YELC is a corpus of 6,572 essays written by pre-college students admitted to the university in 2011. Each participant composed one argumentative essay and one narrative essay, but two essays from one student were excluded because they were identical to those of another student. Thus, 6,570 texts (1,085,465 words) produced by 3,285 participants were used for this study.

Since the YELC is not a part-of-speech (POS) tagged corpus, I used the concordancing program *AntConc* (Anthony 2020) to extract all sentences containing clauses introduced by *that, which*, or *who*. ORCs and SRCs were manually extracted by removing non-relative clauses that had the form [noun + *that/which/who*], for example, noun complement clauses (e.g., *the fact that he passed the exam*) and cleft constructions (e.g., *It was the book that I lost yesterday*). SRCs were restricted to those containing transitive verbs and object NPs because ORCs have been compared with such transitive SRCs in recent corpus studies (Gordon and Hendrick 2005; Roland et al. 2012; Shin 2019, 2020) as well as most experimental studies on RC processing.

L2 learners commonly make grammatical, lexical, and orthographic errors. Errors irrelevant to RC analysis were ignored: e.g., tense and subject-verb agreement errors. The YELC includes metadata on students' proficiency levels based on the Common European Framework of Reference for Languages. Only four RCs were produced by the lowest and highest proficiency level groups (i.e., A1 and C2), and they were filtered out from the data. As a result, the dataset consisted of a total of 3,045 RCs produced by 1,751 participants at seven proficiency levels (A1+, A2, B1, B1+, B2, B2+, and C1).

751 ORCs and 2,294 SRCs were used in the dataset. Each RC was hand-coded for RC type, head noun animacy, embedded noun animacy, and the discourse form or topicality of the embedded NP, as shown in Table 1.

Table 1. Coded examples

Text type	RC type	Head noun	Embedded noun	Topicality	Example sentences
Argumentative	SRC	A	IA	Non-topical	Furthermore, there are *some people who cannot complete the military service.*
Narrative	ORC	IA	A	Topical	*The most interesting thing that I found during the trip* was that …

If nouns contextually denoted living animals or humans, they were coded as animate nouns (= *A*); otherwise, they were coded as inanimate nouns (= *IA*), as illustrated in Table 1. Most animate nouns referred to humans. Following the givenness or topicality hierarchy (Gundel et al. 1993; Mulkern 1996) and the definiteness hierarchy (Aissen 2003), pronouns and proper nouns are more given or topical than definite NPs headed by common nouns and indefinite NPs. For convenience, therefore, demonstrative and personal pronouns and proper nouns were classified as *Topical,* whereas NPs headed by common nouns and indefinite pronouns (e.g., *someone*) were classified as *Non-topical*. This classification is based on the finding that ORCs and SRCs in L1 English corpora exhibit different distributional patterns with respect to these two categories of topicality, as discussed in Section 2.3 (Reali and Christiansen 2007; Shin 2019).[4]

4. Results and discussion

4.1 Effects of subjecthood on RC distributions

RC type was significantly associated with embedded noun animacy ($\chi^2(1) = 1494.4$, $p < .001$, Cramer's V = 0.70) and topicality ($\chi^2(1) = 1429.7$, $p < .001$, Cramer's V = 0.68). Furthermore, as illustrated in Figure 1, embedded noun animacy and topicality had strong pairwise associations with RC type in the L2 data ($\chi^2(3) = 1827.8$, $p < .001$, Cramer's V = 0.77). 86.04% of ORCs contained animate topical NPs (= *A* (*Topical*)), whereas 82.59% of SRCs contained inanimate non-topical NPs

---

4   All statistical analyses were performed using the open-source statistical software (version 4.3.2, R Core Team 2023).

(= *IA* (*Non-topical*)). Embedded noun animacy and topicality are properties of an embedded NP, and, together, they can either increase or decrease the degree of subjecthood and objecthood. ORCs contained NPs that usually showed a high degree of subjecthood (i.e., animate and topical NPs), contrary to SRCs, which predominately contained prototypical objects (i.e., inanimate and non-topical NPs). Since embedded NPs are subjects in ORCs but objects in SRCs, both ORCs and SRCs strongly tend to have unmarked subjects and objects in terms of animacy and topicality.
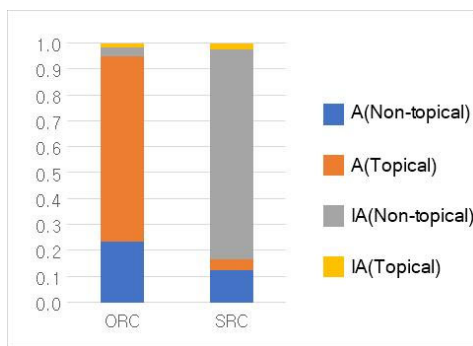

Figure 1. Distribution of embedded NPs (animacy (topicality)) in ORCs and SRCs
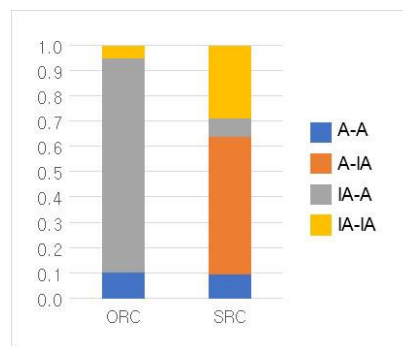
Figure 2. Distribution of animacy configurations (head noun-embedded noun) in ORCs and SRCs

There was a significant association between RC type and head noun animacy ($\chi^2(1)$ = 653.74, p <0.001, Cramer's V= 0.46). As mentioned above, ORCs predominately occurred with inanimate head nouns, whereas SRCs with animate head nouns were slightly more frequent than SRCs with inanimate head nouns. Unlike ORCs, SRCs are not sensitive to the animacy of head nouns. The effect of head noun animacy on RC processing has been claimed to be due to the whole animacy configuration of a head noun and an embedded noun (Traxler et al. 2002, 2005; Gennari and MacDonald 2008, 2009; Baek 2012). Indeed, RC type was very strongly associated with animacy configuration ($\chi^2(3)$ = 1835.5, p < .001, Cramer's V = 0.77). When head nouns and embedded nouns differed in terms of animacy, robust differences were observed. As illustrated in Figure 2, 54.27% of the SRCs had the animate head noun – inanimate embedded noun (= *A-IA*) configuration in which no ORCs appeared at all. On the other hand, 84.82% of the ORCs had the inanimate head noun – animate

embedded noun (= *IA-A*) configuration where SRCs were found the least. This distributional contrast of ORCs and SRCs indicates that both RC types exhibited a strong tendency to have the prototypical animacy configuration of a transitive clause with an animate subject and an inanimate object.

Since embedded noun animacy is closely related to both head noun animacy and topicality, it is necessary to investigate a 3-way non-linear interaction among them. In the linguistic fields, it is common to examine interaction effects by adding interaction terms to logistic regression models, suggesting that interaction effects can be calculated in the same way for all generalized linear models in R (Manning 2007; Levy 2012; Gries 2015).[5] However, there has been a general agreement that the coefficients estimated for interaction terms and their *p*-values are not reliable and misleading in a non-linear model since Ai and Norton (2003) and Norton et al. (2004). Thus, the three factors are combined into a single factor to capture an integrated effect of multiple linguistic factors. The variable *Config* with eight levels was created by combining all possible levels of the three. Each level of Config has the form [head noun animacy - embedded noun animacy (topicality)]: for example, IA-A (Non-topical) means that a RC has a head that is animate and an embedded NP that is animate and non-topical.

A mixed-effects logistic regression model was implemented using the function *glmer* from package *lme4* in R (Bates et al. 2015). I excluded 1,245 SRCs with A-IA (i.e., A-IA (Topical) and A-IA (Non-topical)) in the model to avoid a quasi-complete separation problem, given that A-IA configurations never occurred with ORCs. Thus, the data in the model consists of 751 ORCs and 1,049 SRCs. Text types and head noun lemmas were included as random effect factors. The RC data was extracted from the two corpus types that deal with a very limited number of topics. Despite the absence of writing topic information in the YELC metadata, the specific writing topics frequently lead to the occurrence of a small number of head nouns. Thus, instead of the writing topics, the head noun lemmas were included as a random effect.[6]

---

5   It is possible to use a tree model to investigate non-linear interactions among categorical predictors. A conditional inference tree representation creating using the ctree function (Hothorn and Zeileis 2015; Hothorn et al. 2019) shows a 3-way interaction among embedded noun animacy, head noun animacy, and topicality. However, it is difficult to interpret what the 3-way interaction means, that is, the three non-syntactic factors contribute to the degree of subjecthood or objecthood (Shin 2021).

6   The individual (i.e., the participant who produced RCs) can be considered as another potential random effect variable. The model failed to converge if the individual was included as a random effect predictor.

The mixed effects model with the two random predictors was chosen by comparing the AIC scores of the mixed models with the possible random effects. The mixed effects model (AIC = 1010.6) fitted significantly better than the fixed-effects-only model (AIC = 1094.9). Since there is only one single variable, the intercept is suppressed to get the estimates that are the log odds of the outcome at each level. The estimated variances of the random effects and the estimated fixed effects are summarized in Table 2. Figure 3 visualizes the statistical results in Table 2 using the *effects* package in R (Fox et al. 2022). In Figure 3, the *y*-axis of the plot gives us the predicted probabilities for the RC type: the use of ORCs is favored if the *y*-value is below 0.5, and the use of SRCs is preferred if the *y*-value is above 0.5.

Table 2. The results of the mixed effects logistic regression model

Random effects:					
		Variance		Std.Dev.	
Head noun lemma (Intercept)		1.911		1.382	
Text type (Intercept)		0.141		0.376	
Fixed effects:					
	Estimates	S.E.	OR	95% CI	
				low	high
A-A(Non-topical)	4.328	0.840	75.792	2.682	9.323
A-A(Topical)	-0.399	0.465	0.671	-1.311	-6.802
IA-A(Non-topical)	-0.560	0.334	0.571	-1.216	-7.560
IA-A(Topical)	-3.680	0.416	0.025	-4.494	-13.334
IA-IA(Non-topical)	3.473	0.376	32.220	2.736	-0.650
IA-IA(Topical)	-0.254	0.654	0.776	-1.535	-6.594

Out of 1,751 participants produced the data, only 325 participants (18.56%) used more than one RC, and only 45 participants produced 5 to 9 RCs.
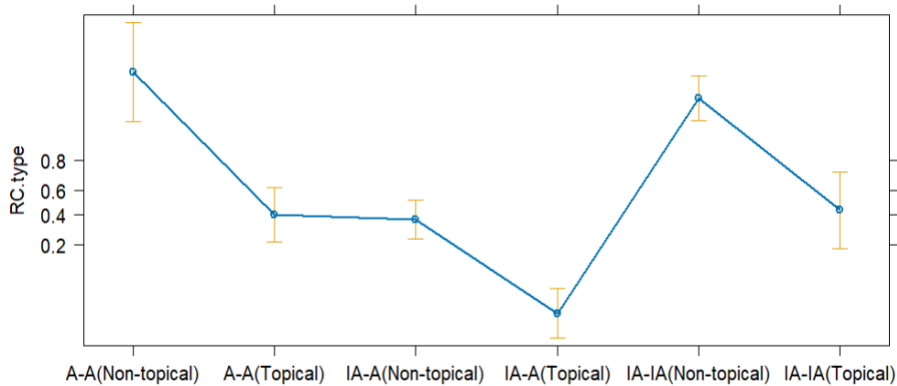
Figure 3. Configuration effect plot of the L2 RC data

First, look at RCs in the A-A and IA-IA configurations, in which the head noun and the embedded noun have the same animacy feature. When embedded NPs are non-topical, the use of SRCs is strongly favored; however, when embedded NPs are topical, ORCs are likely to occur almost as frequently as SRCs. Given that the embedded NP is a subject in an ORC but an object in an SRC, this topicality effect indicates that ORCs and SRCs tend to contain prototypical subjects and objects in terms of topicality.

Consider the RCs with IA-A, where SRCs have an atypical animacy configuration with an inanimate subject and an animate object. In the IA-A (topical) configuration, the use of ORCs is strongly preferred due to their prototypical transitive configuration with inanimate objects and animate topical subjects. However, ORCs are slightly more likely to occur than SRCs in IA-A (Non-topical), that is, when the embedded NP is marked for the subject but unmarked for the object in terms of topicality.

Recall that only SRCs appeared in the A-IA cases that were excluded in Figure 4. This does not mean that ORCs cannot be used in the A-IA configuration, but it suggests that the use of ORCs is strongly disfavored when ORCs have atypical animacy configurations with inanimate subjects and animate objects. Note that SRCs occurred 28.67 times more frequently in A-IA (Non-topical) than in A-IA (Topical). In other words, SRCs strongly tend to have the prototypical transitive configuration with an animate subject and an inanimate non-topical object.

4.2 Comparison of L1 and L2 RC data

The results of the L2 corpus analysis are very similar to those found in Shin's (2020) L1 English corpus data. In both L1 and L2 data, ORCs and SRCs in L2 data showed very different distributional patterns in terms of head noun animacy, embedded noun animacy, and topicality. More specifically, look at Figure 4 and compare it with Figure 3. Figure 4 shows the result of the mixed effects models when the RC data from Shin's L1 corpus study is used. Shin's L1 data were extracted from three spoken corpora and two written corpora consisting of academic essays.[7] The mixed effects model for the L1 data was implemented by including the modes (i.e., spoken and written) and the corpus names as nested random effects. Unlike the L2 data, the L1 data were extracted from corpora dealing with various topics, and hence head lemmas were not included as a random effect factor. In the modal, RCs with A-IA configurations were excluded because no ORCs with A-IA appeared in L1 data, just like in L2 data.
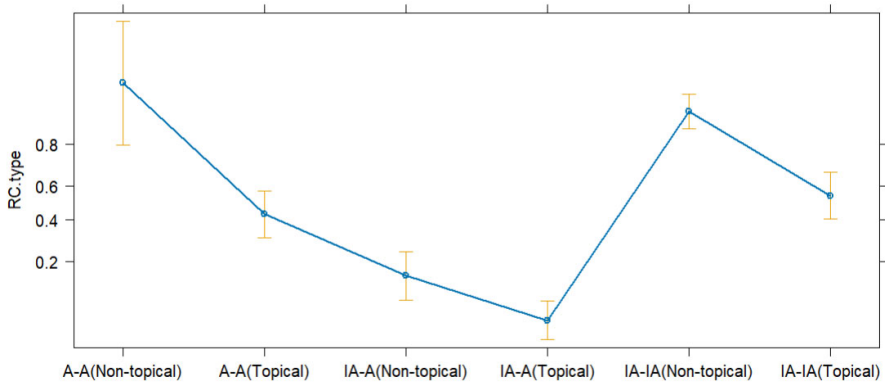

Figure 4. Configuration effect plot for the L1 RC data

In both L1 and L2 data, ORCs and SRCs showed a strong tendency to have animate subjects and inanimate objects: ORCs occurred more frequently than SRCs in IA-A

---

7   The written data consist of BNC (*British National Corpus*) school and university essays and the first-year college students' essays extracted from BAWE (*the British Academic Written English Corpus*). In the case of BAWE, one essay was randomly selected for each student. The spoken data include BNC classroom conversations/lectures and the *Charlotte Narrative and Conversation Collection* narratives.

cases, whereas only SRCs occurred in A-IA cases. When head nouns and embedded nouns are not different in terms of animacy, the use of SRCs is strongly favored unless they contain a topical NP that is marked for an object.

Nevertheless, there is one noticeable difference between Figure 3 and Figure 4. In the IA-A (Non-topical) configuration, the use of ORCs is much more favored in the L1 data, compared to the L2 data: ORCs were 1.17 times more frequent than SRCs in the L2 data, whereas ORCs were 5.45 times more frequent than SRCs in the L1 data. A close examination reveals that 95.42% of the SRCs with IA-A (Non-topical) in the L2 data originated from the argumentative essays, and 21.13% of these modified the same head lemma *experiment*, essentially meaning experiments involving animals. To put it another way, the writing topics made more SRCs with IA-A (Non-topical) in the L2 data. Because of this, ORCs with IA-A (Non-topical) did not outnumber SRCs with IA-A (Non-topical) in the L2 data, as they did in the L1 data. Some slight differences between L1 and L2 data can exist because, unlike the L1 data, the L2 data were extracted from the written texts dealing with a very limited number of topics – presumably less than 10 topics.

In conclusion, both Korean EFL speakers and native speakers of English showed a strong tendency to use ORCs with inanimate subjects and animate topical objects and SRCs that contain inanimate non-topical NPs. SRCs are more likely to occur than ORCs when head nouns and embedded nouns have the same animacy feature and their embedded NPs are non-topical in both L1 and L2 data. These distributional patterns of ORCs and SRCs demonstrate that despite their different word orders, the two types of RCs tend to have the prototypical discourse and semantic structure of a transitive clause.

5. Implications of the corpus study for RC processing

Noun animacy and topicality, as discussed in Section 2.1, have been claimed to be crosslinguistically associated with subjects and objects, although word orders may vary across languages. Some experimental and corpus studies of Korean (Lee 2006, 2010, 2015; Kim 2008) showed that noun animacy and topicality also play a role in whether subjects and objects are case-marked overtly: for example, marked objects in terms of noun animacy and topicality tend to be overtly case-marked. If the experience

of subjecthood and objecthood is universal and influences RC production in L2, it is predicted that Korean EFL learners exhibit native-like sensitivity to the semantic and discourse features of subjecthood and objecthood in producing English RCs. This prediction was confirmed in this L2 corpus study. The results of the L2 corpus analyses also suggest that Korean EFL speakers' L2 experience does not differ from the L1 experience of native English speakers. According to an experience-based approach to RC processing, or the PDC model, Korean EFL learners can use the information of noun animacy and topicality in processing RCs as L1 English speakers do. This section presents the implications of the corpus study for RC processing. In doing so, I also discuss the role of L1-specific experience in processing RCs in L2.

5.1 Frequency interpretation: Input and processing

Previous corpus studies have emphasized the role of input in accounting for why SRCs are not easier to process than ORCs when the embedded NP is pronominal or topical, as in (4) (Reali and Christiansen 2007; cf. Roland et al. 2012) or when the head noun is inanimate, as in (5) (Roland et al. 2007; Gennari and MacDonald 2009).

The corpus study suggests that the integrated effects of noun animacy and topicality should be considered in accounting for L2 RC processing. However, it is misleading to predict the effects of non-syntactic factors on RC processing based solely on the comparative frequencies of ORCs and SRCs in terms of noun animacy and topicality. It has been claimed that the processing differences between ORCs and SRCs are attributed mainly to the embedded argument NP relative to the RC verb (Gibson 1998, 2000; Warren and Gibson 2002; Grodner and Gibson 2005; inter alia). Following this approach, the information of noun animacy and topicality does not play the same role in the incremental processing of ORCs and SRCs because of their different word orders. In an ORC, the embedded subject NP is placed before the RC verb. Thus, the subjecthood information encoded by head noun animacy, embedded noun animacy, and topicality can facilitate the processing of an ORC by allowing comprehenders to anticipate the grammatical roles of the embedded NP as well as the head noun before the parser hits the RC verb. On the other hand, when processing an SRC, the embedded NP does not impose a memory burden because it is placed

right after the RC verb that immediately determines its grammatical role. It is therefore necessary to examine the animacy configuration and topicality that each RC type frequently has in the input to investigate the role of linguistic experience in RC processing.

5.2 Subjecthood effects on RC processing

The subjecthood factors – head noun animacy, embedded noun animacy, and topicality – can account for relative difficulties in processing ORCs. In his experimental study, Baek (2012) discovered that both native English speakers and Korean EFL learners read the ORC with the IA-A (Non-topical) configuration in (4b) faster than the ORC with the A-IA (Non-topical) configuration in (5b), and they processed (4b) as easily as the SRC counterparts in (4a) and (5a).

(4) a. [The movie that pleased the director] received a prize at the film festival.
   b. [The movie that the director watched] received a prize at the film festival.
(5) a. [The director that watched the movie] received a prize at the film festival.
   b. [The director that the movie pleased] received a prize at the film festival.

Given that subjects and objects crosslinguistically tend to be animate and inanimate, respectively, the IA-A configuration in (4b) allows Korean EFL speakers to anticipate an ORC with an inanimate object and an animate subject. Considering that ORCs do not follow the canonical order of English, L2 learners' experience of the target language can also aid in interpreting the subjecthood information correctly and facilitating the incremental processing of ORCs. Indeed, the processing difference of the ORCs in (4b) and (5b) matches the ORC distribution with respect to their animacy configuration. In both the L1 and L2 data, more than 80% of ORCs had IA-A configurations, whereas no ORCs with the A-IA configuration were found.

Recall that ORCs predominately occurred with inanimate head nouns in the L1 and L2 corpora. Nevertheless, in the processing of the ORC with the IA-A configuration in (4b), the effect of head noun animacy emerged when the parser encountered the embedded NP in both L1 and L2 experimental studies (Baek 2012). In other words, the animacy of the head noun can facilitate RC processing when

it interacts with the animacy of the embedded NP. Since the embedded NP is placed after the RC verb in an SRC, it is predicted that noun animacy information has little effect on SRC processing. This prediction was borne out. For example, no processing difference was found between the SRC with the IA-A configuration in (4a) and the SRC with the A-IA configuration in (5a) in L1 and L2 English (Baek 2012). However, SRCs with the IA-A configuration occurred much more frequently than SRCs with the IA-A configuration in the L1 and L2 corpus data. The distributional patterns of SRCs do not mirror their processing similarities because noun animacy as well as topicality do not serve as cues in the processing of SRCs, which is an incremental process (Shin 2019, 2020).

Kim (2016) and Choe and Deen (2020) demonstrated that, just like native speakers of English, Korean adult EFL learners processed ORCs with personal pronouns more easily than those with embedded NPs headed by common nouns.[8] According to Choe and Deen (2020: 111), reading time was significantly shorter in both RC verb and main verb regions in (6a) compared to (6b).

(6) a. The nurse that you ignored last evening answered the phone quickly.
    b. The nurse that the doctor ignored last evening answered the phone quickly.

Since ORCs overwhelmingly contained personal pronouns in both the L1 and L2 corpora, one could argue that, just like native speakers of English, Korean EFL learners read (6a) faster due to the natural discourse structure that L1 and L2 speakers frequently encounter in their inputs.[9] However, because topicality interacts with noun animacy in RC distribution, noun animacy should also be considered to explain the processing ease of (6a) more thoroughly. Recall that ORCs were used predominately with inanimate head nouns. In the L1 and L2 data, ORCs were hardly used with animate head nouns unless they contained animate topical NPs, that is, pronouns

---

8   Kim (2016) used ORCs that modified inanimate head nouns in his experiments: ORCs with IA-A (Topical) and IA-A (Non-topical) configurations. ORCs with the IA-A (Topical) configuration, which is the most frequent ORC form, were processed faster than those with the IA-A (Non-topical) configuration. Choe and Deen (2020) focused on the processing differences between ORCs and SRCs rather than on ORCs with different discourse structures.

9   As pointed out by Roland et al. (2012), most previous experimental studies have used an ORC with the A-A (Non-topical) configuration in an isolated context where the definite NP is unlikely to be interpreted as a discourse-old referent or topic. As a result, unlike a personal pronoun, the embedded definite NP cannot facilitate ORC processing.

usually referring to humans. Thus, the processing difference of ORCs with A-A (Topical) and ORCs with A-A (Non-topical) configurations in (6) matches their frequency difference. In addition, most personal pronouns are inflected for cases; hence, their morphological forms (e.g., *I* and *me*) can overtly encode their grammatical roles. The inflected form of a personal pronoun itself can serve as a very strong cue for the subject role of the embedded NP in an ORC.

5.3 L1 effect on L2 RC processing

Baek (2012) found a processing difference between L1 and L2. Native speakers of English read ORCs more readily when the embedded nouns were animate, as in (7a), compared to when their embedded nouns were inanimate, as in (7b). In other words, they processed ORCs with animate subjects faster. However, Korean EFL learners did not show any significant difference in processing (7a) and (7b).

(7) a. The teacher that the actor surprised was invited to the film festival.
    b. The teacher that the movie surprised was invited to the film festival.

It seems natural that native English speakers read ORCs with A-IA (Non-topical) slower than ORCs with A-A (Non-topical) since inanimate nouns are more marked for subjects than animate ones. Moreover, in the case of ORCs with A-IA, their atypical animacy configuration with animate objects and inanimate subjects can render ORC processing more difficult. The processing difference between ORCs with A-IA (Non-topical) and ORCs with A-A (Non-topical) somewhat mirrors their distribution patterns. While two ORCs with A-A (Non-topical) appeared in the L1 data, no ORCs with A-IA were found.[10]

Baek (2012) proposed that Korean EFL speakers showed no processing difference between ORCs with A-A (Non-topical) and ORCs with A-IA (Non-topical) due to Korean EFL speakers' lack of experience with the target language. However, no

---

10 When we compare the distribution of ORCs with respect to embedded noun animacy in L1 and L2, L2 speakers used ORCs with animate embedded nouns more frequently than native English speakers: the animate embedded nouns accounted for 87.81% of the ORCs in the L1data, and they accounted for 97.34 % of the ORCs in the L2 data. This difference may be due to a more limited range of verbs in the L2 data.

distributional difference between the L1 data and the L2 data was found regarding ORCs with A-IA (Non-topical) and A-A (Non-topical). In the L2 data, no ORCs appeared in A-IA at all, although one ORC appeared in A-A (Non-topical). Furthermore, it is puzzling why Korean EFL learners have trouble using the embedded noun animacy information – but not the head noun animacy information.

One possible explanation for the L1/L2 processing difference can be found in L2 speakers' experience of their native language. Kang (2014), who extracted 136 ORCs and 261 transitive SRCs from Korean newspaper texts in the *Sejong Korean Parsed Written Corpus*, analyzed animacy configurations. Since modes or genres can affect the frequency of ORCs (Roland et al. 2007), ORCs in Kang's L1 Korean data are compared with those in Shin's L1 English written data that consisted of academic essays in Table 3.[11]

Table 3. Animacy configuration frequencies of ORCs in L1 Korean (Kang 2014) and L1 English (Shin 2020) written data

Embedded noun animacy	Animate		Inanimate		Total
Head noun animacy	Animate	Inanimate	Animate	Inanimate	
L1 Korean	4	77	3	52	136
L1 English	17	221	0	65	303

In Kang's L1 Korean corpus data, 40.44% of the ORCs had inanimate embedded subjects, whereas 21.45% of English ORCs contained inanimate embedded subjects. ORCs with inanimate embedded nouns were used more frequently in the L1 Korean data compared to the L1 English data. Moreover, in the L1 English data, all ORCs with inanimate subjects modified inanimate head nouns: no ORCs had an A-IA configuration. In the L1 Korea data, there is almost no frequency difference between ORCs with A-A configurations and ORCs with A-IA configurations. Despite the infrequency of ORCs with A-IA configurations in both L1 and L2 data, Korean speakers are more likely to encounter this atypical ORC configuration in their native language than native English speakers.[12] If this experience of L1 Korean affects how

---

11   Kang (2014) originally analyzed the animacy configurations of 96 ORCs that contained overt NP arguments. Korean is a pro-drop language in which pronouns referring to discourse-old or presupposed referents can be omitted. Table 3 included 40 ORCs without overt subjects. Three of the four ORCs with A-A and all three ORCs with A-IA contained overt subject NPs.

12   In English where a majority of verbs can be passivized, inanimate subjects can be easily avoided using

Korean speakers process ORCs in L2 English, it's possible that unlike native speakers of English, Korean EFL speakers process English ORCs with A-IA (non-topical) just as easily as ORCs with A-A (non-topical).

6. Conclusion

A number of corpus studies have been conducted to investigate the role of linguistic experience or input in sentence processing. Previous studies have used the target language produced by L1 speakers to represent L2 learners' input. However, L2 learners are exposed not only to the target language produced by L1 speakers but also to the L2 language produced by other L2 learners. Therefore, this study analyzed an L2 learner corpus study in an attempt to establish more accurate input for L2 learners whose native language is typologically different from the target language.

The distributional patterns of ORCs and SRCs in the L2 English data were very similar to those in the L1 English data. ORCs overwhelmingly modified inanimate head nouns and had animate topical NPs as their embedded subjects, whereas transitive SRCs predominately had inanimate non-topical NPs as their embedded objects. ORCs and SRCs strongly tend to have prototypical subjects and objects. The distributional differences between ORCs and SRCs were attributed to their different surface orders of subjects and objects. The results of the corpus study indicated that Korean EFL speakers' L2 experience does not differ from native English speakers' L1 experience.

Noun animacy and topicality are crosslinguistically associated with the syntactic roles of subjects and objects. In other words, the experience of subjecthood and objecthood is universal, although the syntactic and morphological realizations of subjects and objects may differ across languages. I have proposed that the universal experience of subjecthood and objecthood influences RC production and processing by L2 speakers whose native language differs from the target language in terms of word order. Korean EFL learners and native English speakers process English ORCs in strikingly similar ways because they both use the information of the subjecthood

---

passivization. In Korean, the passivization process is less productive, compared to English, and some verbs like *ttalakata* 'follow' and *mwutta* 'ask' cannot be used in passive constructions (Kim and Kim 2013). The L1/L2 frequency difference regarding ORCs with A-IA configurations can be partly due to the difference in the productivity of passivization in Korean and English.

and objecthood encoded by the interaction among head noun animacy, embedded noun animacy, and topicality. However, due to the incremental property of the parser, noun animacy and topicality do not serve as cues that facilitate the processing of SRCs. A language-specific experience can also play a role in RC processing. L2 learners' experience in their native language can account for some processing differences between Korean EFL learners and native English speakers This article has further supported an experience-based account of RC processing or a usage-based approach to L2 acquisition.

References

Ai, Chunrung and Edward Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80(1): 123-129.

Aissen, Judith. 1999. Markedness and subject choice in optimality theory. *Natural Language and Linguistic Theory* 17(4): 673-711.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3): 435-483.

Anthony, Lawrence. 2020. AntConc (Version 3.5.8) [Computer Software]. Waseda University. Available at https://www.laurenceanthony.net/software.

Baek, Soondo. 2012. *Processing of English relative clauses by adult L2 learners.* PhD Dissertation. University of Illinois at Urbana-Champaign.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

Bossong, Georg. 1991. Differential object marking in Romance and beyond. In Dieter Wanner and Douglas Kibbee (eds.), *New analyses in Romance linguistics: Selected papers from the XVIII linguistic symposium on Romance languages*, 143-170. Amsterdam; Philadelphia, PA: John Benjamins Publishing Company.

Choe, Jinsun and Kamil Ud Deen. 2020. Pronoun advantage in L2 relative clause processing. In Megan Brown and Alexandra Kohut (eds.), *Proceeding of the 44th Boston University Conference on Language Development*, 105-113. Somerville, MA: Cascadilla Press.

Comrie Bernard. 1989. *Language universals and linguistic typology.* Chicago, IL: University of Chicago Press.

Dixon, Robert. 1979. Ergativity. *Language* 55(1): 59-138.

Dixon, Robert. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Ellis, Nick. 2002. Frequency effect in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2):

143-188.

Ellis, Nick and Laura Collins. 2009. Input and second language acquisition: The roles of frequency, form, and function introduction to the special issue. *Modern Language Journal* 93(3): 329-335.

Ellis, Nick and Stefanie Wulff. 2020. Usage-based Approaches to L2 Acquisition. In Bill VanPatten, Gregory Keating, and Stefanie Wulff (eds.), *Theories in second language acquisition: An introduction*, 63-82. London; New York, NY: Routledge.

Fox, John, Sanford Weisberg, Brad Price, Michael Friendly, Jangman Hong, Robert Anderson, David Firth, Steve Taylor, and R Core Team. 2022. Effect displays for linear, generalized linear, and other models. Available at https://cran.r-project.org/web/packages/effects/effects.pdf.

Gennari, Silvia and Maryellen MacDonald. 2008. Semantic indeterminacy in object relative clauses. *Journal of Memory and Language* 58(2): 161-187.

Gennari, Silvia and Maryellen MacDonald. 2009. Linking production and comprehension processes: The case of relative clauses. *Cognition* 111(1): 1-23.

Gibson, Edward. 1998. Linguistic complexity: Locality and syntactic dependencies. *Cognition* 68(1): 1-76.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil (eds.), *Image, language, brain: Papers from the first mind articulation project symposium,* 95-126. Cambridge, MA: MIT Press.

Gordon, Peter and Randall Hendrick. 2005. Relativization, ergativity, and corpus frequency. *Linguistic Inquiry* 36(3): 456-463.

Gries, Stefan. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1): 95-125.

Grodner, Daniel and Edward Gibson. 2005. Consequences of the serial nature of linguistic input. *Cognitive Science* 29(2): 261-291.

Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2): 274-307.

Hothorn, Torsten and Achim Zeileis. 2015. partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16(118): 3905-3909.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2019. ctree: Conditional inference trees, Vignette R package partykit version 1.1-1, Available at https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf.

Hsiao, Yaling and Maryellen Macdonald. 2013. Experience and generalization in a connectionist model of Mandarin Chinese relative clause processing. *Frontiers in Psychology* 4: 767. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00767.

Kang, Sooyeon and Sun-Young Oh. 2022. A corpus-based study on the use of English relative clauses by Korean EFL learners of different proficiency. *The Journal of Asia TEFL* 19(1):

1-16.

Kang, Soyeon. 2014. The role of syntactic and semantic information in the frequency distribution of relative clauses in Korean: A corpus-based analysis. *Language Information* 19: 5-32.

Keenan, Edward and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1): 63-99.

Kim, Hyunwoo. 2016. The effects of accessibility of English reference in Korean EFL learners' sentence processing. *ITL-International Journal of Applied Linguistics* 167(1): 78-97.

Kim, Kitaek and Hyunwoo Kim. 2013. L1 Korean transfer in processing L2 English passive sentences. In Erik Voss, Shu-Ju Tai, and Zhi Li (eds.), *Selected Proceedings of the 2011 Second Language Research Forum*, 118-128. Somerville, MA: Cascadilla Proceedings Project.

Kim, Taeho. 2008. *Subject and object markings in conversational Korean*. PhD Dissertation. State University of New York at Buffalo.

Lee, Hanjung. 2006. Parallel optimization in case systems: Evidence from case ellipsis in Korean. *Journal of East Asian Linguistics* 15: 69-96.

Lee, Hanjung. 2010. Explaining variation in Korean case ellipsis: Economy versus iconicity. *Journal of East Asian Linguistics* 19: 291-318.

Lee, Hanjung. 2015. Information structure, topic predictability and gradients in Korean case ellipsis: A probabilistic account. *Linguistic Research* 32(3): 749-771.

Lee, Jeongjun and Keun Young Shin. 2017. A syntactic analysis of English relative clauses used by Korean EFL Speakers: A corpus-based study. *English 21* 30(3): 219-245.

Levy, Roger. 2012. Probabilistic methods in linguistics. Lecture 14: Logistic regression. Ms. University of California San Diego.

MacDonald, Maryellen. 1999. Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In Brian MacWhinney (ed.), *The emergence of language*, 177-196. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

MacDonald, Maryellen. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4: 226. https://doi.org/10.3389/fpsyg.2013.00226

MacDonald, Maryellen, Neal Pearlmutter, and Mark Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4): 676-703.

Mak, Willem, Wietske Vonk, and Herbert Schriefers. 2002. The influence of animacy on relative clause processing. *Journal of Memory and Language* 47(1): 50-68.

Mak, Willem, Wietske Vonk, and Herbert Schriefers. 2006. Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language* 54(4): 466-490.

Manning, Christopher. 2007. Logistic regression (with R). Ms. Stanford University.

Monteiro, Kátia, Scott Crossley, and Kristopher Kyle. 2020. In search of new benchmarks: Using L2 lexical frequency and contextual diversity indices to assess second language writing. *Applied Linguistics* 41(2): 280-300.

Mulkern, Ann E. 1996. The game of the name. In Thorstein Frentheim and Jeanette Gundel (eds.), *Reference and referent accessibility*, 235-250. Amsterdam; Philadelphia, PA: John

Benjamins Publishing Company.

Norton, Edward, Hua Wang, and Chunrong Ai. 2004. Computing interaction effects and standard errors in logit and probit models. *Stata Journal* 4(2): 154-167.

Ortega, Lourdes. 2016. Multi-competence in second language acquisition: Inroads into the mainstream? In Vivian Cook and Li Wei (eds.), *The Cambridge handbook of linguistic multi-competence*, 50-76. Cambridge: Cambridge University Press.

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

Reali, Florencia and Morten Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57(1): 1-23.

Roland, Douglas, Frederic Dick, and Jeffrey Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3): 348-379.

Roland, Douglas, Gail Mauner, Carolyn O'Meara, and Hongoak Yun. 2012. Discourse expectations and relative clause processing. *Journal of Memory and Language* 66(3): 479-508.

Shin, Keun Young. 2019. An expectation-based account for the processing difficulty of *it* object-extracted relative clauses. *Korean Journal of Linguistics* 44(4): 807-829.

Shin, Keun Young. 2020. Non-linear interactions of factors influencing relative clause distribution and their implications on relative clause processing. *Korean Journal of Linguistics* 45(4): 919-940.

Shin, Keun Young. 2021. An integrated interaction of multiple linguistic factors in logistic regression models: Comparison with tree models. *The Journal of Studies in Language* 37(3): 291-305.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert Dixon (ed.), *Grammatical categories in Australian languages*, 112-171. Canberra: Australian Institute of Aboriginal Studies.

Sung, HyunMi. 2014. A study on the use of English relative clauses in Korean EFL learners' writing. *English Language Teaching* 26(1): 65-83.

Traxler, Matthew, Rihana Williams, Shelley Blozis, and Robin Morris. 2005. Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language* 53(2): 204-224.

Traxler, Matthew, Robin Morris, and Rachel Seely. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47(1): 69-90.

Warren, Tessa and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85(1): 79-112.

Keun Young Shin
Professor
Department of English Language and Literature

Chonnam National University
77 Yongbong-ro, Buk-gu
Gwangju 61186, Korea
E-mail: kyshin@jnu.ac.kr