

Relation between Frequency and Processing Difficulty of English Relative Clauses by L2 Learners: A Learner Corpus Analysis

Shin, Keun Young

Chonnam National University

ABSTRACT

The Journal of Studies in Language 34.3, 491-504. This paper explores whether frequency can play a role in accounting for difficulties in processing English relative clauses (RCs) by L2 learners. Previous experimental studies showed that the processing difficulty of RCs in L2 is modulated by the syntactic function of the head noun and noun animacy. By analyzing a corpus of essays written by Korean adult learners of English, this paper shows that the distribution patterns of the RCs in the learner corpus are consistent with the experimental findings. Object-extracted RCs almost exclusively occur with inanimate head nouns, while subject-extracted RCs occur about two times more frequently with animate head nouns than inanimate head nouns. It is also found that the object-extracted RCs with inanimate head nouns often contain animate subjects and have the animacy configuration reflecting typical thematic structures of transitive clauses. The results of the corpus analysis provide support to frequency-based approaches for L2 processing, claiming that a higher frequency of a particular RC form makes it easier for L2 learners to acquire and process. (Chonnam National University)

Keywords: relative clause, language processing, second language, frequency, learner corpus

 OPEN ACCESS



<https://doi.org/10.18627/jslg.34.3.201811.491>

pISSN : 1225-4770

Received: October 01, 2018

Revised: October 24, 2018

Accepted: November 14, 2018

This is an Open-Access article distributed under the terms of the Creative Commons Attribution NonCommercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2018 the Modern Linguistic Society of Korea

본인이 투고한 논문은 다른 학술지에 게재된 적이 없으며 타인의 논문을 표절하지 않았음을 서약합니다. 추후 중복게재 혹은 표절된 것으로 밝혀질 시에는 논문게재 취소와 일정 기간 논문게출의 제한 조치를 받게 됨을 인지하고 있습니다.

1. Introduction

Many recent studies of language processing and comprehension have claimed that statistical information plays an important role in modulating the processing difficulty of a particular linguistic form (Cuetos and Mitchell, 1988; MacDonald et al., 1994; Mitchell et al., 1995; Gibson et al., 1996; Gibson and Schütze, 1999; Desmet et al., 2006; inter alia). The general idea behind this frequency-based approach is that it is easier to comprehend or process a particular linguistic form if speakers have encountered it more frequently. Analyses of corpora have been conducted to test the validity of frequency-based approaches of language processing.¹⁾

Recently, Roland et al. (2007) and Gennari and MacDonald (2008, 2009) have

1) A frequency-based approach is known by several different names - constraint-based, expectation-driven, experience-based, usage-based or probabilistic approach.

explored frequency effects in processing relative clauses (RCs). Several experimental studies showed that the relative difficulty of processing RCs is modulated by the grammatical function of the head noun and its animacy (Mak et al., 2002; Traxler et al., 2002, 2005; Gennari and MacDonald, 2008, 2009, *inter alia*). It was also reported that the distributional patterns of English RCs mirror this processing difficulty, supporting the claim that frequency should be considered as one of the main factors influencing language processing (Roland et al., 2007; Gennari and MacDonald, 2009).

It is argued that second language (L2) learners process RCs in quite similar ways as first language (L1) speakers. For example, the experiments of Baek (2012) showed that both English native speakers and Korean learners of English experienced greater difficulty processing RCs when the animate head noun functioned as the object within the RC. One may argue that such striking similarities between L1 and L2 processing are due to the fact that both native speakers and L2 learners are exposed to the same input, that is, the target language produced by native speakers. However, not only input frequencies but also output frequencies are claimed to play roles in facilitating L2 learning and processing (Swain 1995, 1998; Ellis, 2003; Ellis and Collins, 2009; *inter alia*). Thus, it is necessary to analyze learner corpora as well as native speaker corpora in order to investigate frequency effects in L2 acquisition and processing. This paper aims to explore the role of frequency in accounting for the relative ease of RC processing in L2 by examining the distributional properties of RCs in a L2 learner corpus.

Previous L2 corpus studies on RCs (Schumann 1980; Wong 1991; Sung 2014; Lee and Shin 2017; *inter alia*) have focused on syntactic functions of the head nouns that are claimed to influence L2 acquisition and comprehension (Kuno, 1974; Keenan and Comrie, 1977; Hamilton, 1994). Non-syntactic factors such as noun animacy have not yet been considered in L2 corpus studies. By analyzing a corpus of essays written by Korean adult learners of English, this paper investigates the distributional frequencies of RCs with regard to both syntactic and semantic factors that determine the relative ease of RC processing in L2 and provides further support for frequency-based accounts.

2. Previous Studies

In this section, I discuss previous analyses of RC processing in L1 and L2 and show what factors are claimed to influence RC processing and how they are accounted for in different theoretical frameworks.

2.1 Processing Difficulty of RCs by L1 Speakers

Previous studies have mainly dealt with the two types of relative clauses, that is, object-extracted relative clauses (ORCs) and subject-extracted relative clauses (SRCs). The distinction between SRCs and ORCs lies on the different grammatical roles of the so-called gaps, as in examples (1) and (2) used by King and Just (1991). In the psycholinguistic literature, RCs are analyzed to contain a gap, *i.e.*, the empty position from which a phrase is extracted. In (1) and (2), the gap is indicated by *e*, which is co-indexed with the filler *the reporter*.²⁾ In the SRC, the NP [*the reporter*], which is assumed to be extracted from the RC, functions as the subject of the RC, while the same extracted NP functions as the

2) According to the standard syntactic analysis, the relative clause in (1) is an adjunct modifying the head noun *newspaper*, and the sequence *the reporter* does not form an NP constituent. During incremental sentence processing, however, *the reporter* is taken as a filler constituent.

object in the ORC.

- (1) SRC: The reporter₁ [that e_1 attacked the senator] admitted the error.
- (2) ORC: The reporter₁ [that the senator attacked e_1] admitted the error.

It is a well-established finding that ORCs are harder to process or comprehend than SRCs (King and Just, 1991; Just and Carpenter, 1992; Gordon et al., 2001, 2004; Mak et al., 2002; Traxler et al., 2002; Gennari and MacDonald, 2008, *inter alia*).

Recent studies have shown that the processing difficulty of ORCs is affected by the animacy of the head noun (Trueswell et al. 1994, Mak et al., 2002; Traxler et al., 2002, 2005; Gennari and MacDonald, 2008, 2009, *inter alia*). According to Traxler et al. (2002, 2005), who conducted eye-movement monitoring experiments, the processing difficulty of ORCs arises when the head noun is animate and the RC subject is inanimate, as in (3b); ORCs with inanimate head nouns as in (4b) are processed as easily as SRCs as in (3a) and (4a).

- (3) Animate head noun and inanimate noun in the RC
 - 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.
- (4) Inanimate head noun and inanimate noun in the RC
 - 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.

Traxler et al. (2002, 2005) claimed that the processing difficulty of ORCs stems from a process of reanalysis using a structural parsing heuristic called the active filler strategy. According to the active filler strategy, the head NP is forced to be initially interpreted as the subject of the RC, leading to wrongly analyze ORCs as SRCs.³ As a consequence, during processing ORCs, a process of reanalysis must take place to correct this initial misanalysis. Traxler et al. (2002) proposed that it is more difficult to abandon the initial misanalysis of the ORC as the SRC when the head noun is animate, as in (3b), because animate nouns are good candidates for subjects, compared to inanimate nouns. Noun animacy is closely related to a thematic role that is mapped onto a particular syntactic position : AGENT and EXPERIENCER roles which are usually animate by their definitions are more likely to occur in the subject position, and the THEME and LOCATION roles which are often inanimate are more likely to be realized as objects or non-subjects (Fillmore, 1971; Jackendoff, 1990; Dowty, 1991, *inter alia*).⁴ Traxler et al. (2002) claimed that due to the hierarchy of subjecthood, it is easy to reanalyze the inanimate head noun as the object of the ORC.

3) For a detailed explanation of the active filler strategy, refer to Frazier and Clifton (1989), Traxler and Pickering (1996), and Pickering and Traxler (1998).

4) Note that Dowty (1991) originally used the concepts of “Proto-Agent” and “Proto-Patient” rather than distinct thematic roles adopted in this paper. The general hierarchy of subjecthood is illustrated below. The relative ranking of locations with respect to themes differ from authors to authors (cf. Jackendoff, 1972; Grimshaw, 1990).

(i) AGENT > EXPERIENCER > INSTRUMENT > THEME > LOCATION

Traxler et al. (2002, 2005) and Baek (2012) also found in their experimental studies that reliable main effects of animacy emerge before the RC verb is encountered. It means that when they encounter the embedded noun in the ORC, speakers immediately incorporate animacy information into the interpretation of the RC structure by associating the inanimate head noun with the object of the RC verb before its grammatical function or thematic role is determined by the RC verb. It suggests that the effect of noun animacy is attributed to speakers' expectation that the ORC has the inanimate (head) noun - animate (embedded) noun configuration, ultimately matching the animate subject - inanimate object configuration. This raises a fundamental question as to what causes speakers to have such an expectation.

A frequency-based approach has provided a compelling answer to the above question, emphasizing the role of speakers' linguistic experience in processing and learning sentences. This approach claims that speakers experience greater difficulty processing a linguistic form if it occurs less frequently and hence speakers have encountered it less often. Indeed, there has been a growing number of statistical approaches to language processing, most of which dealing with ambiguous resolution of sentences with reduced relative clauses such as *The man examined by the doctor...* (Cuetos and Mitchell, 1988; Desmet et al., 2006; Gibson and Schütze, 1999; Gibson et al., 1996; MacDonald et al., 1994; inter alia).

According to the frequency-based approach, ORCs with animate nouns are harder to process than those with inanimate head nouns or SRCs partly because they have infrequent and unexpected structures. Several corpus studies of RCs produced by English native speakers have supported this view by showing that SRCs outnumber ORCs (Stauble, 1978; Roland et al., 2007, inter alia) and that ORCs with animate head nouns are highly infrequent, compared to ORCs with inanimate head nouns (Roland et al., 2007; Gennari and MacDonald, 2008).⁵⁾⁶⁾ Gennari and MacDonald (2008, 2009) also demonstrated that the effect of noun animacy on the distribution of RCs is due to the general tendency that objects are inanimate nouns carrying THEME roles by analyzing the noun animacy and thematic structures of the ORCs and passive RCs both in production experiments (Gennari and MacDonald, 2008) and in natural productions (Gennari and MacDonald, 2009).⁷⁾ Their analyses revealed that the head nouns of the ORCs were much more frequently associated with THEME roles than with EXPERIENCER or other thematic roles even when they are animate. They suggested that English native speakers have more difficulty processing ORCs when animate head nouns are interpreted as objects bearing EXPERIENCER roles (e.g., *The director that the movie pleased*) because this thematic structure is infrequent, further supporting the frequency-based account.

5) It is a well-known fact that SRCs occurred much more frequently than ORCs in written corpora, but this is not always the case in spoken corpora. Choi and Kim (2009) analyzed the International Corpus of English - the British English Component and found that SRCs occurred more frequently than ORCs in both spoken and written corpora, but the frequency difference between SRCs and ORCs is greatly reduced in the spoken corpus. It is reported that the ORCs outnumbered the SRCs in the Switchboard corpus, a collection of telephone conversations in particular contexts (Roland et al., 2007).

6) Roland et al. (2007) coded a random sample of 200 RCs extracted from the Brown and Switchboard corpora and identified the animacy of the head nouns. They found that animate head nouns often occur with SRCs, while inanimate head nouns are more likely to occur with ORCs.

7) Gennari and MacDonald (2008, 2009) investigated the animacy and thematic role effects by examining the RCs produced by 64 students via sentence completion tasks and by analyzing passive RCs and ORCs after randomly extracting RCs containing a restricted set of verbal predicates taking the agent-theme arguments and theme-experiencer arguments from the British National Corpus.

2.2 Processing Difficulty of RCs by L2 Learners

It is claimed that L2 processing is also affected by the grammatical function of the gap and the animacy of the head noun (Omaki and Arijji 2005; Jackson and Roberts 2010; Baek, 2012). Baek (2012) conducted self-paced reading experiments with both English native speakers and Korean learners of English. The results of his study revealed that adult L1 and L2 speakers of English processed RCs in quite similar ways: both of them found ORCs with animate head nouns and inanimate subjects (e.g., *The director that the movie pleased*) harder to process than ORCs with inanimate head nouns and animate subjects (e.g., *The movie that the director watched*) as well as SRCs. In addition, Baek (2012) observed that L2 learners immediately incorporated the animacy information to construct the structure of the RC, as L1 speakers did. These results indicate that Korean learners of English use the parsing mechanisms employed by native speakers of English. Both L1 and L2 speakers expect the inanimate head noun to be the THEME argument of the RC verb, mapping onto the object position before its thematic role is actually determined by the RC verb.

As discussed earlier, frequency has been recognized as one of the major factors influencing RC processing by L1 English speakers: English native speakers have more difficulty processing ORCs with animate head nouns because they are infrequent in the input. Given that L2 learners are also exposed to the target language produced by native speakers, it can be argued that the similarities between L1 and L2 RC processing is due to the same or very similar input. In L2 acquisition, however, output is another important source of input to the learner (Krashen, 1989). When L2 learners are learning the target language in their home country, they are not only exposed to the language of native speakers but also to the language produced by other L2 learners. Consider the L2 learning environment in Korea. In a classroom, Korean students interact with their Korean peers who share their native language and have similar interlanguages, as well as with their English teachers. In fact, recent L2 acquisition models have emphasized the roles of both input and output frequencies in facilitating L2 learning (Ellis and Collins, 2009; Ellis, 1998, 2002, 2003; inter alia). Frequency information found in L2 production can offer insights into the important factors of L2 acquisition and processing (Swain 1995, 1998).

Several attempts have been made to investigate L2 acquisition and processing of RCs by analyzing L2 corpora (Schumann, 1980; Wong, 1991; Sung, 2014; Lee and Shin, 2017; inter alia). They have focused on testing the validity of the hypotheses claiming that L2 acquisition and processing of RCs are affected by the syntactic functions of fillers and gaps: for example, Keenan and Comrie's (1977) NP accessibility hierarchy hypothesis, Kuno's (1974) perceptual difficulty hypothesis, and Hamilton's (1994) subject-object hierarchy hypothesis. One robust finding is that SRCs, which are easier to comprehend, occur more frequently than ORCs, which are more difficult to comprehend, in L2 corpora. It was also found that RCs are easier to produce and process when they modify nouns in the object position, compared to when they modify nouns in the subject position. These results show that there exists a close relationship between output frequency and acquisition. To my knowledge, however, no L2 corpus analyses have been conducted to investigate the role of frequency in accounting for semantic factors that modulate RC processing in L2, that is, animacy of nouns related to a prototypical thematic structure. This paper addresses this issue by examining the relative frequency of RCs with respect to both syntactic and semantic factors.

3. Corpus Analysis

This section presents the distributional frequencies of RCs found in an L2 corpus and discusses whether they are consistent with those found in previous L1 corpus studies (Roland et al. 2007; Gennari and MacDonald, 2008, 2009), supporting frequency-based accounts of difficulties in processing RCs by L2 learners.

3.1 Materials

I used 159 English essays written by Korean college students who participated in English essay writing contests for L2 learners, held at a university in Gwangju, Korea from 2006 to 2011. The participants were instructed to write argumentative essays within 4 hours after topics were announced.⁸⁾ The texts contained 114,651 words.

The corpus was hand-coded in order to accurately identify syntactic structures of RCs and semantic properties of the head noun and the RC subject. As pointed out by Roland et al. (2007), it is difficult to generate reliable semantic and other meaning-related frequency using automatic parsing. In this study, all types of the relative clauses were extracted from the corpus by hand, and then the restrictive RCs were hand-coded for the syntactic and semantic factors that are claimed to influence the difficulty of RC processing.⁹⁾

In the corpus, there were 1,028 relative clauses, but 142 relative clauses were excluded because they were not restrictive clauses containing a finite verb or they started with *what*, *when*, *where* and *whose*. In addition, 86 relative clauses were excluded in the analysis because they contained major grammatical or lexical errors which made it difficult to analyze the structure of the RC.¹⁰⁾ As shown in Table 1, a total of 795 restrictive clauses were identified in the corpus.¹¹⁾ Table 1 shows the frequency of the RCs by the grammatical function of the gap.

Table 1. Frequency of restrictive RCs (795 tokens): Functions of gaps

Functions of Gap	Subject	Object	Oblique complement
RC counts	666	113	16

The RCs in Table 1 were introduced by the complementizers *which*, *who* and *that* or did not contain any complementizer. Unlike SRCs, ORCs and oblique RCs can be used without any complementizers or occur with the null complementizer, as exemplified in (5).

(5) The student I met passed the exam.

Previous L1 English corpus studies (Roland et al. 2007; Gennari and MacDonald, 2009) extracted all types of SRCs and

8) Although their proficiency in English was not measured, the contest participants were L2 learners with high proficiency who were able to deeply understand social issues and express their opinions in English.

9) Relative clauses were extracted from the corpus by a research assistant. The syntactic and semantic coding was carried out by me.

10) Minor grammatical errors such as subject-verb agreement errors and tense errors were ignored in the corpus analyses.

11) This count excluded 5 cases in which an NP is extracted with a preposition via the so-called pied-piping.

ORCs from corpora, regardless of which complementizer was present. Thus, this study also included all types of SRCs and ORCs to examine the distributional patterns of RCs which are comparable to those found in previous L1 corpus studies.

3.2 Results

3.2.1 Frequency Difference between SRCs and ORCs

Given that 666 SRCs and 113 ORCs were found in the corpus, it seems obvious that Korean learners use SRCs more frequently than ORCs in English. This frequency difference between SRCs and ORCs was reported in previous L2 corpus studies (e.g., Sung, 2014; Lee and Shin, 2017). As pointed out by Gordon and Lowder (2012), however, most previous corpus studies (except Gordon and Hendrick 2005; Choi and Kim, 2009) have a shortcoming in that SRCs with intransitive verbs are included to examine the frequency difference between SRCs and ORCs. Since ORCs only occur with transitive verbs, a higher frequency of SRCs can be simply attributed to the fact that SRCs can occur with all kinds of verbs including intransitive verbs. Moreover, in psycholinguistic or experimental studies (e.g., Traxler et al., 2002, 2005; Baek, 2012), ORCs are compared with SRCs with transitive verbs to explore whether the grammatical function of the head noun plays a role in RC processing. Thus, SRCs should be restricted to those with transitive verbs in order to investigate a correlation between processing difficulty and statistical biases.

Out of 666 SRCs, 375 SRCs contain transitive verbs.¹²⁾ As shown in Table 2, there is still a significant difference in frequency between ORCs and SRCs with transitive verbs.

Table 2. Frequency of SRCs with transitive verbs and ORCs

	SRC	ORC	Total
RC counts	375 (76.8%)	113 (23.2%)	488 (100%)

The SRC, which is easier for L2 learners to process, is much higher in frequency than the ORC, which is more difficult to process. This result therefore provides support for the frequency-based account of L2 processing asymmetry between SRCs and ORCs.

3.2.2 Noun–Animacy Frequency of SRCs vs. ORCs

The ORCs and SRCs in Table 1 were hand-coded for the animacy of the head noun in order to examine the animacy effect on the distribution of SRCs and ORCs. Nouns referring to living entities such as humans and animals were coded as animate nouns; nouns referring to institutions, places, objects or abstract concepts were coded as inanimate nouns (cf. Gennari and MacDonald, 2009).¹³⁾

As shown in Table. 3, the analysis of animacy frequency indicated that almost all the ORCs occurred with inanimate head nouns. In contrast, SRCs occurred with animate head nouns more frequently: SRCs were approximately 2 times

12) I excluded SRCs containing intransitive verbs, copular verbs, or passivized forms of verbs.

13) Gennari and MacDonald (2009) coded institutions as animate nouns.

more likely to occur with animate nouns than inanimate nouns. A χ^2 test comparing the frequency of SRCs and ORCs for animacy indicates there was a significant association between animacy and RC type ($\chi^2 = 185.9$; $p < .0001$).

Table 3. Counts of the SRCs and ORCs by the head noun animacy

	Animate	Inanimate	Total
SRCs	278 (57%)	97 (19.9%)	375 (76.8%)
ORCs	2 (0.4%)	111 (22.7%)	113 (23.2%)
Total	280 (57.4%)	208 (42.6%)	488 (100%)

This result matches the animacy effect on L2 RC processing. Baek (2012) showed that Korean learners of English had greater difficulty processing ORCs with animate head nouns, compared with ORCs with inanimate nouns or SRCs. Therefore, the finding that ORCs rarely occurred with animate head nouns in the L2 corpus indicates that the more frequently a linguistic form is used by L2 learners, the easier it is to process.

3.2.3 Animacy Configuration and Thematic Structure of ORCs

Experimental studies on the animacy effect of RC processing (Traxler et al. 2002, 2005, Gennari and MacDonald 2008, 2009; Baek, 2012) revealed that ORCs with inanimate head nouns and animate subjects processed easier than other animacy configurations, suggesting that it is easier to process ORCs when they have the typical animacy configuration of a transitive clause in which the subject is animate and the object is inanimate. The animacy of the embedded subjects in the ORCs was also analyzed to examine whether this animacy effect matches the frequency of the animacy configurations of nouns involved in ORCs.

While 98.2% of the head nouns in the ORCs were inanimate, 86.7% of the embedded subjects were animate. As shown in Table 4, the analysis of the animacy configuration frequency indicated that ORCs were approximately 6 times more likely to occur in the <inanimate head noun> - <animate embedded subject> configuration (= IA-A configuration) than in any other animacy configurations.

Table 4. Counts of the ORCs by animacy configurations (A: animate, IA: inanimate)

	Head noun animacy - Embedded noun animacy			
	IA-A	IA-IA	A-A	A-IA
ORC counts	97 (85.8%)	14 (12.4%)	1 (0.9%)	1 (0.9%)

This result is consistent with the role of the animacy configuration in modulating L2 processing difficulty of ORCs, as shown in the study of Baek (2012). He found that the IA-A configuration is easier for Korean learners of English to process than the A-A and A-IA configurations. The relative ease of L2 processing with respect to the animacy configuration is strongly correlated with the frequency information found in the L2 corpus.

The effect of noun animacy on RC processing is argued to be due to the thematic structure of a transitive clause (Traxler et al., 2002; Gennari and MacDonald, 2008, 2009). That is, ORCs with the IA-A configuration are easier to process because they match the prototypical <animate subject> - <inanimate object> configuration, which in turn

reflects typical thematic structures in which an AGENT or EXPERIENCER role occupies the subject position and a THEME role occupies the object position (Dowty, 1991).

In order to investigate the relationship between thematic roles and noun animacy, the ORCs were hand-coded for the thematic role of the gap, i.e., the thematic role that is assigned to the head noun by the RC verb. Table 5 shows the frequency of the ORCs by the thematic role of the gap.

Table 5. Counts of the ORCs by the thematic role of the gap

	THEME	LOCATION	INSTRUMENT
ORC counts	111 (98.2%)	1 (0.9%)	1 (0.9%)

The gaps of the ORCs all carried the thematic roles typically associated with inanimate entities: THEME, LOCATION, and INSTRUMENT. The thematic role THEME was assigned to 98.2% of the gaps.¹⁴⁾ This leads us to conclude the strong preference for the inanimate head noun in the ORCs is related to its thematic role.¹⁵⁾ The animacy information interacts with the thematic and syntactic information.

Note that 12.4% of the ORCs (14 tokens) had the IA-IA configuration. This configuration has been hardly discussed in previous studies. A close examination of their thematic structures revealed that the inanimate embedded subjects all referred to institutions such as governments, universities, companies and countries, as illustrated in (6).

(6) The incentive *that the government is trying to promote* is offering benefit to those who are trying to get a job.

Furthermore, they occurred with the transitive verbs assigning an AGENT or EXPERIENCER role to the subject: for example, *announce*, *demand*, *endure*, and *suggest*. The inanimate subjects of the ORCs were metaphorically or metonymically interpreted as animate entities that are sentient or volitional. Ironically, the IA-IA configurations also show that the animacy of the head noun is strongly correlated with its thematic role assigned by the RC verb, supporting the view that ORCs with the IA-A configuration are easier to process and produce because their animacy configuration reflects prototypical thematic structures.

According to Baek's (2012) experimental studies of the ORC processing, the important effect of animacy emerged before the parser encounters the RC verb. The frequency-based approach offers an account for this phenomenon. Comprehenders expect the inanimate head noun to be a THEME argument that is mapped onto the object of a RC verb immediately when they encounter the animate embedded noun because this animacy and thematic role configuration is highly frequent.

14) The two animate head nouns in the ORC data were also assigned the THEME roles by the RC verb, as illustrated in (i) and (ii).

(i) People would pay their efforts to improve themselves and then they get great abilities like *the stars I mentioned before*, finally they would make a profit.

(ii) When Korea was under the IMF crisis, almost all companies tried to cut down on the number of *employees they had*.

15) It may be also worth mentioning that the passive SRCs in the corpus often had inanimate head nouns. Out of 81 SRCs that had passive verb forms, 57 modified inanimate head nouns; 24 modified animate head nouns. This distributional pattern is in contrast with the broad distributional pattern of SRCs as shown in Table 3, which modified animate head nouns more frequently. This is not surprising when we consider the fact that the THEME argument is often placed in the subject position of the passive sentence. The noun animacy effect in processing RCs is therefore closely attributed to the thematic role assigned to the head noun - not simply to its grammatical function.

3.2.4 Distribution of Oblique RCs

As Table 1 shows, there were 16 oblique RCs in which the gap functions as an oblique complement. As exemplified in (7), the structure of oblique RCs is very similar to that of ORCs in that they contain RC subjects and their gaps function as complements following RC verbs.¹⁶⁾

- (7) ... if there are lots of events for information about getting a job, it will solve this problem which local universities suffer from.

The only difference between the oblique RCs and the ORCs is the syntactic form of the complement: unlike the object complement, the oblique complement occurred with a preposition. Given that ORCs and oblique RCs have the same structure before the verb, it is predicted that oblique RCs will be processed like ORCs: oblique RCs will be easier to process when they have the inanimate head and the animate embedded subject. If oblique RC processing is also influenced by frequency, it is further predicted that oblique RCs prefer to occur with inanimate head nouns and have the IA-A configuration. This prediction is borne out, as shown in Table 6.

Table 6. Counts of the oblique RCs by animacy configurations (A: animate, IA: inanimate)

	Head noun animacy - Embedded noun animacy			
	IA-A	IA-IA	A-A	A-IA
Oblique RC counts	15	1	0	0

All the head nouns were inanimate. 15 oblique RCs had the IA-A configuration. The example in (7) was the only case in which the oblique RC had the IA-IA configuration. Note that the inanimate subject in (7) is interpreted as the animate entity that carries a thematic role typically mapped onto the subject position, i.e., EXPERIENCER. Indeed, the animacy effect on the distribution of the oblique RCs is attributed to the thematic roles of the head nouns. The thematic roles assigned to the inanimate head nouns of the oblique RCs were THEMES (11 tokens), SOURCES (3 tokens) and LOCATIONS (2 tokens) that are commonly realized as complements including direct objects.

16) The experimental study of Kim (2016) with English native speakers revealed that oblique RCs were more difficult to process and comprehend than ORCs. Even if we assume that L2 learners experience the similar difficulty in processing oblique RCs, we should not jump to the conclusion that the significantly lower frequency of the oblique RCs in the corpus supports the frequency-based account. Kim's (2016) study dealt with RC verbs taking two complements - direct object and oblique complement - as exemplified below.

- (i) the book [that the boy put on the carton]
(ii) the carton [that the boy put the book on]

In this case, the difference between the two RCs may be due to the distance between the filler and the gap, as Kim (2016) claimed. But all the oblique RCs found in the corpus do not always require the presence of the direct object between the filler and the gap, as in (1). Therefore, further research should be conducted to examine any difference in processing difficulty between ORCs and oblique RCs.

4. General Discussion and Conclusion

A frequency-based approach emphasizes the role of linguistic experience in accounting for processing difficulty and rate of acquisition, claiming that difficulties in learning and processing certain linguistic forms are partly due to their infrequent and unexpected structures. Previous studies (Roland et al., 2007; Gennari and MacDonald, 2009) have supported frequency-based accounts of the processing difficulty of RCs by demonstrating that the distributional frequencies of RCs found in several English native speaker corpora were consistent with the experimental finding that the processing difficulty of RCs is modulated by the grammatical function of the gap and the animacy configuration of the nouns involved.

This paper investigated whether frequency can also play an explanatory role in RC processing in L2. The experiment study of Baek (2012) showed that Korean adult learners of English processed English RCs in similar ways as native speakers. Given that L2 learners are exposed to the language of English native speakers, the results of L1 English corpus analyses can be taken as evidence in favor for frequency-based accounts of L2 learners' processing difficulties. However, not only input frequencies but also output frequencies are claimed to influence L2 language acquisition and processing. This study analyzed the L2 learner corpus collecting essays of Korean adult learners of English and showed that the distribution of the RCs in the L2 corpus is very similar to that found in previous studies of L1 corpora. That is, SRCs occurred much more frequently than ORCs, and ORCs almost exclusively occurred with inanimate head nouns. The results of the learner corpus analysis provided strong support to the frequency-based approach claiming that a higher frequency of a particular RC form makes it easier for L2 learners to acquire and process.

Frequency is regarded as one of the main factors in language processing, and several different types of information can influence sentence processing. It is claimed that noun animacy can facilitate RC processing because it allows a language user to predict its thematic role and ultimately its grammatical function. The closer examination of the learner corpus revealed that the inanimate head nouns of the ORCs carried thematic roles typically mapped onto the object position. Most ORCs had the IA-A configuration that matches prototypical thematic structures in which the object is a THEME, while the subject is an AGENT or EXPERIENCER. These findings suggest that frequency information interacts with syntactic and semantic information in RC processing.

One limitation of this paper is that the size of the corpus is relatively small, and hence the sampled ORCs are small in number. One might argue that the result of the learner corpus analysis is not statistically adequate to support the frequency-based approach. This problem also arises in previous corpus studies dealing with the semantic effects in RC distribution (e.g., Roland et al. 2007; Gennari and MacDonald, 2009). For example, Roland et al. (2007) examined the head noun animacy effect in RC distribution by coding a random sample of 200 RCs from two large corpora - the Brown and Switchboard corpora. In fact, small corpora are commonly used in quantitative studies of grammatical structures especially when extensive manual annotation is needed for semantic and pragmatic information. The corpus analyzed in this paper is a collection of essays written by many different adult L2 learners for 5 years, and the differences in the distribution of the ORCs with respect to animacy configuration and thematic structure are so significant that they are unlikely to arise by chance. Nevertheless, different types of larger learner corpora including spoken corpora need to be analyzed in future studies in order to exclude any potential bias in the distribution of RCs and provide more reliable evidence for frequency-based approaches.

References

- Baek, S. 2012. *Processing of English Relative Clauses by Adult L2 Learners*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Choi, I., and J. Kim. 2009. The Noun Phrase Accessibility Hierarchy and Second Language Acquisition: Corpus Analyses. *Studies in British and American Language and Literature* 92, 317-335.
- Cuetos, F. and D. C. Mitchell. 1988. Cross-Linguistic Differences in Parsing: Restrictions on the Use of the Late Closure Strategy in Spanish. *Cognition* 30, 73-105.
- Desmet, T., C. De Baecke, D. Drieghe, M. Brysbaert, and W. Vonk. 2006. Relative Clause Attachment in Dutch: Online Comprehension Corresponds to Corpus Frequencies When Lexical Variables are Taken into Account. *Language and Cognitive Processes* 21, 453-485.
- Dowty, D. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67, 547-619.
- Ellis, N. C. 1998. Emergentism, Connectionism, and Language Learning. *Language Learning* 48, 631-664.
- Ellis, N. C. 2002. Frequency Effect on Language Acquisition. *Studies on Second Language Acquisition* 24, 143-188.
- Ellis, N. C. 2003. Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In C. Doughty and M. H. Long (eds.), *Handbook of second language acquisition*. Oxford: Blackwell, 33-68.
- Ellis, N. C. and L. Collins. 2009. Input and Second Language Acquisition: The Roles of Frequency, Form, and Function Introduction to the Special Issue. *Modern Language Journal* 93, 329-335.
- Fillmore, C. J. 1971. Some Problems for Case Grammar. In R. J. O'Brien (ed.), *Report of the 22nd Annual Roundtable Meeting on Linguistics and Language Studies*. Washington, DC: Georgetown University Press, 35-56.
- Frazier, L. and C. Clifton. 1989. Successive Cyclicity in the Grammar and the Parser. *Language and Cognitive Processes*, 4, 93-126.
- Gennari, S. P. and M. C. MacDonal. 2009. Linking Production and Comprehension Processes: The Case of Relative Clauses. *Cognition*, 111, 1-23.
- Gennari, S. P. and M. C. MacDonald. 2008. Semantic Indeterminacy in Object Relative Clauses. *Journal of Memory and Language* 58, 161-187.
- Gibson, E. and C. T. Schütze. 1999. Disambiguation Preferences in Noun phrase Conjunction do not Mirror Corpus Frequency. *Journal of Memory and Language* 40, 263-279.
- Gibson, E., C. Schütze, and A. Salomon. 1996. The Relationship between the Frequency and the Processing of Linguistic Structure. *Journal of Psycholinguistic Research* 25, 59-92.
- Gordon, P. and Lowder, M. 2012. Complex Sentence Processing: A Review of Theoretical Perspectives on the Comprehension of Relative Clauses. *Language and Linguistics Compass*, 2012, 403-415.
- Gordon, P. and R. Hendrick. 2005. Relativization, Ergativity, and Corpus Frequency. *Linguistic Inquiry* 36, 456-463.
- Gordon, P., R. Hendrick, and M. Johnson. 2001. Memory Interference during Language Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, 1411-1423.
- Gordon, P., R. Hendrick, and M. Johnson. 2004. Effects of Noun Phrase Type on Sentence Complexity. *Journal of Memory and Language* 51, 97-114.
- Grimshaw, J. 1990. *Argument Structure*. Cambridge: MIT Press.
- Hamilton, R. L. 1994. Is implicational Generalization Unidirectional and Maximal?: Evidence from Relativization Instruction in a Second Language. *Language Learning* 44, 123-157.

- Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.
- Jackendoff, R. 1990. *Semantic Structures*. Cambridge: MIT Press.
- Jackson, C. and L. Roberts. 2010. Animacy Affects the Processing of Subject-Object Ambiguities in the Second Language: Evidence from Self-Paced Reading with German Second Language Learners of Dutch. *Applied Psycholinguistics* 31, 671-691.
- Just, M. A. and P. A. Carpenter. 1992. A Capacity Theory of Comprehension: Individual Differences in Working Memory Capacity. *Psychological Review* 99, 122-149.
- Keenan, E. L. and B. Comrie. 1997. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8, 63-99.
- Kim, C-E. 2016. Processing Direct Object and Oblique Relative Clauses. *Language Research* 52.2, 151-170.
- King, J. and M. A. Just. 1991. Individual Differences in Syntactic Processing: The Role of Working Memory. *Journal of Memory and Language* 30, 580-602.
- Krashen, S. 1989. We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *Modern Language Journal* 73, 440-464.
- Kuno, S. 1974. The Position of Relative Clauses and Conjunctions. *Linguistic Inquiry* 5, 117-136.
- Lee, J. and K. Y. Shin. 2017. A Syntactic Analysis of English Relative Clauses Used by Korean EFL Speakers: A Corpus-Based Study. *English* 21 30, 219-245.
- MacDonald, M., N. Pearlmutter, and M. Seidenberg. 1994. Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review* 101, 676-703.
- Mak, W., W. Vonk, and H. Schriefers. 2002. The Influence of Animacy on Relative Clause Processing. *Journal of Memory and Language* 47, 50-68.
- Mitchell, D. C., F. Cuetos, M. Corley, and M. Brysbaert. 1995. Exposure-Based Models of Human Parsing: Evidence for the Use of Coarse-Grained (nonlexical) Statistical Records. *Journal of Psycholinguistic Research Special Issue: Sentence Processing* 24, 469-488.
- Omaki, A. and K. Arij, K. 2005. Testing and Attesting the Use of Structural Information in L2 Sentence Processing. In L. Dekydtspotter, R. A. Sprouse and A. Liljestr and (eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference*. Somerville, MA: Cascadilla Proceedings Project, 205-218.
- Pickering, M. J. and M. J. Traxler. 1998. Plausibility and Recovery from Garden paths: An Eye-Tracking Study. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 940-961.
- Roland, D., F. Dick, and J. L. Elman. 2007. Frequency of Basic English Grammatical Structures: A Corpus Analysis. *Journal of Memory and Language* 57, 348-379.
- Schumann, J. H. 1980. The Acquisition of English Relative Clause by Second Language Learners. In Scarcella, R.C. and K. D. Krashen (eds.), *Research in Second Language Acquisition*. Rowley, MA: Newbury House, 118-131.
- Stable, A. 1978. *A Frequency Study of Restrictive Relative Clause types and Relative Pronoun Usage in English*. Unpublished manuscript.
- Sung, H. 2014. A Study on the Use of English Relative Clauses in Korean EFL Learners' Writing. *English Language Teaching* 26, 65-83.
- Swain, M. 1995. Three Functions of Output in Second Language Learning. In G. Cook and B. Seidlhofer (eds.), *Principle and Practice in Applied Linguistics*. Oxford: Oxford University Press, 125-144.
- Swain, M. 1998. The Output Hypothesis, Focus on Form, and Second Language Learning. In V. Berry, B. Adamson, and W. Littlewood (eds.), *Applying Linguistics: Insights into Language in Education*. Hong Kong: English Centre, University of Hong Kong, 1-21.

- Traxler, M. J. and M. J. Pickering. 1996. Plausibility and the Processing of Unbounded Dependencies: An Eye-Tracking Study. *Journal of Memory and Language* 35, 454-475.
- Traxler, M. J., R. Morris, and R. Seely. 2002. Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language* 47, 69-90.
- Traxler, M. J., R. S. Williams, S. A. Blozis, and R. K. Morris. 2005. Working Memory, Animacy, and Verb Class in the Processing of Relative Clauses. *Journal of Memory and Language* 53, 204-224.
- Trueswell, J. C., M. K. Tanenhaus, and S. M. Garnsey. 1994. Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language* 33, 285-318.
- Wong, J. 1991. Learnability of Relative Clauses: A Hong Kong Case. *Perspectives* 3, 108-117.

Keun Young Shin, Professor
77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea
English Language and Literature, Chonnam National University
E-mail: kyshin@jnu.ac.kr