



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학박사학위논문

Effects of AI Chatbot-Based Instruction on  
the Learning of English Adjectival Transitive  
Resultative Construction by Korean High  
School Students

한국인 고등학생의 영어 형용사 타동결과구문  
학습에서의 인공지능 챗봇 기반 교수의 효과

2022년 2월

서울대학교 대학원

외국어교육과 영어전공

김 낙 훈

Effects of AI Chatbot-Based Instruction on  
the Learning of English Adjectival  
Transitive Resultative Construction by  
Korean High School Students

한국인 고등학생의 영어 형용사  
타동결과구문 학습에서의 인공지능  
챗봇 기반 교수의 효과

by

Rakhun Kim

A Dissertation Submitted to  
the Department of Foreign Language Education  
in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy  
in English Language Education

At the

Graduate School of Seoul National University

February 2022

Effects of AI Chatbot-Based Instruction on the  
Learning of English Adjectival Transitive Resultative  
Construction by Korean High School Students

한국인 고등학생의 영어 형용사 타동결과구문  
학습에서의 인공지능 챗봇 기반 교수의 효과

지도교수 김 기 택

이 논문을 교육학 박사 학위논문으로 제출함

2022년 2월

서울대학교 대학원  
외국어교육과 영어전공  
김 낙 훈

김낙훈의 박사학위논문을 인준함

2022년 1월

위 원 장 \_\_\_\_\_  
부위원장 \_\_\_\_\_  
위 원 \_\_\_\_\_  
위 원 \_\_\_\_\_  
위 원 \_\_\_\_\_

Effects of AI Chatbot-Based Instruction on the  
Learning of English Adjectival Transitive Resultative  
Construction by Korean High School Students

by

Rakhun Kim

A Dissertation Submitted to the Department of Foreign Language Education in Partial  
Fulfillment of the Requirements for the Degree of Doctor of Philosophy in English  
Language Education at the Graduate School of Seoul National University

FEBRUARY 2022

APPROVED BY THESIS COMMITTEE:

---

HYUNKEE AHN, COMMITTEE CHAIR

---

HEEJEONG KO

---

HYUNWOO KIM

---

MIN-CHANG SUNG

---

KITAEK KIM

# ABSTRACT

## **Effects of AI Chatbot-based Instruction on the Learning of Adjectival Transitive Resultative Construction by Korean High School Students**

Rakhun Kim

Department of Foreign Language Education (English Major)  
Graduate School of Seoul National University

English adjectival transitive resultative constructions (VtR) are notoriously challenging for Korean L2 English learners due to their syntactic and semantic differences from their L1 counterparts. To deal with such a complex structure, like English adjectival VtR, Korean L2 English learners need instructional interventions, including explicit instructions and corrective feedback on the target structure.

Human instructors are virtually incapable of offering adequate corrective feedback, as providing corrective feedback from a human teacher to hundreds of students requires excessive time and effort. To deal with the practicality problems faced by human instructors in providing corrective feedback, numerous artificial intelligence (AI) chatbots have been developed to provide foreign language learners with corrective feedback on par with human teachers. Regrettably, many currently available AI chatbots remain underdeveloped. In addition, no prior research has been conducted to assess the effectiveness of corrective feedback offered by an AI chatbot, a human instructor, or additional explicit instruction via video material.

The current study examined the instructional effects of corrective feedback from an AI chatbot on Korean high school students' comprehension and production of adjectival VtR. Also, the current study investigated whether the corrective feedback generated by the AI chatbot enables Korean L2 English learners to expand their constructional repertoire beyond instructed adjectival VtR to uninstructed prepositional VtR. To investigate these issues, text-based Facebook Messenger AI chatbots were developed by the researcher.

The effectiveness of the AI chatbots' corrective feedback was compared with that of a human instructor and with additional video material. Students were divided into four groups: three instructional groups and one control group. The instructional groups included a chatbot group, a human group, and a video group. All learners in the three instructional groups watched a 5-minute explicit instruction video on the form and meaning pairings of the adjectival VtR in English. After that, learners were divided into three groups

based on their preferences for instructional types. The learners volunteered to participate in the instructional procedures with corrective feedback from a text-based AI chatbot, a human instructor, or additional explicit instruction using a 15-minute video. Moreover, they took part in three testing sessions, which included a pretest, an immediate posttest, and a delayed posttest. The control group students were not instructed, and only participated in the three testing sessions.

Two tasks were used for each test session: an acceptability judgment task (AJT) and an elicited writing task (EWT). The AJT tested participants' comprehension of instructed adjectival VtR and uninstructed prepositional VtR. The EWT examined the correct production of instructed adjectival VtR and uninstructed prepositional VtR.

The results of the AJT revealed that the instructional treatment (e.g., corrective feedback from the AI chatbot or a human instructor, or additional explicit instruction from the video material) was marginally more effective at improving the comprehension of adjectival VtR than was the case with the control group. On the other hand, the instructional treatment on the adjectival VtR failed in the generalization to prepositional VtR which was not overtly instructed. In the EWT, the participants in the corrective feedback groups (e.g., the chatbot and human groups) showed a more significant increase in the correct production of the instructed adjectival VtR more so than those in the video and control groups. Furthermore, the chatbot group learners showed significantly higher production of uninstructed prepositional VtR compared to any other group participants.

These findings suggest that chatbot-based instruction can help Korean high school L2 English learners comprehend and produce complex linguistic structures—namely, adjectival and prepositional VtR. Moreover, the current study has major pedagogical implications for principled frameworks for implementing AI chatbot-based instruction in the context of foreign language learning.

**Key Words:** English transitive resultative construction, AI chatbot for foreign language learning, corrective feedback, foreign language instruction, English syntax, Korean syntax

**Student Number:** 2013-30426

# TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>i</b>
<b>TABLE OF CONTENTS .....</b>	<b>iii</b>
<b>LIST OF TABLES.....</b>	<b>v</b>
<b>LIST OF FIGURES.....</b>	<b>vii</b>
<b>CHAPTER 1. INTRODUCTION.....</b>	<b>1</b>
1.1. Statement of Problems and Objectives .....	1
1.2. Scope of the Research.....	6
1.3. Research Questions.....	9
1.4. Organization of the Dissertation .....	10
<b>CHAPTER 2. LITERATURE REVIEW .....</b>	<b>12</b>
2.1. Syntactic and Semantic Analysis of Korean and English Transitive Resultative Constructions.....	13
2.1.1. Syntactic Analysis of English Transitive Resultative Construction .....	13
2.1.2. Syntactic Analysis of Korean Transitive Resultative Constructions.....	25
2.1.3. Semantic Differences in VtR between Korean and English.....	46
2.1.4. Previous acquisition study on English adjectival and prepositional VtR.....	54
2.2. Corrective Feedback .....	59
2.2.1. Definition of Corrective Feedback .....	59
2.2.2. Types of Corrective Feedback.....	61
2.2.3. Noticeability in Corrective Feedback.....	67
2.2.4. Corrective Recast as a Stepwise Corrective Feedback.....	69
2.3. The AI Chatbot in Foreign Language Learning .....	72
2.3.1. Non-communicative Intelligent Computer Assisted Language Learning (ICALL).....	73
2.3.2. AI Chatbot without Corrective Feedback.....	79
2.3.3. AI Chatbot with Corrective Feedback.....	86
2.4. Summary of the Literature Review.....	92
<b>CHAPTER 3. METHODOLOGY .....</b>	<b>98</b>
3.1. Participants .....	98
3.2. Target Structure .....	102
3.3. Procedure of the Study .....	106
3.4. Instructional Material Shared by the Experimental Group .....	107
3.4.1. General Framework of the Instructional Session .....	108
3.4.2. Instructional Material Shared by Experimental Groups.....	111

3.5. Group-specific Instructional Treatments: Post-Written Instructional Material Activities on Corrective Feedback from Chatbot, Human, and Additional Explicit Instruction via Video .....	121
3.5.1. Corrective Feedback from the AI Chatbot .....	122
3.5.2. Corrective Feedback from a Human Instructor .....	136
3.5.3. Additional Instruction via Video Material .....	139
3.6. Test .....	142
3.6.1. Acceptability Judgment Task (AJT) .....	144
3.6.2. Elicited Writing Task (EWT) .....	150
3.7. Statistical Analysis .....	152
<b>CHAPTER 4. RESULTS AND DISCUSSIONS .....</b>	<b>154</b>
4.1. Results of Acceptability Judgment Task (AJT) .....	154
4.1.1. AJT Results of Instructed Adjectival VtR .....	155
4.1.2. AJT Results of Uninstructed Prepositional VtR .....	160
4.1.3. Discussion .....	164
4.2. Results of Elicited Writing Task (EWT) .....	175
4.2.1. EWT Results for Instructed Adjectival VtR .....	176
4.2.2. EWT Results of Uninstructed Prepositional VtR .....	181
4.2.3. Further Analysis .....	187
4.2.4. Discussion .....	199
<b>CHAPTER 5. CONCLUSION .....</b>	<b>205</b>
6.1. Summary of the Findings and Implications .....	205
6.2. Limitations and Suggestions for Future Research .....	213
<b>REFERENCES .....</b>	<b>217</b>
<b>APPENDICES.....</b>	<b>246</b>
<b>ABSTRACT IN KOREAN .....</b>	<b>297</b>

# LIST OF TABLES

Table 2.1 Verbless Small Clauses by L1 English Children .....	15
Table 2.2 Summary of Syntactic and Semantic Contrasts between English and Korean Transitive Resultative Construction .....	53
Table 2.3 Corrective Feedback Techniques (Lyster & Ranta, 1997) .....	62
Table 3.1 List of Instructional Items for Adjectival VtR .....	103
Table 3.2 Procedures of Instruction and Testing .....	106
Table 3.3 General Framework of Instructional Video and Written Instructional Material Activity .....	109
Table 3.4 Example of Interaction between Learner and AI Chatbot (Elizabeth Ria) .....	130
Table 3.5 Basic Specification of AI Chatbot for Nylon Musk and Elizabeth Ria .....	133
Table 3.6 Sentence Types Used in the Acceptability Judgment Task .....	146
Table 3.7 Linguistic Features of Syntactically and Semantically Ungrammatical Sentences .....	148
Table 3.8 Distribution of Experimental Items in EWT .....	150
Table 4.1 Mean AJT Score of Instructed Adjectival VtR .....	156
Table 4.2 A Repeated Measure ANOVA of AJT on Instructed Adjectival VtR .....	158
Table 4.3 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Immediate Posttest .....	159
Table 4.4 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Delayed Posttest .....	160
Table 4.5 Mean AJT Score of Uninstructed Prepositional VtR .....	161
Table 4.6 A Repeated Measure ANOVA of the AJT on the Comprehension of Uninstructed Prepositional VtR .....	163
Table 4.7 Revised Example of Interaction Between Learner and AI Chatbot (Elizabeth Ria) .....	171
Table 4.8 EWT Results of ANOVA for the Between-tests .....	176
Table 4.9 A Repeated Measure ANOVA of the EWT: Instructed Adjectival VtR .....	178
Table 4.10 Multiple Comparisons of the Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Immediate Posttest .....	180
Table 4.11 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Delayed Posttest .....	181
Table 4.12 EWT Results of ANOVA for the Between-tests: Uninstructed Prepositional VtR .....	183
Table 4.13 Repeated Measure ANOVA of EWT: Uninstructed Prepositional VtR .....	184

Table 4.14 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for the Uninstructed Prepositional EWT: Immediate Posttest .....	186
Table 4.15 Multiple Comparisons of the Chatbot, Human, Video, and Control Groups for Uninstructed Prepositional EWT: Delayed Posttest .....	187
Table 4.16 Frequency of Alternative Syntactic Patterns in EWT .....	294

## LIST OF FIGURES

Figure 2.1. Example of Sasha (Petersen, 2010) .....	88
Figure 3.1. Syntactic and semantic interrelatedness among adjectival VtR, F-VPC, prepositional VtR, and F-VPC .....	105
Figure 3.2. Example taken from the instructional video .....	113
Figure 3.3. Controlled practice (CP) step 1: Guessing the meaning of a figurative particle .....	116
Figure 3.4. Controlled practice (CP) step 2 : Realization of external argument and DOR.....	117
Figure 3.5. Controlled practice (CP) step 3: Generating form and meaning pairing for adjectival VtR.....	118
Figure 3.6. Example of the dialogue completion task (DCT) .....	120
Figure 3.7. Dashboard of Chatfuel .....	123
Figure 3.8. System Overview .....	125
Figure 3.9. Example of corrective feedback from an AI chatbot, Nylon Musk .....	128
Figure 3.10. Example of corrective feedback from the AI chatbot, Elizaveth Ria .....	132
Figure 3.11. Example of Corrective Feedback (elicitation) from a human instructor during the controlled practice (CP) task .....	137
Figure 3.12. Screenshot of additional explicit Instruction via Video Material: No Corrective Feedback .....	140
Figure 3.13. Example of an Acceptability Judgment Task .....	145
Figure 3.14. Example of an Elicited Writing Task .....	151
Figure 4.1. Mean AJT scores of instructed adjectival VtR (error bars represent standard error) .....	157
Figure 4.2. Mean AJT scores of uninstructed prepositional VtR (error bars represent standard error.) .....	163
Figure 4.3. Interaction plot of EWT for Instructed Adjectival VtR .....	177
Figure 4.4. Interaction plot of EWT for uninstructed prepositional VtR .....	182
Figure 4.5. Correct answer rate for the unselected adjectival VtR using <i>run</i> (error bars represent standard errors) .....	189
Figure 4.6. Correct answer rate for the unselected adjectival VtR with a reflexive pronoun in the direct object position (error bars represent standard errors) .....	190
Figure 4.7. Correct answer rate for an unselected prepositional VtR with four relevant test items (error bars represent standard errors) .....	193

# CHAPTER 1. INTRODUCTION

The present study investigates the instructional effects of text-based artificial intelligence chatbots on Korean high school students' learning of English transitive resultative constructions (VtR). The first section introduces the problems and objectives addressed in this dissertation. The scope of the study and the research questions are then presented. This chapter closes with an overview of the organization of the dissertation.

## 1.1. Statement of Problems and Objectives

One of the primary aims of foreign language learning is to attain communicative competence. As widely attested, achieving this requires sufficient target language input. In particular, the input should be 'noticed' to the learners and thus be transformed to 'intake' (Schmidt, 1990).

One of the pedagogical interventions to make the input noticeable for foreign language learners is to provide corrective feedback, one of the linguistic scaffoldings (Long, 1996). That is, corrective feedback helps input to be 'noticed' and transformed to 'intake'. Notably, corrective feedback should be individualized for language learners, since the comprehensibility

in language input differs for each foreign language learner due to their distinct interlanguage developmental stages (Sheen, 2004, 2006). In this respect, individualized corrective feedbacks, which this study employed through AI-chatbot, positively affects foreign language learning.

Unfortunately, it is practically impossible for human instructors to provide foreign language learners with individualized corrective feedback that is specifically adapted to their interlanguage developmental stage. Several studies have demonstrated that providing students with corrective feedback requires a tremendous amount of manual labor (Ferris, 2004; Guenette, 2007; Lee, 1997; Semke, 1984; Truscott, 1999; Valero et al., 2008; Zamel, 1985). To account for this practicality issue, artificial intelligence (AI) has been introduced into foreign language learning contexts (Department for Education, 2019; Hong, 2020; Ma et al., 2014).

The primary objective of AI programs dedicated specifically to foreign language education is not to replace human teachers, but rather to empower them with a pedagogical resource to help foreign language learners attain communicative competence in L2 English alongside their human counterparts (R. Kim, 2018). This supplementary pedagogical resource is designed to provide learners with an optimal type of corrective feedback in a natural conversational context (Ai, 2015; Ayedoun et al., 2015; Baur et al., 2015; Beal et al., 2005; Bouillon et al., 2011; Chiu et al., 2007; Harless et al.,

1999; Hassani et al., 2016; Heift, 2004; Heift & Rimrott, 2008; Johnson, 2007; Kaplan et al., 1998; N. Kim, 2016; K. Lee et al., 2011; S. Lee et al., 2011; S. Lee et al., 2012; Noh et al., 2012; Petersen, 2010; Su et al., 2014; Vlugter et al., 2009; Walker et al., 2011; N. Wang et al., 2008; Wilske & Wolska, 2011).

It is expected that the use of AI programs for foreign language learning may have been as effective as human instructors in the learning of core linguistic properties for L2 English learners. Studies have shown that one of the most difficult argument structures for Korean L2 English learners is the adjectival transitive resultative construction (VtR) (e.g., “Herman hammered the metal *flat*.”). This difficulty has been ascribed to a syntactic and semantic mismatch between Korean and English (Bowers, 1993, 2002; Embick, 2004; Goldberg, 1995; Hong, 2005; Kearns, 2007; Kim & Maling 1997; S. Kim et.al., 2020; Ko 2011, 2014, 2015; Larson, 1988, 1990; Radford, 2009; Shim & den Dikken 2007; Simpson, 1983; Talmy, 1985; Zubizarreta & Oh, 2007). Also, this construction is one of the most productive argument structure constructions and is extendable to a number of interrelated complex argument structure constructions (e.g., prepositional VtR: “The critics laughed the play *off the stage*.”; Goldberg, 1995; Goldberg & Jackendoff, 2004).

In this respect, under the framework of intelligent computer-assisted

language learning (ICALL), the significance of these communicative AI programs, or AI chatbots, lies in that they lead to the customized learning of core English argument structures, namely, VtR. By using AI chatbots for L2 English learning, learners are able to self-correct erroneous utterances with comprehensible corrective feedback from the AI chatbots customized to their interlanguage development stage (Heift, 2004; Heift & Rimrott, 2008; Kerly et al., 2006; Petersen, 2010; Ruan et al., 2021; Wang & Petrina, 2013; Wilske & Wolska, 2011).

Despite its significance in L2 English learning, previous studies have generally failed to investigate the instructional effects of AI chatbots on Korean L2 English learners' comprehension and production of English adjectival VtR. Furthermore, few attempts have been made to examine whether the corrective feedback of adjectival VtR from AI chatbots enables Korean L2 English learners to generalize their knowledge to other interrelated VtR constructions—namely, prepositional VtR. Moreover, no study has yet compared the instructional effectiveness of AI chatbots with that of a human instructor or additional explicit instruction for Korean L2 English learners in the learning of adjectival VtR.

This study addresses these problems by integrating prior research on linguistics, pedagogy, technology, and second language acquisition (SLA; Doughty & Varela, 1998; N. Ellis & Boagart, 2007; S. Kim et al., 2020;

Petersen, 2010; Shim & den Dikken, 2007; H. Sung, 2018; Zubizarreta & Oh, 2007) to construct an AI chatbot for L2 English learning and to demonstrate its effectiveness in comparison to other instructional treatments. Moreover, this study will demonstrate that the corrective feedback from an AI chatbot of a particular construction (e.g., the adjectival VtR) can lead to the generalization to an uninstructed but linguistically inter-related argument structure construction (e.g., prepositional VtR).

Another noteworthy aspect of the present study is that, in terms of the comprehension and production of English VtR, the instructional effects of the corrective feedback generated by the AI chatbot, human instructor, and additional explicit instruction delivered via video material is compared. This approach will provide important evidence that the instructional effect of an AI chatbot's corrective feedback is as effective as human-generated corrective feedback or additional explicit instruction. Finally, this study will suggest the possibility that an AI chatbot may complement the role of a human instructor in foreign language learning contexts.

## 1.2. Scope of the Research

English argument structure constructions (ASCs) are considered a linguistic unit that is used to encode recurrent event types (e.g., the act of giving or receiving, events in which something changes) that are fundamental to everyday human experience (Goldberg, 1995; Hilpert, 2014). ASCs are crucial for Korean L2 English learners, because they provide the linguistic foundation to produce sentence-level utterances without memorization (Yang et al., 2014). ASCs correspond to lemma which is fundamental to producing sentence-level utterances in Levelt's language production model (Barbieri et al., 2010; Bock & Levelt, 1994; Levelt et al., 1999).

Among the various English ASCs, the English transitive resultative construction (VtR) is a hallmark construction that enables Korean language learners to produce sentence-level utterances. One identifying feature of a VtR expression is an adjectival phrase (AP) or prepositional phrase (PP) that occupies the complement position of a verbal argument, such as the italicized parts of (1) and (2). This AP or PP is referred to as result phrase (RP; Hale & Keyser, 2002). In the current study, English formulations with S-V-NP-AP are referred to as adjectival VtR, while those with S-V-NP-PP are called prepositional VtR (Goldberg & Jackendoff, 2004). In this study,

the prepositional VtR will be used as a synonym for a caused-motion construction (Goldberg, 1995).

(1) Herman hammered the metal *flat*. (RP = AP)

(2) The critics laughed the play *off the stage*. (RP = PP)

Figurative and literal VtR share their syntactic and semantic properties with aspectual and lexical verb particle construction (VPC) in (3) and (4), respectively, which constitutes an instance or a subcategory of the transitive resultative constructions (Goldberg, 1995; Gorfach, 2004).

(3) Lee knocked him *out*. [Figurative VPC (F-VPC)]

(Particle *out* = Aspectual with the meaning of disappear)

(4) Juliet stucked her tongue *out*. [Literal VPC (L-VPC)]

(Particle *out* = Literal with the meaning of outside)

Several syntactic categories (e.g., AP, PP, and NP) may appear in the post-verbal result phrase (Carrier & Randall, 1992; Napoli, 1992; M. Sung, 2018).

However, this study focuses only on English VtR co-occurring with AP and PP, as these two constructions—namely, adjectival and prepositional VtR—are the primary source of linguistic variation between Korean and English (Talmy, 1985).

According to several studies (e.g., Diessel & Tomasello, 2005; Goldberg, 1995; Gorfach, 2004; Thim, 2012), prepositional VtR includes a transitive lexical verb particle construction (VPC), whereas adjectival VtR, which is a target construction for instruction, takes a transitive aspectual VPC (e.g., “The teacher woke him *up*.”) as an instance of the adjectival transitive resultative construction (M. Sung, 2018).

### 1.3. Research Questions

The current study is guided by the following research questions:

(1) Is chatbot-based instruction as effective as human-based instruction (e.g., corrective feedback from a human instructor or additional explicit instruction of video material) in improving Korean high school English learners' comprehension of English transitive resultative constructions?

(1a) Is chatbot-based instruction as effective as human-based instruction in improving the Korean high school English learners' comprehension of instructed English adjectival transitive resultative construction?

(1b) Is chatbot-based instruction as effective as human-based instruction in improving the Korean high school English learners' comprehension of uninstructed English prepositional transitive resultative construction?

(2) Is chatbot-based instruction as effective as human-based instruction (e.g., corrective feedback from a human instructor or additional explicit instruction of video material) in improving Korean high school English learners' production of English transitive resultative constructions?

(2a) Is chatbot-based instruction as effective as human-based instruction in improving Korean high school English learners' production of instructed English adjectival transitive resultative constructions?

(2b) Is chatbot-based instruction as effective as human-based instruction in improving Korean high school English learners' production of uninstructed English prepositional transitive resultative constructions?

#### **1.4. Organization of the Dissertation**

This dissertation is organized into five chapters. Chapter 1 introduces the motivation and research questions of the present study. Chapter 2 reviews the theoretical framework of the transitive resultative construction and examines theoretical and experimental issues surrounding the target construction (i.e., the transitive resultative construction). This chapter also explores the technological background of artificial intelligence chatbots for language learning and their practical application in the foreign language learning context. Chapter 3 describes the research method of the main experiment, including participants, target structures, instruction, tests, and data analysis. Chapter 4 reports on the quantitative and qualitative results of the study, highlighting significant between-group differences, and discusses

the findings of the study and crucial issues in the learning of transitive resultative construction with an artificial intelligent (AI) chatbot, providing answers to the study's research questions. Chapter 5 summarizes the major findings and concludes the study with pedagogical implications, limitations, and suggestions for future research.

## **Chapter 2. Literature Review**

The objective of this research is to investigate the usefulness of corrective feedback from an artificial intelligence chatbot in supporting Korean L2 English learners' acquisition of adjectival VtR. The first section of this chapter is devoted to the linguistic analysis of the cross-linguistic variation in the syntactic and semantic features of English and Korean VtR expressions, with a particular emphasis on their adjectival and prepositional sub-constructions. In addition, the key findings of foreign language research on English VtR acquisition are presented. Section 2.2 deals with the theoretical foundations and practical applications of corrective feedback in foreign language learning environments. The following section covers the technological foundations of artificial intelligence chatbots and their pedagogical applications in the context of intelligent computer-assisted language learning (ICALL). The concluding section summarizes the literature review and presents the predictions of the present study.

## **2.1. Syntactic and Semantic Analysis of Korean and English Transitive Resultative Constructions**

This section examines the syntactic and semantic nature of Korean and English transitive resultative constructions (VtR). Specifically, this section demonstrates the cross-linguistic variation between Korean and English in the syntactic formulation of result-denoting and manner-denoting components of VtR expressions, which contributes to the severe challenges Korean L2 English learners face when learning English VtR.

### **2.1.1. Syntactic Analysis of English Transitive Resultative Construction**

English VtR has been syntactically analyzed to have (i) a result-denoting component and (ii) a manner-denoting component. As Talmy (1985) explained, English adjectival VtR is syntactically correlated with its prepositional counterpart. This section will investigate such linguistic ubiquity between adjectival and prepositional VtR formulations with a special emphasis on the linguistic configuration of result- and manner-

denoting components and their syntactic relationship.

Resultative constructions are an intensively researched area of English grammar, with many studies focusing on the distinction between depictives and resultatives (Huddleston & Pullum, 2002; Quirk et al., 1985; Simpson, 1983), as shown in (1).

- (1) a. John *hammered* the metal *flat*. [Result Phrase (RP)]
- b. John *ate* the meat *raw*. [Depictive Phrase (DP)]
- c. \*John *hammered* the metal *flatly*. [Resultative – adverbial]

Although the constructional formulations in (1a) and (1b) seem to have the same surface structure (VP + NP + AP), they have different syntactic properties, where the result phrase (RP) is a complement of the verb in (1a) and the depictive phrase (DP) in (1b) is an adjunct (C. Lee, 1996). Moreover, (1b) semantically differs from (1a) in that (1b) does not denote end-state results that depend on the action expressed by the main verb. Accordingly, as can be seen in (1c), replacing an end-state result denoting an RP-adjective (e.g., *flat*) with an adjunct adverbial expression (e.g., *flatly*) leads to severe ungrammaticality.

Such an end-state result (or, caused-event semantics) is represented syntactically by a small clause, consisting of the adjectival or prepositional expression predicated of the NP object (Bowers, 1993, 2002; see Table 2.1). According to several English L1 studies (Radford, 1988; Tomasello, 1992), the English small clause is a two-word predicate in which the earliest clausal structure is expressed by L1 English children. As Table 2.1 demonstrates, the English small clause is usually realized as an [NP AdjP] or [NP PP] schema (Chomsky, 1981; Kayne, 1985; Stowell, 1981). A small clause is a propositional constituent that does not contain a verb (Matushansky, 2019).

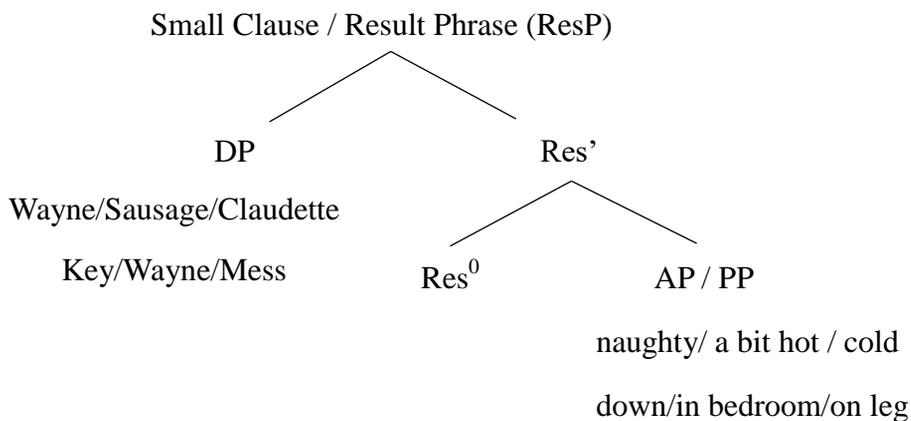
**Table 2.1 Verbless Small Clauses by L1 English Children**

<b>[NP AdjP] Schema</b>	<b>[NP PP] Schema</b>
(i) Wayne naughty	(i) Key down
(ii) Sausage a bit hot	(ii) Wayne in bedroom
(iii) Claudette cold	(iii) Mess on legs

The internal structure of an English small clause consists of the subject of the small clause (e.g., Wayne, Sausage, Claudette, Key, Wayne, Mess) and its predicate. As (2) illustrates, the internal structure of the English small clause is mediated by a result-head ( $\text{Res}^0$ ) whose semantic property denotes

a transition to a result or a place (Folli & Harley, 2020).<sup>1</sup> Res<sup>0</sup> takes an adjectival phrase (AP: naughty, a bit hot, cold) or prepositional phrase (PP: down, in bedroom, on leg) as its complement.

## (2) Internal Structure of English Small Clause



The English small clause provides a syntactic scaffold for more complex sentence structures. As (3) shows, when formulating an English transitive resultative construction, the small clause is selected by a verb denoting a causative head (e.g.,  $v_{[CAUSE]}$ ) to express a transition to a result or a location

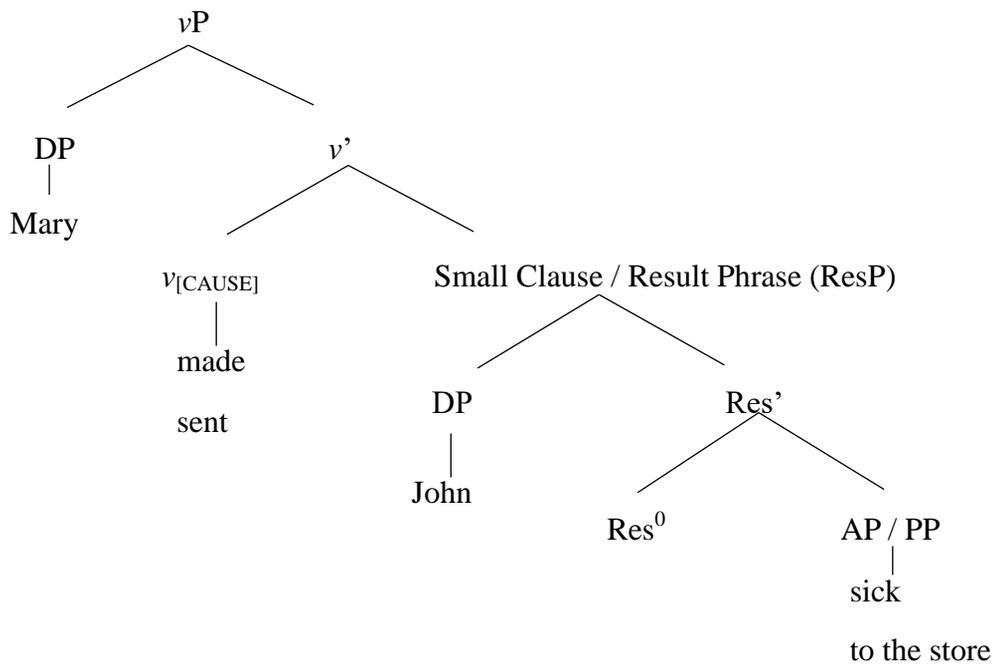
---

<sup>1</sup> There has been intense discussion about the internal structure of a small clause (SC) and its syntactic status as a lexical or functional phrase. In this dissertation, I will put these technical issues aside. Instead, following Harley (2005), the SC will be treated as the syntactic category of a result phrase (ResP; Ramchand, 2001; Ramchand & Svenius, 2002). See Matushansky (2019) for a relevant discussion of the linguistic status of SC.

syntactically realized as an AP or PP, respectively.<sup>2</sup> The English transitive resultative construction (VtR) is realized in the following structure in (3), in which the small clause is directly selected by the causative verb (Folli & Harley, 2020; Harley, 2005).

(3) Structure of English Transitive Resultative Construction with a CAUSE Verb Embedded by a Small Clause

- a. Mary made John sick.
- b. Mary sent John to the store.



<sup>2</sup> There are at least three semantic types of *v* or 'light verb': BECOME, DO, and CAUSE. According to Folli and Harley (2004), the causative verbal head selects SC as its syntactic complement, while others do not. See another discussion of the syntactic nature or the strength/weakness of *v* or light verb in Chomsky (2001).

In addition, as demonstrated in (4), in order to express the manner of transition to a result or location, the  $v_{[CAUSE]}$  is modified by another verbal root whose meaning in itself has no causal component. The interpretation of the adjoined verbal root ( $\sqrt{RUN}$ ,  $\sqrt{FLOAT}$ ) as a manner adverbial is syntactically realized in the structure of (4). The sister-to- $v'$  position represented in the tree structure is generally the classic locus for adverbial modification to denote manner semantics (e.g., by-Manner; Carnie, 2012; Levinson, 2007; Pylkkänen, 2002; Ramchand, 2008). In English, this manner denoting adverbial verbal root (e.g.,  $\sqrt{RUN}$ ,  $\sqrt{FLOAT}$ ) on the sister-to- $v'$  position is lowered to the head of the  $v_{[CAUSE]}$  and is incorporated into the matrix verb (e.g., *ran*, *floated*). This syntactic operation is referred to as ‘manner incorporation’ (e.g., Folli & Harley, 2004, 2006; Harley 2005; McIntyre 2004). In this syntactic construal, the manner-denoting element is realized as a matrix verb, and the result-denoting component is realized as a small clause or ResP.

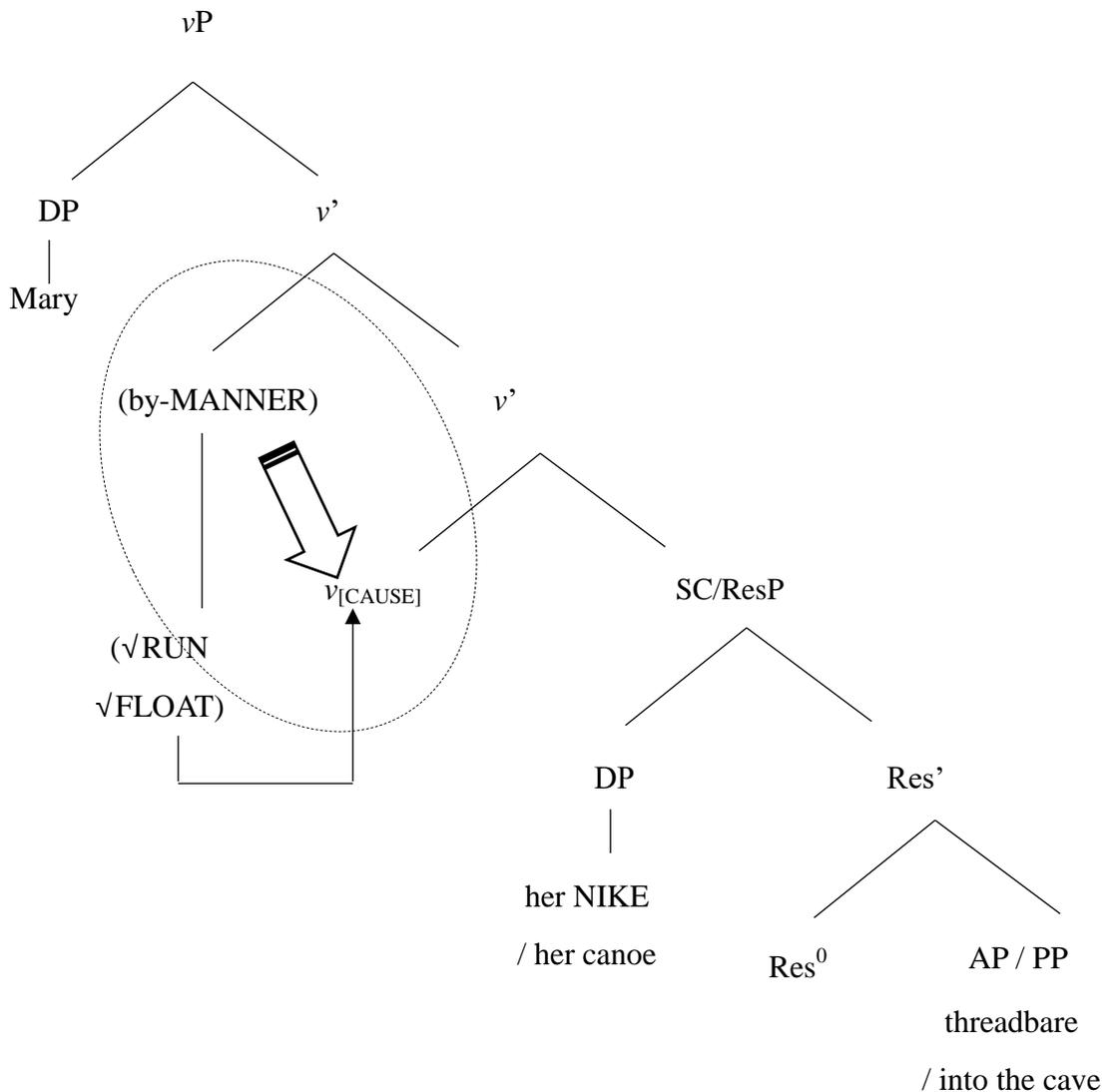
(4) English VtR Embedded by a Manner Verb: Manner Incorporation

- a. [<sub>vP</sub> Mary [<sub>v'</sub> (by RUNNING) [<sub>v'</sub> CAUSE [<sub>SC</sub> her NIKE threadbare]]]].

Mary ran her Nike threadbare.

- b. [<sub>vP</sub> Mary [<sub>v'</sub> (by FLOATING) [<sub>v'</sub> CAUSE [<sub>SC</sub> her canoe into the cave]]]].

Mary floated her canoe into the cave.



The resultative construction includes a wide variety of syntactic subtypes. Despite these variations, numerous previous studies have by and large treated resultatives as a relatively unified phenomenon (Goldberg, 1995; Goldberg & Jackendoff, 2004; Jackendoff, 1990). One identifying characteristic of a transitive resultative construction is that it must adopt an accusative-marked theme argument as the postverbal NP element (or as a subject of the small clause), followed by either an adjectival phrase (AP) or a prepositional phrase (PP). This study refers to the former as an adjectival transitive resultative construction (adjectival VtR) and the latter as a prepositional transitive resultative construction (prepositional VtR).

(5) a. Herman hammered the metal *flat*.

(RP = AP; Adjectival VtR)

b. The critics laughed the play *off the stage*.

(RP = PP; Prepositional VtR)

According to M. Sung (2018), adjectival and prepositional VtR parallels figurative (or aspectual) and literal verb particle construction (VPC) in (6), respectively.

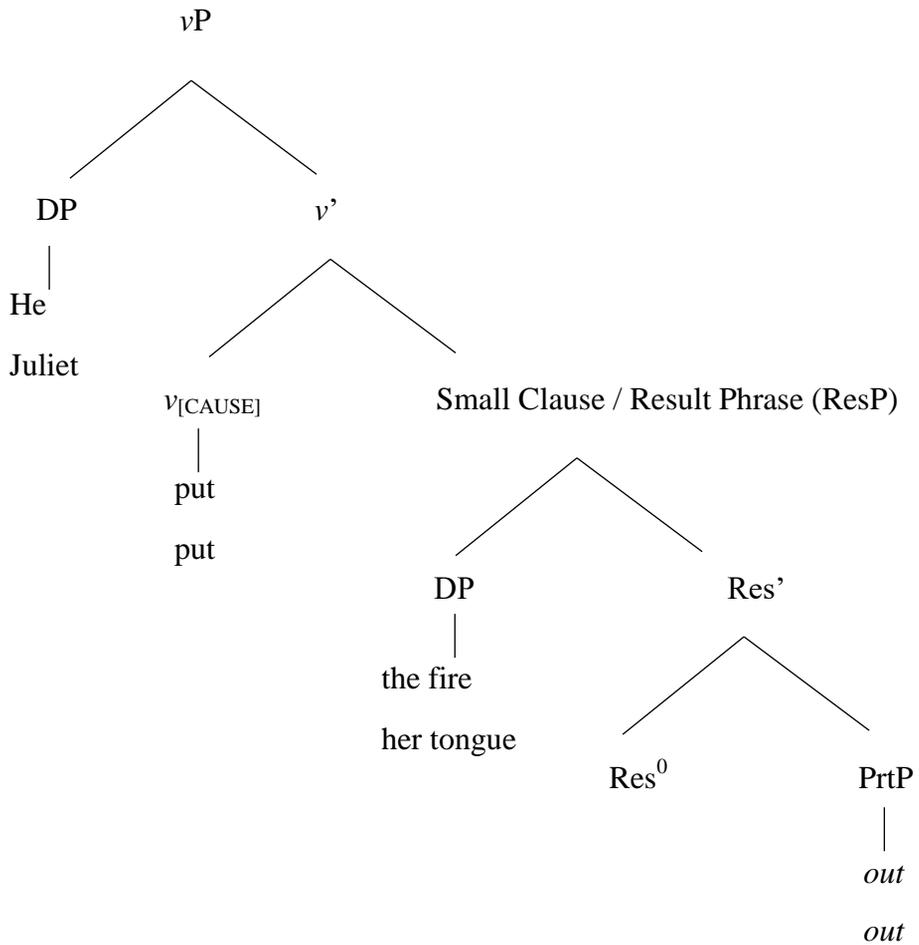
(6) a. He put the fire *out*.

(RP = Figurative Particle)

b. Juliet put her tongue *out*.

(RP = Literal Particle)

c.



Many studies (Side, 1990; M. Sung, 2018) have concluded that figurative VPC is distinct from its literal counterpart in that the former denotes less transparent or more idiomatic information (e.g., the figurative meaning of “out” for “displacement” in (6a)), whereas the latter denotes the transparent meaning of the particle (e.g., literal meaning of “out” for “into open, away from” in (6b)). The relevant syntactic structure is rendered in (6c).<sup>3</sup>

In some VtR formulations, the verb selects the object NP independently; in others, it does not. The former is referred to as *selected* transitive resultatives (Example (7)), whereas the latter is referred to as *unselected* transitive resultatives (Example (8)).

(7) Selected transitive resultatives

a. The gardener watered the flowers flat. (RP = AP)

(cf. The gardener watered the flowers.)

b. Bill broke the bathtub into pieces. (RP = PP)

(cf. Bill broke the bathtub.)

---

<sup>3</sup> Transitive VPC has been analyzed as occurring in two types: V+OBJ+Prt and V+PRT+OBJ (Gorlach, 2004). In this dissertation, only the former will be treated as a subtype of VtR, as Prt remains at SC to denote resultative meaning. In this vein, V+OBJ+Prt (e.g., John opened the window *up*) is distinct from V+OBJ (e.g., John opened the window) in that the resultative meaning of the former remains at the particle level in the SC, while that of the latter is attached to the matrix verb. See more discussion in Folli and Harley (2020) or in footnote 10 of this dissertation.

(8) Unselected transitive resultatives with transitive verb

a. They drank the pub dry. (RP = AP)

(cf. \*They drank the pub.)

(*paraphrase*: They made the pub dry by drinking.)

b. The professor talked us into a stupor. (RP = PP)

(cf. \*The professor talked us.)

(*paraphrase*: The professor drove us into a stupor by talking.)

For example, in (7a) and (7b), the object NPs (i.e., *the flowers* in (7a) and *the bathtub* in (7b)) are selected by the verbs (i.e., *water* in (7a) and *broke* in (7b)). In (8a-b), however, the object NPs (i.e., *the pub* in (8a) and *us* in (8b)) are not selected by the verbs (i.e., *drank* in (8a) and *talked* in (8b)).

As the example sentences in (9) demonstrate, an unselected transitive resultative construction is paired with an intransitive verb that normally does not take an NP as its direct object.

(9) Unselected transitive resultatives with intransitive verb

a. The dog barked him awake. (RP = AP) (Rappaport & Levin, 2001)

(cf. \*The dog barked him.)

(*paraphrase*: The dog made him awake by barking.)

b. He sneezed the paper off the table (RP = PP)

(cf. \*He sneezed the paper.)

(*paraphrase*: He dropped the paper off the table by sneezing.)

The sentences in (9a) and (9b) are ungrammatical without the result phrase (i.e., *awake* in (9a) and *off the table* in (9b)), since the verbs in (9a) and (9b) are intransitive (i.e., *barked* in (9a) and *sneezed* in (9b)), which cannot assign an accusative case to its following NPs.

A special case of unselected transitive resultatives is the fake reflexive provided in (10) (Simpson, 1983). It has a reflexive object that cannot alternate with other NPs. Specifically, replacing the reflexive pronouns in (10a) and (10b) with the proper noun (i.e., *Harry* in (10a) and *us* in (10b)) leads to this instance of ungrammaticality.

(10) Unselected Fake reflexive resultatives

a. We yelled ourselves hoarse. (RP = AP)

(\*We yelled ourselves. / \*We yelled *Harry* hoarse.)

(*paraphrase*: We made ourselves hoarse by yelling)

b. Harry coughed himself into insensibility. (RP = PP)

(\*Harry coughed himself. / \*Harry coughed us into insensibility.)

(*paraphrase*: Harry made himself into insensibility by coughing)

To summarize, in formulating English adjectival VtR expressions, the result-denoting element is syntactically realized as a small clause while the

manner-denoting component is realized in the matrix verb via manner incorporation (Folli & Harley, 2020; Harley, 2005). English adopts a mono-clausal strategy to syntactically encode various subtypes, such as (i) adjectival or prepositional VtR, (ii) selected or unselected VtR, and (iii) unselected VtR with an intransitive verb or a reflexive pronoun.

### **2.1.2. Syntactic Analysis of Korean Transitive Resultative Constructions**

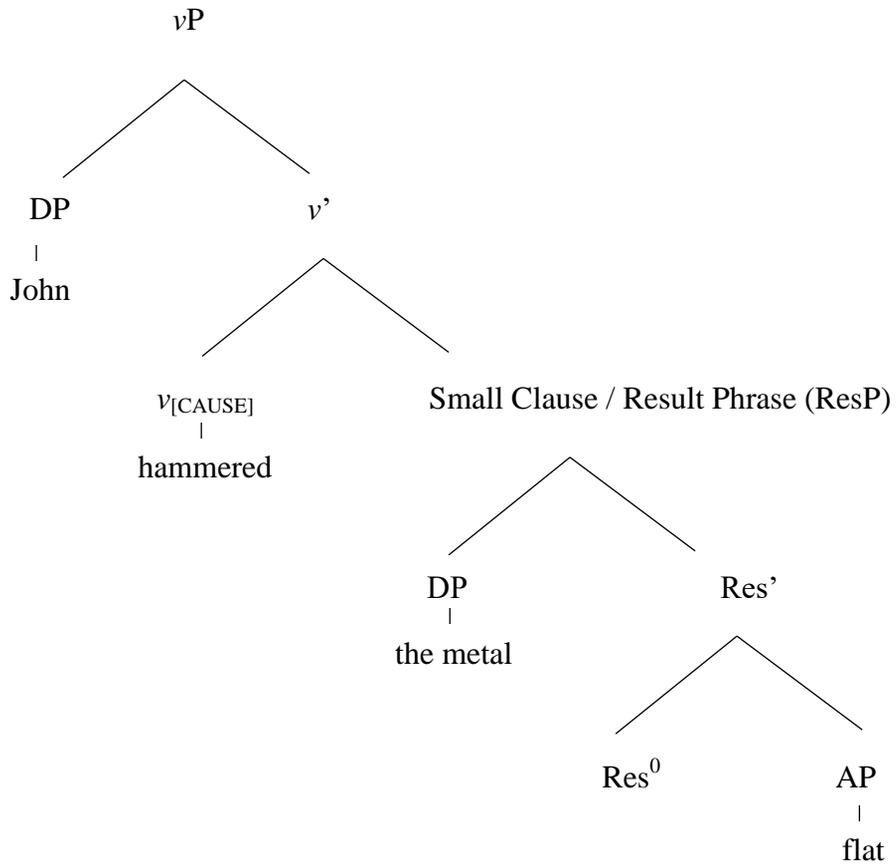
In the previous section we reviewed how English adjectival and prepositional transitive resultative construction syntactically shares a manner-denoting and result-denoting component in a mono-clausal formulation via manner incorporation. However, unlike English, Korean adjectival and prepositional VtR have their own unique syntactic construction, which is deficient in the syntactic ubiquity between the two constructions. This section discusses the syntactic analysis of Korean adjectival VtR in Section 2.1.2.1, and prepositional VtR in Section 2.1.2.2.

### **2.1.2.1. Korean Adjectival Transitive Construction**

Most previous studies converge on the complementhood of the small clause in English VtR (Bowers 1993, 2002; Carrier & Randall, 1992; Embick, 2004; Larson, 1988, 1990; Radford, 1988; Simpson, 1983). Specifically, as (11) and (12) demonstrate, in constituting English VtR, the matrix verb takes the adjectival small clause as its complement. Moreover, although the matrix verb itself cannot take an NP complement, it can still take the whole small clause as its complement, as (11) shows. In this respect, the only complement of the matrix verb is the small clause, which denotes the end-state of the resultative event (Folli & Harley, 2020; S. Kim, 2016).

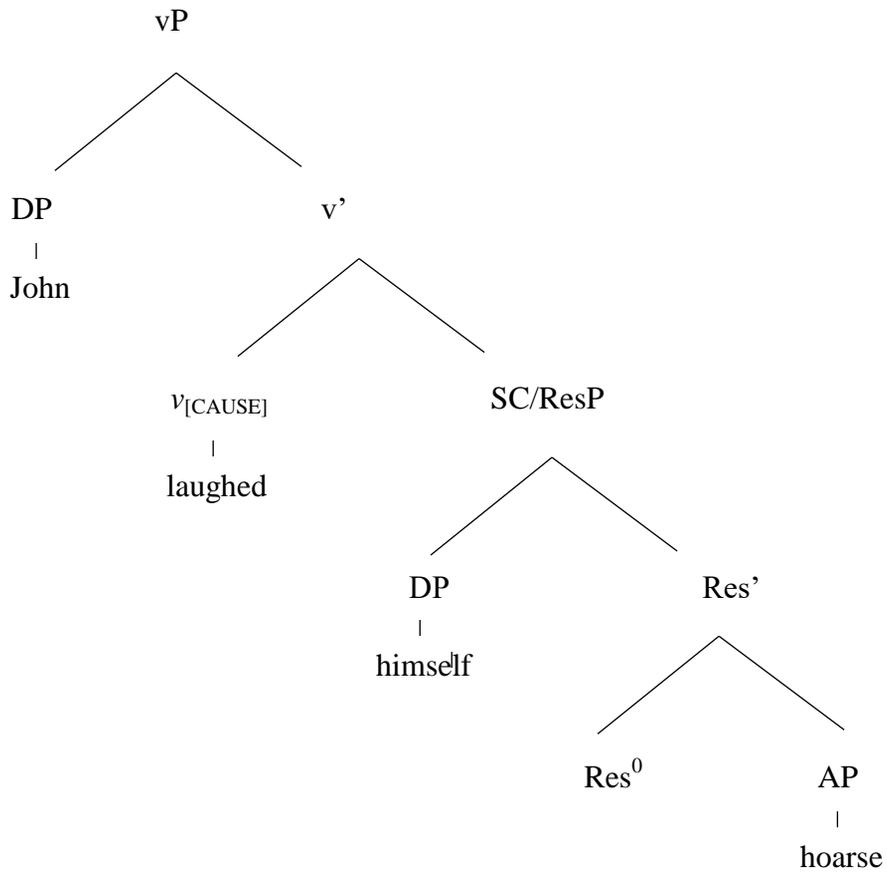
(11) Syntactic Configuration of *John hammered the metal flat*.

(cf. John hammered the metal.)



(12) Syntactic Configuration of *John laughed himself hoarse*.

(cf. \* *John laughed himself*)



However, most studies have shown that the adjectival phrase in Korean adjectival VtR expressions is not a complement, but rather an adjunct to the matrix verb. This is primarily because Korean and English adjectival VtR differ in terms of (i) the syntactic relationship between the result-denoting component and the matrix verb (Hong, 2005; Hwang, 2011; Ko, 2011, 2014; Oh, 2010; Shim & den Dikken 2007), and (ii) the internal structure of the result-denoting component (Ko, 2015; Shim & den Dikken, 2007).

The result-denoting component of a small clause in Korean adjectival VtR has a different underlying structure from that of English SC/ResP.<sup>4</sup> Similar to English, Korean adjectival VtR can take the SC-subject in the accusative case. Even so, the internal structures of Korean small clause and its syntactic status is significantly distinct from those of English. For example, Ko (2015) noted that Korean adjectival VtR takes an adjunct RP headed by *-key*. She also demonstrated that the distribution of an RP-predicate for the Korean adjectival VtR is crucially affected by the presence or absence of an overt subject of SC, where the former takes the postverbal NP as the accusative case (Example (13)) and the latter takes the postverbal

---

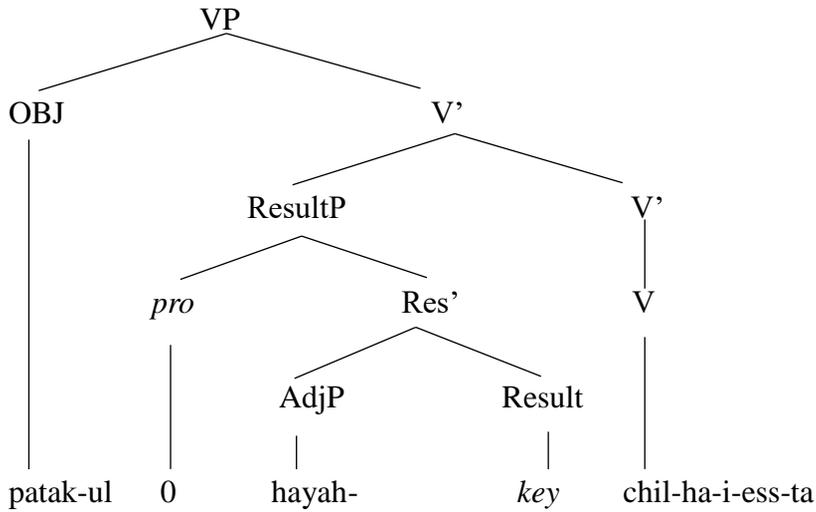
<sup>4</sup> Ko (2015) demonstrated four types of Korean transitive resultative constructions, headed by *-key* and *-lo*. Mostly, *-key* resultative co-occurs with adjectival VtR, and *-lo* with nominal VtR, which is by and large a translation equivalent to English INTO-causative (e.g., *The father pounded the beans into powder / A magician turned water into wine*). Each *-key* and *-lo* VtR was further subdivided depending upon whether or not each of them has an overt subject in the small clause.

NP as the nominative case (Example (14)).

(13) Adjunct Small Clause with null SC-Subject (Accusative-marked)

(Ko, 2015, p. 369, (52))

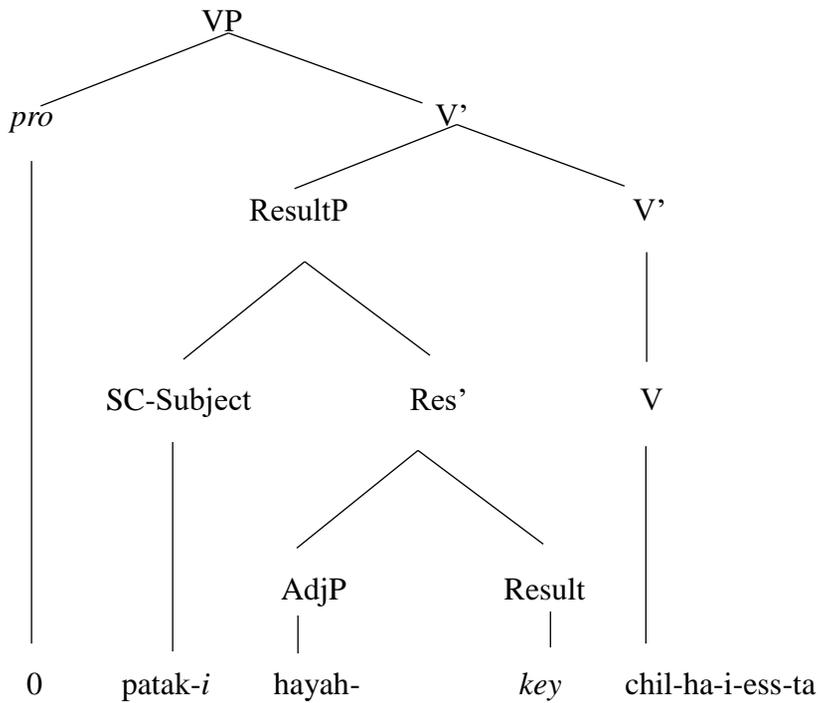
John-un     patak-*ul*     hayah-*key*     chil-ha-i-ess-ta  
 John-TOP   floor-ACC   white-RES   paint-do-LK-PST-DECL  
 ‘John painted the floor white.’



(14) Adjunct Small Clause with overt SC-Subject (Nominative-marked)

(Ko, 2015, p. 369, (53))

John-un    patak-*i*        hayah-*key*    chil-ha-i-es-sta  
 John-TOP   floor-NOM    white-RES    paint-do-LK-PST-DECL  
 ‘John painted the floor white.’



The grammaticality contrasts given in (15a) and (15b) confirm that the internal structure of the result-denoting small clause of Korean adjectival VtR differs from English (Shim & den Dikken, 2007). The English SC-subject always assumes an overt SC-subject, whereas the Korean SC-subject expression was either covert or governed by a discourse-bound *pro*. As (15a)

demonstrates, the nominative-marked NP is licensable, considering that it is the default case in Korean (Jang, 1991). However, the ungrammaticality in (15b) directly results from the fact that the accusative-marked NP (i.e., *mok-ul*) cannot get its case licensed from an intransitive verb (i.e., *wul-ess-ta* ‘cried’). This is clear evidence that Korean adjectival VtR does not permit an intransitive verb to take a whole small clause as its direct complement.

- (15) a. Jim-i                      mok-*i*                      swi-*key*                      wul-ess-ta  
           Jim-NOM                  throat-NOM                hoarse-RES                cry-PST-DECL  
           ‘Jim cried himself hoarse.’
- b. \*Jim-i                      mok-*ul*                      swi-*key*                      wul-ess-ta  
           Jim-NOM                  throat-ACC                hoarse-RES                cry-PST-DECL  
           ‘Jim cried himself hoarse.’

The result-denoting small clauses in Korean adjectival VtR constitute an adjunction relationship with the matrix verbs. For example, Shim and den Dikken (2007) argued that the nominative case marking in Korean adjectival VtR reflects that the subject of the SC/ResP was adjoined to the matrix verb, which is strongly disallowed in English.

These syntactic differences between Korean and English are evidenced by direct object restriction (DOR). DOR was proposed by Levin and Rappaport Hovav (1995) to capture the fact that a resultative predicate in English can

be predicated of the direct object, but not of the subject the oblique argument in the main event. Thus, the direct object needs to be the overt subject of an RP. English VtR has been reported to conform to DOR. As shown in (16a), an RP in English can be the predicate of the direct object, but not of the subject (e.g., *the metal = flat* / *\*John = flat*). In addition, an unaccusative subject or a *deep object* can be understood as the subject of an RP, as demonstrated in (16b) (e.g., *the pond = solid*). In contrast, an unergative subject cannot be interpreted as the subject of an RP, as in (16c) (e.g., *\*John = hoarse*). To demonstrate such predicational relationship, a pseudo-reflexive must be used, as in (16d).

(16) a. John hammered the metal flat. [Transitive]

b. The pond froze solid. [Unaccusative]

c. \*John laughed hoarse. [Unergative]

d. John laughed himself hoarse. [Unergative/Pseudo-reflexive]

(S. Kim et.al., 2020)

In contrast, DOR does not hold in Korean adjectival transitive resultative construction (Hong, 2005, 2011; Kim & Maling 1997; Ko, 2011; Shim & den Dikken 2007; Son, 2008; Yeo, 2006): In (17a), both the subject,

Susanna, and the direct object, Jim, can be interpreted as the subject of the result phrase, unlike in English (16a). In Korean, an unaccusative subject (e.g., *Kang-i*) can be comprehended as the subject of the result phrase (17b), but the subject of an unergative verb (e.g., *John-i*) also can be the subject of the result phrase in (17c). In (17c), the Korean result phrase *mok-i swi-key* (hoarse) does not co-occur with a fake reflexive *caki* ‘himself’. However, a fake reflexive is grammatically required in English unergative resultatives, as in (16d). The contrast between (16) and (17) clearly shows that DOR does not hold in Korean RC.

(17) a. Susana-ka            Jim-ul     aphu-key     ttayli-ess-ta. [transitive]

Susanna-NOM   Jim-ACC   in pain-RES   hit-PST-DECL

‘Susana<sub>i</sub> hit Jim<sub>j</sub> so that she<sub>i</sub>/he<sub>j</sub> was in pain.’ (Ko, 2011, p. 23, (50))

b. Kang-i                    tantanha-key     el-ess-ta. [unaccusative]

river-NOM                solid-RES                freeze-PST-DECL

‘The river froze solid.’ (Park & Lakshmanan, 2007, p. 330)

c. John-i                    mok-i swi-key     kohamcil-ess-ta. [unergative]

John-NOM     hoarse-RES                shout-PST-DECL

‘John shouted himself hoarse.’ (Ko, 2011, p.24, (52))

This cross-linguistic evidence demonstrates that there is a syntactic difference between English and Korean in the adjectival VtR. Specifically, the English adjectival VtR is realized by the syntactic operation of complementation between result-denoting and manner-denoting components, in which the result-denoting small clause is selected by the manner-denoting matrix verb via manner incorporation. In contrast, Korean adjectival VtR is realized by the syntactic strategy of adjunction, in which the manner-denoting verbal component and the result-denoting *-key* phrases are simply stacked on top of each other.

### 2.1.2.2. Korean Translation Equivalent Expressions of English Prepositional Transitive Resultative Constructions

In the previous section, we observed that Korean adjectival VtR takes an adjoined small clause, violating DOR. This section analyzes the syntactic distribution of Korean translation equivalent to English prepositional VtR with respect to whether the manner- or result-denoting components are syntactically encoded in a mono-clausal formulation.

The mono-clausal formulation of prepositional VtR is related to the direct causative construction.<sup>5</sup> In essence, a direct causative has a tight temporal integration of the causing and caused event, whereas an indirect causative is deficient in the temporal overlap between the causing and caused events (Goldberg & Jackendoff, 2004).<sup>6</sup>

---

<sup>5</sup> Based on Shibatani and Pardeshi (2002), Ambridge et al. (2020) who summarized the distinction between more direct causation and less direct causation as follows:

•*More-direct causation*: “entails a spatiotemporal overlap of the causer’s activity and the caused event, to the extent that the two relevant events are not clearly distinguishable.” (Shibatani & Pardeshi, 2002, p. 89)

•*Less-direct causation*: “entails an event in which both the causing and the caused event enjoy some degree of autonomy ... The caused event ... may have its own spatial and temporal profiles distinct from those of the causing event.” (Shibatani & Pardeshi, 2002, p. 89)

<sup>6</sup> According to Haiman (1985, p.783), “the conceptual distance between cause and result will correspond to the formal distance between cause and result.” A similar idea (e.g., scale of compactness) was discussed in Dixon (2000).

- (18) a. *Sue made Bill die on Thursday*  
by poisoning his breakfast on Wednesday.
- b. \* *Sue killed Bill on Thursday*  
by poisoning his breakfast on Wednesday.

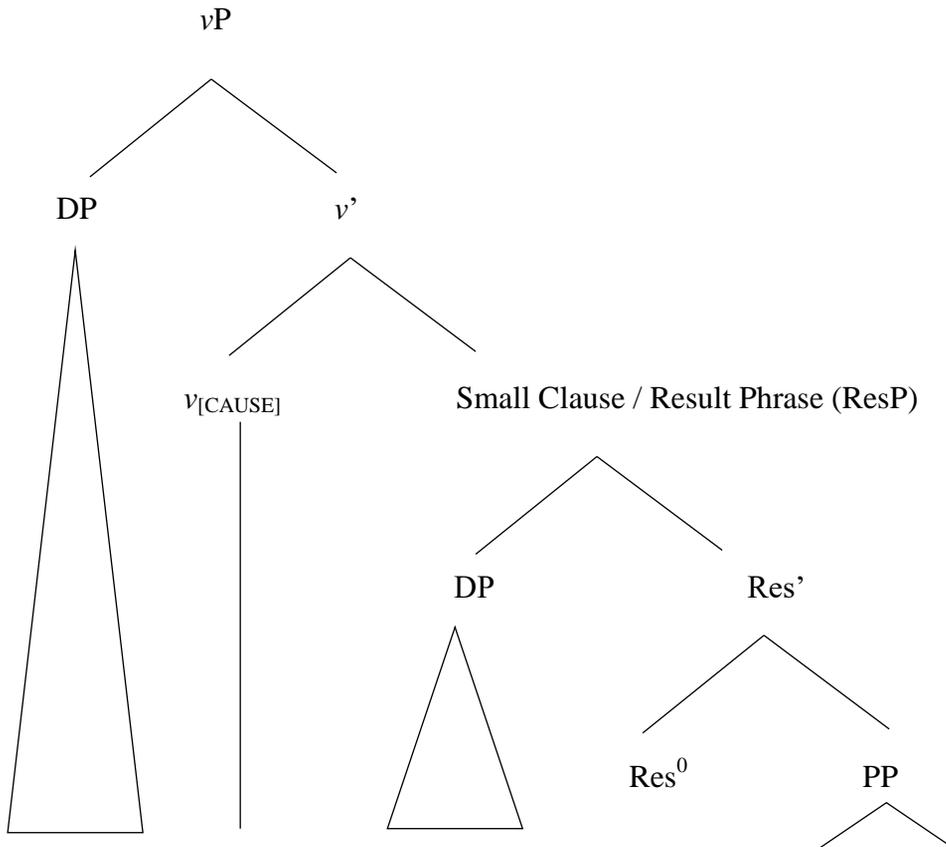
For example, Fodor (1970) observed that, as (18a) illustrated, bi-clausal causatives permit a temporal delay between manner- and result-denoting components. He reported that mono-clausal causatives such as (18b) do not allow a temporal overlap between manner-denoting and result-denoting events.

- (19) a. Bill rolled the ball up the hill. (Accompanying)
- b. Mary whistled Rover to her side. (Non-accompanying)

The evidence in (19) suggests that the causing component is co-temporal with the caused element or at least the former partially overlaps with the inception of the caused component in the English prepositional VtR (Folli & Harley, 2006; Goldberg & Jackendoff, 2004). Specifically, (19a) demonstrates that there is a total temporal overlap between the manner-denoting component (e.g., *rolled*) and the result-denoting element syntactically packaged in a mono-clausal English prepositional VtR (e.g., *the ball up the hill*). (19b) also illustrates that there is at least a partial

temporal overlap between the causing event (e.g., Mary’s whistling action) and the caused event (e.g., Rover’s movement to her side) in (19b).

(20) Co-temporality in an English Prepositional VtR



Bill	rolled	the ball	up the hill
<i>Manner</i>	Rolling	→	
<i>Result</i>	Up the hill	→	

Mary	whistled	Rover	to her side
<i>Manner</i>	Whistling	→	
<i>Result</i>	To her side	→	

To summarize, similar to the English adjectival VtR, the English prepositional counterpart has a temporal overlap between the causing and caused event. The English prepositional VtR has a tight semantic relationship (or co-temporality) between the manner-denoting and result-denoting component (Goldberg, 1995), thereby constituting direct causation in a mono-clausal formulation.

On the other hand, the Korean translation equivalent of the English prepositional VtR shows a clear disparity from its English counterpart in that the former does not permit mono-clausal packaging of manner- and result-denoting components, whereas the latter does. First, similar to the adjectival VtR (see the grammatical contrast in (15a) and (15b)), the Korean translation equivalent expressions to the English prepositional VtR does not allow mono-clausal prepositional VtR formulation to be embedded in an intransitive verb. Instead, the Korean translation equivalent of English prepositional VtR expressions employs bi-clausal formulation (e.g., conjunctive clause) to syntactically represent the manner–result component.<sup>7</sup>

---

<sup>7</sup> Similar to English, the result-denoting component may remain in the small clause, not in the matrix verb, in Korean adjectival and prepositional VtR. However, Rappaport Hovav and Levin (1998, 2010) observed that there is a ban against the dual lexicalization of both the manner- and result-denoting component in a single verb. This restriction is referred to as Manner/Result Complementarity (Folli & Harley, 2020). The result feature of the English VtR is weak and thus remains at the head of the small clause, reserving the matrix verb to contain a manner component. However, the result feature of the Korean VtR is strong, thereby moving to its higher matrix verb. In this case, the matrix verb is saturated for the

Instead, Korean uses alternative translation equivalent expressions for the English prepositional VtR with a serial verb construction (SVC). There are at least two strategies for formulating such SVC expressions.<sup>8</sup> By relying on SVC, the second verbal element (V2) expresses a result-denoting component, while the first verbal element (V1) denotes a manner-denoting one. Firstly, Korean translation equivalent expressions for English prepositional VtR are realized with an SVC with a transitive- and a transitive-verb complex.

(21) John-i	khathu-lul	cengwon-ey
John-NOM	the cart-ACC	garden-PATH
mil-e	ponay-ess-ta.	
push-Lk	pass-Pst-DECL	

‘John caused the cart to go to the garden by pushing it.’

(Zubizarreta & Oh, 2007, p.100, (291b))

---

result-denoting component, so the Korean VtR may need to have another verbal element to denote the manner-denoting component (H. Kim, 2017). In this respect, the main goal of Korean L2 English learners when learning English is reduced to (i) unlearning L1 transferred bi-clausal construction and (ii) learning the L2 specific manner incorporation strategy.

<sup>8</sup> It is reasonable to ask whether Korean caused-motion verbs, like *noh-ta* ‘put’ corresponds to English translation equivalent counterpart. However, it is questionable whether Korean *noh-ta* takes a result-denoting small clause like its English counterpart. See Kageyama (2003, p. 358) for the syntactic discussion that Korean does not have morphological inventories to denote a ‘change of (result) location’ equivalent to English (e.g., preposition, particle)..

The SVC in (21), ‘*mil-e ponay-ess-ta*’, the first transitive verb (*mil-* ‘push’) is linked to the second transitive verb (*ponay-* ‘send’) by a morphological linking marker *-e*.

Secondly, Korean translation equivalent expressions to English prepositional VtR are composed of an SVC with a transitive verb and an intransitive verb. As there are no Korean transitive deictic verbs comparable to English *take* and *bring*, the Korean translation equivalent of the English prepositional VtR is syntactically realized in a simultaneous serial verb construction (SSVC) with an intransitive deictic verb (e.g., *kata* ‘go’ or *ota* ‘come’, which is modified by a transitive verb, *kaci-* ‘have’ or *teyli-* ‘carry’ (Choi & Bowerman, 1991). Thus, in the example of *Inho-ka chayk-ul cip-ey kacy-e ka-ss-ta*, Y. Kim (1995) asserts that this is not a transitive clause, but rather an intransitive clause, denoting a spontaneous motion, which is consistent with the argument for SSVC.<sup>9</sup> Contrasted with (21), the nominative-marked NP in (22) moves along with the accusative-marked moved theme NP (e.g., *chayk-ul* ‘book-ACC’), which corresponds to an accompanying English prepositional VtR in (19a).

---

<sup>9</sup> Thus, Y. Kim (1995) states that the appropriate translation of the sentence should be ‘Inho went home, having a book’ rather than ‘John took a book home’.

(22) John-i	chayk-ul	cip-ey
John-NOM	book-ACC	home-PATH
kacy-e	ka-ss-ta.	
have-LK	go-PST-DECL	

‘John went to the library, having a book.’

(Choi & Bowerman, 1991, p.92)

(22) shows that when someone takes or brings something while moving, this can be expressed with the intransitive deictic verb, *kata* ‘go’, to which the transitive verb of *kacy-e* ‘have-LK’ is added. In this formulation, the matrix verb is the intransitive deictic verb, and the modifying verb is the transitive verb.

Korean translation equivalent expressions for English prepositional VtR are readily compatible with a bi-clausal formulation in which temporal simultaneity between the manner- and result-denoting component is not obtained. As demonstrated in (23), the first verbal phrase for describing a manner-denoting component is connected to the second verbal phrase, which denotes a result state represented by the intransitive deictic motion verb. The two verb phrases are mediated by a conjunctive marker *-(e)se* (Choi, 2020): The conjunctive marker is required in this syntactic formulation, because the absence of it leads to ungrammaticality.

(23) Syntactic Configuration of Korean Translation Equivalent of English prepositional VtR in a Bi-clausal Formulation

John-i caychayki-hay-*ese*

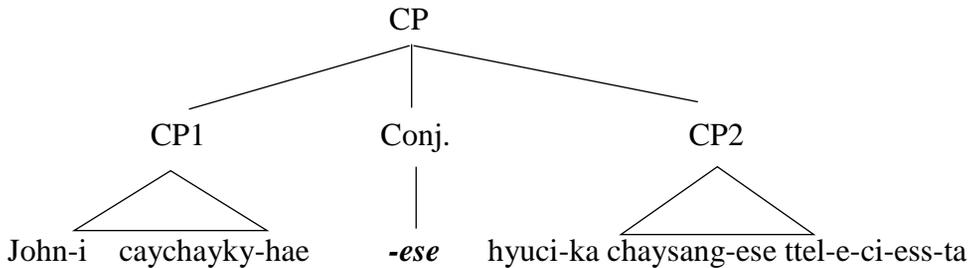
John-NOM sneeze-DO-*ese*

hyuci-ka chayksang-ese *ttel-e-ci-ess-ta.*

tissue-NOM desk-LOC *fall-LK-PASS-PST-DECL*

‘John sneezed **and then** the tissue fell off the table.’

(= ‘John sneezed the tissue off the table’)



(23) demonstrates that John’s action of sneezing (causing event) carries over to the tissue’s falling event (caused event; Shibatani & Pardeshi, 2002) in a bi-clausal construal. Again, these manner- and result-denoting components in (23) are temporally separated by the *-se* marker (and then). As (24) illustrates, the *-(e)se* morpheme has been reported as a conjunctive marker representing a temporal sequence (and then) between the two events (Choi, 2020; Sohn, 2009).

- (24) Emma-ka          cip-ey          o-*a-se*,          yatanchi-lkke-ya?  
 mother-NOM      home-LOC      come-LK-(*E*)*SE*      scold-FUT-Q  
 ‘Will mommy come home *and then* scold me?’

In this respect, the Korean prepositional VtR is compatible with the *-(e)se* marker,<sup>10</sup> which makes explicit that the verbs in the prepositional VtR have a temporal separation between the causing and caused event in a bi-clausal formulation (Choi, 2020; Kang, 1993; Li, 1993).

- (25) Kong-i          kkocpyong-ul          chi-e-*se*  
 ball-NOM          vase-ACC          hit-LK-(*E*)*SE*  
 kkocpyong-i          nemeci-ess-ta.  
 vase-NOM          fall-PST-DECL  
 ‘The ball hit the vase, *and then* the vase fell down.’  
 (= ‘The ball knocked the base off.’)

As (25) illustrates, *vP1* (*chi-e* ‘hit-LK’) and *vP2* (*nemeci-ess-ta* ‘fall’) in the Korean translation equivalent of English prepositional VtR expressions are separated by the *-(e)se* morpheme. The compatibility with the *-(e)se*

---

<sup>10</sup> The *-(e)se* marker also represents the manner-denoting component. See Choi (2020).

morpheme suggests that there is a temporal delay between  $\nu$ P1 and  $\nu$ P2. Thus, in (25), the moment of hitting the vase must precede the moment of its falling. With the conjunctive marker *-(e)se*, the causing event in  $\nu$ P1 is a pre-condition for the caused-event in  $\nu$ P2,<sup>11</sup> in that the caused event takes place in the state produced by the event of the conjunctive clause headed by *-(e)se* (Choi, 2020; Sohn, 2009). For example, in (25), the speaker states that the vase fell down when the ball hit it, which means that the ball must have hit the vase first for the vase's falling to take place (Choi, 2020). Such a temporal delay found in Korean may interact with the relatively weak causal relationship, packaged loosely in a bi-clausal formulation.<sup>12</sup>

---

<sup>11</sup>  $\nu$ P1 can be a presupposition for the event that  $\nu$ P2 may occur without an explicit causal relationship between the two verbal elements (cf. consequential SVC in Zubizarreta & Oh, 2007). See Ko and Sohn (2015) regarding the H-SVC without the causative morpheme.

<sup>12</sup> Further evidence may be found in Choi (2018) from the perspective of utterance density (UD) with subtypes of the Korean prepositional VtR. She found that the UD of the Korean prepositional VtR is much looser than its English counterpart.

### 2.1.3. Semantic Differences in VtR between Korean and English

The syntactic analyses of English adjectival or prepositional VtR argue for complementhood between the matrix verb and its small clause. An RP or small clause is tightly related to a verb, as a sister of V in the syntactic structure. The result events of the changed state or changed location on the object are expressed by a small clause, which can make explicit a result state for the action demonstrated by the verb (Aske, 1989; Hoekstra & Mulder, 1990; Levin & Rappaport Hovav, 1995; Talmy, 1985).

This complementhood of the English VtR results in a severe selectional restriction on the choice of a result phrase, which is related to the telicity of the event.<sup>13</sup> These semantic restrictions on English resultative constructions have been widely discussed with respect to the semantic restrictions for the telicity of VtR—namely, the measuring-out constraint (Tenny, 1994), end-of-scale constraint (Goldberg, 1995), and boundedness restriction on resultatives (Wyngaerd, 2001). These semantic restrictions are presented in (30).

---

<sup>13</sup> In Tenny's (1994) terminology, telicity is relevant to the boundedness that imposes the bounds on the properties of an affected patient. A telic predicate describes a bounded eventuality, and the boundedness indicates the presence of a culmination point. According to Wechsler (2001), "If there is any aspect of resultatives that is completely uncontroversial, it is that they are 'telic': they describe events with a definite endpoint" (p. 4).

(30) Semantic restrictions on the telicity of English adjectival VtR

(S. Kim, 2016)

a. Tenny's (1994) Measuring-Out Constraint:

Resultative constructions require a MEASURE role to be licensed, since the relevant constructions introduce a temporal end-point, or measuring-out, to the event.

b. Goldberg's (1995) End-of-Scale Constraint

The type of adjective that occurs in RP or Small Clause in English adjectival VtR is highly restricted. They are non-gradable adjectives which have a clear delimited lower-bound. These non-gradable adjectives cannot co-occur with a quantifying modifier.

(e.g., ? *a little flat/dead/awake*)

c. Wyngaerd's (2001) Boundedness Restrictions on Resultatives

Resultative predicates denote a bounded scale in which an unbounded modifier (i.e., *very*) cannot co-occur in RP, while bounded scale modifiers (i.e., *half/almost/completely*) can.

(e.g., *Max shouted himself \*very/half/completely/almost hoarse.*)

The most distinctive selectional restrictions of the abovementioned telicity is that they commonly represent the end-point or delimited lower-boundedness of an event. In Tenny's (1994) terminology, telicity also is relevant in measuring-out constraint and boundedness that impose the

bounds on the properties of an affected patient. A telic predicate describes a bounded eventuality, and this boundedness indicates the presence of a culmination point. According to Wechsler (2001, p. 4), “If there is any aspect of resultatives that is completely uncontroversial, it is that they are ‘telic’: they describe events with a definite endpoint.”

Telicity has three different characteristics at the syntactic and semantic interface. First, the argument structure construction determines the telicity of the event. The verb is not the only factor in the establishment of the telicity of the event; other variables such as verb argument structures or adjuncts also play a key role in determining telicity. As Tenny (1994) identified, even when a delimiting component (e.g., result phrase) is added to an atelic verb, the event type of the whole predicate can be changed into a telic category.<sup>14</sup>

(31) a. He pushed the door. (atelic)

b. He pushed the door open. (telic)

Second, telicity is established between the manner-denoting component

---

<sup>14</sup> There are at least three primary components to delimit the event type: verb particles, resultatives, and the dative argument in double object creation (Tenny, 1994).

and result-denoting complex events that have been analyzed into inner and outer events (Jackendoff, 1990). Outer events are associated with causation and agency (process), and inner events are associated with telicity and change of state (result) (Comrie, 1976; Croft, 1998; Dowty, 1979; Jackendoff, 1983). In English, the former is linguistically realized as a matrix verb, whereas the latter, as a small clause (Aske, 1989; Hoekstra & Mudler, 1990; Levin & Rappaport Hovav, 1995; Talmy, 1985).

Finally, there is a significant link between telicity and the affectedness/transitivity in the direct object.<sup>15</sup> Kratzer (2004) postulated a link between telicity and the accusative case in German and English. Winkler (1997) hypothesized that, to establish telicity, the postverbal NP must be an affected theme. Hovav and Levin (2001) discovered that such affectedness and telicity are required for the resultative construction. According to Wagner (2006), children use transitivity as structural evidence for telicity.

Telicity has traditionally been tested by the compatibility of unbounded time adverbial modification (e.g., *for ten minutes*) with the event structure (Baker, 1995; Fleischhauer, 2013; Vendler, 1957) and by the availability of the negation of the end-state (Fleischhauer 2013; M. Park 2011). As telicity

---

<sup>15</sup> Telic predicates are compatible with definite objects, and atelic predicates are compatible with plural or bare object NPs.

is mainly defined in terms of a culmination point (or an end-point) (Kearns, 2007, p. 47), English adjectival and prepositional VtR expressions are modifiable only with a bounded time adverbial (e.g., *in ten minutes*), not an unbounded time adverbial (e.g., *for ten minutes*), and they strongly disallow the negation of the end-state in the RP. As the example sentences in (32) and (33) show, English adjectival and prepositional VtR expressions denote telic interpretation: English adjectival and prepositional VtR expressions are not modified by an unbounded time adverbial (32), and the RP in the small clause is never allowed to be cancelled (33).

(32) Incompatibility with For-modification

- a. Adjectival VtR: John ran the shoes threadbare in/#for 5 minutes.
- b. Prepositional VtR: John waltzed Matilda into the bedroom in 5/#for 5 minutes. (Folli & Harley, 2006)

(33) Cancellability Test: non-cancellable RP

- a. Adjectival VtR: ?? He hammered the metal flat, but the metal is not flat.
- b. Prepositional VtR: #I pulled/dragged the box to the door, but stopped before I got there. (Hovav & Levin, 2008).

On the other hand, the Korean adjectival or prepositional VtR denotes atelic or durative semantics. Thus, as the example sentences in (34) demonstrate, Korean adjectival and prepositional VtR expressions allow both (i) for-modification, and (ii) cancellable expressions (S. Kim, 2016).

(34) a. Compatibility with For-modification

John-i	<i>o-pwun-maney</i>	/	<i>-tongan</i>
John-NOM	five-minute-in	/	-for
sinpal-i	talh-key		ttwi-ess-ta.
shoes-NOM	threadbare-KEY		run-PST-DECL

‘John ran the shoes threadbare *in/for five minutes.*’

b. Cancellability Test: cancellable RP

Keu-ka	chelphan-ul	pyengpyengha-key	<i>twutulki-ess-una</i>
he-NOM	metal-ACC	flat-KEY	pound-PST-BUT
<i>chelpan-un</i>	<i>pyeongpyeonghay-chichi</i>		<i>anh-ass-ta</i>
metal-TOP	flat-become		NEG-PST-DECL

‘He hammered the metal flat, *but it did not become flat.*’

In addition, the Korean translation equivalent of English prepositional VtR expressions is atelic due to its syntactic density. Thus, the Korean translation equivalent of English prepositional VtR expressions is compatible with both for-modification (35a) and cancellable expressions (35b).

(35) a. Compatibility with For-modification

Keuni-ka	<i>sip-pwun-maney</i>	/	<i>-tongan</i>
Keuni-NOM	ten-minute-in	/	-for
yaksumul-ul	pyeong-eh		chaewu-ess-ta.
water-ACC	bottle-PATH		fill-PST-DECL

‘Keuni filled the water into the bottle *in/for ten minutes.*’

(Nam, 2003, p.26)

b. Compatibility with Cancellable RP

John-i	cip-pakk-ulo		kong-ul
John-NOM	house-outside-toward		ball-ACC
cha-ss-una	kong-un	wumciki-cianh-ass-ta	
kick-PST-BUT	ball-TOP	move-NEG-PST-DECL	

‘John kicked the ball toward the outside of the house

*but the ball didn't move.*’ (Zubizarreta & Oh, 2007, p.93 (268))

To recap, the evidence demonstrated above strongly suggests that Korean resultative constructions represent the degree of achievement and do not necessarily entail a change of state.

A summary of syntactic and semantic contrasts between English and Korean VtR construction is provided in Table 2.2.

**Table 2.2 Summary of Syntactic and Semantic Contrasts between English and Korean Transitive Resultative Construction**

	Syntax	Semantics	
	Syntactic relation between a matrix verb and a small clause	For-phrase Modification	Cancellability Test
English VtR	Complement	X	X
KoreanVtR	Adjunct	O	O

In summary, Korean resultatives are not equated with English resultatives since the two languages have distinct semantic features, let alone syntactic ones: English resultatives signify a telic event, but Korean resultatives can denote an atelic event. This means that Korean L2 English learners must acquire syntactic and semantic interface of English transitive resultative constructions which underpin the telicity.

#### **2.1.4. Previous Acquisition Studies on English Adjectival and Prepositional VtR**

English L1 acquisitional studies have extensively demonstrated that English VtR is syntactically developed from the formulation of a verbless small clause, or verb particle construction (VPC), to sentential transitive resultative construction (VtR). According to several studies (Diessel & Tomasello, 2005; Radford, 1988; Tomasello, 1987), the complete formulation of adjectival or prepositional VtR is preceded by a V-O-Prt VPC, and the production of VPC is followed by the production of a small clause.

From a typological standpoint (Talmy, 1985), the English-specific small clause structure causes serious learning difficulties for Korean L2 English learners whose native language allows both manner- and result-denoting components to be linguistically realized as separate verbal phrases. Thus, recent L2 acquisition studies have consistently reported that Korean EFL students have difficulty learning English adjectival or prepositional VtR. (H. Lee, 2019; J. Lee & H. Kim, 2011; S. Kim et al., 2020; Rah, 2014; Shin, 2010; H. Sung, 2018; M. Sung, 2012; M. Sung & Yang, 2016). Indeed, adjectival and prepositional VtR in English appear late in the developmental sequence for Korean L2 English learners (J. Lee & H. Kim, 2011).

Moreover, Korean L2 English learners usually fail to formulate English-like small clauses either by omitting the SC-subject (or direct object) or by dropping the entire result-denoting component of adjectival phrases or prepositional phrases in formulating relevant formulations (Rah, 2014; H. Sung, 2018).

It has been widely reported that foreign language learners generally accept grammatically incorrect L2 expressions whose translation equivalent formulations are accepted in their first language (Cabrera & Zubizarreta, 2005). Thus, novice-level Korean L1 learners may generate L1 transferred syntactic constraints by rejecting grammatical L2 formulations. For example, R. Kim and Yang (2017) replicated Robenalt and Goldberg's (2016) study and investigated whether Korean and Chinese L2 English learners could correctly comprehend English prepositional VtR expressions. It has been widely reported that English prepositional VtR expressions share linguistic commonalities with Chinese translation-equivalent expressions, as the two languages allow an intransitive verb to be incorporated as a matrix verb in a mono-clausal construal (e.g., *The lifeguard swam the children to the shore*).

Moreover, to produce prepositional VtR expressions in English, Korean adopts a bi-clausal strategy to denote the relevant expressions, but Chinese uses a mono-clausal strategy, as in English. Considering these crosslinguistic variations, Chinese L2 English learners were expected to

accept unselected English prepositional VtR expressions more so than Korean L2 English learners. The results showed that these expectations proved accurate: Chinese L2 English learners, whose L1 prepositional VtR corresponds to L2 English, generally accepted English prepositional VtR expressions embedded by intransitive verbs. On the other hand, Korean L2 English learners, whose L1 does not correspond to translation equivalent L2 expressions, rejected syntactically grammatical prepositional VtR expressions in English (e.g., *The lifeguard swam the children to the shore*), and generated alternative expressions with intransitive motion expressions (e.g., *The lifeguard swam to the shore*).

In this vein, it is extremely difficult for Korean L2 English learners to produce mono-clausal English prepositional VtR expressions embedded in intransitive verbs. H. Sung (2018) conducted a production task (Sentence Completion Task [SCT]) to investigate how Korean L2 English learners generate English prepositional VtR expressions. Given the cross-linguistic variations between the two languages, she expected Korean L2 English learners to fail to correctly produce prepositional VtR expressions embedded in an intransitive verb. This prediction was supported by the results of the SCT, as Korean L2 English learners generally failed to provide mono-clausal prepositional VtR expressions embedded in intransitive verbs. Instead, they produced multiclausal construction (e.g., \*He *made* Matilda to

*go* into the room *by dancing* with her / He danced Matilda into the room), intransitive motion construction (e.g., \*He sneezed *at* the tissue off the table / He sneezed the tissue off the table), or other alternative constructions, whose translation-equivalent expressions are grammatically accepted in their L1 Korean syntax. This result is consistent with Choi and Bowerman (1991) in that Korean prepositional VtR expressions are bi-clausal constructions where the result state is expressed via the intransitive deictic verb construction (e.g., *teli-ko-gata* “HAVE-conjunctive marker-GO”).

S. Kim et al. (2020) demonstrated that Korean L2 English learners also have difficulty comprehending and producing English adjectival VtR. Their study investigated how syntactic and semantic differences between Korean and English influence the comprehension and production of L2 English adjectival VtR. To determine the L1 transfer effects, two different experiments—the acceptability judgment task (AJT) and the elicited writing task (EWT)—were employed. The results strongly confirmed that the L1 transfer effects were persistent in Korean L2 English learners. Specifically, the results of the AJT demonstrated that Korean L2 English learners failed to reject syntactically and semantically ungrammatical expressions that are affected by L1 properties. Furthermore, even advanced-level Korean L2 English learners generally had difficulty understanding adjectival VtR expressions. The results of the EWT also revealed that more than half of the

Korean L2 English learners relied heavily on L1 transferred bi-clausal constructions (e.g., adjunct, coordination, serial verb, and object omission) for English adjectival VtR expressions. This suggests that Korean L2 English learners fail to use the complementation structure (e.g., English-like small clause) and assign a telic meaning to English RCs, regardless of proficiency level.

These cross-linguistic variations in the syntax and semantics of adjectival VtR between Korean and English raise a question about how Korean L2 English learners can learn the relevant construction in L2 English. One of the ways to resolve this tension is to provide learners with construction-based instruction (H. Kim & Rah, 2021). However, explicit instruction alone may not help Korean L2 English learners comprehend and produce L2 English adjectival VtR expressions. According to S. Kim (2016), problems with L1-Korean influence on the comprehension and production of L2-English adjectival VtR are addressed by providing learners with corrective feedback. Similarly, Lightbown and Spada (2011) suggested that it is important to draw students' attention to the cross-linguistic contrasts between the two languages by providing various types of corrective feedback. In the following section, the pedagogical nature of corrective feedback and its practical application will briefly be reviewed.

## **2.2. Corrective feedback**

This section covers corrective feedback and its use in the learning of foreign languages. Sections 2.2.1 and 2.2.2 describe ‘corrective feedback’ and discuss its subtypes. Following that, the noticeability of corrective feedback will be addressed (section 2.2.3), with a particular emphasis on its practical application, and corrective recast (section 2.2.4).

### **2.2.1. Definition of Corrective Feedback**

The term ‘corrective feedback’ has been interchangeably used with ‘negative evidence’, ‘negative feedback’, or ‘error correction’. Given this conflation of terms, and to eliminate ambiguity, it is therefore necessary to define each terminology.

The first term to be defined is ‘negative evidence’, which is frequently used with its antonym, ‘positive evidence’. While positive evidence can be defined as that which provides learners with grammatically correct examples in their target language, negative evidence is defined as that which provides learners with information about what is unacceptable and grammatically incorrect in the target language (Long, 1996).

Furthermore, negative evidence can be given either indirectly or directly. Indirect negative evidence signals to learners that a certain constructional formulation is impossible in L2 because such a formulation is unavailable in the distribution of input (Chomsky, 1981; Goldberg, 2019). In contrast, direct negative evidence connotes, for example, a teacher's reaction to a student's error which attracts the learner's attention to his or her mistake.

Negative feedback is a subset of such direct negative evidence. According to Chaudron (1977), negative feedback is "any reaction of the teacher which clearly transforms, disapprovingly refers to, or demands improvement of the learner's utterance" (p. 31). When corrective feedback is defined as "any indication to the learner [by the teacher] that his/her use of the target language is incorrect" (Lightbown & Spada, 2011, p. 197), then one may use 'negative feedback' and 'corrective feedback' interchangeably. Thus, negative or corrective feedback are both explicit (e.g., metalinguistic feedback, overt error correction) or implicit (e.g., confirmation checks and recasts).

It is critical to distinguish negative or corrective feedback from error correction which is defined as an attempt to repair learners' grammatically incorrect formulations (Chaudron, 1977). It is worth noting that corrective feedback only indicates the presence of a mistake; it does not always result in learners' speaking errors being corrected. Thus, in this study

corrective feedback is defined operationally as any effort by instructors or any attempt of AI chatbot agents to alert learners to the existence of their mistakes (Carroll & Swain, 1993). Notably, the instructor's objective must be to enable students to self-correct their erroneous utterances.

### **2.2.2. Types of Corrective Feedback**

Recasts, explicit feedback, elicitation, metalinguistic feedback, repetition, and clarification requests all are different corrective feedback techniques that have been exploited to refer to the ways how corrective feedback is provided. Originally, these techniques were coded by Chaudron (1977). Later, the six feedback techniques were organized by Lyster (2004) into three different categories of (1) recasts, (2) explicit correction, and (3) prompts. Then, Lyster and Ranta (1997) categorized again into 'reformulations' and 'prompts' (see Table 2.3). For instance, in Lyster and Ranta's (1997) framework, recasts and explicit correction were grouped under the classification of 'reformulations' because they supply the learner with either implicit or explicit information about the reformulation of erroneous utterances. Thus, the corrective feedback techniques in the

reformulation category are input-providing, and the corrective intent of these techniques is implicit.<sup>16</sup>

**Table 2.3 Corrective Feedback Techniques (Lyster & Ranta, 1997)**

Technique	Definition
Recasts	Teacher's reformulation of all or part of a students' utterance minus error.
Explicit Correction	Explicit provision of the correct form by the teacher.
Metalinguistic Feedback	Contains comments, information, or questions related to the well-formedness of the students' utterance, without explicitly providing the correct form.
Elicitation	Teachers either (i) elicit completion of their own utterance by strategically pausing to allow students to fill in the blank, or (ii) use questions
Repetition	Teacher's repetition, in isolation, of the students' erroneous utterance generally with a change in intonation (rising, in most cases)
Clarification Requests	Indicates to students either that their utterance has been misunderstood by the teacher or that the utterance is ill-formed in some way. Thus, a repetition or reformulation is required. Usually, a repetition or a reformulation is needed.

Teacher-Corrects

Student-Corrects

REFORMULATION

PROMPT

<sup>16</sup> This classification, however, is not definite, since studies have shown that both types of feedback may vary (Ellis & Sheen, 2006; Ellis et al., 2006; Egi, 2007). Alternatively, R. Ellis et al. (2006) attempted to categorize the corrective feedback types into input-providing (recasts and explicit correction) and output-pushing (prompts) (Lyster & Ranta, 1997).

Specifically, recasts refer to the teacher's reformulation of all or part of a learner's utterances, minus the error. Example (36) illustrates an example of recast.

(36) S: \*Why you don't like Marc?

T: Why don't you like Marc? (Lightbown & Spada, 2011, p. 46)

According to Tarone and Swierzbin (2009), when dealing with an erroneous utterance (e.g., '*More ease*. I feel, to learn ...'), recast is further divided into four subtypes: Isolated recast (e.g., '*Easier*'), Regular recast (e.g., 'So, you said it's *easier* now. '), Integrated recast (e.g., 'It's *easier* now. How difficult was it before?'), and Interrogative recast (e.g., 'How *easier* is it now?').

The relationship between the different properties of recasts and the rate of successful uptake (or self-repair) was reported by Sheen (2006): Her findings demonstrated that recasts that are short and readily recognizable (e.g., declarative, short, repeated, with a single error-focus, or involve substitutions instead of deletions or additions) for the learners are more likely to lead to successful uptake and self-repair. Her findings empirically support Lyster's (1998) claim that recasts focusing on a single error in a learner's incorrect utterances are more likely to lead to successful self-repair

(e.g., Chaudron, 1977). Furthermore, reduced recasts lead to more efficient noticing, a conclusion supported by more recent studies (e.g., Kartchava & Ammar, 2014).

Explicit correction provides learners with correct form with an explicit indication that their utterances were incorrect. (37) describes the example of explicit correction.

(37) S: The dog run \*fast~~ly~~.

T: 'Fastly' doesn't exist. 'Fast' does not take *-ly*.

You should say 'quickly'. (Lightbown & Spada, 2011)

Another innovation on the part of Lyster and Ranta (1997) concerns how they used the term 'prompt'. Traditionally, 'prompt' was used to be correlated with 'negotiation of form' (Lyster & Ranta, 1997). For Lyster and Ranta (1997), however, the prompt category is composed of four corrective techniques: (a) metalinguistic feedback, (b) repetition, (c) clarification request, and (d) elicitation. Most commonly, these techniques are used to force learners to recognize the corrective intent of feedback and to motivate the learners to make independent corrections for their errors. Thus, the types of corrective feedback within the prompt category are those pertaining to learners' output-promoting, and the corrective intent of these techniques is

explicit. Examples from (38) to (40) illustrate the subtypes of prompt (Lightbown & Spada, 2011, p.126-127).

(38) Metalinguistic Feedback

S: \*We look at the people yesterday.

T: What's the ending we put on verbs when we talk about the past?

S: /e-d/

(39) Repetition

S: \*I have three new toy.

T: Three new toy?

(40) Clarification Request

T: How often do you wash the dishes?

S: Fourteen.

T: Excuse me. (Clarification Request)

S: Fourteen.

T: Fourteen what? (Clarification Request with Repetition)

S: \* Fourteen for a week.

T: Fourteen times a week? (Reformulation: Interrogative Recast)

Special attention needs to be paid to elicitation, which refers to the use of corrective feedbacks that language instructors frequently employ to prompt students' self-repair. Lyster and Ranta (1997) defines elicitation as follows: (i) pausing (e.g., '*It's a ...*'), (ii) asking questions (e.g., '*How do we say X in English?*'), and (iii) asking students to reformulate their own utterances (see example (41)).

(41) Elicitation (asking students to reformulate their own utterances)

S: \*My father cleans the plate.

T: (Excuse me), he cleans the \_\_\_\_\_ ?

S: Plates?

These prompting strategies are used to encourage learners to capture the corrective intent of feedback and to assist the learners to self-correct erroneous component in their utterances. Thus, the corrective feedback techniques in the prompt category are intended to be output-promoting; therefore, the corrective intent of these techniques is explicit.

### **2.2.3. Noticeability in Corrective Feedback**

Corrective feedback (both implicit and explicit) is an extremely effective method for raising learners' awareness of their errors when they are learning a foreign language. By and large, prior research indicates that providing corrective feedback is more efficient for language learners than not offering it (Han, 2002; Russell & Spada, 2006; Mackey & Goo, 2007; McDonough & Mackey, 2006; Lyster & Saito, 2010).

Furthermore, research indicates that the degree of awareness, or 'noticeability' of corrective input, positively correlates with the efficiency with which a foreign language is acquired (Long, 1996; Long & Robinson, 1998; Schmidt, 1990, 2001). Schmidt (2001) asserted that "people learn about the things that they attend to and do not learn much about the things they do not attend to" (p. 30). In an ideal world, learners would be equally aware of the corrective intent of explicit and implicit strategies. However, since learners are more likely to detect explicit than implicit feedback, it is possible to jump to the hasty conclusion that explicit corrective feedback (e.g., recast) is more efficient than the implicit form (e.g., clarification request) (De Bot, 1996).

Nonetheless, such a speculative approach is likely to spark substantial controversy. For example, in classroom-based research, prompts are reported to be more effective than recasts (Ammar & Spada, 2006; Lyster, 2004; Loewen & Nabei, 2007; Yang & Lyster, 2010). Yet, in laboratory-based studies, recasts are found to be just as effective as prompts (Lyster & Izquierdo, 2009; McDonough, 2007). In another study, however, recasts were reported as more effective than prompts (Long, Inagaki, & Ortega, 1998; Mackey & Philip, 1998).

To summarize, prior studies suggest that explicit corrective feedback is more beneficial for the language learners than implicit form of corrective feedback. Nonetheless, numerous studies have also reported that both prompts and recasts are effective. To resolve such tensions, Lyster and Saito (2010) argued that the implementation of a variety of corrective feedback is likely to be more effective than the consistent use of only one type. In the following section, corrective recast—an example of the combination of both explicit and implicit corrective feedback—will be explored.

#### 2.2.4. Corrective Recast as a Stepwise Corrective Feedback

A potentially ideal corrective feedback strategy may be a *hybrid* corrective attempt, consisting of a prompt and a recast (Li et al., 2016). As illustrated below in Example (42), this hybrid feedback strategy (referred to as ‘corrective recast’) is the category of recast used in this study.

According to Doughty and Varela (1998), a *corrective recast* consists of a prompt that performs the dual function of drawing learners’ attention to an erroneous part in their speech, while also eliciting a student’s self-correction. For example, a teacher may repeat a student’s error by turning it into a question such as, *I think that the worm will go under the soil?* In addition, a corrective recast also includes a reformulation of the student’s error, but without eliciting a self-correction. For instance, the teacher might say, instead, *I thought that the worm would go under the soil.*

(42) Example of Corrective Recast (Doughty & Varela, 1998, p.124)

Jose: I think that the worm will go under the soil.

Teacher: *I think* that the worm *will go* under the soil? → Prompt

Jose: (no response)

Teacher: *I thought* that the worm *would go* under the soil. → Recast

Jose: *I thought* that the worm *would go* under the soil.

As shown in (42), the corrective recast combines the advantages of output-prompting and input-providing corrective feedback strategies (Lyster, 2004). Furthermore, it is important to note that the instructional effects of corrective recasts have been reported frequently in several relevant studies (Erlam & Loewen, 2010; Li et al., 2016; Quinn, 2014; Zhao & Ellis, 2020).

To cite one such study, Doughty and Varela (1998) reported that a corrective recasting during a lab experiment demonstrated substantial gains in learning: They investigated the instructional effects between corrective feedback and no-feedback on the acquisition of English past tense by adolescent ESL learners who attended a compulsory science course. The students were asked to produce written or oral reports about experiments in which they had participated. While those in the corrective recast group were provided with corrective feedback on their written or oral reports, the participants in the control group did not. The participants' knowledge of the target structure was evaluated three times using either a written or an oral task three times: before the instructional intervention (pretest), immediately after the instructional treatment (immediate posttest), and two months after that (delayed posttest). The results of the two posttests generally showed that the recast group attained significant gains, particularly in oral reporting. However, the progresses shown by the control group, which received no

focus on form instruction, was much less significant than those demonstrated by the corrective recast group.

Corrective recast is more efficient when it is provided immediately (e.g., ‘online’, or as soon as an error occurs) after erroneous utterances were produced.<sup>17</sup> For example, Li et al. (2016) investigated the comparative effects of online (i.e., immediate) and offline (i.e., delayed) corrective recasts in the learning of the English past passive construction by Chinese middle school English L2 learners. Participants were randomly assigned to an immediate feedback group, a delayed feedback group, a task-only group, and a control group. The two feedback groups—namely, immediate and delayed—received either immediate or delayed corrective feedback in the form of a corrective recast. Their performances were assessed using an oral elicitation task to measure their implicit L2 knowledge, and with a grammaticality judgment task to measure their explicit L2 knowledge. While both immediate and delayed corrected feedback groups revealed significant gains on the grammaticality judgment test, the immediate feedback group demonstrated a certain degree of advantage compared to the delayed feedback group.

---

<sup>17</sup> This is consistent with previous studies (Aljaafreh & Lantolf, 1994; Arroyo & Yilmaz, 2018; Long, 2015; Quinn, 2014) that immediate corrective feedback is more efficient than the delayed alternative.

To recap, corrective feedback can be classified as implicit or explicit. The degree to which corrective feedback is noticeable determines its instructional effect. Combining implicit and explicit corrective feedback can make such responses more noticeable. This hybrid strategy is referred to as ‘corrective recast’; it is most successful when offered immediately following student’s erroneous utterances. Given that the primary objective of this study is to examine the effects of chatbot-based corrective feedback on the L2 learning of English transitive resultative constructions, the following section will investigate previous research on the implementation and design of AI chatbots that provide corrective feedback, specifically corrective recast.

### **2.3. The AI Chatbot in Foreign Language Learning**

This section will examine the pedagogical characteristics of the AI chatbot and its instructional effectiveness for foreign language acquisition. The properties of non-communicative intelligent computer-assisted language learning (ICALL) will be discussed in Section 2.3.1. The subsections that follow will examine the instructional effectiveness of the communicative ICALL or AI chatbot for foreign language learning, depending on whether corrective feedback is absent (Section 2.3.2) or available (Section 2.3.3).

### **2.3.1. Non-communicative Intelligent Computer Assisted Language Learning (ICALL)**

According to Levy (1997), Computer-assisted Language Learning (CALL) is defined as “the search for and study of applications of the computer in language teaching and learning” (p.1). CALL includes a wide range of technological endeavors in teaching and learning of foreign languages, ranging from traditional drill-and-practice to more modern pedagogical manifestations, such as (a)synchronous computer-mediated communication (SCMC) and intelligent computer-assisted language learning (ICALL).

The majority of CALL research has focused on the investigation of computer-mediated communication (CMC). CMC studies are usually conducted on spoken or written discourse that occurs as learners engage in human-to-human interaction using communication platforms. In Asynchronous CMC (ACMC), learners are usually involved in written L2 tasks that utilize asynchronous communication technologies, such as email, blogs, social networking services, or discussion board, to promote human-to-human L2 conversation (Poza, 2005). In Synchronous CMC (SCMC), synchronous communication tools for text/voice chat or instant messaging services are used to elicit synchronous human-to-human L2 interaction

(Blake, 2009; Jepson, 2005; Satar & Özdener, 2008; Sykes, 2005; Volle, 2005; Wang, 2004).

Distinct from CMC, Intelligent Computer Assisted Language Learning (ICALL) involves human-computer interaction (HCI) by the practical application of computer science technologies for the teaching and learning of a second or foreign language (Gamper & Knapp, 2002; Thomas et al., 2013). ICALL integrates artificial intelligence technologies (e.g., ASR, NLP, machine learning) into CALL systems to produce foreign language learning software that interacts intelligently with students, responding flexibly and dynamically to their utterances or learning progress (Sentence, 1993).

Still, most ICALL systems are incapable of maintaining dyadic communicative discourse, thus constituting non-communicative ICALL. The nature of interaction by non-communicative ICALL systems is incomparable with oral interaction. That is, these systems do not provide opportunities for spontaneous conversational interaction. Instead, they provide learners with a modeling example that might best be characterized as an enhanced extension of traditional workbook practice. The following is an example of a non-communicative ICALL, 'E-Tutor' (Heift, 2004).

(43) Example Interaction with Heift's E-Tutor (Petersen, 2010, p.238)

**Sample Template: Anne / aus / Amerika / sein**

Learner: \* Anne aus America sein.

[Translation: \*Anne from Amerika is.]

System: *Anne aus **America** sein.* There is a spelling mistake with America.

*Suggestions: Amerika*

Learner: \*Anne aus Amerika sein.

[Translation: \*Anne from America is.]

System: *Anne ... aus Amerika **sein**.* Careful! There is a problem in word order. How about moving **sein**?

*See help on: Word-order*

Learner: \*Anne sein aus Amerika.

[Translation: \*Anne from is Amerika.]

System: *Ann sein aus Amerika.* You made a mistake with the verb **sein**. You need an inflected form of the present tense here.

*See help on: Verb, Sein*

Learner: Anne ist aus Amerika.

[Translation: Anne is from America.]

System: Prima!

Corrective feedback is one of the central features of non-communicative ICALL systems for foreign language learning. Few studies have

investigated the relationship between corrective feedback from non-communicative ICALL and L2 development; there is no consensus that corrective feedback technique is significantly more efficient than others. In this subsection, the instructional benefits of elicitation and metalinguistic feedback in a non-communicative ICALL system will be briefly examined.

First, elicitation was reported to be statistically more efficient than metalinguistic feedback or explicit correction for Korean L2 English learners (I. Choi, 2016; I. Choi & S. Jang, 2020). Based on the comparative analysis of corrective feedback, they created a learning-oriented diagnostic assessment (LODA). The LODA program was designed to provide three types of corrective feedback, namely, elicitation, metalinguistic feedback, and explicit correction. To maximize the instructional effect, LODA was based on a process-oriented method. In the first step, learners responded to a question, minus corrective feedback. In the second step, elicitation that employed a color-highlight was immediately provided. When the learners' utterances contained correct words, those were colored blue. When they exhibited incorrect words, those were colored red. If all the words were grammatically correct, they were colored purple. Next, metalinguistic feedback was provided utilizing the abovementioned color elicitation. Finally, the learners could check the correct answer (explicit correction), which was colored purple. The results strongly demonstrated that gains in

elicitation were statistically more significant than those taken from metalinguistic corrective feedback. A comparative investigation of the frequency and percentage of linguistic errors revealed that sentential structure was one of the most difficult constructs.

Second, several studies have reported that metalinguistic feedback is more efficient than other corrective feedback techniques. For example, Robo-sensei or Banzai developed by Nagata (1997, 1998, 2002) was a non-communicative ICALL system providing metalinguistic feedback that covered basic grammatical structures. Robo-sensei, whose core activity consisted of sentence production exercises, was used as a supplement to a textbook. In Nagata's study, learners were provided with a description of conversational backgrounds in English, as well as an English translation equivalent expression of what they were required to produce in Japanese. The system then provided immediate corrective feedback to learners' responses.

Nagata (1997) examined the instructional effects of explicit metalinguistic feedback and L1 translations delivered by Robo-sensei or Banzai. The feedback was informative in that it located not only the erroneous part but also provided an explanation for grammatical rules that had been violated. Nagata (1997, 1998, 2002) commonly reported that

metalinguistic feedback was statistically more efficient than L1 translations in the learning of Japanese particle expressions.

Similar to Robo-sensei, 'E-tutor', which was developed by Heift (2002, 2003, 2004) also employed a non-communicative ICALL system to provide metalinguistic feedback for foreign language learners' erroneous utterances. One of the characteristics of E-tutor is that its feedback delivery system maintains extended form-focused exchanges that can lead learners, step-by-step toward target-like production. In this study, the task design of E-tutor consistently required learners to formulate grammatically correct German sentences from syntactic templates that provided learners with lexical items (Heift, 2002, 2003, 2004). Learners had to use all the words in the displayed template to construct a morpho-syntactically correct sentence. For each practice item, learners could make an unlimited number of attempts at the target form until their respective utterance was deemed target-like. Utilizing this E-tutor system, Heift (2004) explored the relationship between various forms of corrective feedback and the number of learners' self-repair patterns. Based roughly on Lyster and Ranta's (1997) six categorizations for corrective feedback techniques, she identified two distinct feedback techniques generated by the E-tutor system: One was metalinguistic feedback and the other a repetition of learners' output, highlighting the erroneous part. The results of Heift (2004) clearly demonstrated that explicit

metalinguistic feedback may be more effective in promoting self-correction or uptake than is the case for highlighted repetitions.

To recap, LODA, Robo-sensei (or Banzai system), and E-tutor were briefly described as examples of non-communicative ICALL systems that provide learners with grammatical activities and detailed corrective feedback on form-related errors minus any communicative contexts. These systems are partly relevant for the present study since they illustrate state-of-the-art technologies regarding the form-related corrective feedback. They also provide evidence for the effects of different types of corrective feedback in a non-communicative ICALL system. Since the present study aims to provide corrective feedback in the communicative ICALL interaction context, the practical application and instructional effect of corrective feedback strategies in a communicative ICALL system will be described in the next subsection.

### **2.3.2. AI Chatbot without Corrective Feedback**

Most frequently-used communicative ICALL systems are artificial intelligence (AI) chatbots (or chatter bots), which interact with humans in

text-chat mode.<sup>18</sup> AI chatbot is a computer program that can perform a dyadic conversation (or chat) with humans in natural languages (Shawar & Atwell, 2007). One of the most well-known examples of a chatbot is ELIZA, which simulates a human psychiatrist (Weizenbaum, 1966). ELIZA was technically based upon rather simple pattern matching algorithms (e.g., Naïve-Matching, Automata Algorithm, Rabin-Karp Algorithm, Knuth-Morris Pratt (KMP) Algorithm, and Boyer-Moore Algorithm) to generate a response to human utterances (Deng & Liu, 2018; Goodfellow et al., 2016). Despite the simplicity of the underlying algorithms, ELIZA can maintain a coherent conversation. Still, the AI chatbots, including ELIZA, at times exhibited a peculiar behavior which was due to the technological shortcomings of that era. Due to these technological shortcomings, the development of chatbots dramatically declined during the so-called ‘dark ages of AI (1974–1993)’ (Goodfellow et al., 2016).

The development of chatbots has been accelerated with the advancement of deep learning technologies in the early twenty-first century. In addition, a rising number of AI chatbots that use state-of-the-art Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) engines are

---

<sup>18</sup> As Bibauw et al., (2019) noted, research on the practical use or instructional effect of AI chatbots for foreign language learning should be related to the general research agenda of ICALL. Thus, in this study, the terminology of the AI Chatbot and communicative ICALL will be interchangeably used.

currently on the market in the domain of foreign-language learning. However, because deep learning technology is still in its infancy, many existing chatbots for foreign language learning perform strangely and appear to be non-human agents. The following is an example of *Jabberwacky*, an AI Chatbot created for an AI assistant who would typically be uninterested in learning a foreign language.

(44) Example Interaction with *Jabberwacky* (Bibauw et al., 2019)

User: Let's talk about character!

System: No, let's talk about politics.

User: I think that I have a character hating to lose.

System: Everyone has a bad period from time to time.

User: And I am very careless. Are you?

System: No.

User: Good. There are many careless people in my friend.

System: There are many problems too.

Recent studies of AI chatbots' instructional impact on foreign language acquisition have mostly focused on communicative ICALL systems, which largely ignore linguistic errors in learner input and instead focus on preserving conversational flow. This is done to provide learners with as many opportunities as possible to negotiate meaning in the target language (N. Kim, 2017). In particular, most chatbots provide feedback about meaning that focuses on (i) a confirmation check, (ii) a comprehension check, (iii) a clarification request, (iv) repetition, and (v) a reformulation (or paraphrasing a preceding utterance) (Long, 1983).

Two representative communicative ICALL systems, the Military Tactical Tutor (MILT) and the Tactical Language Training System (TLTS), will be reviewed. Additionally, the instructional effect of communicative ICALL systems or AI chatbots lacking corrective feedback on L2 development also will be examined.

First, MILT was a traditional communicative ICALL system developed by the Army Research Institute for training US soldiers in Modern Standard Arabic (Kaplan et al., 1998; Holland et al., 1999). This system relied on a gamified 3D virtual microworld in which learners could control a virtual agent by using written or spoken commands. The objective of the interaction between the learners and the system was defined by the problem to be solved (i.e., '*Where will the enemy attack?*').

MILT had two versions of the system, namely, text and speech. In the speech-based system, it assessed pronunciations accuracy by making learners read pre-defined sentences. In the text-based system, the learner could freely produce Arabic expressions, but the system could only understand commands related to the objectives of a particular task. By employing syntactic and semantic analysis, the system could provide feedback to learners via a help-window (Kaplan et al., 1998) or through a reaction by a virtual character. If the command from the learner could be interpreted and correctly formulated, the virtual character behaved as intended. Otherwise, the character acted unexpectedly or might say like, “*I don't understand.*” MILT deployed many interesting pedagogical activities. With regard to the instructional effects, however, the developers (Kaplan et al., 1998; Holland et al., 1999) did not report what type of feedback learners had received prior to mispronouncing words or sentences.

Second, and similar to MILT, TLCTS was developed for instruction in Arabic, Persian, and other languages relevant to U.S. Armed Forces (Johnson & Valente, 2009). This system mainly teaches cultural knowledge of the relevant regions (e.g., the Middle East) in addition to language skills. The system incorporated two complementary components, the skill builder, and the practice environment. According to Johnson and Valente (2009), initially, in the skill builder stage, learners were provided with form-focused

instruction and individual feedback on pronunciation and grammatical errors. Next, learners participated in a practice environment stage, which is a virtual world with 3D landscapes. In this pedagogical stage, learners interacted with animated characters in localized virtual 3D villages to accomplish their respectively assigned mission.

Learners' production of the target language was not constrained; they could freely speak what they wanted to say or what they had learned in the skill builder stage. No grammatical corrective feedback was provided to the learners. Instead, feedback provided in the practice environment was meaning-based; if the learner's utterance was unintelligible or inappropriate, they would not be understood by the villager character. Learners also were supported by a particular pedagogical-assistant character that offered hints that acted to advance the game. The hints are designed to be specific to the stage and knowledge of the learner.

Many studies presented encouraging results regarding the effectiveness of communicative ICALL systems without corrective feedback. For example, the participants who took part in the Tactical Iraqi TLCTS demonstrated significant mean score increase in both linguistic and cultural knowledge, with considerable learning gains in the holistic oral proficiency (Surface et al., 2007). Also, using a free talking CSIEC (Computer Simulator in Educational Communication) chatbot, middle school students demonstrated

major improvements to their regular English examination scores (Jia et al., 2013). N. Kim (2016) reported that the application of Indigo, another free-talking chatbot, was as effective as a human interlocutor in improving holistic speaking proficiency. She also observed that the use of a free talking chatbot, Replika, showed a statistically significant improvement on the enhancement of holistic English grammatical skills, a level of improvement beyond that of the human instructor (N. Kim, 2016).

These insightful findings are not without caveats: According to Bibauw et al. (2019), research on artificial intelligence chatbots for foreign language learning has completely overlooked the importance of goal-oriented dialogues (i.e., accomplishing a task or providing correct answers to a set of questions). Thus, the research on AI chatbots for foreign language learning is relatively disconnected from mainstream ICALL research trends (Bibauw et al., 2019). This disregard for pedagogical considerations in research on AI chatbots for foreign language learning could be due in part to a dearth of evidence demonstrating a direct link between human-chatbot interaction and L2 development. To establish this connection, Mackey (1999) emphasized the importance of developing experimental designs that allow for direct assessment of the target linguistic structures that may occur during interaction. In this vein, it is critical to create experimental tasks that may expose learners to target structures in a variety of communicative

situations. Unfortunately, few studies on communicative ICALL have attempted to go beyond describing the characteristics of human-chatbot interaction and to investigate its effect on the L2 development of the target structure.

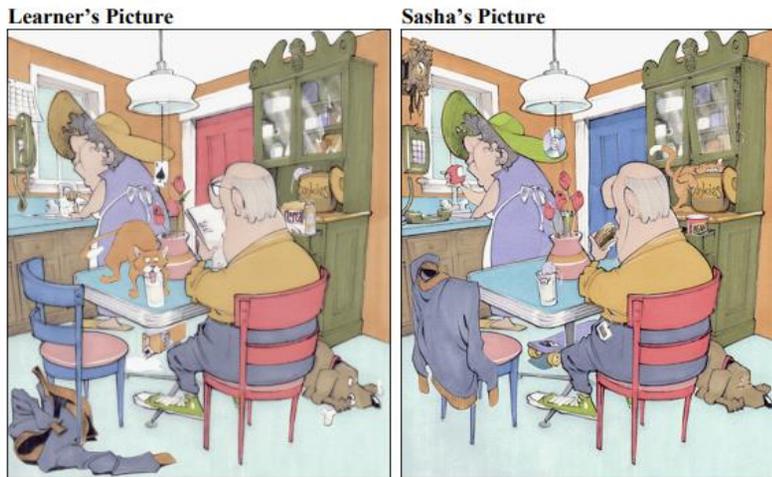
### **2.3.3. AI Chatbot with Corrective Feedback**

This subsection examines three representative examples of AI chatbots providing corrective feedback in terms of their instructional effects. The first system that incorporates a communicative approach with form-related feedback is *Te Kaitito*, which is used for the instruction of the pronoun system in Maori (Knott & Vlugter, 2008; Vlugter et al., 2009). *Te Kaitito* is a bilingual communicative ICALL system or AI chatbot in the sense that it not only simulates the learners in a Maori language conversation, but also provides metalinguistic instructions in English. The conversational topics were organized into lessons, which were associated with particular sets of grammatical forms to be learned during that specific lesson. The unpredictability from the open-endedness of the conversations was reduced by asking questions of the user, and successfully provided corrective feedback (Vlugter et.al., 2009). Corrective feedback is given in the form of explicit corrections of the form “*I think you mean X.*”

The instructional efficiency of *Te Kaitito* is evaluated regarding how it affects a learner's knowledge about Maori pronouns (Vlugter et al., 2009). The instructional effects were compared to those of learners who received regular human teacher-based instruction and a control condition of learners who did not receive instruction. The results reveal that the instructional effects of the *Te Kaitito* were as efficient as those of the human instructor, and significantly outperformed the control group (Vlugter et al., 2009).

Another example is an AI Chatbot, *Sasha*, developed by Petersen (2010). This study examined the developmental effects on the learning of English question construction and morphosyntactic accuracy. As seen in Figure 2.1, Petersen (2010) compared the instructional effects of text-based recasts delivered by an ICALL system with those delivered in spoken-based recasts provided by a human teacher or an AI system. Fifty-six high school ESL learners participated in a series of communicative tasks—namely, spot-the-difference—with either a native English speaker or an intelligent virtual interlocutor, *Sasha*. He found a positive instructional effect for English learners' question construction development in both modes, but no difference between the spoken and textual modes. Furthermore, he suggested that the implicit corrective feedback (recasts) had equally positive instructional effects on foreign language development in a text-based

human-computer interaction as was the case in a face-to-face oral conversation with a native speaker.



**Learner:** Is he reading the news?  
**Sasha:** I don't think that there is a news in this picture.  
**Learner:** \*Does the cat drinking the milk?  
**Sasha:** *Is the cat drinking the milk?*  
**Sasha:** No, it is not.  
**Learner:** Is there a jacket on the floor?  
**Sasha:** No, there is not.  
**Learner:** \*Does the men have green shoes?  
**Sasha:** *Do the men have green shoes?*  
**Sasha:** Yes, they do.

**Figure 2.1** Example of *Sasha* (Petersen, 2010, p.239)

The final example is the communicative ICALL system providing corrective feedback, which was developed by Wilske (2015). The system was intended to teach German dative prepositional phrases and causal subordinate clauses to GSL (German as a Second Language) learners. Wilske (2015) compared form-focused activities (e.g., fill-in-the gap activity)

to elicit learners' utterances to supply a grammatical target form with task-based dialogues (e.g., giving-direction task, setting-appointment task). For the latter, the instructional effects between implicit (e.g., recast) and explicit (e.g., metalinguistic feedback) types of feedback also were compared, as shown in (45).

(45) Examples of corrective feedback (Wilske, 2015, p.149)

a. Recast and Prompt

S: Gehen Sie vor das Café nach links.

'Turn left, in front of the coffee-shop.'

AI: Okay, [ vor dem Café nach links, ] Recast

[ und dann? ] Prompt

'Okay, left in front of the coffee-shop, and then?'

b. Metalinguistic Feedback

S: Gehen Sie vor das Café nach link.

'Turn left, in front of the coffee-shop.'

AI: ['das' in 'das Café' ist nich richtig.] Metalinguistic Feedback

'das in 'das café' is not correct.'

[Bitte noch einmal!] Prompt / [Nimm Dativ!] Prompt

'Please try again!' / 'Use the dative!'

The findings of her study revealed that constrained, explicit form-oriented instruction (e.g., fill-in-the gap activity) resulted in greater immediate learning gains. On the other hand, when freer, more implicit, and meaning-oriented instruction was given, more delayed instructional effects were found. Similarly, by comparing the learning gains between implicit recast and metalinguistic feedback, no significant differences in the learning of the language were found between the two groups in the immediate posttest. However, recast feedback led to increased learning gains in the delayed posttest.

These three systems all imply that providing corrective feedback (e.g., recast, metalinguistic feedback, and explicit correction) via an AI Chatbot helps learners to acquire a particular morphosyntactic or grammatical construction. Despite the pedagogical potential of AI chatbots for foreign language learning, they have several practical limitations because of insufficient integration of innovative but still premature state-of-the-art artificial intelligence technology into the traditional SLA/CALL framework (Thomas et al., 2013). First, previous research has neglected to emphasize the importance of teaching English argument structure construction which is typologically distinct from the learners' native language. Additionally, previous research has concentrated on providing learners with a single piece of corrective feedback but has overlooked providing them with a loop of

corrective feedback (e.g., Corrective Recast), which has been shown to be more effective than providing learners with a single piece of feedback. Finally, prior research on AI chatbots has focused exclusively on the instructed target structure, ignoring the pedagogical value of extending the instructed construction to its structurally related but uninstructed sentence structure.

Noting that previous attempts to develop a communicative ICALL system or AI chatbot have suffered from the aforementioned shortcomings, the current study will devote systematic attention to the development of a novel AI chatbot by triangulating recent findings from technological considerations, linguistic theories, and empirical evidence from SLA studies (N. Ellis, & Bogart, 2007; M. Sung et al., 2014). This novel AI chatbot is expected to be as effective as—if not more so—than a human instructor by providing learners with a loop of corrective feedbacks (or corrective recasts) to (a) improve their comprehension and production of adjectival transitive resultative constructions, which have been observed to be particularly difficult for Korean L2 English learners, and (b) to assist learners in generalizing such knowledge to unfamiliar prepositional transitive resultative constructions.

## 2.4. Summary of the Literature Review

It has been well established that the syntactic contrasts in adjectival and prepositional VtR between English and Korean are straightforward. The former allows the matrix verb to take a complete Small Clause as its complement in a mono-clausal formulation, but the latter allows more diverse syntactic options. To produce adjectival or prepositional VtR expressions in Korean, Korean native speakers rely on bi-clausal formulation to denote manner-result interpretation. Korean native speakers also may use a mono-clausal adjectival or prepositional VtR with a serial verb construction in which English-like manner-result interpretations are not available. Moreover, the matrix verb in Korean mono-clausal adjectival or prepositional VtR expression has an adjunct relationship with a Small Clause which is devoid of an overt NP subject, leading to the violation of direct object restriction (DOR).

Considering syntactic mismatches between Korean and English, Korean L2 English learners are predicted to have difficulty comprehending and producing mono-clausal English adjectival and prepositional VtR expressions. Korean L2 English learners may have difficulty in constituting a complementhood between a matrix verb and a complete small clause with

an overt subject. In comprehending English adjectival and prepositional VtR, Korean L2 English learners may have a difficulty in accepting grammatically correct English expressions, and rejecting grammatically incorrect ones. Similarly, in producing English adjectival and prepositional VtR, Korean L2 English learners are expected to generate persistent errors (e.g., violating DOR) especially for unselected VtR expressions to be realized in mono-clausal formulations which are manifested as bi-clausal expressions in their L1 system.

The syntactic divergences between Korean and English in adjectival and prepositional VtR lead to a semantic contrast with respect to the interpretation of telicity. Simply put, English resultatives denote a telic event, but Korean resultatives may denote an atelic interpretation. This strongly suggests that pedagogical intervention is required for Korean L2 English to acquire not only the syntactic structure but also the semantic conditions of English resultatives which imply telicity. Thus, Korean L2 English learners are predicted to have difficulty rejecting atelic English adjectival and prepositional VtR expressions.

L1 transfer errors by Korean L2 English learners are predicted to be minimized through instructional intervention. According to Lightbown and Spada (2011), it is important to draw students' attention to crosslinguistic contrasts by means of explicit instruction of the contrastive information

between L1 and L2. Another noticing technique, corrective feedback, also is predicted to be effective in eliciting learners' correct use of the target language (Kartchava & Ammar, 2014).

Although many AI chatbots have been invented and introduced for foreign language learning, they still have limitations. First, many of them merely provide learners with mechanical pattern drills devoid of communicative contexts. Second, most of them do not aim to help learners acquire particular linguistic structures, such as the adjectival VtR referred to in this study. Finally, they could not provide learners with corrective feedback in a way that the learners could self-correct their errors. This study, however, overcame such limitations with the development of an AI chatbot for foreign language learning, by triangulating an operational ontology with recent findings in linguistics, second language acquisition, and computer science. Thus, this study is designed to test the following hypotheses on the research question (RQ) outlined in Chapter 1:

**Research Question 1:** Is chatbot-based instruction as effective as human-based instruction (e.g., corrective feedback from a human instructor or additional explicit instruction of video material) in improving Korean high school English learners' comprehension of English transitive resultative constructions?

(1a) Is chatbot-based instruction as effective as human-based instruction in improving Korean high school English learners' comprehension of instructed English adjectival transitive resultative construction?

(1b) Is chatbot-based instruction as effective as human-based instruction in improving Korean high school English learners' comprehension of uninstructed English prepositional transitive resultative construction?

**Hypothesis 1 for RQ 1:** This study hypothesizes that with the instructional intervention of English adjectival VtR, Korean L2 English learners are expected to comprehend adjectival VtR expressions in a more target-like way (Lightbown & Spada, 2011). Also, the instructional treatment of English adjectival VtR is expected to be generalized to its interrelated construction, prepositional VtR.

**Research Question 2:** Is chatbot-based instruction as effective as human-based instruction (e.g., corrective feedback from a human instructor or additional explicit instruction of video material) in improving Korean high school English learners' production of English transitive resultative constructions?

(2a) Is chatbot-based instruction as effective as human-based instruction in improving Korean high school English learners' production of instructed English adjectival transitive resultative constructions?

(2b) Is chatbot-based instruction as effective as human-based instruction in improving Korean high school English learners' production of uninstructed English prepositional transitive resultative constructions?

**Hypothesis 2 for RQ 2:** This study hypothesizes that Korean L2 English learners transfer their L1 syntax-semantics representations in their production of English adjectival and prepositional VtR. Thus, it is predicted that Korean L2 English learners will have difficulty producing English adjectival and prepositional VtR with correct syntactic and semantic knowledge. Therefore, Korean L2 English learners may produce alternative constructions available in their L1 to express a resultative meaning in English. Specifically, they may experience a significant difficulty producing

unselected VtR in English, since Korean VtR does not require the matrix verb to take a subject-filled small clause as its complement. In receiving the instructional treatment for English adjectival VtR, Korean L2 English learners are expected to produce selected and unselected adjectival VtR in a more target-like way, which also will be generalized to selected and unselected prepositional VtR.

This study aims to contribute to the discussion by including both a comprehension and a production task in the experiment, and subsequently investigating potential asymmetry between the comprehension and production of learning English adjectival and prepositional VtR. The details of the instructional treatment (e.g., chatbot, human, and video-based instruction) and the experimental procedures (e.g., the Acceptability Judgment Task and the Elicited Writing Task) will be dealt with in the following chapter.

## **Chapter 3. Methodology**

This study aimed to investigate the effects of AI chatbot-based instruction on Korean high school L2 English learners' acquisition of English transitive resultative constructions (VtR). To this end, this study measured learners' ability to comprehend and produce target constructions by using different types of tasks. This chapter presents a general overview of the methodological considerations and research design employed in the study. It begins with a description of the participants, target structures, and procedures, followed by the instructional instruments, tests, and methods for data analysis.

### **3.1. Participants**

A total of 128 volunteer participants from three 10<sup>th</sup>-grade classes in a boys' high school were recruited for the present study during the spring semester of 2021. Initially, 139 students volunteered, but 11 were excluded from the analysis because they failed to complete the instructional procedures, or the tasks given in the tests. Their English language learning experiences at public school began when they were in the third grade of

elementary school, but almost two-thirds of the participants ( $n = 77$ , 60.1%) started learning English earlier than this through private education (onset age of L2 English learning:  $M = 9.09$ ,  $SD = 1.36$ ). Three of the participants had previously lived in English-speaking countries (i.e., the USA, England, and the Philippines), with a mean length of stay of about 1.5 years. However, the pretest revealed that these participants had limited knowledge of English transitive resultative constructions, as they made similar errors to those who had not lived abroad. Their proficiency was measured by employing the Michigan test, and all participants were judged to be in the low proficiency group as their scores indicated fewer than 22 correct answers out of 48 test items (S. Kim, 2016).

All the participants were receiving English language instructions at school at the time of the study. In the regular school classes, the participants did not receive any instruction on English VtR expressions. Many of them reported that they attended cram schools. The participants were divided into a control group and three treatment groups. Participants in the control group did not receive any instructional treatments. Students in the treatment group commonly received explicit instructions for a total of 12 lessons. They commonly watched an explicit 5-minute instruction video on English transitive resultative construction with an adjectival result phrase. The purpose of watching the 5-minute instruction video was to provide the

participants in the treatment groups with syntactic and semantic information on English adjectival VtR (e.g., form and meaning pairings of adjectival VtR, presentation of telic event, and the (un)availability of the modification with time adverbial expressions). The participants were then further divided into three groups depending on which type of instruction they chose to receive: a chatbot-based instruction group in which corrective feedback was provided via a text-based artificial intelligence (AI) chatbot (henceforth, ‘chatbot group’), a human instructor-based group in which corrective feedback was delivered by a human instructor (henceforth, ‘human group’), or a video group in which students were given additional explicit instruction by watching a 15-minute-video (henceforth, ‘video group’).<sup>19</sup> All corrective feedback and additional explicit instruction mainly were focused on the adjectival VtR and figurative VPC; no prepositional VtR expressions were dealt with.

The selection of this particular group of participants (i.e., 10<sup>th</sup>-grade Korean high school students) was intended to address a serious gap in instructional research regarding text-based AI chatbot instruction by which the majority of the participants in previous studies that examined the

---

<sup>19</sup> The instructional effect of corrective feedback from AI and a human instructor is investigated by means of a comparison between the chatbot group [CF from AI chatbot] and the human group [CF from human]. Also, the instructional role in the availability of corrective feedback is comparable between the chatbot/human group [+corrective feedback] and the video group [-corrective feedback].

instructional effects of AI chatbots for the foreign language learning were limited to elementary school students (Choo & Min, 2019; Hong et al., 2021; Im et al., 2019; S. Kim & S. Lee, 2020; K. Lee et al., 2011; S. Lee et al., 2011; Noh et al., 2012; M. Sung, 2020; Yang et al., 2019) or university students (e.g., N. Kim, 2016; Petersen, 2010; Wilske & Wolska, 2011). In addition, few studies have investigated whether English adjectival VtR can be learned by secondary school English learners via an AI chatbot.

This research gap seems particularly problematic, given that adjectival VtR is so complex that the persistent and systematic provision of corrective feedback is required to learn the target structures (S. Kim, 2016). Therefore, this dissertation takes L2 secondary school learners as its participants to identify how effectively transitive resultative constructions can be taught to these L2 secondary school students through AI chatbot-based instruction. Participants' knowledge of English adjectival VtR before the instructional treatment was measured by a pretest that consisted of two tasks: the Acceptability Judgment Task (AJT) and the Elicited Written Task (EWT). See Appendix 1 for the complete list of the test items.

The results indicate that the four groups were homogeneous regarding their knowledge of English transitive resultative constructions, as one-way ANOVA tests found no significant differences among the four groups: in AJT for adjectival VtR ( $M = 3.85$ ,  $SD = 1.29$ ,  $F(3, 124) = .752$ ,  $p = .521$ ),

prepositional VtR ( $M = 3.83$ ,  $SD = 1.36$ ,  $F(3, 124) = .416$ ,  $p = .742$ ) and in EWT for adjectival ( $M = 2.89$ ,  $SD = 2.06$ ,  $F(3, 124) = .086$ ,  $p = .771$ ) and prepositional VtR ( $M = 3.71$ ,  $SD = 2.38$ ,  $F(3, 124) = .003$ ,  $p = .959$ ).

### **3.2. Target Structure**

English Adjectival VtR has been reported to be notoriously difficult for Korean L2 English learners. According to the National English Education Curriculum (2015), English adjectival VtR is introduced mainly for middle school English learners. S. Kim (2016) discovered that Korean L2 English learners in high-school or university responded that almost 80% of Korean high school English learners had encountered adjectival VtR expressions in school textbooks. However, several studies have reported that Korean L2 learners remain insensitive to the syntactic and semantic properties of the adjectival VtR even after explicit instruction. This may require Korean L2 English learners to acquire the relevant construction via corrective feedback in dyadic interactional contexts.

In the present study, the students in the three experimental groups (e.g., the chatbot, human, and video groups) were instructed on the adjectival VtR. Table 3.1 illustrates the instructional items for adjectival VtR. The

instructional items include (i) selected adjectival VtR with a figurative particle complement (figurative verb particle construction, henceforth F-VPC), (ii) selected adjectival VtR with an adjectival complement, and (iii) an unselected adjectival VtR with an adjectival complement.

**Table 3.1 List of Instructional Items for Adjectival VtR**

Result Phrase		Matrix Verb	Selected adjectival VtR	Unselected adjectival VtR
Figurative Particles	Apart		Push/rip NP apart	
	Completion (Up)		Make/set/shut NP up	
	Stop /Termination (Off)		Close/shut/turn NP down	
	Abstract Entity (In)		Bring/put/take NP in	<i>n.a.</i>
	Displacement /Removal (Out)		Cross/put/turn NP out	
	Continue (On)		Carry/keep/pass NP on	
	Start (Active) (Off)		Set/start/trigger NP off	
Adjectives	Awake		Kick/Kiss NP awake	
	Black		Paint NP black	
	Clean		Polish NP clean	
	Deaf		Strike NP deaf	
	Dry		Bleed/suck NP dry	
	Empty		Bang/knock NP empty	
	Hoarse			Shout oneself hoarse
	Ill			Worry oneself ill
	Insane		Drive NP insane	
	Open		Bust/cut/wrench NP open	
	Shut		Snap/squeeze/zip NP shut	
	Silly			Laugh oneself silly
	Smooth		Rub NP smooth	
	Sore		Scratch NP sore	
Thin		Roll/spread NP thin		
Unconscious		Batter NP unconscious		

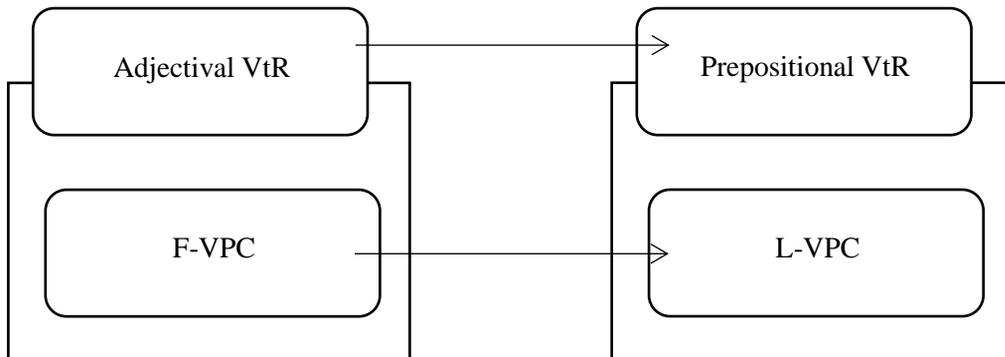
*Note.* *n.a.* represents that the relevant linguistic items were excluded from the instruction.

The instructional items used in the lessons were inspired by numerous studies, as the English adjectival VtR is one of the most difficult constructions for Korean L2 English learners; it is a construction that rarely is covered in English textbooks used in Korean secondary educational contexts (e.g., Bowerman et al., 1995; Choi & Bowerman, 1991; S. Kim, 2016; Olson, 2013; H. Sung, 2018; M. Sung, 2018; Tomasello, 1987, 1992).

There is a syntactic and semantic relationship among figurative verb particle construction (F-VPC), literal verb particle construction (L-VPC), adjectival VtR and prepositional VtR.<sup>20</sup> Several previous studies have reported that the F-VPC is syntactically or semantically correlated with adjectival VtR construction and prepositional VtR construction (Diesessl & Tomasello, 2005; Golarch, 2004; Goldberg, 1995, 2015; Thim, 2012; Goldberg, 1995, 2015). Although F-VPC is semantically distinct from literal verb particle construction (L-VPC), it is certain that the two constructions share syntactic and semantic properties, as adjectival and prepositional VtR constructions do. In this respect, as is illustrated in Figure 3.1, adjectival VtR is to prepositional VtR what F-VPC is to L-VPC (M. Sung, 2018). Considering the interrelatedness among the constructions, Figure 3.1 illustrates the interrelatedness among prepositional VtR, adjectival VtR, and F-VPC.

---

<sup>20</sup> Refer to Section 1.2 for the definition of each construction.



**Figure 3.1 Syntactic and Semantic Interrelatedness Among Adjectival VtR, F-VPC, Prepositional VtR, and F-VPC**

Figure 3.1 demonstrates that learning of adjectival VtR and F-VPC may affect the learning of prepositional VtR and L-VPC. Or it is highly likely that learning the constructional properties of adjectival VtR may extend it to a generalized understanding of prepositional VtR with the scaffold of F-VPC.

### 3.3. Procedure of the Study

Data were collected from 3<sup>rd</sup> May to the 16<sup>th</sup> of July 2021 during afterschool hours. In the first week, participants took a pretest that included the AJT and the EWT. From the second to the fifth weeks, all groups except the control group completed 15 instructional sessions on English adjectival VtR. During these three weeks, learners were provided with instructional material five times a week. Each lesson took approximately 25 minutes to compare. In the fifth week, all students took an immediate posttest. A delayed posttest was administered in the ninth week, with a 4-week interval from the immediate posttest (cf. Rah, 2014; M. Sung, 2018). All the pretest items were minimally adjusted for immediate and delayed posttests. The overall procedure is illustrated in Table 3.2.

**Table 3.2 Procedures of Instruction and Testing**

Week	Instructional Session	Chatbot Group	Human Group	Video Group	Control Group
1			Pretest (AJT, EWT)		
2 – 4	1~15	<b>Chatbot-based instruction</b>	<b>Human-based instruction</b>	<b>Video-based instruction</b>	<b>No instructional treatment</b>
5			Immediate posttest (AJT, EWT)		
6-8			Interval		
9			Delayed posttest (AJT, EWT)		

It should be noted that the chatbot group learners received a series of corrective feedback without time delay for a single erroneous part. The human group participants also received corrective feedback, but it was provided once, and there was a time delay between the production of erroneous utterances and the provision of corrective feedback from the human instructor. Finally, the video group participants did not receive any corrective feedback, but received additional explicit instruction which may or may not have been relevant to their production of errors.

### **3.4. Instructional Material Shared by the Experimental Group**

This section primarily describes the overall structure of educational sessions and the arrangement of instructional materials that were uniformly provided to the instructional group learners (i.e., the chatbot, human, and video groups). Section 3.4.1. describes the general framework of the instructional session. In the next two sections, the video material and the written instructional material will briefly be reported.

### **3.4.1. General Framework of the Instructional Session**

All experimental group participants received uniform or group-universal pedagogical activities with an instructional video or written instructional material. The participants in the three treatment groups participated in three tasks distributed over the course of the instructional sessions. Table 3.3 briefly illustrates the general framework of the 15 instructional sessions. Approximately 6 hours and 15 minutes (25 minutes  $\times$  15 sessions) of instruction on adjectival VtR were implemented for the experimental group learners (i.e., chatbot, human, and video groups). One instructional session involved 5 minutes for watching an explicit instruction video and 5 minutes for completing the task, followed by 15 minutes of participating in post-written instructional material activities. The post-written instructional material activities include participating in the pedagogical activities of receiving corrective feedback from an AI chatbot, submitting what the participants performed in the written instructional material activities and having errors corrected by the human instructor, or watching an additional explicit instructional video. After finishing the post-written instructional material activities, learners proceeded to the next instruction session. The description of the post-written instructional material activities is presented in Section 3.3.

**Table 3.3 General Framework of Instructional Video and Written Instructional Material Activity**

Target Structure	Unit in Written Instructional Material	No. of Session	Instructional Video	Written Instructional Material Activity	
				Controlled Practice & Dialogue Completion Task	Drafting a school newspaper article
Selected adjectival VtR with Figurative Particle Complement	I	1	•	•	
	II	2	•	•	
	III	3	•	•	
Selected adjectival VtR with Adjectival Complement	IV	4	•	•	
		5			•
	V	6	•	•	
		7			•
		8	•	•	
Unselected adjectival VtR with Adjectival Complement and Fake Reflexive	VI	9			•
	VII	10	•	•	
		11			•
	VIII	12	•	•	
		13			•
	IX	14	•	•	
		15			•
<b>Total</b>			9	9	9

*Note.* • represents that the relevant instructional material (e.g., instructional video and textbook activity) was provided to language learners for instructional session. After finishing the written instructional material activities, learners began post-written instructional material activities by participating in AI chatbot activity, receiving corrective feedback from the human instructor, or watching additional instructional videos.

As Table 3.3 indicates, nine instructional videos were provided along with the controlled practice or dialogue completion tasks in the written instructional material a total of nine instructional session (e.g., Session 1, 2, 3, 4, 6, 8, 10, 12, 14). When the instructional video was not provided (e.g., Session 5, 7, 9, 11, 13, 15), learners were given tasks to draft school newspaper articles. For example, in instructional session 1–3, when the experimental group learners were exposed to figurative VPC, only instructional video and controlled practices were provided. Subsequently, in instructional sessions 4-15, when learners were exposed to selected or unselected adjectival VtR, they were provided with controlled practices and drafted school newspaper articles every other day.

Participants in this study were instructed using an instructional video for English adjectival VtR to provide the participants in the treatment groups with the syntactic and semantic information of English adjectival VtR (e.g., introduction of new vocabulary, presentation of form and meaning pairings of adjectival VtR, demonstration of telic event, and (un)availability of the modification with time adverbial expressions). They were then asked to complete three written instructional material activities: controlled practice (CP), a dialogue completion task (DCT), and drafting school newspaper (DSN). A more detailed description of the instructional material will be presented in the following sections.

### **3.4.2. Instructional Material Shared by Experimental Groups**

#### **3.4.2.1. Instructional Video**

The primary objective of the instructional video is to provide learners with the presentation of (i) a list of new vocabulary, (ii) form and meaning pairings in adjectival VtR expressions, (iii) visual stimuli for telic events, and (iv) compatibility between telic events and the modification of time adverbial expressions.

As previously described, the pedagogical objectives of watching explicit instructional videos commonly shared by the three treatment groups were to provide participants with form and meaning pairings of English adjectival VtR. The instructional videos were composed of six components: (i) presentation of lexical items, (ii) presentation of constructional events of adjectival VtR with visual stimuli, (iii) presentation of constructional schema for adjectival VtR, (iv) presentation of a telic event, (v) presentation of positive evidence of telicity in adjectival VtR (i.e., grammatical *in-modifications*) and (vi) presentation of negative evidence of telicity in adjectival VtR (i.e., incompatibility with cancellable expressions or for-modifications). The instructional video was 5 minutes in length.

Instruction was provided in learners' first language (i.e., Korean) and was

narrated by the researcher, who was personified as *Bird-Head*, the main character in the YouTube cartoon titled *Beast Friends*. Bird-Head is both a humorous and genius-level pigeon who speaks Korean. As the character is well known to the students, he was introduced to the instructional video as a way to enhance their learning motivation. The license to use the character for academic purposes was granted by its original creator.

In the stage of presentation of lexical items, new words were presented to learners. In the stage of presentation of constructional events of adjectival VtR with visual stimuli, learners were provided with a visual cue. Guided by this visual guidance, learners were instructed on semantic roles, such as the causer in a subject, the affected theme in an object, the change of state or location in a result phrase, and the manner of action in a main verb. In the next phase, learners learned how to package English adjectival VtRs that conformed to constructional schema. Based upon this activity, learners were required to identify whether the scene described telic interpretation with a focus on an adjectival result phrase. Learners also were tested on how *in-* and *for-*modification affected the grammaticality of a sentence. After watching the video, learners were required to complete practice tasks in provided written instructional material. The pedagogical features of the instructional video reviewed thus far are briefly summarized in Figure 3.2.

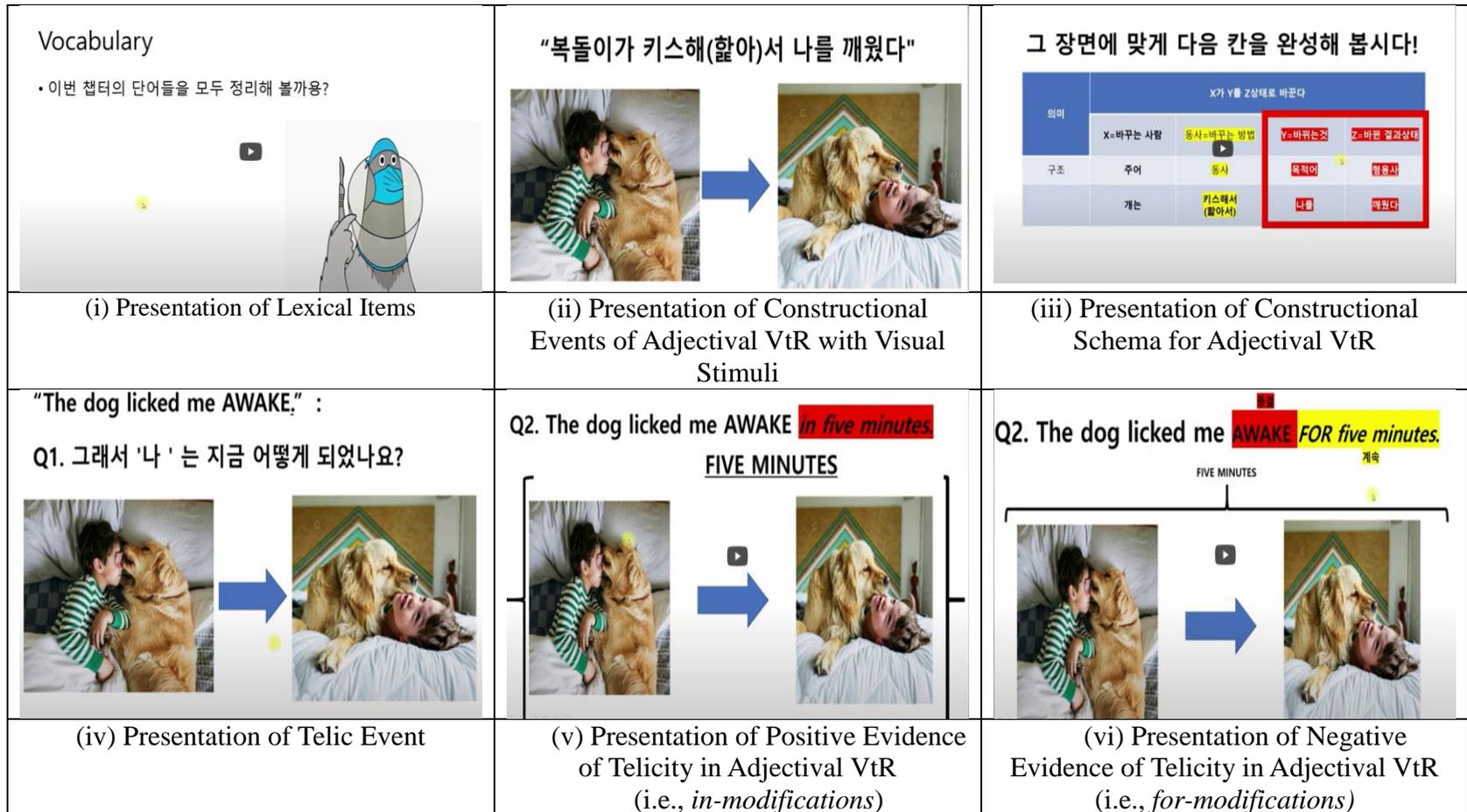


Figure 3.2. Example taken from the instructional video

### **3.4.2.2. Written Instructional Material**

Written instructional material activities included Controlled Practice (CP), a Dialogue Completion Task (DCT), and Drafting School Newspaper (DSN). The main objective of CP is to provide learners with several pattern drills to internalize the form and meaning pairings of the adjectival VtR expressions, with a special emphasis on the syntactic realization of the small clause in the correct word order. The pedagogical aim of DCT is to guide the participants in listening to adjectival VtR expressions in a spoken dialogue and fill in the blanks with relevant expressions. Thus, the nature of CP and DCT are rather linguistic and controlled. The main goal of DSN is to allow learners to produce adjectival VtR in a freer and more contextualized setting. Learners' main task in DSN is to complete the real-life task of drafting a school newspaper by using adjectival VtR expressions. These three types of activities (CP, DCT, and DSN) are distributed over Units I–III (3.4.2.2.1) and Units IV–IX (3.4.2.2.2). In the former part of the written instructional material activity, only CP was implemented. CP, DCT, and DSN are incorporated into the latter part of the written instructional material.

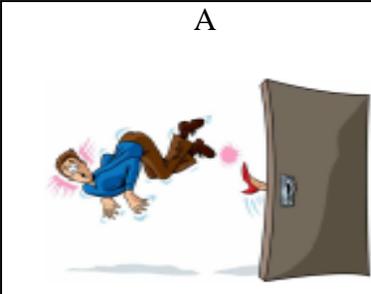
***3.4.2.2.1. Written Instructional Material Activities for Aspectual VPC:  
Controlled Practices***

In Units I–III, aspectual particle expression was used for the RP in adjectival VtR, with a special emphasis on figurative meaning. Each unit consisted of 8 to 13 items designed exclusively for the controlled exercises. These controlled practices include three stages of learning.

In the first step of CP, learners were invited to guess the meaning of the figurative particle (Figure 3.3). Specifically, the learning objective of this stage was to help learners extend beyond the literal particles to the figurative ones. In this task, learners were asked to guess the meaning of figurative particles with the basis of their literal counterpart. For example, as Figure 3.3 illustrates, learners were required to link the meaning of the figurative particle (displacement or removal) with that of its literal counterpart (outside).

Q. Try to guess the meaning of ‘*Out*’ on B  
(Hint: It is similar to the meaning on A)

A



B



Figurative meaning of *Out* = displacement

**Figure 3.3. Controlled practice (CP) step 1  
: Guessing the meaning of a figurative particle**

In the second step of CP (see Figure 3.4), two scenes related to the semantics of transitive or intransitive resultative constructions were introduced with a special focus on the figurative meaning of the particle *out*. This task involved two questions. The first question dealt with learners making distinctions between intransitive and transitive resultative constructions, depending on the availability of the conceptualization of the external causer. For the second question, learners were asked to describe the scene with a theme + particle sequence or a telegraphic construction (Tomasello, 1992) to help them acquire the direct object restriction (DOR) for adjectival VtR (Simpson, 1983). As the theme argument that undergoes change of state is obligatorily manifested in English transitive resultative

construction (Simpson, 1983), learners were expected to package the theme argument and the particle (e.g., *the oil out*, or *the fire out*) together as a small clause.

**※ Look at the pictures below and answer the two questions (Q1–Q2)**



(1) \_\_\_\_\_ (2) \_\_\_\_\_

**Q-1. Fill in the blanks below the pictures by following the directions:**

- Mark X→Y if the change of the state was caused by somebody.
- Otherwise, mark X

**Q-2. Describe the X→Y scene with TWO words only.**

(Hint: Focus on what went through the change of the state)

(1) \_\_\_\_\_

(2) \_\_\_\_\_

**Figure 3.4. Controlled practice (CP) step 2  
: Realization of external argument and DOR**

The last phase of the exercise expanded learners' interlanguage system from telegraphic speech (e.g., small clause) to sentential adjectival VtR expressions. As indicated in Figure 3.5, to provide learners with opportunities to produce adjectival VtR, learners were asked to describe an image using a constructional template of the form and meaning pairing for the adjectival VtR. This is accomplished through two cognitive processes: (a) determining the semantics of the images and (b) mapping meaning to the shapes.

<b><u>Meaning: X causes Y to become Z</u></b>				
				
Form	Subject	Verb	Object	Particle
+	↓	↓	↓	↓
Meaning	Changer	Manner	Theme	Result
↓	↓	↓	↓	↓
<b><u>Sentence:</u></b>	<b>He</b>	<b>Put</b>	<b>the fire</b>	<b>out</b>

**Figure 3.5. Controlled practice (CP) step 3  
: Generating form and meaning pairing for adjectival VtR**

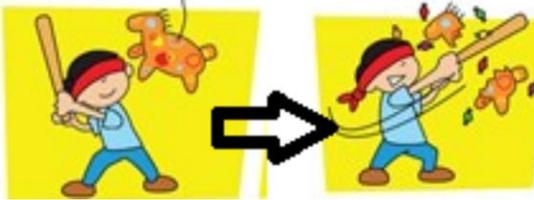
*3.4.2.2.2. Written instructional material activities for selected and unselected adjectival VtR: Dialogue completion tasks (DCT) and drafting school newspaper (DSN)*

In Unit IV–IX, both selected and unselected adjectival VtR expressions are covered during the 12 instructional sessions. The six instructional units implemented three types of activities: CP, DCT, and DSN. Learners were asked to perform the CP and DCT using the video materials. DSN was provided in another instructional session without the instructional video. CP in Units IV–IX were identical to those provided in Units I–III.

As Figure 3.6 illustrates, in DCT, participants were asked to complete the dialogue by filling in the blanks after listening to a bidirectional dialogue. Each DCT corresponded to three to four sentences. All the blanks in Figure 3.6 were targeted at producing adjectival VtR, but the required productions still existed at the fragmented level (e.g., I was lucky enough to hit the toy horse open).

W: What did you do in the party? Did you hit the toy horse?

M: Yes, I just swing without knowing where the toy horse is. I was lucky enough \_\_\_\_\_  
\_\_\_\_\_. It was very fun! \_\_\_\_\_!

The illustration shows a boy in a blue shirt and red headband swinging a wooden bat. In the first panel, the bat is raised and a toy horse is visible in the air. In the second panel, the bat is lowered and the toy horse is shown in a state of being hit, with small pieces flying off. A large black arrow points from the first panel to the second.

**Figure 3.6. Example of the dialogue completion task (DCT)**

Finally, in the DSN task, learners received contextual cues written in Korean. These learners were then asked to compose a school newspaper article. Topics for this task were similar to those of the DCT. The former was more distinct than the latter in that learners were required to write a narrative using full sentences rather than fragmental expressions. They also were asked to write a school newspaper article in English that included five to six sentences. Across all DSN tasks, more than 50% of the stimuli asked learners to produce an adjectival VtR, whether a particle or an adjectival.

### **3.5. Group-specific Instructional Treatments: Post-Written Instructional Material Activities on Corrective Feedback from Chatbot, Human, and Additional Explicit Instruction via Video**

The previous section described the video and text materials commonly employed by the experimental group participants. This section covers group-specific feedback provided by the AI chatbot, the human instructor and additional video materials.

Corrective feedback in the chatbot group was distinct from that of the human group with respect to (i) feedback timing (instant versus delayed) and (ii) feedback type (implicit + explicit as well as implicit or explicit). In the chatbot group, the corrective feedback was simultaneously provided to learners when they produced erroneous expressions. When learners first produced errors, implicit corrective feedback was provided; if they produced errors repeatedly, explicit corrective feedback was offered. In the human group, there was a time delay between the timing of learners' production of erroneous expressions and the provision of corrective feedback. In addition, learners in the human group received either implicit or explicit corrective feedback.

The corrective feedback groups (i.e., the chatbot and human group) were

differentiated from the video group in that the former provided adaptive feedback based on learners' production of erroneous utterances whereas the latter did not. The feedback provided from the human group was organized prior to the learners' production of errors. Section 3.5.1 describes the corrective feedback from the AI chatbot. Corrective feedback from the human instructor and additional explicit instruction from video material is briefly described in Sections 3.5.2 and 3.5.3, respectively.

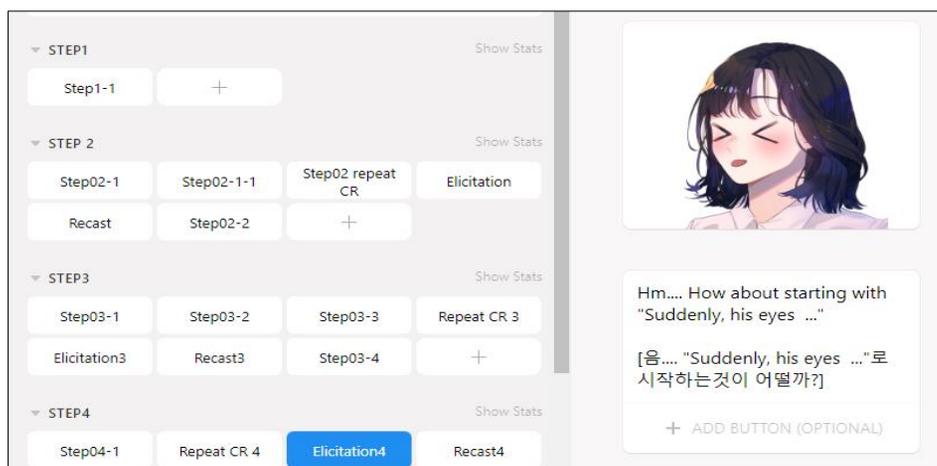
### **3.5.1. Corrective Feedback from the AI Chatbot**

All the chatbot-based interactions in this study were carried out using an intelligent computational system developed by the author to generate corrective feedback in response to Korean L2 English learners' errors and to manage instructional procedures.<sup>21</sup> All of these functions were performed

---

<sup>21</sup> The intelligent computational system employed in the current study encompasses a wide range of artificial intelligence technologies. Artificial intelligence includes practical application, such as the automatization of routine labor and the recognition of speech or images (Goodfellow et al., 2016). Artificial intelligence is categorized into two types: strong and weak forms of AI (Searle, 1980). Strong AI is an intelligence system that is indistinguishable from that of a human being. However, the technical viability of strong AI has been questioned by many experts. In contrast, weak AI is a computational system designed to complete a specific task (e.g., playing a chess game, recognizing human language, etc.). Practically all AI systems currently available on the market are weak AI. In this respect, all chatbot systems are narrow AI systems specifically designed to perform routinized communicative tasks, such as making a reservation for a hair salon. The

by minimally embodied conversational agents. Chatbots were developed on *Chatfuel*, one of the most famous text-based Facebook Messenger chatbot-building platforms.<sup>22</sup> Thus, the instructional chatbots were deployed via Facebook Messenger.



**Figure 3.7. Dashboard of Chatfuel**

Figure 3.7 illustrates the Chatfuel dashboard. Similar to *Dialogflow*, which is a speech-based AI chatbot building platform, the conversational flow is hand-made by a human chatbot builder. Thus, as depicted in Figure 3.6, the human chatbot builder is required to manually set the conversational

---

aim of the AI chatbot developed for the present study was not to have a conversation with the system, but to perform more pedagogically driven communicative activities to help learners perform a real-life task, and to provide them with a set of corrective feedback that goes against conventional dialogue flows.

<sup>22</sup> Available online at <https://chatfuel.com>.

flow in Step 2, constituting Step02-1, Step02-1-1, Step02 repeat CR, Elicitation, recast, and Step02-2. Unlike *Dialogflow*, Chatfuel can provide human chatbot users with visual stimuli (e.g., still images, videos, or gif images), which are engaging for chatbot users. Moreover, the text-based AI chatbot created by Chatfuel can deal with chatbot users' utterances in a manner irrelevant to conversational flow. The "Set Up AI" function in Chatfuel spots deviant messages by chatbot users and helps them to return to a normal conversation step. It also serves as the technical backbone for providing learners with sequential corrective recasts for learners' erroneous utterances.

In Chatfuel, the conversation between the AI agent and human learners is manually produced by the human chatbot developer, meaning only pre-planned conversation is available. Thus, the conversation between the AI chatbot and human language learners is mainly goal oriented, with the conversational goal of the AI chatbot being to provide learners with corrective feedback while performing linguistic or non-linguistic tasks with an AI agent. Figure 3.8 illustrates the general algorithm carried out by the AI chatbots utilized in this study.

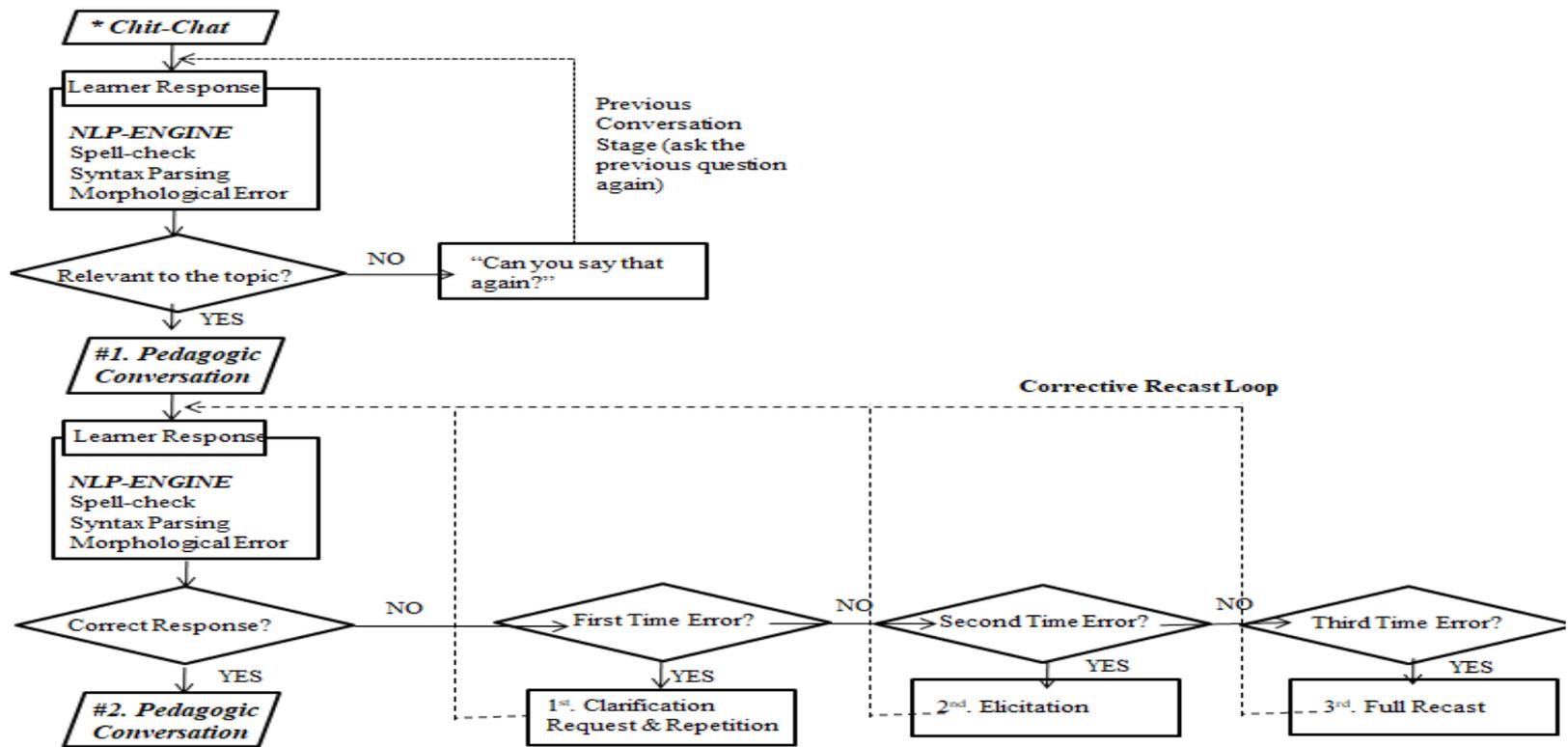


Figure 3.8. System Overview

As Figure 3.8 illustrates, the conversation prompted by the AI chatbot consists of two stages. One is the chit-chat stage, the other a pedagogic task stage. In the chit-chat stage, the AI agent can welcome the learners, present learning objectives, describe background information and provide information about what they are going to do in the task. At this stage, learners' output is not required, but sometimes learners may say something (e.g., 'What is your favorite color, Eliz Ria?'). Then the AI system evaluates their responses in the NLP engine in Chatfuel and judges whether the response is relevant to the conversation flow. If learners' responses are computed to be acceptable in the conversation they are directed to the next stage. Otherwise, they are directed to the "*can you say that again?*" slot and are re-directed to the previous conversation stage.

The core property of the AI system is entering into the pedagogic conversation stage. In this stage, learners' erroneous utterances are treated within the corrective recast loop.<sup>23</sup> Suppose that a learner made a mistake. Then, if it is the first such mistake, an implicit form of corrective feedback (i.e., clarification request and repetition) is provided. If the learner fails to provide the correct form despite receiving the implicit form of corrective feedback, a more explicit type of corrective feedback (i.e., elicitation) is

---

<sup>23</sup> Categorizing learners' utterances into correct or incorrect one is dependent upon the NLP engine of Chatfuel.

provided. If at that point the learner cannot fix the erroneous form, the most explicit type of corrective feedback (i.e., full recast) is provided. Since providing a full recast is equivalent to providing a correct answer, the learners were directed to the next pedagogic conversation. See Appendix 2 for the practical realization of the system. The chatbot worked well since 85.7 percent of the learners' utterances were appropriately processed.

Two personified chatbot agents were employed with human-like backgrounds (e.g., age, name, personality, physical appearance; K. Lee et al., 2019; L. Li et al., 2021; Lopatovska et al., 2018; Purington et al., 2017; Sanon et al., 2018), having two different conversational goals. One chatbot agent is Nylon Musk. Nylon Musk is an artifact adult male who is a parody of Elon Musk (See Figure 3.9).<sup>24</sup> In this context, Nylon Musk was characterized as a humorous male English teacher who spoke Korean. The primary objective of Nylon Musk was to perform the CP and DCT with language learners, which required them to participate in controlled practices.

---

<sup>24</sup> Nylon Musk was created by a Korean comedian who uploaded his videos on YouTube. Similar to Bird-Head, the license for academic use was granted by the original creator. Original works are available online at [https://www.youtube.com/watch?v=huQdbOTL\\_2A&list=PLX0uqwxLwywMAIJQcZ9fc\\_17KD8UaORBu](https://www.youtube.com/watch?v=huQdbOTL_2A&list=PLX0uqwxLwywMAIJQcZ9fc_17KD8UaORBu)



As Figure 3.9 illustrates, automatic corrective feedback was continually provided until the learner produced the correct response. For example, when a learner made a linguistic error, like “I laughed silly”, the AI chatbot provided a request for repetition. If the learner failed to provide a correct response, the AI chatbot provided a more explicit form of corrective feedback—namely, elicitation (e.g., “Let’s start with ‘I laughed ...’”). Finally, when the participant repeated another error, Nylon Musk presented the learner with metalinguistic feedback on the thematic role of subject, verb, object, and adjectival phrase. In addition, when a fake reflexive was required, a hint for such syntactic information was provided.

The other chatbot agent was Elizabeth Ria, an original chatbot agent developed by the researcher. Elizabeth Ria is personified as a cheerful teenage female student who speaks English. The primary aim of Elizabeth Ria is to perform DSN with human language learners. Distinct from Nylon Musk, Elizabeth Ria relies more on the human-like interaction, as the main objective of Elizabeth Ria is to make learners perform real-life tasks, such as writing school newspaper articles. Table 3.4 provides a sample interaction between learners and the AI chatbot. To make the interaction more communicative, her background was established as a school newspaper journalist. The objective of the communication between Elizabeth Ria and the learner was to complete the writing of a school newspaper article.

**Table 3.4 Examples of Interaction between Learner and AI Chatbot (Elizabeth Ria)**

Step	Dialogue	CF Type
Conversation Initiation	<p>Chatbot: I am going to write an article for the school newspaper. I heard that something special happened to Jack, but I don't know exactly what happened to him. I heard that Jack sent you some pictures about what happened to him. Please tell me what happened to Jack.</p> <p>Student: Sure!</p> <p>&lt;&lt; Picture #1 sent: He sang himself hoarse &gt;&gt;</p> <p>Chatbot: OK, tell me what happened.</p>	-
	<p>Student: [picture description] He sang hoarse. (Target: He sang himself hoarse)</p> <p>Chatbot: Excuse me? {{ He sang hoarse }}? Can you say that again?</p>	Repetition Clarification Request
Corrective Feedback	<p>Student: He sang hoarse.</p> <p>Chatbot: He sang ...</p>	Elicitation
	<p>Student: He sang hoarse.</p> <p>Chatbot: Oh! He sang himself hoarse? Please say it again!</p>	Full Recast
Proceeding to Next Picture	<p>Chatbot: Great! Tell me more about it!</p> <p>&lt;&lt; Picture #2 sent &gt;&gt; : Conversation proceeds</p>	-

As shown in Table 3.4, Elizabeth Ria initiates the conversation by stating the goal of the conversation (i.e., to write an article for the school

newspaper). Specific details of the relevant events for the school newspaper article are presented via visual cues (i.e., Picture #1). Learners were required to produce four sentences; therefore, four pictures were provided. After they saw the pictorial cue, learners produced one sentence corresponding to it. If the sentence was grammatical, they proceeded to the next picture by receiving the second one (i.e., Picture #2). If that sentence was not grammatical, they received a series of corrective feedback or corrective recast (Doughty & Varela, 1998).

As is shown in Figure 3.10, learners' erroneous sentence formulations were responded to a gradual increase in the explicitness of corrective feedback. For example, when they produced an initial mistake in the formulation of adjectival VtR expressions (e.g., "I don't know"), the AI chatbot repeated what they said (e.g., "I don't know?") and asked for a clarification request (e.g., "Excuse me? Can you say that again?"). When the learners made another error (e.g., "He shook his body dry"), the explicitness of the corrective feedback generated from the AI chatbot increased beyond the previous stage by using the elicitation technique with two-word hints (e.g., "Hmm... How about starting with '*Olaf shook ...*'? "). Finally, unlike Nylon Musk, learners' third repetition of an error (e.g., "He rubbed his eye and it be sore") was answered with full recast (e.g., "Oh! '*Thomas rubbed his eyes sore!*' Please say it again!").

### (1) Repetition + Clarification Request

모름



모름???  
Excuse me? Can you say that again?  
[모름???  
미안한데, 다시 말해줄래?]

### (2) Elicitation

He shoke his body dry



Hm.... How about starting with  
"Olaf shook ..."?  
[음.... "Olaf shook ..."로 시작하  
는 것이 어떨까?]

### (3) Full Recast

He rubbed his eyes and it be sore.



Oh! "Thomas rubbed his eyes  
sore" Please say it again.  
[아! "Thomas rubbed his eyes  
sore" 라구? 다시 말해줘]

Figure 3.10. Example of corrective feedback from the AI chatbot, Elizabeth Ria

To enhance the emotional engagement between the chatbots and learners (Liu et al., 2015), each chatbot agent expressed varying emotional states in the course of conversation. As shown in Figure 3.10, the chatbot’s facial expressions varied according to the learners’ responses; when the learners generated correct answers, they created positive facial expressions, but otherwise they showed negative facial expressions. Table 3.5 briefly summarizes the basic specification of AI chatbots, Nylon Musk and Elizabeth Ria. The researcher manually created conversation flows, set correct answers, and generated predictable wrong answers.

**Table 3.5 Basic Specification of AI Chatbots for Nylon Musk and Elizabeth Ria**

(1) Building Communicative Context	(N) Talking with a humorous English teacher about language (E) Writing a school newspaper article
(2) Conversation Flow Management	
a. Aim: to make learners adhere to the dialogue flow	
b. Method: when learners make irrelevant utterances,	
(i) say, “I don’t know what you said.”	
(ii) return to the previous conversation step	
(3) Target Language	English (with Korean translation)
(4) Corrective Feedback Provision	
a. Aim: to provide learners with step-wise corrective feedback	
b. Method: when learners make erroneous utterances,	
(i) provide implicit to explicit corrective feedback until they can provide correct answers	
(ii) augment the AI categorization by manually stacking probable errors	

\* (N) represents Nylon Musk, (E) represents Elizabeth Ria.

\*\* When (N) and (E) are not explicitly marked, the relevant specification is shared by both Nylon Musk and Elizabeth Ria.

As summarized in Table 3.5, these chatbots' conversational flow included generalized scenarios. These include such methods as providing conversational context (e.g., writing a school newspaper article) or managing conversation flow by dealing with learner responses that were irrelevant to the questions given by the chatbot agent. For example, when the chatbot outputs, "I cannot understand what you said," then the conversation flow algorithm takes the learners to the previous conversational step.

The target language produced by the chatbot agent was one of the most significant concerns in drafting the scenarios. As outlined above, Elizabeth Ria was an English-speaking bot. However, the absence of L1 translation may have contributed to an insufficient understanding of the conversational flow. Thus, two versions of Elizabeth Ria were developed. In the first version, Elizabeth Ria wrote only in English; in the second, her message was followed by a Korean translation. Learners were asked to choose whichever version they preferred. The learners unanimously selected the second version.

The automatic categorization of a learner's response—whether correct or incorrect—was computationally processed via text similarity calculation techniques, such as cosine similarity, Jaccard similarities, or TF-IDF (Deng & Li, 2018; Yu & Deng, 2016). By using such computational techniques, the

degree of text similarity between the learners' response and the correct answer set is calculated: If a learner's response is calculated to be similar to the example sentences in the correct answer set, the response is categorized as correct. Otherwise, it is categorized as incorrect (Schütze et.al., 2008). The AI/NLP engine on Chatfuel calculated text similarity between learners' responses and the correct answers. To augment the performance of automatic error categorization, example sentences in the incorrect answer set for argument structure-related errors (e.g., errors in word order or the omission of NP, VP, or PP) were manually provided to the Chatfuel system by the researcher.

If learners' responses were computed to be numerically closer to the correct answer rather than the incorrect ones, the AI/NLP engine on Chatfuel (e.g., Set Up AI function) judged them to be correct, and the AI/NLP engine on Chatfuel advanced the learners to the next stage of the conversation. Otherwise, corrective feedback was provided to treat erroneous responses. When such errors were repeatedly generated for the same question, a variety of corrective feedback was provided, starting with a relatively implicit form (e.g., clarification request, or elicitation; Doughty, 2001; Doughty & Varela, 1998; Havranek, 2002; R. Ellis, 2009). If the provision of implicit corrective feedback did not enable the learners to self-correct their erroneous formulations, then the chatbot agent provided more

explicit forms of corrective feedback (e.g., metalinguistic feedback or full recast) to the learner. When a learner could not self-repair an erroneous response after receiving three instances of corrective feedback, the corrective feedback loop was terminated and the correct formulation was explicitly provided.

### **3.5.2. Corrective Feedback from a Human Instructor**

One Korean female English teacher, who had 20 years teaching experience at the time of the study, was employed to check learners' assignments and provide corrective feedback for them. At that time she was in her fifties, and her first language was Korean. Her English proficiency was considered at an advanced level, and she remained in Australia for five years when she was in her thirties while pursuing an M.A. degree in English literature.

Figure 3.11 illustrates how human-based corrective feedback was provided. When the learners finished the CP or DCT with both visual and written stimuli, they were asked to provide answers on Naver Band, which is an internet-based learning management system (LMS). In response to incorrect expressions, explicit corrections were provided to the learners.

### 1) Learners' Responses:

구조	주어	동사	목적어	첨용사
(1)	Ken	made	desk	bang
(2)	Bu	rolled	반죽을	-thin
(3)	Judy	left	개지불	sofe

### 2) Teacher's Feedback:

**평가 의견** 반죽: dough 허벅지: thigh (1) Ken banged ... 로 시작해 봅시다. (3) Judy scratched ... 로 시작해 봅시다. (4) I laughed myself ... 로 시작해 봅시다.

**Figure 3.11. Example of corrective feedback (elicitation) from a human instructor during the controlled practice (CP) task**

As demonstrated in Figure 3.11, the learner produced incorrect formulations in (1) (e.g., \*Ken made desk bang / Target Sentence: Ken banged the desk off). To deal with such an error, the instructor provided corrective feedback in Korean that is similar to elicitation (e.g., “How about starting with ‘Ken banged ...?’”). To deal with the lexical errors in (2) and (3), the instructor provided learners with explicit correction for them (e.g., \*Bu rolled *banjook-ul* (Korean translation for ‘dough’) thin / Target Sentence: Bu rolled the dough thin). Other than the types of the errors shown in Figure 3.9, when the participants produced micro-level syntactic or morphological errors (e.g., \*He piled the cushions highly / Target Sentence: He piled the cushions high), recast was provided (e.g., “Highly?”

Did you say *Highly*? Please think about it again”). An implicit form of corrective feedback, like a clarification request, also was provided to the learners.

Corrective feedback for the human group of learners was more restricted than for the chatbot group learners with respect to (i) the time delay in the provision of corrective feedback and (ii) the type of corrective feedback. Due to Naver Band’s technical limitations, learners’ assignments were returned to the students after all Naver Band members had finished them.

First, there was a time delay in the provision of corrective feedback between the moment of production of errors and the moment of providing corrective feedback. In the early phase of the experiment, an average of 10 minutes was spent providing corrective feedback on a single assignment. However, as the experiment unfolded, the teacher could not individually deal with assignments, as the teacher had other work to do. In the later phase of the instruction, the teacher raised the issue of the excessiveness in providing corrective feedback. She complained that it was too much for her to provide corrective feedback in addition to her ordinary workload.

For the CP and DSN tasks, the participants were asked to take a photo of the assignment and to submit it via Naver Band. For a single assignment, there was an average interval of one week between the day that the instruction video was made available, and the day students received

corrective feedback on the assignment. After receiving feedback, students were required to review the teachers' feedback for a total of 15 minutes. Thus, learners received asynchronous feedback from the human teacher.

Second, the type of corrective feedback provided in the human group was restricted to either implicit or explicit corrective feedback. Unlike the Chatbot group, it was practically impossible for the learners in the human group to receive a series of corrective feedback from implicit to explicit. Instead, only one type of corrective feedback, either implicit or explicit, was provided in response to learners' erroneous utterances after a single attempt.

### **3.5.3. Additional Instruction via Video Material**

Learners in the video group did not receive any corrective feedback but received additional explicit instruction for all the practice items from the written instructional material, irrespective of their responses. After learners finished the written instructional material exercises (i.e., CP, DCT, and DSN), the video group learners were asked to watch a 15-minute video.

An item-by-item explanation was provided to participants. Syntactic and semantic explanations were given to the learners for the CP, DCT, and DSN. The explanation was explicit, followed by the correct answer for each



action), theme (i.e., object), and result state (i.e., adjectival result phrase). For example, the instructor provided an explanation for (1) in Figure 3.9, as follows: First, he explained that in the subject position, *Simpson* should be filled, as he is the one who rubbed the glass. Next, the instructor provided syntactic comments on the manner of the change of state with the verb *rubbed*. The instructor subsequently provided syntactic accounts that the theme underwent in a specific result state (i.e., *the glasses clean*). Finally, the correct formulation of the adjectival VtR was explicitly provided to the learners (i.e., *Simpson rubbed the glasses clean*). With respect to its semantic specification, the question in the written instructional material dealt with how the theme underwent a change of state by linking the theme (the glasses) to the result state (clean) with the use of the verb *become* (the glasses become clean).

An additional instructional video was provided to the learners via Naver Band. The video was provided to the learners every other day. After the learners completed all the tasks related to the additional instructional treatment, learners' written output was submitted to the researcher, who determined whether they performed the tasks correctly. Their output product was returned to them after the researcher examined their written output. To cope with incorrect answers, a list of correct answers was provided to learners.

### 3.6. Test

This study included three test sessions — one before and two after the instruction took place, i.e., pretest, immediate posttest, and delayed posttest. Each test consisted of two assessment tasks; an Acceptability Judgment Task (AJT) and an Elicited Written Task (EWT).<sup>25</sup> Knowledge of transitive resultative constructions was assessed via the AJT, while the ability to accurately produce transitive resultative construction was measured by the EWT.

The main objective of AJT was to assess learners' comprehension of grammatical and ungrammatical formulations of English VtR for target-likeness judgments for test items.<sup>26</sup> On the other hand, the primary aim of EWT was to evaluate Korean English learners' correct production of English VtR expressions.

All the pre-test items in the AJT and EWT were minimally adjusted for

---

<sup>25</sup> Although the learners participated in distance-learning classes during the COVID-19 pandemic, they took the on-site test when they were allowed to attend classes, which enhanced the reliability of the tests.

<sup>26</sup> According to Robenalt and Goldberg (2016), most non-native English learners, but especially the most advanced ones, are more likely to judge ungrammatical sentences as grammatical formulations. This suggests that the learning of ungrammatical properties of a certain construction may not readily be reflected in the AJT. In this respect, AJT results for the ungrammatical formulations should be carefully interpreted.

immediate and delayed posttests to ensure that practice effect and memorization would not interfere with learners' performance (See Appendix 1 for the complete list of the test items across all tests). The modification was mainly targeted at nominal expressions, since they were irrelevant to the major syntactic or semantic properties in the adjectival or prepositional VtR. Verb–adjective pairs in the adjectival VtR expressions remained consistent across the three tests, since the two expressions are highly likely to co-occur. In addition, the sequence of test items of AJT and EWT was randomized across the three test sessions. Moreover, to avoid the familiarity effect, the pictures used in the EWT were visually re-designed and differed from those used in each test session.

Each task was administered over the course of two days, with AJT taking place on Day1 and EWT occurring on Day2, so that learners could not guess the objective of each task and therefore would not be skewed by their performance on the preceding task. The overall framework of the testing tasks employed in this study replicated those used by S. Kim (2016).

### **3.6.1. Acceptability Judgment Task (AJT)**

The AJT required participants to judge the acceptability of the sentences. This task was conducted to investigate Korean EFL learners' comprehension of English transitive resultative constructions. Participants were asked to judge the acceptability of the test sentences. The test items were ranked on a 6-point Likert-type scale, plus a 'not sure' option (1 = completely unacceptable, 2 = unacceptable, 3 = slightly unacceptable, 4 = slightly acceptable, 5 = acceptable, and 6 = completely acceptable). The 'not sure' option was intended to prevent learners from selecting answers in an arbitrary manner. All the responses for the 'not sure' option were excluded from the analysis.

Participants also were given vocabulary support by being presented with the meanings of some unknown words at the end of a given sentence. The test design was mainly based on Park and Lakshmanan (2007) and S. Kim et al. (2020), including the provision of preceding contexts in English. Figure 3.13 illustrates an example of an Acceptability Judgment Task included in the current research.

굵은 글씨로 된 문장만의 문법적인 정확성을  
아래와 같이 6점 만점으로 판단하여  
해당 점수의 칸에 **동그라미** 하세요.

1점                  2점                  3점                  4점                  5점                  6점  
 ←—————→  
 완전히 어색함    어색함    약간 어색함    약간 자연스럽게    자연스럽게    완전히 자연스럽게

만약 판단하기 어렵다면, 맨 끝의 “모르겠음”에 표시하세요.

번호	문 장	점수						모르 겠음
		1	2	3	4	5	6	
		⊗					⊗	
5	A fire broke out in a tall building, so the shopkeeper hollered the tourists out of the building.							

**Figure 3.13. Example of an Acceptability Judgment Task**

The AJT was designed to explore how different instructional treatments for Korean English learners affected their linguistic representation of the target-likeness of English transitive resultative constructions. For this reason, the AJT included sentence types that were not only grammatical, but also ungrammatical in terms of syntax or semantics (Kim et al., 2020). Thus, four types of English resultative sentences were presented (Table 3.6). Type 1 sentences include syntactically grammatical resultative complement phrases that would be realized using a different syntactic structure in Korean. Type 2 sentences include semantically grammatical transitive resultative expressions. Type 3 sentences include syntactically ungrammatical result phrases that would be grammatical in Korean. Type 4 sentences have

semantically incongruous resultatives, which are unacceptable in English but would be acceptable in equivalent expressions via Korean translation.

**Table 3.6 Sentence Types Used in the Acceptability Judgment Task**

Conditions Type		Instructed Condition (Adjectival RP)	Uninstructed Condition (Prepositional RP)
Grammatical Formulation	Type1 Syntactically Grammatical	Mary shook him awake.	The shopkeeper hollered the teenagers out of the building.
		I cut her hair short.	Terry's horse hopped her straight out of the saddle.
		The fan beat him unconscious.	The lifeguard backstroked the children to the shore.
		Let's scrub the pot shiny.	He waltzed Matilda around and around the room.
	Type2 Semantically Grammatical	John wiped the table clean in three minutes.	John loaded the hay onto the truck in ten minutes.
		The dog ate the flowerbed bare in ten minutes.	they ran the dog over the bridge in 30 minutes.
		The dog barked the chicken awake, and it became awake.	We launched the rocket to the moon, and the rocket arrived there.
		He tore her free, so she became free.	I threw the ball to Julian, and the ball arrived at him.
Ungrammatical Formulation	Type3 Syntactically Ungrammatical	* You can make the meat <i>tenderly</i> .	* I urged Josh (into) the room.
		She piled the cushions highly.	I helped him (into) the hospital yesterday.
		Aliza wiggled (her tooth) loose.	he hit (the ball) into the hole.
	Type4 Semantically Ungrammatical (Possible in Korean)	Jim washed (the jacket) clean.	Sam frightened (Bob) out of the house.
		* She will drive you mad <i>for a few minutes</i> .	* Bill pushed Harry off the sofa <i>for hours</i> .
		Dana tugged the rope loose for twenty seconds.*	# John walked Matilda to his new apartment for 20 minutes.
		He dyed his hair red but it did not become red.	Jane dragged the suitcase to the room, but stopped before she got there.
		She rolled the dough thin, but it did not become thin	Tom pulled the cart to the door, but stopped before he got there.

To be specific, Type 1 and Type 2 stimuli constitute *grammatical* formulations of adjectival and prepositional VtR. Specifically, Expressions in Type 1 are syntactically correct, and those in Type 2 are semantically correct since they denote telicity interpretation. It should be noted that the Type 1 stimuli for prepositional VtR are composed only of unselected prepositional VtR embedded by an intransitive verb. Type 3 and Type 4 stimuli are *ungrammatical* in English for different reasons: Type 3 stimuli are ungrammatical because the result phrases have an incorrect grammatical category or violate Direct Object Restriction (DOR). Type 4 stimuli, on the other hand, are ungrammatical since the expressions are related with the atelic interpretation of English VtR. The linguistic features reflected in ungrammatical sentence types (Type 3 and Type 4) for both conditions are summarized in Table 3.7.

**Table 3.7 Linguistic Features of Syntactically and Semantically Ungrammatical Sentences**

Type		Instructed Condition (Adjectival RP)	Uninstructed Condition (Prepositional RP)
Type3 (Syn)	Adverbial result phrase	make the meat <i>tenderly</i>	N.A.
	No Direct Object Restriction(DOR)	sang (no object) hoarse	hit (no object) into the hole
	Preposition dropping	N.A.	urge Josh (no preposition) the room
Type4 (Sem)	Allowance of a <i>for</i> -phrase	drive you mad <i>for a few minutes</i>	pushed Harry off the sofa <i>for hours</i> .
	Cancellability	dyed his hair red <i>but it did not become red</i>	carried the box to the council meeting, <i>but it did not arrive there</i> .

A total of 40 sentences were presented, including 32 experimental sentences (4 types \* 4 tokens of items \* 2 VtR types) and 8 fillers (see Appendix 1.1 for the complete list of items in AJT). The purpose of the fillers was to distract participants' attention from the test items to avoid biased responses. Test items were evenly distributed as Types 1 – 4. The items were further balanced by the condition of the availability of instructional treatment (e.g., instructed *adjectival* transitive resultative construction vs. uninstructed *prepositional* transitive resultative construction/caused-motion construction).

The primary objective of the AJT was to assess learners' judgment of target-likeness for English VtR sentences. Since learners' responses to the acceptability of the grammatical formulations correspond to the target-

likeness of the sentence, their responses remained constant for the data coding. On the other hand, the raw data of learners' acceptability scores for the ungrammatical formulations *negatively* reflected the target-likeness of the sentence.<sup>27</sup> Thus, participants' acceptability ratings for an ungrammatical formulation were *reversely* coded from 1 (completely non-target-like) to 6 (completely target-like) which was originally 6 (completely acceptable) to 1 (completely unacceptable).

It is hypothesized that, irrespective of their grammaticality, Korean L2 English learners would generally *accept* the testing stimuli. Indeed, novice-level English L2 learners often make overgeneralization errors (Ambridge et al., 2020; Goldberg, 2019).

Participants were given 40 minutes to complete the task, meaning that on average, 60 seconds were assigned to each test item. Within this time limit, they were asked to read not only the target sentences, but also the preceding context and calculate their syntactic and semantic compatibility (S. Kim et al., 2020). This pressurized format was designed to elicit linguistic judgments based on implicit knowledge, by preventing participants from relying upon explicit knowledge (R. Ellis, 2009; R. Ellis & Barkhuizen, 2005).

---

<sup>27</sup> This is mainly because the target-like response score in the raw-data for the ungrammatical sentence is one.

### 3.6.2. Elicited Writing Task (EWT)

The Elicited Writing Task (hereafter referred to as EWT) examined how Korean English learners produce transitive resultative constructions. The EWT consisted of thirty items, including twenty experimental items (ten for *instructed* adjectival VtR expressions and ten for *uninstructed* prepositional VtR expressions) and ten distracters. In this task, participants were provided with a preceding context in Korean, a picture, a partial English sentence, and two English cue words to complete the sentence. The distribution of experimental items in the EWT is reported in Table 3.8 (see Appendix 1.2 for the complete items in AJT).

**Table 3.8. Distribution of Experimental Items in EWT**

[ <i>Instructed</i> ] Adjectival Transitive Resultative Construction (n=10)	[ <i>Uninstructed</i> ] Prepositional Transitive Resultative Construction (n=10)
1) Subj <u>hammered</u> NP flat 2) Subj <u>danced</u> REFL dizzy. 3) Subj <u>ran</u> NP threadbare. 4) Subj <u>sliced</u> NP thin. 5) Subj <u>shot</u> NP dead. 6) Subj <u>knocked</u> NP unconscious. 7) Subj <u>kicked</u> NP open. 8) Subj <u>painted</u> NP blue. 9) Subj <u>nailed</u> NP shut. 10) Subj <u>get</u> NP dirty.	1) Subj <u>blinked</u> Obj off LOC 2) Subj <u>stared</u> Obj out of LOC 3) Subj <u>sneezed</u> Obj off LOC 4) Subj <u>sarled</u> Obj into LOC 5) Subj <u>pulled</u> Obj out of LOC 6) Subj <u>broke</u> Obj into LOC 7) Subj <u>put</u> Obj on LOC 8) Subj <u>threw</u> Obj on LOC 9) Subj <u>rolled</u> Obj out of LOC 10) Subj <u>pushed</u> Obj up LOC

\* REFL represents reflexive pronoun



task was motivated by the need to elicit the use of the target construction (Wu, 2011). The cue words were presented in random order throughout the task to avoid having the presentation order (e.g., verb-adjective sequence or preposition-verb sequence) function as a cue for the completion task.

In the EWT, the participants were provided with a contextual cue written in Korean. This stimuli design was adopted based upon Ionin and Zizyk's (2014) suggestion that in order to make sure that the context is clearly understood by low-proficiency L2 language learners, the contextual cues should be written in the participants' L1.

### **3.7. Statistical Analysis**

Quantitative analysis of the data was conducted using R, a free language and environment for statistical computing (R Development Core Team, 2017). First, descriptive statistics were calculated to report on general patterns across different groups and in different test sessions. To compare the instructional effects among the four groups and across the three tests, four-by-three repeated measure ANOVA was conducted on all mean scores for the AJT and EWT tasks, including Group as the between-group variable (chatbot, human, video, and control groups) and Test as the within-group

variable (pretest, immediate posttest, and delayed posttest). For any meaningful interaction between Group and Test, Tukey HSD comparisons, which included Group as the within-group variable for each of the immediate and delayed posttests, were conducted to compare group performance for each test period. In turn, this statistical analysis was conducted for each instructed and uninstructed transitive resultative construction. The statistical alpha level was set at .05 in all statistical analyses.

## **Chapter 4. Results and Discussions**

This chapter reports the results of the Acceptability Judgment Task (AJT), and the Elicited Writing Task (EWT). The first section presents results of the AJT on the transitive resultative constructions, including both adjectival and prepositional, and discusses the major findings taken from the comprehension test. The second section reports the findings from the EWT on both adjectival and prepositional transitive resultative constructions and presents a discussion of the major findings of the production test.

### **4.1. Acceptability Judgment Task**

This section is devoted to answering the first research question (RQ(1)) as to whether the corrective feedback from the AI chatbot was as effective as that of the human instructor and video-based instruction in the Korean high school English learners' comprehension of English transitive resultative constructions. Section 4.1.1 reports on the results of the AJT to determine whether the corrective feedbacks from the AI chatbot was as effective as that of the human instructor or additional instructions from video material in the comprehension of instructed adjectival VtR. In the subsequent section,

AJT results of the uninstructed prepositional VtR will be supplied to determine the instructional effectiveness of chatbot-based corrective feedback. Section 4.1.3 briefly discusses major findings from the comprehension test. Appendix 1 presents for the item-by-item results of the learners' acceptability rating for the adjectival and prepositional VtR.

#### **4.1.1. AJT Results of Instructed Adjectival VtR**

To answer RQ1(a), an analysis of the AJT on the adjectival VtR (e.g., Tom hammered the metal *flat*) was carried out to determine whether the four groups (e.g., chatbot, control, human, and video) differed in their ability to comprehend the adjectival VtR sentences in a target-like way<sup>28</sup>. Table 4.1 reports the mean score of the AJT for adjectival VtR. It demonstrates that, prior to instruction all participants had relatively similar scores in the comprehension of adjectival VtR expressions (chatbot: 3.80, human: 3.90, video: 3.85, control: 3.85).

---

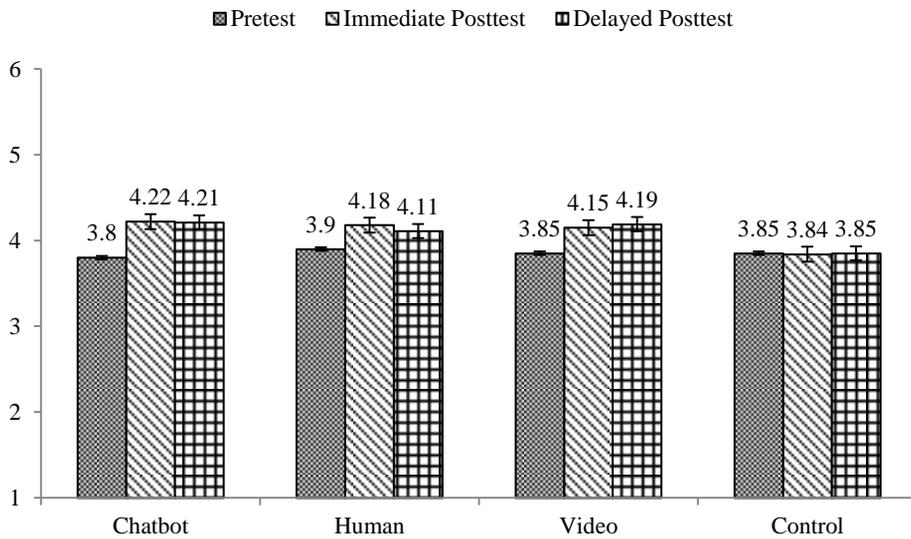
<sup>28</sup> 1 = completely non target-like, 2 = non-target-like, 3 = slightly non-target-like, 4= slightly target-like, 5 = target-like, 6 = completely slightly target-like

**Table 4.1 Mean AJT Score of Instructed Adjectival VtR**

<b>Group</b>	<b>Pretest</b>		<b>Immediate Posttest</b>		<b>Delayed Posttest</b>	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Chatbot	3.80	1.21	4.22	1.38	4.21	1.33
Human	3.90	1.42	4.18	1.50	4.11	1.46
Video	3.85	1.35	4.15	1.31	4.19	1.28
Control	3.85	1.19	3.84	1.21	3.85	1.13

In the two posttest sessions, the chatbot group showed the highest mean score. After the instructional treatment was provided, the participants in the instruction groups (i.e., chatbot, human, and video) demonstrated the AJT scores in the two posttest sessions were higher than those of the control group, showing better comprehension of the adjectival VtR.

Figure 4.1 illustrates the mean differences between the three tests across the four groups. When the mean scores of all test sessions were plotted, each group showed different patterns. Although every group except the control group showed a mean increase in the immediate posttest, the mean increase of the chatbot group (+0.42) was greater than that of the human (+0.28) or video (+0.3) groups. Between the immediate and delayed posttest, all the participants in the corrective feedback groups showed a moderate mean score decrease (chatbot: -0.01, human: -0.07), while the video group showed a slight mean score increase (+0.04).



**Figure 4.1. Mean AJT scores of instructed adjectival VtR (error bars represent standard error)**

To investigate the statistical significance of the instructional effects within and between the groups, a repeated-measure ANOVA was conducted. The results of this process are presented in Table 4.2.

**Table 4.2 A Repeated Measure ANOVA of AJT on Instructed Adjectival VtR**

	Source	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Sig</i>
Within-Subjects	Test_session	2	8.8	5.05	.006 **
	Test_session * Group	6	3.35	1.93	.07
	Error (Time)	248	1.74		
Between-Subjects	Group	3	23.58	13.56	< .001 ***
	Error	124	10.24		

Statistical significance also was found in the test session ( $F(2, 248) = 5.05, p = .006$ ) and in the treatment groups ( $F(3, 124) = 13.56, p < .001$ ). However, there was marginal significance for the interaction effect of the four groups with the test sessions ( $F(6, 248) = 1.93, p = .07$ ).

The mean differences were found in the three test sessions across the instructional groups—namely, the chatbot ( $F(2, 31) = 33.8, p < .001$ ), human ( $F(2, 31) = 14.59, p = .001$ ), and video ( $F(2, 31) = 24.36, p < .001$ ). No statistical difference was found in the control group ( $F(2, 31) = 1.81, p = .299$ ). This strongly indicates that the instructional treatments are more effective for persistent improvements in the comprehension of adjectival VtR than would be the case if no such treatments had occurred.

The result of the one-way ANOVA for the between-group analysis of the pretest demonstrated that the four groups were homogenous prior to instruction ( $F(3, 124) = .075, p = .521$ ). This result also showed a statistical

difference among the four groups in the immediate posttest ( $F(3, 124) = 11.48, p < .001$ ). A Tukey post-hoc test was analyzed to explore if the mean score differences of the AJT in the immediate posttest scores are significant among the four groups. The post-hoc Tukey test on the data presented in Table 4.3 revealed a significant difference among the instructional groups (i.e., chatbot, human, and video groups) and the non-instructional group (i.e., the control group). This implies that the instructional intervention positively affected the comprehension of English adjectival VtR.

**Table 4.3 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Immediate Posttest**

(I) Group	(J) Group	Mean Difference (I-J)	Sig
	Control	.38	< .001***
Chatbot	Human	.04	.99
	Video	.07	.81
Control	Human	-.30	< .001***
	Video	-.34	< .001***
Human	Video	.03	.91

The one-way ANOVA for the between-group analysis also showed a statistical difference among the four groups in the delayed posttest ( $F(3, 124) = 12.99, p < .001$ ). A Tukey post-hoc test was analyzed to explore if the

mean score differences of the AJT in the delayed posttest scores were statistically significant across the four groups. The post-hoc Tukey test on the data presented in Table 4.4 reported that, like the immediate posttest results, a significant statistical difference emerged between the instructional groups and the non-instructional group. This reveals the possibility that the provision of instructional treatments improved the comprehension of English adjectival VtR.

**Table 4.4 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Delayed Posttest**

(I) Group	(J) Group	Mean Difference (I-J)	<i>Sig</i>
	Control	.36	< .001***
Chatbot	Human	.10	.67
	Video	.03	.99
Control	Human	-.26	.011*
	Video	-.34	< .001***
Human	Video	-.08	.67

#### 4.1.2. AJT Results of Uninstructed Prepositional VtR

To answer RQ1(b), this section investigates whether instructional treatment on adjectival VtR affected the generalization to its related

construction, prepositional VtR (e.g., *The man sneezed the tissue off the table*). Thus, statistical analyses for the comprehension of the uninstructed prepositional VtR expressions were undertaken.

Table 4.5 reports the mean score of the AJT for the prepositional VtR. It clearly demonstrates that learners in the instruction groups performed better than the non-instruction group across all the test sessions.

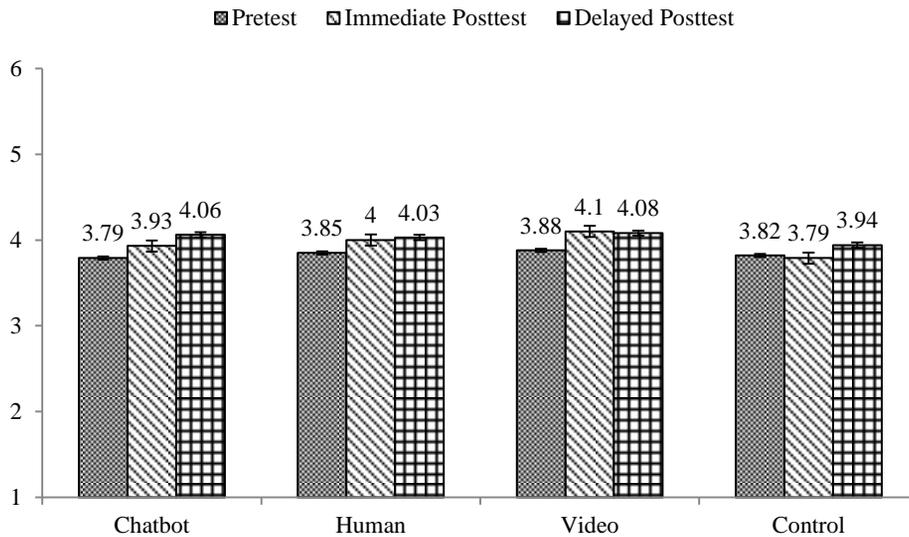
**Table 4.5 Mean AJT Score of Uninstructed Prepositional VtR**

Group	Pretest		Immediate Posttest		Delayed Posttest	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Chatbot	3.79	1.30	3.93	1.34	4.06	1.24
Human	3.85	1.51	4.00	1.47	4.03	1.38
Video	3.88	1.26	4.10	1.29	4.08	1.20
Control	3.82	1.38	3.79	1.26	3.94	1.16

In the pretest session, all participants demonstrated less target-like behaviors in the comprehension of prepositional VtR expressions (chatbot: 3.79, human: 3.85, video: 3.88, control: 3.82). In the two posttest sessions, the participants in the instruction groups comprehended the adjectival VtR in a more target-like way than in the pretest session. However, the comprehension of the uninstructed prepositional VtR by the control group

remained underdeveloped across all test sessions (pretest: 3.82, immediate posttest: 3.79, delayed posttest: 3.94).

Figure 4.2 illustrates the mean differences among the three tests across the four groups. Each group showed a different pattern in the comprehension of prepositional VtR expressions. The mean difference between the pretest and the immediate posttest for the chatbot group (+0.14) was as much as the human group (+.15) and video group (+0.13). Between the immediate and delayed posttest, all participants in the corrective feedback groups (i.e., chatbot and human groups) showed a slight decrease in the mean score (chatbot: +0.13, human: +0.03), while a slight mean score decrease was observed for the video group (-0.03). However, improvements were not present in the control group.



**Figure 4.2. Mean AJT scores of uninstructed propositional VtR (error bars represent standard error.)**

To investigate the instructional effects within and between the groups, a repeated-measure ANOVA was conducted. The results of this process are presented in Table 4.6.

**Table 4.6 A Repeated Measure ANOVA of the AJT on the Comprehension of Uninstructed Prepositional VtR**

	Source	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Sig</i>
Within-Subjects	Test_session	2	1.04	0.6	0.55
	Test_session * Group	6	2.77	1.59	.145
	Error (Time)	248	1.74		
Between-Subjects	Group	3	18.87	10.86	< .001 ***
	Error	124	.07		

Statistical significance was only found in the treatment groups ( $F(3, 248) = 10.86, p < .001$ ). Statistical significance was not found for either the test sessions ( $F(2, 124) = 0.6, p = .55$ ) or the interaction effect of the four groups with the test sessions ( $F(6, 248) = 1.59, p = .145$ ). This strongly indicates that, despite the instructional treatment on the adjectival VtR, the participants in the instructional treatment groups were unable to generalize the linguistic representation of the adjectival VtR to the comprehension of the prepositional VtR.

#### **4.1.3. Discussion**

The first research question of the present study (RQ1(a)) is whether chatbot-based instruction was as effective as human-based instruction (i.e., corrective feedback from a human instructor or additional explicit instruction of video material) in Korean L2 English learners' comprehension of English adjectival VtR expressions, which were explicitly instructed. To address this research question, the quantitative results of the AJT were analyzed in Section 4.1.1.

As the results demonstrate, the overall mean scores of adjectival VtR for the instructional groups significantly increased throughout the test sessions.

In particular, statistical differences in the immediate posttest were found between the instructional groups (i.e., chatbot, human, and video) and the control group. The mean score increase of adjectival VtR by the chatbot group was the greatest in the immediate and delayed posttests, followed by the human-based instruction groups (i.e., human and video groups) and the control group. The between-group differences in the AJT demonstrated similar patterns. The mean scores of every group in the pretest were not much different from one another, but the immediate posttest showed statistically significant differences between the instructional groups (i.e., chatbot, human, and video) and the control group.

These results indicate that the corrective feedback from the AI chatbot for adjectival VtR is at least as effective as human-based instruction groups (i.e., human and video groups). The effectiveness of the corrective feedback from the AI chatbot may be attributed to its instant provision of multiple corrective feedbacks with varying degrees of explicitness. In contrast to the corrective feedback generated from the human instructor, the corrective feedback from the AI chatbot was instantly provided to language learners. In addition, multiple instances of corrective feedback from the AI chatbot were provided to learners until they produced correct responses, varying the explicitness of corrective feedback from implicit (e.g., repetition) to explicit techniques (e.g., elicitation, metalinguistic feedback, or full recast).

The second research question of the present study (RQ1(b)) concerns whether chatbot-based instruction was as effective as human-based instruction (i.e., corrective feedback from the human instructor or additional explicit instruction of video material) in Korean L2 English learners' comprehension of English prepositional VtR expressions that were not explicitly instructed. To address this research question, quantitative results of the AJT were analyzed, as reported in Section 4.1.2.

The results of AJT for the uninstructed prepositional VtR indicated that the learners in the chatbot group failed to perform significantly better than the human-based instruction groups (i.e., human and video) and the non-instructional control group. In the immediate posttest, every group showed a small mean score increase. Similarly, the chatbot and human groups demonstrated a very small mean score increase between the immediate and delayed posttest, with the human group losing a little of its earlier gains. The between-group differences in the AJT for the uninstructed prepositional VtR demonstrated similar patterns. The mean scores of all participants on the pretest were not much different from one another. On the immediate posttest, significant differences in the mean score of the AJT for the uninstructed prepositional VtR were not found between the instructional groups (i.e., chatbot, human, and video) and the control group. The mean score of prepositional VtR by the chatbot group was the greatest on the immediate

posttest. This difference was little changed on the delayed posttest.

The test results for prepositional VtR by Korean L2 English learners may indicate that the instructional intervention for adjectival VtR employed in this study was not effective enough to facilitate learners in making generalizations for the comprehension of uninstructed prepositional VtR. This empirical result raises an immediate question of why Korean L2 English learners in the chatbot group could not make linguistic generalizations for the comprehension of the uninstructed prepositional VtR.

The video group participants' failure to make generalizations for the uninstructed English prepositional VtR construction may be ascribed to the absence of corrective feedback for one of the most complex argument structure constructions of English, adjectival VtR. Also, the human group participants' failure to make generalizations to English prepositional VtR construction may be accounted for by their delayed provision of an insufficient amount of corrective feedback. Notably, the chatbot group participants' failure to make generalizations for the English prepositional VtR formulation may be attributed to the loosely-design instructional framework of the corrective feedback provided by the AI chatbot. The corrective feedback from the AI chatbot mainly focused on the production of VtR expressions, so less attention has been paid to their comprehension. If more pedagogical effort had been made for the comprehension of the VtR,

the participants could have offered successful generalizations for the comprehension of prepositional VtR.

The chatbot group participants' incomplete learning of English prepositional VtR formulation also may be attributed to weak saliency in their corrective feedback, especially by recast. Recasts are reported to be most effective when learners are explicitly made aware of their own erroneous forms. Thus, the more salient recasts are, the more they lead to uptake and successful repair (Sheen, 2004). Samples (1) and (2) illustrate examples of explicit recast, which contrasts with (4), an example of implicit recast.

(1) Explicit Recast #1: Intonational Focus

S: Korean is more faster.

T: Is faster.

S: Is faster than English.

(2) Explicit Recast #2: Partial Recast

S: I think she'll spend a wonderful time in Australia and East South Asia.

T: South east Asia.

S: South east Asia.

(3) Implicit Recast #3

T: Okay, it's good. You wanna tell us one?

S: Eh...: Kaii convention.

T: What kind of convention? (recast)

S: Kaii convention...eh...some people...

According to Sheen (2004), the recasts in (1) and (2) were much more explicit and provided more uptake opportunities than (3). This finding lends support to Doughty's (2001) claim that recasts are more effective when they take the form of the isolated interrogative recast shown in (4) below (Lyster, 1998) or the corrective recast in (5) (Doughty & Varela, 1998).

(4) Isolated interrogative recast

Student: He was rob on the way home.

Teacher: He was robbed?

(5) Corrective recast

Student: I think that the worm will go under the soil.

Teacher: I *think* that the worm *will* go under the soil?

S: (No response)

T: I *thought* that the worm *would* go under the soil.

S: I *thought* that the worm *would* go under the soil.

The saliency or the explicitness in these two samples (4) and (5) are obtained using intonational focus. Rising intonation is used in the isolated interrogative recast. In the corrective recast, the repetition of the student's incorrect utterance also includes rising intonation (Teacher: I *think* that the worm *will* go under the soil?), while the recast has a falling intonation (Teacher: I *thought* that the worm *would* go under the soil).

However, the corrective feedback provided by the AI chatbot employed in the present study may not have been salient enough to help Korean L2 English learners acquire the syntactic and semantic information from the prepositional VtR expressions. The AI chatbot employed in the present research was text-based, so providing intonational prominence was unavailable. In addition, due to technological limitations, the location of the error was not explicitly provided to learners. These factors may be partly responsible for the full recast employed in the present study to be analyzed as positive evidence devoid of corrective forces (Leeman, 2003; Loewen & Philip, 2006; Sheen, 2006). Thus, the corrective feedback and recast employed in the present study require further modification.

To enhance the saliency, simultaneity, and straightforwardness of chatbot-based corrective feedback, conversational flows for the provision of corrective feedback employed in the present study should be slightly modified in three ways (see Table 4.7).

**Table 4.7 Revised Example of Interaction Between Learner and AI Chatbot (Elizabeth Ria)**

Step	Dialogue	CF Type	
Conversation	<p>Chatbot: I am going to write an article for the school newspaper. I heard that something special had happened to Jack, but I don't know exactly what happened to him. I heard that Jack sent you some pictures about what happened to him. Please tell me what happened to Jack.</p> <p>Student: Sure!</p> <p>&lt;&lt; Picture #1 sent: He sang himself hoarse&gt;&gt;</p> <p>Chatbot: OK, tell me what happened.</p>	-	
	<p>Student: [picture description] He sang hoarse. (Target: He sang himself hoarse)</p>		
	<p>Chatbot: Excuse me? {{ He sang hoarse }}? Can you say that again?</p>		<p>Repetition Clarification Request</p>
	<p>Student: He sang hoarse.</p>		
Corrective Feedback	<p>Chatbot: He sang ...</p>	Elicitation#1	
	<p>Student: He sang...</p>		
	<p>Chatbot: He sang <i>himself</i> ...</p>	<u>Elicitation#2</u>	
	<p>Student: He sang himself ...</p>		
	<p>Chatbot: OK. Tell me the result. What happened to HIM? (start with 'him', with two words)</p>	<u>Metaling.F#1</u>	
	<p>Student: Him hoarse. Chatbot: Great. Then, he sang ...</p>		
<p>Student: He sang him hoarse.</p>			
<p>Chatbot: When the object refers to the same</p>	<u>Metaling.F#2</u>		

	person as the subject, change the object into – self (e.g., himself)	
	Student: He sang him hoarse.	
	Chatbot: You mean he sang himself hoarse?	Full Recast
	Student: Yes.	
	Chatbot: Full sentence.	<u>Request for Self-Repair</u>
	Student: He sang him hoarse	
	Chatbot: He sang himself hoarse?	
	Student: Yes.	
	Chatbot: Full sentence.	
	Student: He sang him hoarse.	
	Chatbot: He sang himself hoarse. Copy it.	<u>Explicit Correction</u>
	Student: He sang himself hoarse.	
	Chatbot: Great! Tell me more about it!	
Conversation	<< Picture #2 sent >>	-
	: Conversation proceeds	

\* Underlined CF Type (i.e., Elicitation #2, Metaling.F#1~4, Request for Self-Repair, and Explicit Correction) represents the corrective feedback strategies that are suggested to be added for the corrective feedback delivered by Elizabeth Ria. Refer to Table 3.4 for the original version of the corrective feedback incorporated into the conversational flow employed in the current study.

First, a three-word elicitation including subject, object, and direct object, should be applied to make learners aware of the relevant syntactic restriction (e.g., DOR) that the object position (or overt subject in the small clause) should be filled to express the semantics of English VtR. Thus, as Table 4.7 demonstrates, the conversational flow of the chatbot incorporated Elicitation

#2 (i.e., He sang *himself* ...) to make learners aware of DOR by presenting a three-word hint.

Second, it would be necessary to provide learners with a more densely-designed instructional framework for the provision of gradual metalinguistic feedback. Considering the cross-linguistic differences between the Korean and English languages, Korean L2 English learners may have difficulty formulating English small clauses to express result-denoting components. They also may find it difficult to project a manner-denoting element into the matrix verb position. Moreover, they may have great difficulty formulating an adjectival VtR with a reflexive object (e.g, He sang *himself* hoarse).

To help learners deal with such difficulties in learning English prepositional VtR, stepwise metalinguistic feedback should be presented (Metaling.F#1~2). As described in Table 4.7, learners are asked to formulate an English-type small clause with an overt subject followed by a result phrase (e.g., Metaling.F#1: *Tell me the result. What happened to HIM? (start with 'him', with two words). I will call it A*). The learners are then requested to fill in the slot for the matrix subject and matrix verb (e.g., Then, he sang ...). By doing so, they are prompted to combine the two parts. Still, they made a grammatical error on the use of the reflexive pronoun (e.g., “He sang him hoarse”). Finally, learners are asked to check the grammaticality in

the reflexive object (e.g., Metaling.F#2: *When the object refers to the same person as the subject, change the object into –self (e.g., himself)*).

As a last remark, learners would be required to “self-repair” until they can successfully provide a correct answer. If learners are unable to give correct phrases despite several attempts, the system needs to present the correct answer and demand that they copy the expression (i.e., explicit correction).

To conclude, the corrective feedback of instructed adjectival VtR provided by the AI chatbot was effective in the comprehension of the relevant construction. This instructional effect is associated in a stepwise fashion in the instant provision of corrective feedback. On the other hand, the corrective feedback from the AI chatbot failed to make generalizations for the comprehension of the uninstructed prepositional VtR. This insufficient instructional effect is partly attributed to the instructional design in the corrective feedback provided by the AI chatbot, which lacks noticeability.

## **4.2. Elicited Writing Task (EWT)**

This section answers RQ2. An analysis of the picture description task was carried out to see whether the three groups differed in their ability to map the visual stimuli and constructional cues onto the corresponding transitive resultative constructions. The high increase in scores indicated learners' improvement in the accurate production of the transitive resultative constructions.

Section 4.2.1 deals with whether the corrective feedback from the chatbot is as effective in improving Korean L2 English learners' production of instructed adjectival VtR as was the case for the human and video groups. Section 4.2.2 investigates whether corrective feedback from the chatbot is as effective in improving Korean L2 English learners' production of uninstructed prepositional VtR as in the other treatment groups. Section 4.2.3 discusses the major findings of the current study. Appendix 2 provides item-by-item results of the learners' acceptability rating for the production of adjectival and prepositional VtR.

### 4.2.1. EWT Results for Instructed Adjectival VtR

To answer RQ2(a), this chapter presents the results for the question of whether chatbot-based instruction is as effective as human-based and video-based instruction in improving Korean high school English learners' production of instructed English adjectival VtR.

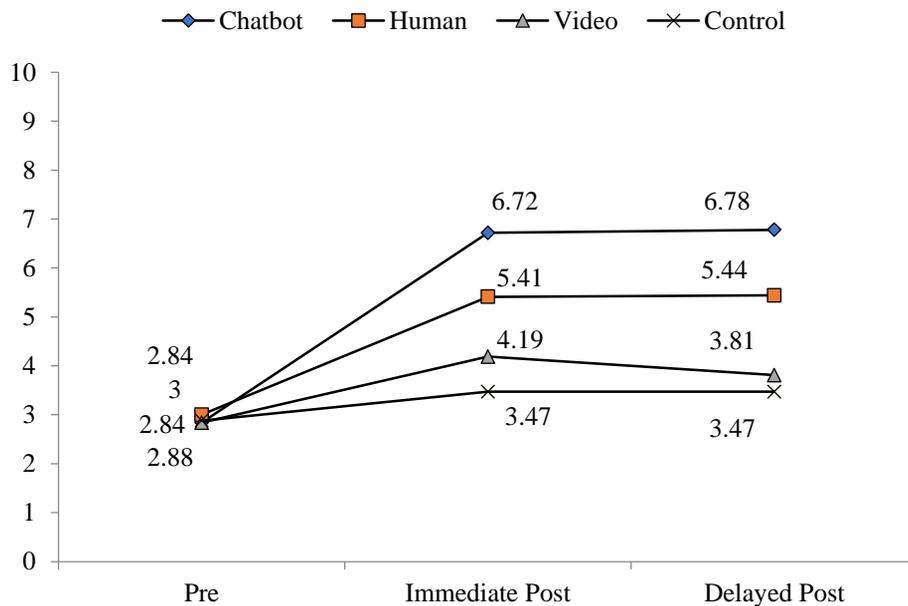
Table 4.8 illustrates that learners in the corrective feedback groups (i.e., chatbot and human) performed better than the explicit instruction group (i.e., the video group) and the non-instructed group (i.e., the control group) across all test sessions. A series of post hoc Tuckey tests were conducted for each group to examine whether significant differences occurred between the test sessions.

**Table 4.8 EWT Results of ANOVA for the Between-tests**

Group	N	Pretest		Immediate Posttest		Delayed Posttest	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Chatbot	32	2.84	2.26	6.72	2.61	6.78	2.89
Human	32	3.00	2	5.41	2.96	5.44	2.77
Video	32	2.84	2.23	4.19	2.81	3.81	2.96
Control	32	2.88	1.76	3.47	2.06	3.47	1.81

\* The maximum mean score is 10.

As Figure 4.3 indicates, in the intermediate and delayed posttests, the chatbot group learners outperformed all the other treatment group learners. Among the instructional treatment groups, the chatbot group participants showed the biggest difference between their mean score in the pretest and immediate posttest (+3.88). This difference was followed by human group (+2.42) and video group learners (+1.35). Between immediate and delayed posttest sessions, all learners, except those in the video group, maintained improvement in the immediate posttest.



**Figure 4.3. Interaction plot of EWT for Instructed Adjectival VtR**

A repeated-measure ANOVA was conducted to investigate the instructional effects both within and between the groups for the EWT. The results are presented in Table 4.9.

**Table 4.9 A Repeated Measure ANOVA of the EWT: Instructed Adjectival VtR**

	Source	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Sig</i>
Within-Subjects	Test_session	2	174.17	72.84	< .001***
	Test_session * Group	6	23.45	9.81	< .001***
	Error (Time)	248	2.39		
Between-Subjects	Group	3	93.75	6.97	< .001***
	Error	124	13.45		

There were statistically significant results for the interaction effect of the treatment group with the test sessions,  $F(6, 248) = 23.45, p < .001$ . Statistical significance was also found in Time ( $F(2, 248) = 72.84, p < .001$ ) and in Group ( $F(3, 124) = 6.97, p < .001$ ). This indicates that all groups changed in different ways during the course of the study.

The mean differences among the three test sessions across all four groups demonstrated a statistically significant difference in the corrective feedback groups (chatbot:  $F(2, 31) = 24.1, p < .001$ ; human:  $F(2, 31) = 9.19, p < .001$ ). However, across the tests, no statistical significance was found in the video group ( $F(2,31) = 2.14, p = .124$ ) and the control group ( $F(2,31) = 1.06, p$

= .35). This strongly indicates that the instructional treatment using corrective feedback, as with the human and chatbot groups, is more effective for persistent improvement in the production of adjectival VtR than without using corrective feedback, as observed in the video and control groups.

For the between-group analysis of the pretest and its descriptive statistics of the mean EWT scores of the grammatical formulation of the adjectival VtR, the result of the one-way ANOVA demonstrated that the three groups were homogenous prior to instruction ( $F(3, 124) = .041, p = .989$ ). The result of the between-group analysis of the immediate posttest for the formulation of adjectival VtR in the immediate posttest demonstrated that the chatbot group performed best ( $M = 6.72, SD = 2.61$ ), followed by the human group ( $M = 5.41, SD = 2.96$ ), then the video group ( $M = 4.19, SD = 2.81$ ), and finally the control group ( $M = 3.47, SD = 2.06$ ). The result of the one-way ANOVA for the immediate posttest demonstrated that the four groups were not homogenous ( $F(3, 124) = 9.41, p < .001$ ).

A Tukey post-hoc test was analyzed to explore if the differences in EWT scores are significant. The post-hoc Tukey test on the data presented in Table 4.10 revealed a significant difference between the corrective feedback groups (i.e., chatbot and human groups) and the non-corrective feedback groups (i.e., video and control groups). This implies the possibility that the

degree of score improvements varied depending upon whether corrective feedback was provided. That is, the provision of corrective feedback positively improved the production of English adjectival VtR.

**Table 4.10 Multiple Comparisons of the Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Immediate Posttest**

(I) Group	(J) Group	Mean Difference (I-J)	Sig
	Control	3.25	< .001***
Chatbot	Human	1.31	.195
	Video	2.53	.01**
Control	Human	-1.94	.01**
	Video	-.72	.694
Human	Video	1.22	.254

The result of the between-group analysis of the delayed posttest for the formulation of the adjectival VtR in the EWT delayed posttest reported that the chatbot group outperformed all other groups. The one-way ANOVA demonstrated that the four groups were not homogenous prior to the instructional treatment ( $F(3,124) = 10.73, p < .001$ ). To explore whether a difference exists across the four groups, a post-hoc Tukey test was performed.

Table 4.11 shows the result of the post-hoc Tukey test. Similar to the immediate posttest, the chatbot group learners were significantly better at

producing adjectival EWT than the non-corrective feedback groups (i.e., video and control groups). The human group learners also performed significantly better than non-instruction group (i.e., control group) learners. However, the mean score of the human group was higher than that of the video group, with no statistical significance. This suggests the possibility that corrective feedback from the chatbot was more persistent and more effective than that from a human instructor.

**Table 4.11 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for Instructed Adjectival EWT: Delayed Posttest**

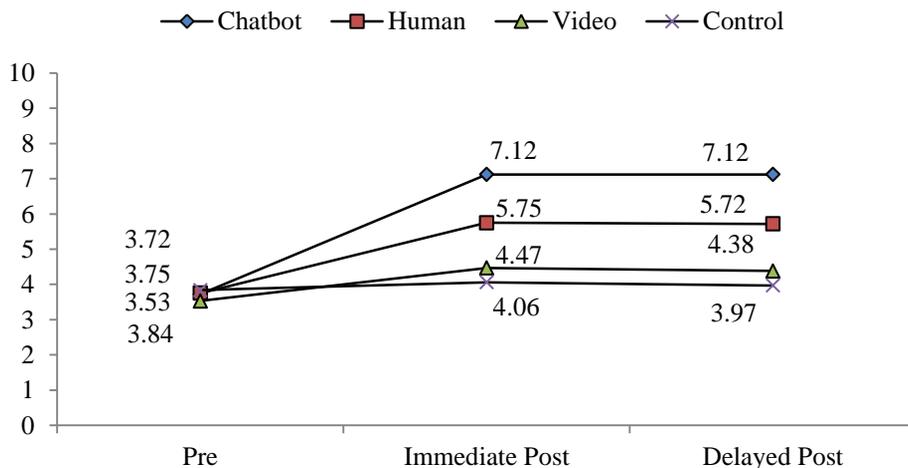
(I) Group	(J) Group	Mean Difference (I-J)	Sig
	Control	3.31	< .001***
Chatbot	Human	1.34	.183
	Video	2.97	< .001***
Control	Human	-1.97	.02*
	Video	.34	.954
Human	Video	1.63	.07

#### 4.2.2. EWT Results of Uninstructed Prepositional VtR

To answer RQ2(b), this section reports on the results that demonstrate whether chatbot-based instruction is as effective as or more effective than

human-based instruction in improving Korean high-school English learners' production of uninstructed English prepositional VtR.

As seen in Figure 4.4, the chatbot group learners outperformed all the other treatment group learners on the intermediate and delayed posttests. Among the instructional treatment groups, the chatbot group participants showed the largest mean score difference between their mean score on the pretest and immediate posttest (+3.40). This difference was followed by the human group (+2.00) and the video group learners (+0.94), respectively. Between the intermediate and delayed posttest sessions, all learners seemed to maintain improvement in the immediate posttest; a mean score decrease was found in all participant groups except for the chatbot group (chatbot: 0, control: -0.09, human: -0.03, video: -0.09).



**Figure 4.4.** Interaction plot of EWT for uninstructed prepositional VtR

Table 4.12 reports the mean EWT scores of the uninstructed prepositional VtR expressions. It demonstrates that the learners in the corrective feedback groups (i.e., chatbot and human) performed better than the explicit instruction group (i.e., video group) and the non-instructed group (i.e., control group) across all test sessions. A series of post hoc Tuckey tests were conducted for each group to examine whether there were significant differences between the test sessions.

**Table 4.12 EWT Results of ANOVA for the Between-tests:  
Uninstructed Prepositional VtR**

Group	N	Pretest		Immediate Posttest		Delayed Posttest	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Chatbot	32	3.72	2.34	7.12	3.07	7.12	3.16
Human	32	3.75	2.38	5.75	3.19	5.72	2.84
Video	32	3.53	2.63	4.47	3.25	4.38	3.34
Control	32	3.84	2.20	4.06	2.24	3.97	2.47

\* The maximum mean score is 10.

A repeated-measure ANOVA was conducted to investigate the instructional effects within and between the groups for the EWT. The results are presented in Table 4.13.

**Table 4.13 Repeated Measure ANOVA of EWT:  
Uninstructed Propositional VtR**

	Source	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Sig</i>
Within-Subjects	Test_session	2	111.14	39.12	< .001***
	Test_session * Group	6	21.18	7.46	< .001***
	Error (Time)	248	2.84		
Between-Subjects	Group	3	84.89	4.80	< .001***
	Error	124	17.68		

There were statistically significant results for the interaction effect of the treatment groups in the test sessions,  $F(6, 248) = 7.46, p < .001$ . Statistical significance also was found in the test sessions ( $F(2, 248) = 39.12, p < .001$ ) and in the treatment groups ( $F(3, 124) = 4.8, p < .001$ ). This indicates that all groups changed in different ways during the study.

Table 4.15 demonstrates the mean score differences between the three test sessions across all four groups. This data indicates that only the corrective feedback groups (i.e., chatbot and human groups) showed a statistical significance between the pretest and the immediate posttest (chatbot:  $p < .001$ ; human:  $p < .001$ ), and between the pretest and the delayed posttest (chatbot:  $p < .001$ ; human:  $p < .001$ ). However, no statistical significance was found in either the explicit instruction group (i.e., the video group) or the non-instruction group (i.e., the control group). This strongly suggests that the instructional treatment using corrective feedback is more effective

for persistent improvements in the production of English prepositional VtR than without relying on corrective feedback, as was true of the video and control groups.

Table 4.16 illustrates the result of the between-group analysis of the pretest and its descriptive statistics of the mean EWT scores of the grammatical formulation of the adjectival VtR. The result of the one-way ANOVA confirmed that the three groups were homogenous prior to the instructional treatment ( $F(3,124) = .096, p = .962$ ).

The result of the between-group analysis of the immediate posttest for the prepositional VtR showed that the chatbot group performed best ( $M = 7.12, SD = 3.07$ ), followed by the human group ( $M = 5.75, SD = 3.19$ ), then the video group ( $M = 4.47, SD = 3.25$ ), and finally the control group ( $M = 4.06, SD = 2.24$ ). The results of the one-way ANOVA for the immediate posttest session demonstrate that the mean EWT scores of the four groups were statistically different ( $F(3,124) = 6.96, p < .001$ ).

A post-hoc Tukey test was used to investigate whether the differences in EWT scores were significant across the groups. The post-hoc Tukey test on the data presented in Table 4.14 revealed a significant EWT mean score difference between the chatbot group and the non-corrective feedback groups (i.e., video and control groups). However, the mean EWT scores of

the human group were not significantly different from those of the chatbot group or the non-corrective feedback groups.

**Table 4.14 Multiple Comparisons of Chatbot, Human, Video, and Control Groups for the Uninstructed Prepositional EWT: Immediate Posttest**

(I) Group	(J) Group	Mean Difference (I-J)	Sig
Chatbot	Control	3.06	< .001***
	Human	1.37	.253
	Video	2.65	< .001***
Control	Human	-1.69	.109
	Video	-.41	.947
Human	Video	1.28	.313

The result of the between-group analysis of the delayed posttest for the prepositional VtR demonstrated that the chatbot group performed better at producing target-like prepositional VtR expressions than all the other groups regarding the EWT mean score. The one-way ANOVA demonstrated that between-group differences in the delayed posttest were statistically different ( $F(3,124) = 7.41, p < .001$ ). To find out whether a difference existed across the four groups, a post-hoc Tukey test was performed.

As reported in Table 4.15, the result of the post hoc Tukey analysis for the delayed posttest revealed that, similar to the results from the immediate posttest, a significant difference existed between the chatbot group and the

non-corrective feedback group (i.e., the video and control groups). However, statistical significance was not observed between the chatbot and the human groups. In addition, the EWT score of the human group was not statistically distinct from that of the video group.

**Table 4.15 Multiple Comparisons of the Chatbot, Human, Video, and Control Groups for Uninstructed Prepositional EWT: Delayed Posttest**

(I) Group	(J) Group	Mean Difference (I-J)	Sig
Chatbot	Control	3.15	< .001***
	Human	1.40	.236
	Video	2.74	< .001***
Control	Human	-1.75	.091
	Video	-0.41	.947
Human	Video	1.34	.273

#### 4.2.3. Further Analysis

In the previous section we investigated whether the corrective feedback from the AI chatbot was as effective as or better than other instructional interventions in the production of instructed adjectival VtR or uninstructed prepositional VtR expressions. The result clearly demonstrated that corrective feedback from the AI chatbot was as effective as that of the human instructor, and more effective than explicit instruction.

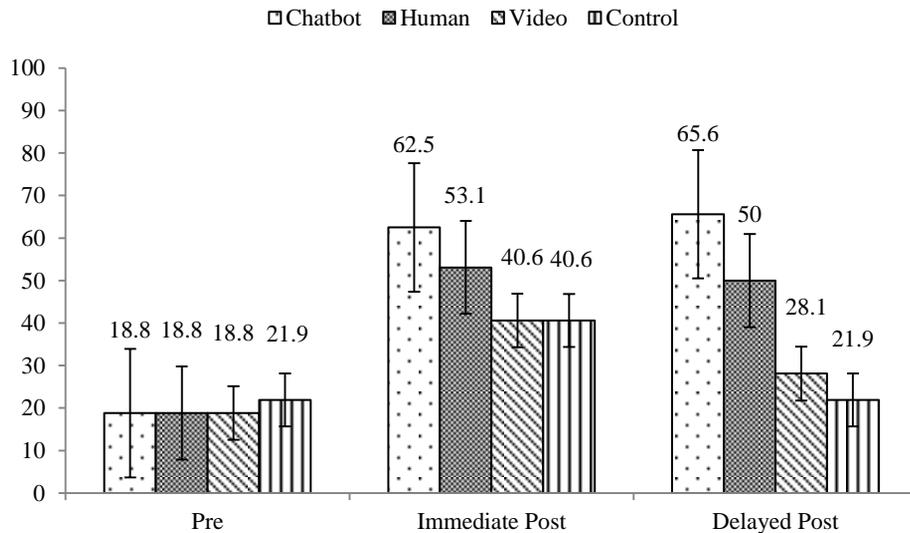
Improvements in the production of adjectival and prepositional VtR are proportional to the unlearning of L1 Korean, which is evidenced by the correct provision of unselected VtR embedded by an intransitive verb. English unselected VtR is composed of a small clause with an overt subject selected by an intransitive verb, which is unavailable in Korean syntax. In this respect, this section will briefly report the findings on the improvements in the production of unselected adjectival or prepositional VtR by Korean L2 English learners.

First, corrective feedback from the AI chatbot was effective in the correct production of an unselected adjectival VtR expression. As reported in Section 2.1.2, Korean adjectival VtR does not have any corresponding structure equivalent to English unselected formulations. Thus, in this study, the Korean L2 English learners were predicted to produce English adjectival and prepositional VtR expressions by unlearning their L1 properties via instructional intervention. This prediction was proven by participants' behavior on the production of unselected adjectival VtR embedded by the intransitive verb *run* (e.g., Target sentence: *The man ran the shoes threadbare*).<sup>29</sup> The general tendency plotted in Figure 4.5 demonstrates that corrective feedback from either the AI chatbot or a human instructor was more effective than the explicit instruction provided by the video material in

---

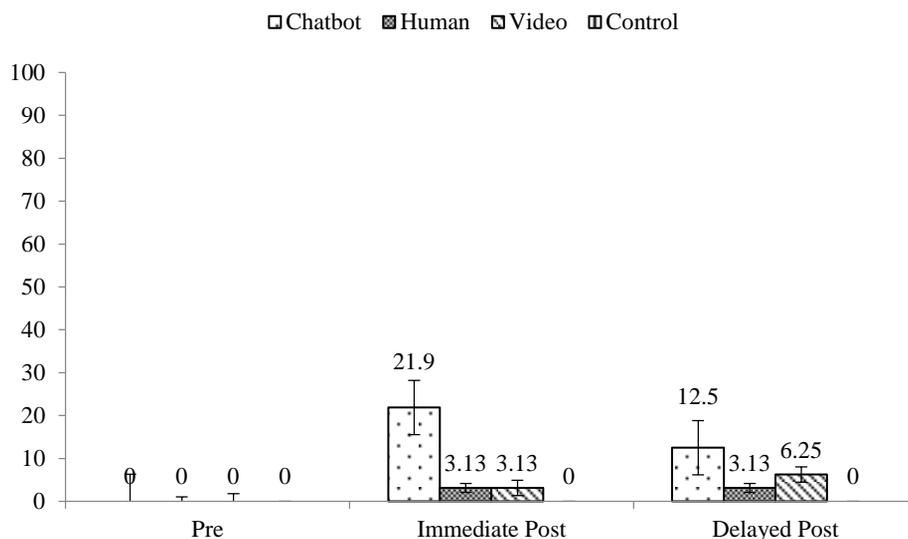
<sup>29</sup> Only one example was relevant to this category, so the generalization of results should only be done with care.

the production of unselected adjectival VtR (e.g., *He ran his shoes threadbare*).



**Figure 4.5. Correct answer rate for the unselected adjectival VtR using *run* (error bars represent standard errors)**

Second, although corrective feedback from the AI chatbot or human instructor seems effective for Korean L2 English learners' mastery of unselected adjectival VtR, they showed a considerable difficulty in producing unselected adjectival VtR with a fake reflexive object (e.g., Target Sentence: *She danced herself dizzy*).

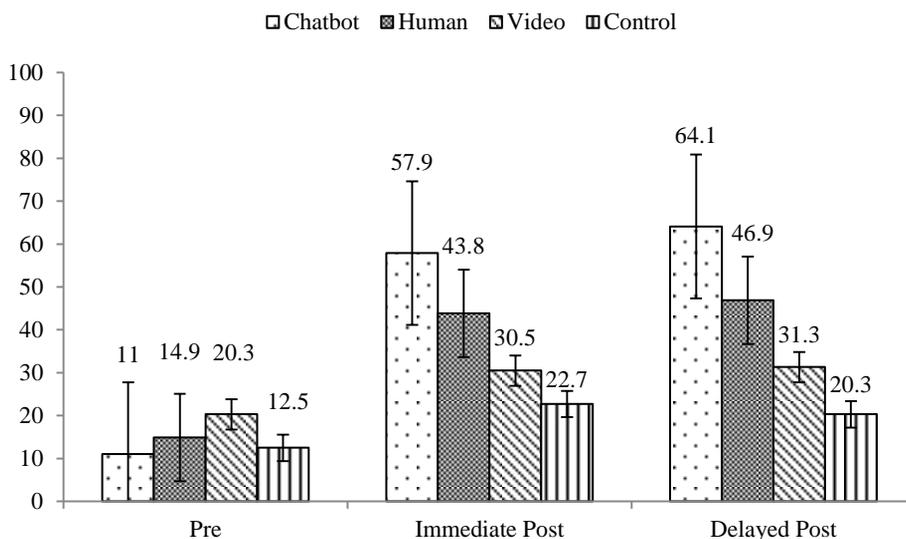


**Figure 4.6. Correct answer rate for the unselected adjectival VtR with a reflexive pronoun in the direct object position (error bars represent standard errors)**

As is illustrated in Figure 4.6, on the pretest none of the participants succeeded in producing an unselected adjectival VtR with a reflexive object. After the instruction was provided, 20% of the participants in the chatbot group produced the correct formulation of the unselected adjectival VtR with a reflexive object in the immediate posttest session, which decreased by 10% in the delayed posttest session. Moreover, no significant differences were found for the human and video groups, where in less than 10% of participants provided correct answers.

Considering L1–L2 differences, English unselected adjectival VtR with a reflexive pronoun in the direct object position is an extremely marked construction for Korean L2 English learners. As investigated in Section 2.1.2, participants’ L1 does not allow an intransitive verb to be used in a transitive resultative construction that takes a subject-filled small clause as its complement. Moreover, there is a cross-linguistic variation in the reflexive pronoun between English and Korean. In English, the reflexive pronoun (e.g., *herself*) should be c-commanded by its nominal antecedent within the syntactically local domain (Chomsky, 1980, 1981). However, a Korean reflexive pronoun (e.g., *caki-casin*) allows both local binding and long-distance binding (J. Kim & Yoon, 2009). Thus, in producing an unselected English adjectival VtR with a reflexive pronoun in the direct object position, Korean L2 English learners are required to cope with the asynchronous syntactic and semantic properties in the formulation of unselected adjectival VtR. Moreover, they should deal with syntactic and semantic ambiguity caused by the different distribution of the reflexive pronoun between the two languages. These cross-linguistic variations may result in cognitive demand in producing the unselected adjectival VtR expression paired with a reflexive pronoun in the direct object position (Argyri & Sorace, 2007; Hulk & Müller, 2000; Sorace, 2011, 2016; Yip & Matthews, 2000).

Third, the corrective feedback from the AI chatbot was generalized to Korean L2 English learners' production of unselected prepositional VtR expressions, which were uninstructed. Four test items (e.g., Target structure: (i) *She blinked the dust off her eyes*, (ii) *My mom stared Tom out of the room*, (iii) *He sneezed the paper off the table*, (iv) *Jack snarled the gangs into the room*) were relevant for this category, and the participants' correct production rate is plotted in Figure 4.7. Although no explicit instruction was provided to the participants on the formulation of prepositional VtR embedded by an intransitive verb, across all test sessions, the chatbot group participants showed the largest increase in the correct production of the unselected prepositional VtR, followed in order by the human, video, and control groups. To summarize, these results suggest that corrective feedback from either the AI chatbot or human instructor is more effective than explicit instruction in formulating unselected VtR expressions embedded by intransitive verbs.



**Figure 4.7. Correct answer rate for an unselected prepositional VtR with four relevant test items (error bars represent standard errors)**

Finally, due to L1 transfer effects, Korean L2 English learners could not formulate an English-like small clause with an overt NP subject as firmly as their native English counterparts. Table 4.16 reports on the frequency of the alternative syntactic patterns in EWT by Korean L2 English learners. Specifically, Korean adjectival and prepositional VtR expressions rely on either an atelic adjunct small clause without an overt subject or a bi-clausal strategy. Considering cross-linguistic variation, Korean L2 English learners' production of alternative patterns was grouped into the following categories: (1) incorrect formulation of Small Clause (SC), (2) serialization or coordination (V1\_V2), (3) adjunction (AdjNt), (4) preposition-related errors

(Pre.), (5) others, and (6) no response. The incorrect formulation of a small clause is further subdivided into three types: (a) missing direct object, (b) missing result phrase, and (c) missing manner verb. Sentences from (1) to (5) are examples of each pattern observed in the EWT.

**Table 4.16 Frequency of Alternative Syntactic Patterns in EWT**

	Chatbot (%)			Human (%)			Video (%)			Control (%)		
	Pre	IP	DP									
<b>Correct Response</b>	<b>32.81</b>	<b>69.06</b>	<b>69.37</b>	<b>33.75</b>	<b>55.62</b>	<b>55</b>	<b>31.87</b>	<b>43.12</b>	<b>40.94</b>	<b>33.59</b>	<b>37.5</b>	<b>37.19</b>
<b>Incorrect Response</b>												
<b>(1) SC</b>	<b>19.84</b>	<b>13.75</b>	<b>14.2</b>	<b>17.97</b>	<b>17.19</b>	<b>18.28</b>	<b>19.85</b>	<b>18.13</b>	<b>22.19</b>	<b>16.25</b>	<b>14.85</b>	<b>17.34</b>
-Missing Direct Object	11.09	9.22	9.84	9.06	12.9	10.5	11.7	10.8	16.88	8.75	8.91	10.31
- Missing Result Phrase	4.69	2.34	2.66	4.84	2.19	4.06	3.75	4.06	2.81	4.53	2.5	3.13
- Missing Manner Verb	4.06	2.19	1.7	4.07	2.1	3.72	4.4	3.27	2.5	2.97	3.44	3.9
<b>(2) V1_V2</b>	<b>2.97</b>	<b>5</b>	<b>4.38</b>	<b>14.38</b>	<b>5.78</b>	<b>5.78</b>	<b>10.16</b>	<b>7.19</b>	<b>8.91</b>	<b>13.75</b>	<b>11.25</b>	<b>9.69</b>
<b>(3) AdjNt</b>	<b>12.19</b>	<b>2.97</b>	<b>2.5</b>	<b>16.88</b>	<b>5.47</b>	<b>5.78</b>	<b>9.38</b>	<b>4.69</b>	<b>5.63</b>	<b>20.94</b>	<b>17.19</b>	<b>16.72</b>
<b>(4) Prep.</b>	<b>4.06</b>	<b>2.5</b>	<b>2.81</b>	<b>2.34</b>	<b>3.13</b>	<b>2.5</b>	<b>3.13</b>	<b>3.28</b>	<b>1.88</b>	<b>3.44</b>	<b>5.63</b>	<b>5.63</b>
<b>(5) Others</b>	<b>4.22</b>	<b>3.13</b>	<b>3.13</b>	<b>3.75</b>	<b>4.53</b>	<b>3.91</b>	<b>9.22</b>	<b>8.12</b>	<b>5.14</b>	<b>3.59</b>	<b>3.28</b>	<b>3.74</b>
<b>(6) No Response</b>	<b>23.91</b>	<b>3.39</b>	<b>3.59</b>	<b>10.93</b>	<b>8.28</b>	<b>8.75</b>	<b>16.39</b>	<b>15.47</b>	<b>15.31</b>	<b>8.44</b>	<b>10.3</b>	<b>9.69</b>
<b>Total</b>	<b>100</b>											

[(1) Incorrect Formulation of Small Clause (Small Clause)]

- a. Missing Direct Object: *He danced dizzy.*

(Target Structure: He danced himself dizzy.)

- b. Missing Result Phrase: *He brushed the teeth.*

(Target Structure: He brushed the teeth clean.)

- c. Missing Manner Verb: *He shut the window.*

(Target Structure: He nailed the window shut.)

[(2) Serialization or Coordination (V1\_V2)]

*He sneezed the tissue fall off.*

(Target Structure: He sneezed the tissue off.)

[(3) Adjunction(AdjNt)]

*He kicked the door to open.*

(Target Structure: He kicked the door open.)

[(4) Preposition (Prep.)]

*She blinked out dust from eyes.*

(Target Structure: She blinked the dust out of her eyes.)

[(5) Others]

*He was sliced the potato thin.*

(Target Structure: He sliced the potato thin.)

Table 4.21 shows that the instructional treatment was effective for all instructional group participants with respect to the adjunction and preposition-related errors in the chatbot, human, and video groups between the pretest and the immediate test and between the pretest and the delayed posttest. However, the chatbot group participants generally produced the lowest percentage of such errors in the immediate posttest and delayed posttest session, except for the prepositional errors in the delayed posttest session.

The only error type for which the chatbot group failed to show a decrease in terms of occurrence frequency was in the serialization and coordination error. There was an increase in the error frequency for the serialization and coordination error between the pretest and the immediate posttest, and between the pretest and the delayed posttest. Still, the percentage of errors by the chatbot group was smaller than for any other instructional groups in the immediate posttest session and in the delayed posttest session.

Table 4.16 demonstrates that the corrective feedback from the AI chatbot notably decreased the frequency of no response occurrences between the pretest and the immediate posttest and between the pretest and the delayed posttest. However, such a developmental pattern was not found for other instructional group participants.

A more detailed analysis of the errors in the formulation of the English small clause is warranted. As demonstrated in Table 4.21, the errors in the formulation of the English small clause were among the most persistent errors by all participants across all groups. Although there are many positive aspects to instructional intervention, the error frequency in the formulation of the small clause did not decrease significantly in the human, video, and control groups. Only the chatbot group participants decreased the relevant error frequency between the pretest and the immediate posttest and between the pretest and the delayed posttest.

Most of the errors on the formulation of the small clause were related to learners' failure to provide an overt NP subject in the small clause, which is categorized as an object omission error (i.e., a missing direct error), as displayed in Table 4.21. Also, as Table 4.16 demonstrates, the object omission error was one of the most prominent errors for all participants in the formulation of English VtR across the three test sessions. S. Kim (2016)'s study supports this finding, namely, that the direct object omission error was one of the most frequent errors committed by Korean L2 English learners in formulating English adjectival VtR expressions. She also reported that this object omission error reflects the fact that Korean adjectival VtR does not obey DOR.

To deal with the object omission errors, this study provided several instructional treatments for the learners, while these instructional interventions were less effective in decreasing the error rates for relevant errors than was the case for corrective feedback from the AI chatbot. Specifically, even after the instructional treatment was provided, the frequency of the object omission errors in the human and video groups increased between the pretest and the immediate posttest and between the pretest and the delayed posttest. However, a decrease in the object omission error was only observed in the chatbot group between the pretest and the immediate posttest, and between the pretest and the delayed posttest.

To sum up, the corrective feedback from the AI chatbot provides Korean L2 English learners with opportunities to unlearn their L1 properties in the production of adjectival or prepositional VtR expressions in English. Specifically, the corrective feedback from the AI chatbot guided the learners to produce unselected adjectival and prepositional VtR expressions by generating a target-like small clause with an overt subject embedded by an intransitive verb. However, such instructional effects were rarely found in other human-based instruction groups (e.g., the human and video groups) or the control group.

#### 4.2.4. Discussion

The third research question, RQ2(a), is whether the chatbot-based instruction was as effective as human-based instruction (e.g., human and video groups) for Korean L2 English learners' production of English adjectival VtR expressions for which they received explicit instruction. To address this research question, the relevant data for EWT were provided in Section 4.2.1. The results clearly confirmed that corrective feedback from the AI chatbot positively affected the production of prepositional VtR that was explicitly instructed in the instructional sessions.

The between-test differences in the EWT for the instructed adjectival VtR demonstrated a statistically significant mean score difference for the chatbot group, between the pretest and the immediate posttest, and between the pretest and the delayed posttest. Similarly, significant differences were found in the mean score for the human group between the pretest and the immediate posttest, and between the pretest and the delayed posttest. However, across tests, no statistical differences were found for the human group or control group. A between-group analysis of the EWT revealed that learners in the chatbot group were much better regarding production of adjectival VtR. In the immediate posttest of the EWT, every group that received an instructional intervention demonstrated a mean score increase

for the instructed adjectival VtR. However, the highest mean score was found in the chatbot group, followed in order by the human, video, and control groups. A statistically significant mean score difference was reported between the chatbot group and the video, and between the chatbot group and the control group. The mean score in the human group was statistically different from that of the control group. Similar patterns were found in the delayed posttest session, with the highest mean score for the chatbot group, followed by the human, video, and control groups. However, no statistical difference was found between the chatbot and human group, which suggests that chatbot-based instruction was as effective as human-based instruction.

The final research question of the present study, RQ2(b), is whether chatbot-based instruction was as effective as human-based instruction for Korean L2 English learners' production of uninstructed prepositional VtR expressions that were not explicitly taught. To address this research question, the relevant data for EWT are provided in Section 4.2.2. The results clearly demonstrate that corrective feedback from the AI chatbot was generalized to the production of prepositional VtR that was not explicitly taught in the instructional sessions.<sup>30</sup>

---

<sup>30</sup> After the instructional intervention, the rate of correct formulation of adjectival VtR out of *correct* formulation of prepositional VtR was 62.1%. On the other hand, the percentage of correct provision of adjectival VtR out of *incorrect* formulation of prepositional VtR was only 34.7%. Thus, it is reasonable to suggest that there is

The results of the between-test analysis showed a notable EWT mean score difference for the uninstructed prepositional VtR. Across the test sessions, the largest mean score increase was found in the chatbot group. For this group there was a statistical difference between the pretest and immediate posttest, and between the pretest and delayed posttest. In addition, a notable difference was found in the human group between the pretest and immediate posttest and between the pretest and delayed posttest. However, there was no statistical difference across the test sessions between the video and control groups.

The results of the between-group analysis clearly demonstrated a notable EWT mean score difference for the uninstructed prepositional VtR in the immediate posttest, between the chatbot group and the video and control groups. In addition, in the delayed posttest, there was a mean score difference between the chatbot group and the other participant groups, except that the human group exhibited a statistical significance. In both tests, no statistical difference was found between the chatbot and human groups.

There seem to be two plausible explanations for these positive interlanguage changes in the chatbot group. First, the relative instructional effectiveness of the corrective feedback from the chatbot may be attributed

---

a certain relationship between the learning of adjectival VtR and prepositional VtR.

to its consistent and multiple provisions of various types of corrective feedback. The chatbot-based corrective feedback in the present study focused on the consistent provision of error corrections to make learners notice the gap between their erroneous utterances and the target forms. In this study, a loop of corrective feedback composed of (i) repetition and clarification request, (ii) elicitation, and (iii) full recast was constantly provided to the learners, enabling them to notice the linguistic gap between their utterances and the target forms (Doughty & Varela, 1998).

Second, the effectiveness of corrective feedback from the AI chatbot also may be ascribed to its spontaneity. Corrective feedback from the chatbot was distinct from human-based corrective feedback in that the former provided more instant and spontaneous feedback to the learners than the latter. As several studies have demonstrated (Doughty, 2001; Schmidt, 1990), providing immediate corrective feedback to induce learners' instant modified output constitutes a "window of opportunity" (Doughty, 2001, p. 257) that invites learners to recognize a linguistic gap between (i) their communicative intents and (ii) a grammatical formulation of the target form. In this way, knowledge of the form and meaning pairings in the adjectival VtR was more likely to be retained and generalized, even when the learners were required to produce uninstructed prepositional VtR expressions during the 4-week interval between the immediate and the delayed posttest.

Indeed, the instructional effects of the chatbot group were obtained as a reflection of unlearning L1 properties in the production of adjectival and prepositional VtR expressions in English. The results covered in Section 4.2.3 demonstrate that the corrective feedback from the AI chatbot affected the correct production of unselected adjectival or prepositional VtR expressions. In addition, the corrective feedback from the AI chatbot decreased the mean frequency of the generation of L1-transferred errors (e.g., a missing direct object in the formulation of the English small clause) in the production of adjectival or prepositional VtR expressions. Although several interlanguage errors are still untreated by corrective feedback coming from the AI chatbot (e.g., V1\_V2, or errors in the formulation of a reflexive object for the unselected adjectival VtR), these developmental issues will be dealt with if the noticeability of the corrective feedback from the AI chatbot is enhanced (see Table 4.23 for the relevant information).

Overall, the instructional effect of the AI chatbot observed in this study mainly stems from customized learning. Depending upon their linguistic proficiency, learners received corrective recast. For example, more proficient language learners could self-correct their errors with a less explicit type of corrective feedback (i.e., clarification request and repetition), but less proficient language learners could do that with a more explicit type of corrective feedback (i.e., elicitation and full recast). This allowed them to

have more input exposure and more opportunities for language practices and further 'notice' the target language input, which helped them better 'intake' the input.

## Chapter 5. Conclusion

This chapter draws conclusions for the present research based on the results of the data analysis. The first section summarizes the major findings of the current study and proposes pedagogical implications; the second section discusses the limitations of this study and provides suggestions for further research.

### 5.1. Major Findings and Pedagogical Implications

The present study explored the effects of instructional intervention on students' learning of adjectival English transitive resultative construction (VtR). Three types of instruction—corrective feedback from a chatbot or a human instructor and explicit instructions from video material—were provided to Korean high school L2 English learners. The results presented in the previous chapter strongly suggest that chatbot-based instruction was as effective or more effective than human-based instruction (e.g., human or video groups) in Korean high-school English learners' comprehension and production of English transitive resultative constructions.

The first major finding of the current study is that chatbot-based instruction was as effective as human-based instruction (e.g., corrective feedback from a human instructor or additional explicit instruction stemming from video material) for Korean L2 English learners' comprehension of English adjectival VtR expressions that were explicitly instructed. In the immediate posttest of AJT, greater mean score increases were identified for the instructional treatment groups—namely, chatbot, human, and video—resulting in statistically significant mean scores that were different from those of the non-instruction group. These mean score differences remained in the delayed posttest session, indicating a positive long-term effect of corrective feedback coming from the AI chatbot, which was as effective as other human-based instructional interventions for the comprehension of instructed adjectival VtR expressions. However, across all test sessions, the instructional treatment of the adjectival VtR was not generalized to the prepositional VtR that was not explicitly instructed. This finding may simply imply that the corrective feedback from the AI chatbot was as effective as other human-based instructional interventions in the comprehension of English adjectival VtR expressions offered by Korean L2 English learners.

The second major finding of the present study is that chatbot-based instruction was more effective than the other types of human-based

instruction in the production of adjectival English VtR by Korean L2 English learners. A statistical analysis of the immediate posttest of EWT revealed that the chatbot group had higher mean score increases than the other groups (i.e., human, video, and control) in the correct formulation of English adjectival VtR expressions. A statistical difference was found between the chatbot group and the video and control groups. This difference held true in the delayed posttest session, suggesting a long-term effect from the instructional intervention of corrective feedback from the AI chatbot. Considering the cross-linguistic variation between Korean and English in the formulation of English adjectival VtR, the gradient provision in the explicitness of corrective feedback from the AI chatbot positively affected Korean L2 English learners' formulation of English adjectival VtR expressions without relying on their L1 syntactic properties, except for the unselected VtR expressions embedded by an intransitive verb and a fake reflexive object (e.g., *He danced himself dizzy*). This finding may indicate that corrective feedback from the AI chatbot is beneficial for Korean L2 English learners' production of the typologically distinct English adjectival VtR, which was explicitly instructed.<sup>31</sup>

---

<sup>31</sup> More prominent learning gains observed in EWT rather than in AJT may be ascribed to the activity types in which the learners participated. Specifically, the chatbot group learners were mainly asked to produce correct formulations of adjectival VtR. If the chatbot activities included comprehension-wise activities, such as forced-choice tasks or error correction tasks, it might be possible that their

The last major finding is that corrective feedback from the AI chatbot was generalizable to the production of prepositional VtR expressions that were not explicitly instructed. Greater mean score increases were found in the immediate posttest of EWT for the chatbot group participants than for the other instructional group participants. A statistical difference was found between the chatbot group and the video group, and between the chatbot group and the control group. The mean score differences remained in the delayed posttest session, which indicates a more persistent instructional effect due to the corrective feedback from the AI chatbot rather than from other human-based instructional interventions with regard to the production of uninstructed prepositional VtR expressions. Corrective feedback enabled approximately 60% of the chatbot participants to produce correct formulations of prepositional VtR expressions that were embedded not only by transitive verbs (e.g., *He broke an egg into a bowl*) but also by intransitive verbs (e.g., *She stared Tom out of the room*). On the contrary, such instructional effects were not apparently observed in the human, video, and control groups. This finding may suggest the possibility that corrective feedback from the AI chatbot on the adjectival VtR expressions is effective for the linguistic generalization of its interrelated argument structure construction—namely, the prepositional VtR.

---

AJT scores improved. However, future studies are called for to test this possibility.

The findings of the present study offer meaningful pedagogical implications for the teaching and learning of English via an AI chatbot in an EFL context. The first pedagogical implication is that linguistic consideration provides a theoretical foundation for the effective design of an AI chatbot for foreign language learning. It often has been a mistaken notion that English argument structure constructions (e.g., transitive construction, ditransitive construction, adjectival VtR, and prepositional VtR) constitute separate syntactic and semantic entities that are linguistically irrelevant to each other. Most AI chatbots currently on the market ignore the possibility that teaching one argument structure may lead to the generalization of other constructions. The present study investigated whether teaching one argument structure construction affects the learning of other syntactically- and semantically-related argument structure constructions. Thus, in the current research, corrective feedback via an AI chatbot was provided only regarding adjectival VtR. Adjectival VtR expressions comprise one of the most marked argument structure constructions. Previous studies have found that learners are able to acquire marked and unmarked constructions when they are provided with marked construction from an implicational hierarchy, not the other way around (Doughty, 2001; Ozeki & Shirai, 2007). The results of the present study support this finding; the instructional effect of the AI chatbot for the adjectival VtR was generalized to the production of a

prepositional VtR that was not explicitly instructed. Thus, in the design of the AI chatbot for foreign language learning, such interrelatedness must be considered in order to maximize the instructional effects and reduce the burdens of learning a separate argument structure construction.

The second pedagogical implication is that corrective feedback from the AI chatbot on the formulation of English VtR is a viable alternative to corrective feedback stemming from human instructors. As an alternative to a human instructor, an AI chatbot can provide Korean L2 English learners with corrective feedback to deal with their persistent erroneous production of English adjectival VtR. As reviewed in Chapter 2 (Section 2.1), the Korean small clause is devoid of an overt NP subject whereas the English small clause is not. The small clause also is selected by either a transitive verb or an intransitive verb in English, while it is not selected by an intransitive verb in Korean. These cross-linguistic differences lead to the availability or unavailability of DOR between English and Korean, resulting in Korean L2 English learners' deviant production of adjectival and prepositional VtR expressions. To address such developmental errors, the present study proposed the use of an AI chatbot in the provision of corrective feedback for English adjectival VtR. The results of the EWT in the present study clearly demonstrate that corrective feedback from the AI chatbot contributes to the correct formulation of adjectival or prepositional

VtR in English being more effective than corrective feedback originating from a human instructor. Thus, corrective feedback from an AI chatbot is a practical supplement for corrective feedback from a human instructor.

The final pedagogical implication of the present study is that corrective feedback from an AI chatbot must be as noticeable as possible. Although explicit instruction and corrective feedback were fully provided, the chatbot group participants demonstrated limited generalization for the comprehension of prepositional VtR expressions. The researcher admits that the corrective feedback from the AI chatbot employed in this study may not be so noticeable for language learners that it leads to self-repair of their errors. Thus, the corrective feedback generated by AI chatbots for foreign language learning must be as explicit as possible in order to help learners notice their linguistic gaps and engage in self-repair their erroneous utterances.

To conclude, AI chatbots for foreign language learning purposes should be designed based upon the triangulation of technological, linguistic, and pedagogical considerations (N.Ellis & Bogart, 2008). The present study demonstrated that corrective feedback from an AI chatbot for a more marked construction (e.g., adjectival VtR) positively affects the comprehension and production of relevant constructions and is further generalizable to the production of less marked construction (e.g.,

prepositional VtR). In this respect, instructional focus should be placed on the provision of corrective feedback for adjectival VtR.

Unfortunately, the general practices of building AI chatbots for foreign language learning have demonstrated that linguistic or pedagogical insights from linguistic and SLA research have frequently been overwhelmed by technological considerations, such as automatic speech recognition (ASR), natural language processing (NLP), or computer vision (CV). As a result, relatively little attention has been paid to the linguistic properties in the organization of the language curriculum for AI chatbot-based language learning. To the author's knowledge, no currently available AI chatbots have been designed with serious consideration of linguistic principles, such as the previously mentioned constructional network or cross-linguistic differences between L1 Korean and L2 English. This is one reason why the instructional effect of an AI chatbot for foreign language learning has reportedly had a short-term effect; after the novelty effect of an AI chatbot disappears, no constant instructional effect is reported.

Contrary to these previous practices, the present study reports that an AI chatbot designed with a special emphasis on the consideration of linguistic properties can have a permanent instructional effect. Thus, the interdisciplinary efforts for the AI chatbot design for foreign language learning should be guided more by linguistic or pedagogical insights rather

than by cutting-edge technological issues from the field of computer science.

## **5.2. Limitations and Suggestions**

This study's practical limitations are worth mentioning to provide meaningful suggestions for future studies. First, the current research was conducted with Korean high school students in Seoul, and it is therefore too early to generalize this finding to other foreign language contexts. Only when similar observations are found in other foreign language settings, such as those involving different first languages, student ages, or regions, will it be possible to draw the conclusion that corrective feedback from an AI chatbot is effective when teaching transitive resultative constructions to foreign language learners.

Second, the AI chatbot-based instruction in the present research did not explore the question of whether the acquisition of transitive resultative construction leads to the development of other argument structure construction. For example, the pedagogical intervention provided the overall form and meaning pairings of an adjectival transitive resultative construction and asked the learners to extend this understanding to its prepositional counterpart. However, the study did not investigate whether

learners could extend this understanding to other interrelated constructions, such as intransitive resultative construction, intransitive motion construction, intransitive literal verb particle construction, intransitive figurative verb particle construction, and ditransitive construction. In addition, this study could not explore the extent to which chatbot-based instruction on the adjectival VtR could contribute to the development of learners' oral or written English proficiency, measured by Complexity, Accuracy, and Fluency (Bygate, Skehan & Swain, 2013; Skehan, 2001).

Lastly, the recast from the AI chatbot employed in the present study did not provide learners with an exact source of errors. As recast is a kind of implicit negative feedback in which learners' focus is attracted to the mismatches between input and output (Long, 1996), the exact identification of erroneous forms may lead learners to clearly recognize and correct them (Gass, 2003). Unfortunately, the recast of the AI chatbot employed in the current research only provided learners with the indirect information that their utterances included an erroneous portion. This stems from a technological restriction that makes it practically impossible for the AI chatbot to classify error types from learners and provide them with exact cues for the location of such errors. This caused inconveniences for the learners, led to learner dissatisfaction with the system, and obscured the actual effect of recasting from the AI chatbot. Therefore, a technologically

more mature design of a recast providing system (e.g., (i) spotting exact error sources, (ii) categorizing error types automatically through a machine learning (ML) or deep-learning (DL) framework, or (iii) predicting a future learning path in a ML/DL-based adaptive learning framework) will more reliably clarify the advantages of recast feedback (Oh et al., 2020).

Notwithstanding these limitations, the current research is significant, as it is the first empirical effort to develop and administer chatbot-based instruction designed by linguistic and pedagogical principles and show its relative effectiveness in the production and comprehension of English transitive resultative constructions put forth by Korean high-school learners. Furthermore, this novel approach, which emphasizes linguistic and pedagogic principles over technological ones, provides an alternative to the traditional model of AI chatbot design for foreign language, whose consideration is mainly restricted to technological concerns. Consequently, the present study sheds new light on the important issue of how to make English learning more efficient in EFL settings by means of optimizing the design of an AI chatbot for foreign language learning.

Future research considering and overcoming these limitations will provide a more comprehensive design for the practical application of an AI chatbot in foreign language learning. For instance, it would be meaningful to investigate the instructional effect of an AI chatbot on the learning of

literal VPC, and its extension to other constructions (e.g., figurative VPC, adjectival VtR, prepositional VtR), or to the development of oral proficiency (i.e., complexity, accuracy, and fluency, Robinson, 2001). In addition, as Kim and Rah (2020) have indicated, it is important to investigate the instructional effects of the AI chatbot, whose corrective feedback is sequenced by the linguistic complexity of the English argument structure construction. Finally, long-term research from quantitative and qualitative standpoints that focus on the process of foreign language acquisition via human-chatbot interaction could be expected to provide insightful results in this type of acquisition study.

## REFERENCES

- Ai, H. (2015). *Promoting second language development with concept-based language instruction and intelligent computer-assisted language learning*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78(4), 465-483.
- Ambridge, B., Maitreyee, R., Tatsumi, T., Doherty, L., Zicherman, S., Pedro, P. M., ... & Mendoza, M. J. (2020). The crosslinguistic acquisition of sentence structure: Computational modeling and grammaticality judgments from adult and child speakers of English, Japanese, Hindi, Hebrew and K'iche'. *Cognition*, 202. Advance online publication. <https://doi.org/10.1016/j.cognition.2020.104310>
- Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in second language acquisition*, 28(4), 543-574.
- Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, 10, 79-99.
- Arroyo, D. C., & Yilmaz, Y. (2018). An open for replication study: The role of feedback timing in synchronous computer-mediated communication. *Language Learning*, 68(4), 942-972.
- Aske, J. (1989). Path predicates in English and Spanish: A closer look. In K. Hall, M. Meacham, & R. Shapiro (Eds.), *In Proceedings of the Berkeley Linguistics Society* (vol. 15, pp.1–14). UC Berkeley: Berkeley, CA.
- Ayedoun, E., Hayashi, Y., & Seta, K. (2015). A conversational agent to

- encourage willingness to communicate in the context of English as a foreign language. *Procedia Computer Science*, 60, 1433-1442.
- Barbieri, E., Basso, A., Frustaci, M., & Luzzatti, C. (2010). Argument structure deficits in aphasia: New perspective on models of lexical production. *Aphasiology*, 24(11), 1400-1423.
- Baur, C., Rayner, E. & Tsourakis, N. (2015). What motivates students to use online CALL systems? A case study. In *Proceedings of the 9th International Technology, Education and Development Conference (INTED 2015)* (pp. 2331– 2338). Madrid: IATED.
- Beal, C. R., Johnson, W. L., Dabrowski, R., & Wu, S. (2005). Individualized feedback and simulation-based practice in the tactical language training system: an experimental evaluation. In *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 747-749). Amsterdam: IOS Press.
- Baker, C. L. (1995). *English Syntax*. Cambridge: The MIT Press.
- Baker, M., and Stewart, O.T. (1999). On Double-Headedness and the Anatomy of the Clause. Unpublished manuscript, Rutgers University. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.6340&rep=rep1&type=pdf>
- Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, 32(8), 827-877.
- Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *Modern Language Journal*, 93(2), 227-240.
- Boas, H. C. (2003). *A Constructional Approach to Resultatives*. Stanford:

CSLI publications.

- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 945–984). New York: Academic Press.
- Bouillon, P., Rayner, E., Tsourakis, N., & Qinglu, Z. (2011). A student-centered evaluation of a web-based spoken translation game. In *Proceedings of the SLaTE Workshop*, Venice, Italy.
- Bowers, J. (1993). The syntax of predication. *Linguistic Inquiry* 24(4). 591-656.
- Bowers, J. (2002). Transitivity. *Linguistic Inquiry* 33(2). 183-224.
- Bygate, M., Skehan, P., & Swain, M. (2001). (Eds.). *Researching pedagogic tasks: Second language learning, teaching, and testing*. London: Longman.
- Cabrera, M. and Zubizarreta, M. (2005). Overgeneralization of Causatives and Transfer in L2 Spanish and L2 English. In D. Eddington (Ed.) *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 15-30). Somerville, MA: Cascadilla Proceedings Project.
- Carnie, A. (2012). *Syntax: A generative introduction*. (3rd ed.). Malden, MA: Wiley-Blackwell.
- Carrier, J., & Randall, H. (1992). The argument structure and syntactic structure of resultatives. *Linguistic Inquiry* 23(2). 173-234.
- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in second language acquisition*, 15(3), 357-386.
- Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, 27(1), 29-46.
- Chiu, T.L., Liou, H.C., & Yeh, Y. (2007). A study of web-based oral

- activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning*, 20(3), 209-233.
- Choi, I.C. (2016). Efficacy of an ICALL tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, 29(2), 334-364.
- Choi, I. C., & Jang, S. (2020). Efficacy of Learning-Oriented Diagnostic Assessment based on Process-oriented Automated Error Analysis. *Multimedia-Assisted Language Learning*, 23(3), 11-44.
- Choi, S. & Bowerman, M. (1991). Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. *Cognition* 41, 83-121.
- Choi, S. (1997). Language-specific input and early semantic development: Premises, predictions, and evidence. *Language and Cognitive Processes*, 16, 113-142.
- Choi, S. (2018). Language-specificity in motion expression: Early acquisition in Korean compared to French and English. In M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Sources of Variation in First Language Acquisition: Languages, Contexts, and Learners (Trends in Language Acquisition Research)* (pp. 103-122). Amsterdam: John Benjamins.
- Choi, S. (2020). Development of Clause Chaining in Korean. *Frontiers of Psychology*. 11:256. doi: 10.3389/fpsyg.2020.00256.
- Chomsky, N. (1980). On binding. *Linguistic Inquiry* 11(1), 1-46.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Eds.), *Ken Hale: A Life in Language* (pp. 1-52). Cambridge, MA: MIT Press.
- Choo, S., & Min, D. K. (2019). A study of using task-based artificial

- intelligence (AI) chatbot for further interaction in English and the analysis of students' production. *Primary English Education*, 25(2), 27-52. [Written in Korean]
- Chung, C., & Kim, J.B. (2008). Korean serial verb constructions: A construction-based approach. *Studies in Generative Grammar*, 18(4), 559-582.
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems* (Vol. 2). Cambridge, UK. Cambridge University Press.
- Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: University of Chicago Press.
- Cormack, A., & Smith, N. (1999). Why are depictives different from resultatives? *UCL working papers in linguistics*, 11, 251-284.
- Croft, W. (1998). Linguistic evidence and mental representations. *Cognitive Linguistics*, 9(2), 151-174.
- De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language learning*, 46(3), 529-555.
- Deng, L., & Liu, Y. (Eds.). (2018). *Deep learning in natural language processing*. Singapore: Springer.
- Department for Education. (2019). *Realising the potential of technology in education: A strategy for education provider and the technology industry*. Retrieved June 2, 2020, from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/791931/DfE-Education\\_Technology\\_Strategy.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/791931/DfE-Education_Technology_Strategy.pdf).
- Diessel, H. and Tomasello, M. (2005) Particle placement in early child language: a multifactorial analysis. *Corpus Linguistics and Linguistic Theory* 1(1). 89-112.

- Dixon, R. M. W. (2000). A typology of causatives: Form, syntax and meaning. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Changing valency: Case studies in transitivity* (pp. 30-83). Cambridge, UK: Cambridge University Press.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). New York: Cambridge University Press.
- Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 206–257). New York: Cambridge University Press.
- Dowty, D. R. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTO*. Dordrecht: Reidel.
- Egi, T. (2007). Interpreting recasts as linguistic evidence: The roles of linguistic target, length, and degree of change. *Studies in Second language acquisition*, 29(4), 511-537.
- Ellis, N. C., & Bogart, P. S. (2007). Speech and language technology in education: the perspective from SLA research and practice. In *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education* (pp. 1–8). Farmington, PA: Carnegie Mellon University and ISCA Archive.
- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 1(1), 3-18.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, 28(2), 339-368.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford:

Oxford University Press.

- Ellis, R., & Sheen, Y. (2006). Re-examining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28, 575-600.
- Embick, D. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3). 355-392.
- Erlam, R., & Loewen, S. (2010). Implicit and explicit recasts in L2 oral French interaction. *Canadian Modern Language Review*, 66(6), 877-905.
- Fleischhauer, J. (2013). Interaction of telicity and degree gradation in change of state verbs. In B. Arsenijević, B. Gehrke, R. Marín (Eds.) *Studies in the Composition and Decomposition of Event Predicates* (pp. 125-152). Springer: Netherlands.
- Fodor, A. (1970). Three reasons for not deriving 'kill' from 'cause to die'. *Linguistic Inquiry* 1.429-38.
- Folli, R., & Harley, H (2004). Consuming Results in Italian & English: Flavors of *v*. In P. Kempchinsky & R. Slabakova (Eds.), *Aspect Book Title* (pp.1-25). Holland: Kluwer Academic Publishers.
- Folli, R., & Harley, H. (2006). On the licensing of causatives of directed motion: Waltzing Matilda all over. *Studia Linguistica*, 60(2), 121-155.
- Folli, R., & Harley, H. (2020). A head movement approach to Talmy's typology. *Linguistic Inquiry*, 51(3), 425-470.
- Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning*, 15(4), 329-342.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E., & Jackendoff, R. (2004). The English resultative as a

- family of constructions. *Language*, 532-568.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, A. E. (2019). *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2017). *Deep learning*. Cambridge, MA: MIT press.
- Gorlach, M. (2004). *Phrasal constructions and resultativeness in English*. Aesterdam and Philadelphia: John Benjamins.
- Haiman, J. (1985). *Natural syntax: Iconicity and erosion*. Cambridge: Cambridge University Press.
- Hale, K., and Keyser, S, J. (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- Hampe, B. (2012). Transitive phrasal verbs in acquisition and use: A view from construction grammar. *Language Value* 4(1). 1-32.
- Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL quarterly*, 36(4), 543-572.
- Harless, W. G., Zier, M. A., & Duncan, R. C. (1999). Virtual dialogues with native speakers: The evaluation of an interactive multimedia method. *Calico Journal*, 16(3), 313-337.
- Harley, H. (2005). How do verbs get their names? Denominal verbs, manner incorporation and the ontology of verb roots in English. In N. Erteschik-Shir & T. Rapoport (Eds.), *The syntax of aspect* (pp. 42-64). Oxford: Oxford University Press.
- Hassani, K., Nahvi, A., & Ahmadi, A. (2016). Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, 24(1), 252-271.

- Havranek, G. (2002). When is corrective feedback most likely to succeed? *International Journal of Educational Research*, 37(3-4), 255-270.
- Heift, T. (2002). Learner Control and Error Correction in ICALL: Browsers, Peekers and Adamants. *CALICO Journal*, 19(3), 295-313.
- Heift, T. (2003). Multiple learner errors and meaningful feedback: A challenge for ICALL systems. *CALICO Journal*, 20(3), 533-548.
- Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, 16(2), 416-431.
- Heift, T., & Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL. *System*, 36(2), 196-213.
- Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Hoekstra, T., & Mulder, J. (1990). Unergatives as copular verbs: Locational and existential predication. *Linguistic Review* 7, 1-79.
- Holland, V. M., Kaplan, J. D., & Sabol, M. A. (1999). Preliminary tests of language learning in a speech-interactive graphics microworld. *Calico Journal*, 16(3), 339-359.
- Hong, S. (2005). *Exceptional Case-marking and resultative constructions*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Hong, S. (2020). Artificial Intelligence and EduTech in School Education. Korea Institute for Curriculum and Evaluation. RRI 2020-2
- Hong, S., Yoon, T., Lee, S., & Oh, E. J. (2021). An Analysis of Using Dialogue-Based Chatbot in Elementary English Education. *The Journal of Korea Elementary Education*, 31(5), 31-55. [Written in Korean]
- Hovav, M. R., & Levin, B. (2001). An event structure account of English

- resultatives. *Language* 77(4), 766-797.
- Hovav, M. R., & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of Linguistics*, 44(1), 129-167.
- Huang, Y.T., Lin, Y.L., Yang, J.C., & Wu, Y.C. (2008). An English dialogue companion system for supporting conversation practice. In *Proceedings of the 16<sup>th</sup> International Conference on Computers in Education (ICCE 2008)* (pp. 43-48). Taipei: APSCE.
- Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: language and cognition*, 3, 227-224.
- Huddleston, R. and Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge UK: Cambridge University Press.
- Hwang, J. (2011). The meanings of the Korean resultatives. *Korean Semantics* 36. 531- 556.
- Im, W., Kim, S., Cho, S., Lee, S., Kim, R. (2019). *Action Research Report: Developing AI Conversation-based English Tutor Mobile App for Elementary School Korean L2 English Learners* (Report Number: SER 2019-42). Seoul: Seoul Metropolitan Office of Education.  
<https://www.serii.re.kr/fus/MI0000000000000000487/board/BO00000361/ctgynone/view0010v.do>. [Written in Korean]
- Ionin, T., & Zyzik, E. (2014). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics*, 34, 37-64.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jang, Y. (2002). Small Clause and Default Case. In I-H. Lee et al. (Eds.) *Proceedings of the 16th Pacific Asia Conference on Language, Information, and Computation (PACLIC)* (pp. 123-134). Seoul:

Korean Society for Language and Information.

- Jia, J. (2009). An AI framework to teach English as a foreign language: CSIEC. *AI Magazine*, 30(2), 59-59.
- Jia, J., & Ruan, M. (2008). Use chatbot CSIEC to facilitate the individual learning in instruction: A case study. *Intelligent Tutoring Systems*, LNCS 5091, 706-708.
- Jeon, H., Chung, H., Kang, B., & Lee, Y. (2021). Survey of Recent Research in Education based on Artificial Intelligence. *Electronics and Telecommunications Trends*, 36(1), 71-80.
- Jepson, K. (2005). Conversations – and negotiated Interaction – in text and voice chatrooms. *Language Learning & Technology*, 9(3), 79–98.
- Johnson, W. L. (2007). Serious use of a serious game for language learning. In Dimitrova, Mizoguchi, du Boulay, & A. Graesser (Eds.) *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (pp. 67-74). Berlin: Springer-Verlag.
- Johnson, W. L. & Valente, A. (2009). Tactical Language and Culture Training Systems: Using AI to teach foreign languages and cultures. *AI Magazine*, 30(2), 72. doi:10.1609/aimag.v30i2.2240.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An introduction to speech recognition, computational linguistics and natural language processing*. Upper Saddle River, NJ: Prentice Hall.
- Kang, S. (1993). Serial verb constructions in Korean and their implications. *Studies in Generative Grammar* 3(1) 79-109.
- Kaplan, J. D., Sabol, M. A., Wisher, R. A., & Seidel, R. J. (1998). The military language tutor (MILT) program: An advanced authoring system. *Computer Assisted Language Learning*, 11(3), 265-287.
- Kartchava, E. (2019). *Noticing oral corrective feedback in the second language classroom: background and evidence*. London: Rowman

& Littlefield.

- Kartchava, E., & Ammar, A. (2014). The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research, 18*(4), 428-452.
- Kayne, R. (1985). Principles of particle constructions. In J. Gutron et al. (Eds.), *Grammatical Representation*, Dordrecht: Foris Publication.
- Kearns, K. 2007. Telic senses of deadjectival verbs. *Lingua 117*(1). 26-66.
- Kerly, A., Hall, P., & Bull, S. (2006). Bringing chatbots into education: Towards natural language negotiation of open learner models. In R. Ellis, T. Allen, & A. Tuson (Eds.), *Applications and Innovations in Intelligent Systems XIV* (pp. 179-192). London: Springer.
- Kim, H. (2017). The Argument and Event Structure Types of Serial Verb Constructions in Korean: The Case of ‘*V-e ka-ta/ o-ta*’, ‘*V-e cwu-ta*’. Unpublished doctoral dissertation, Seoul National University, Seoul. [Written in Korean]
- Kim, H., & Rah, Y. (2016). Effects of verb semantics and proficiency in second language use of constructional knowledge. *The Modern Language Journal, 100*(3), 716-731.
- Kim, H., & Rah, Y. (2021). Applying constructionist approaches to teaching English argument structure constructions to EFL learners. *TESOL Quarterly, 55*(2), 568-592.
- Kim, J., & James Yoon. (2008). An experimental syntactic study of binding of multiple anaphors in Korean. *Journal of Cognitive Science 9*(1), 1–30.
- Kim, N.-Y. (2016). *Effects of different voice-chat conditions on Korean EFL learners’ speaking ability, oral interaction, and affective factors*. Unpublished doctoral dissertation, Ewha Womans University, Seoul, Korea.

- Kim, R. (2018). A critical review of the impact of the fourth industrial revolution on the development of the basic communicative competence of Korean EFL learners. *Multimedia-Assisted Language Learning*, 21(3), 115-148. [written in Korean]
- Kim, R., & Yang, H. (2017). Why do nonnative English learners perform L2 statistical preemption less than native counterparts: The role of different repertoires for L1 and L2 constructions. Poster session at the 42<sup>nd</sup> Annual Boston University Conference on Language Development, Boston, 3-5 November 2017.
- Kim, S. (2016). *L1 influence on L2 learning of English resultative constructions: The syntactic and semantic structure of Korean students' interlanguage*. Unpublished Doctoral dissertation. Seoul National University, Seoul, Korea.
- Kim, S., Ko, H., & Yang, H.-K. (2020). Telicity and mode of merge in L2 acquisition of resultatives. *Language Acquisition*, 27(2), 117-159.
- Kim, S., & Lee, S. (2020). An analysis of a primary school English speaking lesson using an AI-powered conversational English learning application: A focus on student engagement as well as learners' and teachers' perceptions of the class. *Primary English Education*, 26(3), 177-202.
- Kim, S., & Maling, J. (1998). A crosslinguistic perspective on resultative formation. In Ralph Blight & Michelle Moosally (eds.), *Texas Linguistic Forum 38: The Syntax and Semantics of Predication* (pp. 189-204). Austin: University of Texas.
- Kim, Y. (1995). Verb lexicalization patterns in Korean and some issues of language acquisition. *Language Research*, 31(3), 501-543.
- Knott, A., & Vlugter, P. (2008). Multi-agent human-machine dialogue: issues in dialogue management and referring expression semantics.

- Artificial Intelligence*, 172, 69-102.
- Ko, H. (2011) Predication and edge effects. *Natural Language Linguist Theory* 29. 725-778.
- Ko, H. (2014). *Edges in syntax: scrambling and cyclic linearization (Vol. 50)*. Oxford: OUP.
- Ko, H. (2015). On the typology of Small Clauses: null subject and mode of merge in resultatives. *Studies in Generative Grammar* 25 (2). 347-375.
- Ko, H., & Sohn, D. (2015). Decomposing complex serialization: The role of *v*. *Korean Linguistics* 17(1). 78-125.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.
- Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. Harlow: Longman.
- Kratzer, A. (2004). Telicity and the meaning of objective case. In J. Guéron & J. Lecarme, (eds.), *The Syntax of Time* (pp. 398–425). Cambridge, MA.: MIT Press.
- Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry* 19(3). 335-391.
- Larson, R. K. (1990). Double objects revisited: Reply to Jackendoff. *Linguistic Inquiry* 21(4). 589-632.
- Lee, C. S. (1996). *A Predicational Approach to the Resultative and the Depictive Construction in English*. Unpublished Doctoral Dissertation, Seoul National University, Seoul.
- Lee, H. (2019). *Effects of Input Frequency Distribution Types on Korean EFL Middle School Students Learning of English Caused-Motion Construction*. Unpublished Master's Thesis. Seoul National University, Seoul.

- Lee, J.-H., & Kim, H. M. (2011). The L2 developmental sequence of English constructions and underlying factors. *Korean Journal of English Language and Linguistics*, 11(3), 577-600.
- Lee, K., Kweon, S.O., Lee, S., Noh, H., Lee, J., Lee, J., Kim, H.R., & Lee, G.G. (2011). Effects of language learning game on Korean elementary school students. *Speech and Language Technology in Education*,
- Lee, K., Lee, K.Y., & Sheehan, L. (2019). Hey Alexa! A magic spell of social glue?: Sharing a smart voice assistant speaker and its impact on users' perception of group harmony. *Information Systems Frontiers*, 22(3), 563-583.
- Lee, S., Noh, H., Lee, J., Lee, K., Lee, G.G., Sagong, S., & Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL*, 23(1), 25-58.
- Lee, S., Noh, H., Lee, J., Lee, K., & Lee, G.G. (2012). Foreign language tutoring in oral conversations using spoken dialog systems. *IEICE transactions on information and systems*, 95(5), 1216-1228.
- Lee, K., Kweon, S.O., Lee, S., Noh, H., & Lee, G.G. (2014). POSTECH immersive English study (POMY): Dialog-based language learning game. *IEICE transactions on information and systems*, 97(7), 1830-1841.
- Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in second language acquisition*, 25(1), 37-63.
- Levelt, W. J. M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38.
- Levinson, L. (2007). *The roots of verbs*. Unpublished doctoral dissertation,

New York University, New York.

- Levin, B. (1993). *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, B. (2020). Resultatives and constraints on concealed causatives. In Elitzur Bar-Asher Siegal & Nora Boneh (Eds.), *Perspectives on causation: Selected papers from the Jerusalem 2017 workshop* (pp. 185-217). Dordrecht: Springer.
- Levin, B., Hovav, M.R., & Keyser, S.J. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT press.
- Levshina, N. (2016). Finding the best fit for direct and indirect causation: a typological study. *Lingua Posnaniensis*, 58(2), 65-82.
- Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford: Oxford University Press.
- Li, L., Lee, K.Y., Emokpae, E., & Yang, S.B. (2021). What makes you continuously use chatbot services? Evidence from Chinese online travel agencies. *Electronic Markets*, 3, 1-25.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language learning*, 60(2), 309-365.
- Li, S., Zhu, Y., & Ellis, R. (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *Modern Language Journal*, 100(1), 276-295.
- Li, Y. (1993). *Structural head and aspectuality*. *Language* 69(3), 480-504.
- Lightbown, P. M., & Spada, N. (2011). *How Languages are Learned*. Oxford: Oxford University Press.
- Liu, K., Tolins, J., Tree, J.E.F., Neff, M., & Walker, M.A. (2015). Two techniques for assessing virtual agent personality. *IEEE Transactions on Affective Computing*, 7(1), 94-105.
- Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective

- feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 361-377). Oxford: Oxford University Press.
- Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *Modern Language Journal*, 90(4), 536-556.
- Long, M.H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). San Diego, CA: Academic Press.
- Long, M.H. (2015). *Second Language Acquisition and Task-Based Language Teaching*. Malden, MA: Wiley Blackwell.
- Long, M.H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, 82(3), 357-371.
- Long, M.H., & Robinson, P. (1998). Focus on form: Theory, research, process. In C.J. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 15-41). New York: Cambridge University Press.
- Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language learning*, 48(2), 183-218.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in second language acquisition*, 26(3), 399-432.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in second language acquisition*, 19(1), 37-66.

- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265-302.
- Ma, W., Adesope, O.O., Nesbit, J.C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901.
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557-587.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407–452). Oxford: Oxford University Press.
- Mackey, A., & Philip, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82(3), 338-356.
- Matushansky, O. (2019). Against the PredP Theory of Small Clauses. *Linguistic Inquiry*, 50(1), 63-104.
- McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, 56(4), 693-720.
- McIntyre, A. (2004). Event paths, conflation, argument structure, and VP shells. *Linguistics* 42, 523–571.
- Muysken, P. (1988). Parameters for serial verbs. In V. Manfredi (Ed.), *Niger-Congo syntax and semantics* (vol. 1, pp. 65-75). Boston: Boston University African Studies Center.
- Nagata, N. (1997). The effectiveness of computer-assisted metalinguistic instruction: A case study in Japanese. *Foreign Language Annals*, 30(2), 187-200.

- Nagata, N. (1998). Input vs. output practice in educational software for second language acquisition. *Language Learning & Technology*, 1(2), 23-40.
- Nagata, N. (2002). BANZAI: An Application of Natural Language Processing to Web based Language Learning. *CALICO Journal*, 19(3), 583-599.
- Nam, S. (2003). Lexical semantic Structures and argument alternations of movement verbs in Korean. *Language Research* 39(1), 111-145.  
[Written in Korean]
- Namaghi, S.A.O., Safaee, S.E., & Sobhanifar, A. (2015). The effect of shyness on English speaking scores of Iranian EFL learners. *Journal of Literature, Language and Linguistics*, 12, 22-28.
- Napoli, J. (1992). Secondary resultative predicates in Italian. *Journal of Linguistics* 28(1). 53-90.
- Noh, H., Ryu, S., Lee, D., Lee, K., Lee, C., & Lee, G.G. (2012). An example-based approach to ranking multiple dialog states for flexible dialog management. *IEEE journal of selected topics in signal processing*, 6(8), 943-958.
- O'Grady, W.D. (1991). *Categories and case: The sentence structure of Korean*. Amsterdam: John Benjamins.
- Oh, Y.R., Park, K., Jeon, H.B., & Park, J.G. (2020). Automatic proficiency assessment of Korean speech read aloud by non-natives using bidirectional LSTM-based speech recognition. *ETRI Journal*, 42(5), 761-772.
- Oh, C.Y. (2010). A study on category of Korean resultative. *The Journal of Studies in Language* 26(3). 595-619.
- Olson, A. (2013). *Constructions and Result: English phrasal verbs as analysed in Construction Grammar*. Unpublished master's thesis,

Trinity Western University, Langley, BC.

- Ozeki, H., & Shirai, Y. (2007). Does the noun phrase accessibility hierarchy predict the difficulty order in the acquisition of Japanese relative clauses? *Studies in Second Language Acquisition*, 29(2), 169-196.
- Park, K.S., & Lakshmanan, U. (2007). The unaccusative-unergative distinction in resultatives: Evidence from Korean L2 learners of English. In A. Belikova et.al., (Eds.), *Proceedings of the 2<sup>nd</sup> conference on generative approaches to language acquisition North America (GALANA)* (pp. 328-338). Somerville, MA: Cascadilla Proceedings Project.
- Park, M.K. (2011). Case in 'ECM' resultatives. *Linguistic Research* 28(1), 1-18.
- Petersen, K.A. (2010). *Implicit corrective feedback in computer-guided interaction: Does mode matter?* Unpublished doctoral dissertation, Georgetown University, Washington, DC.
- Pi, C.Y.T., & Stewart, O.T. (1998). Micro-events in two serial verb constructions. In *Proceedings from Semantics and Linguistic Theory (SALT)* 8, (pp. 202-214). Cornell: CLC publications.
- Poza, M. (2005). *The effects of asynchronous computer voice conferencing on learners' anxiety when speaking a foreign language*. Unpublished doctoral dissertation, West Virginia, West Virginia University.
- Purinton, A., Taft, J.G., Sannon, S., Bazarova, N.N., & Taylor, S.H. (2017). "alexa is my new bff": Social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2853-2859). New York, USA: ACM.
- Pylkkänen, L. (2002). *Introducing arguments*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

- Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English*. Unpublished doctoral dissertation. University of Toronto.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Radford, A. (1988). Small children's Small Clauses. *Transactions of the philological society*, 86(1), 1-43.
- Radford, A. (2009). *Analysing English sentences: A minimalist approach*. Cambridge: Cambridge University Press.
- Rah, Y. (2014). *Effects of Construction-grammar-based Instruction on the Sentence Production Ability of Korean College Learners of English*. Unpublished Doctoral Dissertation. Seoul National University, Seoul.
- Ramchand, G. (2001). First phase syntax. Ms., University of Tromsø.
- Ramchand, G. (2008). *Verb meaning and the lexicon*. Cambridge: Cambridge University Press.
- Ramchand, G., & Svenonius, P. (2002). The lexical syntax and lexical semantics of the verb-particle construction. In Mikkelsen, Line & Christopher Potts (Eds.), *Proceedings of WCCFL 21* (pp. 387-400), Somerville, MA: Cascadilla Press.
- Rappaport Hovav, M., and Levin, B. (1998). Building verb meanings. In M. Butt and W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 97-134). Stanford, CA: CSLI Publications.
- Rappaport Hovav, M., and Levin, B. (2010). Reflections on manner/result

- complementarity. In E. Doron, M. Rappaport Hovav, and I. Sichel (Eds.), *Syntax, lexical semantics, and event structure* (pp. 21-38). Oxford: Oxford University Press.
- Robenalt, C., & Goldberg, A.E. (2016). Nonnative speakers do not take competing alternative expressions into account the way native speakers do. *Language learning*, 66(1), 60-93.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge, England: Cambridge University Press.
- Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G.M., Brunskill, E., & Landay, J.A. (2021). EnglishBot: An AI-Powered Conversational System for Second Language Learning. In *Proceedings of IUI 21: 26th International Conference on Intelligent User Interfaces* (pp.434-444). New York, NY: ACM.
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133-164). Amsterdam: Benjamins.
- Salteh, M.A., & Sadeghi, K. (2012). Teachers' corrective feedback in L2 writing revisited: Concerns against and suggestions for its employment. *World Applied Sciences Journal*, 17(3), 375-383.
- Sannon, S., Stoll, B., DiFranzo, D., Jung, M., & Bazarova, N.N. (2018). How personification and interactivity influence stress-related disclosures to conversational agents. In *Proceedings of the Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp.285-288). ACM, New York, NY.

- Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R.W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). New York: Cambridge University Press.
- Schütze, H., Manning, C.D., & Raghavan, P. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press.
- Searle, J.R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3, 417-457.
- Semke, H.D. (1984). Effects of the Red Pen. *Foreign Language Annals*, 17(3), 195-202.
- Shawar, B.A., & Atwell, E. (2007). Different measurements metrics to evaluate a chatbot system. In F. Weng (Ed.), *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies* (pp. 89-96). New Brunswick, NJ: Association for Computational Linguistics.
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language teaching research*, 8(3), 263-300.
- Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language teaching research*, 10(4), 361-392.
- Shibatani, M., & Chung, S.Y. (2002). Japanese and Korean causatives revisited. *Japanese/Korean Linguistics* 10, 32-49.
- Shibatani, M., & Pardeshi, P. (2002). The causative continuum. In M. Shibatani (Eds.), *The grammar of causation and interpersonal manipulation*, (pp. 85-126). Amsterdam: Benjamins.
- Shim, J., & Den Dikken, M. (2007). The tense of resultatives: The case of

- Korean. In S. Anisa, M. Walkow & M. Abdurrahman (Eds.), *Proceedings of NELS (Vol. 38)* (pp. 1-22). Amherst, MA: GLSA Publications.
- Shin, G.H. (2010). On the contribution of argument structure constructions to sentence meaning for Korean learners of English. *English Teaching*, 65(4), 209-227.
- Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal*, 44(2), 144–152.
- Simpson, J. (1983). Resultatives. In Levin, Rappaport, and Zaenen (Eds.), *Papers in Lexical-Functional Grammar* (pp.143-157). Bloomington, Indiana: Indiana University Linguistics Club (IULC).
- Slabakova, R. (2000). L1 transfer revisited: the L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics*, 38(4), 739–770.
- Sohn, H. (2009). The semantics of clause linking in Korean. In R. M. W. Dixon & A. Aikhenvald (Eds.), *The Semantics of Clause Linking: A Cross-linguistic Typology* (pp. 285-317). Oxford: Oxford University Press.
- Son, M. (2007). Directionality and resultativity: The cross-linguistic correlation revisited. *Tromsø Working Papers on Language and Linguistics: Nordlyd*, 34(2), 126–164.
- Son, M. (2008). Resultatives in Korean revisited: complementation versus adjunction. *Tromsø Working Papers on Language & Linguistics: Nordlyd* 35, 89-113.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1-33.
- Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguistic Approaches to Bilingualism*, 6(5), 669-684.

- Stowell, T. 1981. *Origins of phrase structure*. Unpublished doctoral dissertation, MA: MIT.
- Su, P.H., Wu, C.H., & Lee, L.S. (2014). A recursive dialogue game for personalized computer-aided pronunciation training. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(1), 127-141.
- Surface, E.A., Dierdorff, E.C. & Watson, A.M. (2007). *Special operations language training software measurement of effectiveness study: Tactical Iraqi study final report* (Technical report No. 2007010602). SWA Consulting. Retrieved from [https://rdl.train.army.mil/catalog/view/100.ATSC/56F12B52-936B-4B85-A11D-44D3F590E2DF-1274550929321/BCKStest/ussocom\\_tactical\\_iraqi\\_study\\_final\\_report\\_swa\\_20070501.pdf](https://rdl.train.army.mil/catalog/view/100.ATSC/56F12B52-936B-4B85-A11D-44D3F590E2DF-1274550929321/BCKStest/ussocom_tactical_iraqi_study_final_report_swa_20070501.pdf)
- Sung, M.C., & Yang, H.K. (2016). Effects of construction-centered instruction on Korean students' learning of English transitive resultative constructions. In S.D. Knop & G. Gliquin (Eds.), *Applied construction grammar* (pp. 89–114). Berlin: Walter de Gruyter.
- Sung, M.C. (2012). *Effects of Instruction on the Learning of the English Transitive Resultative Construction by Korean Secondary School Students*. Unpublished master's thesis, Seoul National University, Seoul.
- Sung, M.C. (2018). *Effects of Construction-Grammar-Based Instruction on the Learning of English Verb-Particle Constructions by Korean Middle School Students*. Unpublished doctoral dissertation, Seoul National University, Seoul.
- Sung, M.C. (2020). Pre-Service Primary English Teachers' AI Chatbots. *Language Research*, 56(1), 97-115.

- Suzuki, W. (2018). Negotiation of meaning versus negotiation of form. *The TESOL encyclopedia of English language teaching*, 1-7.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language Typology and Semantic Description: Grammatical Categories and the Lexicon* (vol. 3, pp. 36–149). Cambridge: Cambridge University Press.
- Tarone, E., & Swierzbinska, B. (2009). *Exploring Learner Language*. Oxford: Oxford University Press.
- Thim, S. (2012). *Phrasal verbs: The English verb-particle construction and its history*. Berlin: Walter de Gruyter.
- Thomas, M., Reinders, H., & Warschauer, M. (2013). CALL in language education: Section introduction. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary Computer-Assisted Language Learning* (pp. 245-248). Huntington, UK: Bloomsbury Academic.
- Tenny, C.L. (1994). *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, UK: Cambridge University Press.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- Vendler, Z. (1957). Verbs and times. *Philosophical Review*, 66(2), 143-160.
- Vlugter, P., Knott, A., McDonald, J., & Hall, C. (2009). Dialogue-based CALL: A case study on teaching pronouns. *Computer Assisted Language Learning*, 22(2), 115-131.
- Wagner, L. (2006). Aspectual bootstrapping in language acquisition: telicity and transitivity. *Language Learning and Development*, 2(1), 51-76.
- Walker, N.R., Trofimovich, P., Cedergren, H., & Gatbonton, E. (2011).

- Using ASR technology in language training for specific purposes: A perspective from Quebec, Canada. *Calico Journal*, 28(3), 721-743.
- Wang, N., & Johnson, W.L. (2008). The Politeness Effect in an intelligent foreign language tutoring system. In B.P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the International Conference on Intelligent Tutoring Systems (ITS'08)* (pp.270-280). Berlin: Springer.
- Wang, Y.F., & Petrina, S. (2013). Using learning analytics to understand the design of an intelligent language tutor–Chatbot Lucy. *Editorial Preface*, 4(11), 124-131.
- Wechsler, S. (2001). An analysis of English resultatives under the event-argument homomorphism model of telicity. In *Proceedings of the 3rd Workshop on Text Structure* (pp. 13-15). Austin: Department of Linguistics, University of Austin.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, 9(1), 36-45.
- Whong-Barr, M. (2005). *Morphology, derivational syntax and second language acquisition of resultatives*. Unpublished doctoral dissertation, Durham University, United Kingdom.
- Wilske, S. (2015). *Form and meaning in dialog-based computer-assisted language learning*. Unpublished doctoral dissertation, Saarland University, Saarbrücken.
- Wilske, S., & Wolska, M. (2011). Meaning versus Form in Computer-assisted Task-based Language Learning: A Case Study on the German Dative. *J. Lang. Technol. Comput. Linguistics*, 26(1), 23-37.
- Winkler, S. (1997). *Focus and Secondary Predication*. Berlin: Mouton de

Gruyter.

- Wu, S. (2011). Learning to express motion events in an L2: The case of Chinese directional complements. *Language Learning*, 61(2), 414-454.
- Wyngaerd, G. (2001). Measuring events. *Language* 77(1). 61-90.
- Xu, W. (2016). The influence of language transfer and L2 proficiency on the acquisition of argument structure constructions: An ERP study. In A. Healey, R. Napoleão de Souza, P. Pešková & M. Allen (eds.), *Proceedings of the High Desert Linguistics Society Conference* (vol 11, pp. 393-416). New Mexico: New Mexico University.
- Yang, I.S. (1972). *Korean Syntax: Case Markers, Delimiters, Complementation, and Relativization*. Seoul: Paekhap Publishing Company.
- Yang, I.S. (1974). Two causative forms in Korean. *Language Research*, 10(1), 83-117.
- Yang, I.S. (1976). Semantics of Korean causation. *Foundations of Language*, 14(1), 55-87.
- Yang, H.K., Kim, R., & Sung, M.C. (2014). Basic communicative competence and sentential utterance production. *SNU Journal of Education Research*, 23, 97-116.
- Yang, H., Kim, H., Shin, D., & Lee, J.H. (2019). A Study on Adopting AI-Based Chatbot in Elementary English-Speaking Classes. *Multimedia-Assisted Language Learning*, 22(4), 184-205.
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32(2), 235-263.
- Yeo, S. (2006). Some notes on resultatives in Korean. *Studies in Generative*

*Grammar* 16(4), 687-706.

- Yip, V., & Matthews, S. (2000). Syntactic transfer in a Cantonese–English bilingual child. *Bilingualism: Language and cognition*, 3, 193-208.
- Yotsuya, A., Asano, M., Koyama, S., Suzuki, K., Shibuya, M., Iwagami, E., Endo, K., Ono, M., Takeda, K., & Hirakawa, M. (2014). Crosslinguistic effects in L2 acquisition: Strong/weak resultatives and the directional/locational interpretation of PPs in L2 English by Japanese speakers. In Ryan T. Miller et al. (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum* (pp. 89-100). Somerville, MA: Cascadilla Press.
- Yu, D., & Deng, L. (2016). *Automatic Speech Recognition: A Deep Learning Approach*. London: Springer.
- Zhao, Y., & Ellis, R. (2020). The relative effects of implicit and explicit corrective feedback on the acquisition of 3rd person -s by Chinese university students: A classroom-based study. *Language Teaching Research*. <https://doi.org/10.1177/1362168820903343>
- Zubizarreta, M., & Oh, E. (2007). *On the Syntactic Composition of Manner and Motion*. Cambridge, MA: MIT Press.

## APPENDICES

### APPENDIX 1. Testing Materials (IRB No. 2104/004-031)

1.1. Acceptability Judgment Task (AJT) .....	247
1.1.1. AJT in Pretest .....	247
1.1.2. AJT in Immediate Posttest.....	250
1.1.3. AJT in Delayed Posttest .....	253
1.2. Elicited Writing Task (EWT) .....	256
1.2.1. EWT in Pretest .....	256
1.2.2. EWT in Immediate Posttest.....	267
1.2.3. EWT in Delayed Posttest.....	279

### APPENDIX 2. Chatbot Interaction Data Sample

2.1. A Conversation sample with AI chatbot for intransitive motion construction, <i>Changsoo biked to XX mountain to volunteer</i> .....	290
2.2. A Conversation sample with AI chatbot for adjectival transitive resultative construction with a particle complement, <i>Changsoo stomped the fire out</i> .....	294

## Appendix 1. Testing Materials (IRB No. 2104/004-031)

### 1.1. Acceptability Judgment Task (AJT)

#### 1.1.1. AJT in Pretest

IRB No. 2104/004-031      유효기간: 2022년 04월 25일

\*\*문법성 판단 평가(Acceptability Judgment Test)

CODE NUMBER: \_\_\_\_\_

<이름/학번을 절대로 기입하지 말고, 부여된 CODE NUMBER만을 기입하세요>

굵은 글씨로 된 문장만의 문법적인 정확성을  
아래와 같이 6점 만점으로 판단하여  
해당 점수의 칸에 **동그라미** 하세요.

1점	2점	3점	4점	5점	6점
완전히 <u>어색함</u>	<u>어색함</u>	약간 <u>어색함</u>	약간 <u>자연스러움</u>	<u>자연스러움</u>	완전히 <u>자연스러움</u>

만약 판단하기 어렵다면, 맨 끝의 "모르겠음"에 표시하세요.

\* 본 시험은 총 **40문항**으로 구성됩니다. 풀고 싶은 시간 만큼 문제를 풀기 바랍니다.

\*\*\* 모르는 단어는 같이 나누어 준 단어 리스트를 확인하시기 바랍니다. 이에 관한 질문은 받지 않습니다.

\*\*\*\* 시험 종료 후 본 시험지를 제출하기 바랍니다.

번호	문 장	점수						모르 겠음
		1	2	3	4	5	6	
P1	Look at that! <b>His teeth the man brush.</b>	(1)						
P2	The school opened. <b>The girl went to the school.</b>							
P3	I have a lovely cat. <b>The cat drank the milk.</b>							
1	John wanted to show his new apartment to his mother, who was visiting is city, so <b>John walked his mother to his new apartment for 20 minutes.</b>							
2	Aliza's dad promised to give her 100 dollars when she lost her first tooth, so <b>She wiggled loose.</b>							
3	Jim was golfing. When it was his turn to hit the ball, <b>he hit into the hole.</b>							
4	It is weird that my dog, Dogee, loves to eat flowers. One day, <b>the dog ate the flowerbed bare in ten minutes.</b>							
5	A fire broke out in a building, so <b>the shopkeeper hollered the teenagers out of the building.</b>							
6	Christina had a hamster that she loved. <b>Christina cried when her hamster died.</b>							

Version: 1.1(2021.04.26.)

번호	문 장	점수						모로 점수
		1	2	3	4	5	6	
7	Mary was playing baseball. When she got the ball, <b>she threw the ball to the other side of the playground, and it got there just fine.</b>	Ⓜ					Ⓜ	
8	Jim's jacket was dirty, but he was supposed to give it to Mary. Mary wanted a clean jacket, so <b>Jim washed clean.</b>							
9	I heard that he made the same mistakes again and again but <b>I don't think him stupidly.</b>							
10	The suitcase was really important for the business meeting between Jack and IBM. <b>Jack dragged the suitcase to business meeting, but stopped before he got there.</b>							
11	Dana found that the rope was too tight, so <b>she tugged the rope loose for twenty seconds.</b>							
12	After a long walk, Jack felt tired. To get some rest, <b>the chair unfolded he.</b>							
13	My dogs got up earlier than my chicken. In the morning, <b>the dogs barked the chicken awake, and it became awake.</b>							
14	It was a busy day, and John had a lot of hay to deliver to Boston. <b>He loaded the hay onto the truck in ten minutes.</b>							
15	My grandfather broke his leg in the car accident, so <b>I helped him the hospital yesterday.</b>							
16	My daughter hates to tie her hair up during hot summer days, so <b>I cut her hair short.</b>							
17	I don't like sunbathing on the beach because <b>it makes me to get a sunburn.</b>							
18	Terry tried to tame a wild horse, but <b>the horse hopped her straight out of the saddle.</b>							
19	The baseball fans were mad at the player, so <b>the fans beat him unconscious.</b>							
20	Jack got a letter from the company saying that he was fired, so <b>he ripped the paper.</b>							
21	Jane is a worker at a cushion factory. Her boss told her to pile up the cushions. She piled them up and up and up, and finally, <b>she piled the cushions highly.</b>							
22	Grandmother asked Kim to get the cart to the house. So, <b>Kim pulled the cart to the door, but stopped before he got there.</b>							
23	Kim and Mary's dog got too fat, so <b>they ran the dog over the bridge in 30 minutes.</b>							

번호	문 장	점수						모로 점수
		1	2	3	4	5	6	
24	The pot is dirty. <b>Let's scrub the pot shiny.</b>							
25	She is a nagging person who always complains about everything. <b>She will drive you mad for a few minutes.</b>							
26	Sam wanted to look fancy. <b>He dyed his hair red, but it did not become red.</b>							
27	Mary's dancing teacher believes that Mary needs more exercise. <b>He waltzed Mary around and around the room for hours.</b>							
28	I didn't have time to go to a library and borrow the book. <b>So, Tom brought me the book.</b>							
29	John was half-dozing in English class. <b>Mary shook him awake.</b>							
30	Lee was a volleyball player. She scored a very important point when <b>She tossed the ball to the other side of the court, and it arrived there.</b>							
31	My teacher is a warm and friendly person. <b>She always speaks kindly to everyone.</b>							
32	The poor animal was tied to the tree, so <b>the rescuers tore the animal free, so it became free.</b>							
33	Bill hates seeing his son, Harry, lying on the sofa, idling his precious time away, so <b>Bill pushed Harry off the sofa for hours.</b>							
34	Sam did not want Bob to stay in his house. Sam knew that Bob hated spooky things. <b>Finally, Sam frightened out of the house!</b>							
35	The children were floundering in the sea. <b>The lifeguard backstroked the children to the shore.</b>							
36	Marinate the meat and pound it thoroughly. <b>You can make the meat tenderly.</b>							
37	It was time to start the class, but Josh was out of the classroom. <b>I urged Josh the room.</b>							
38	Mom made a pizza for lunch. <b>She rolled the dough thin, but it did not become thin.</b>							
39	I had been waiting for the elevator for ten minutes but it was still at the second basement floor. <b>Finally, the elevator wheezed upward.</b>							
40	John is a waiter in that pizza house. <b>He wiped the table clean in three minutes.</b>							

\*\*\*\* 문법성 판단 평가가 모두 끝났습니다. 수고하셨습니다. \*\*\*\*

### 1.1.2. AJT in Immediate Posttest

IRB No. 2104/004-031      유효기간: 2022년 04월 25일

\*\*문법성 판단 평가(Acceptability Judgment Test)

CODE NUMBER: \_\_\_\_\_

<이름/학번을 절대로 기입하지 말고, 부여된 CODE NUMBER만을 기입하세요>

**굵은 글씨로 된 문장만의 문법적인 정확성을  
아래와 같이 6점 만점으로 판단하여  
해당 점수의 칸에 **동그라미** 하세요.**

1점                  2점                  3점                  4점                  5점                  6점  
 완전히 어색함    어색함            약간 어색함    약간 자연스러움    자연스러움    완전히 자연스러움

단약 판단하기 어렵다면, 맨 끝의 "모르겠음"에 표시하세요.

\* 본 시험은 총 **40문항**으로 구성됩니다. 풀고 싶은 시간 만큼 문제를 풀기 바랍니다.

\*\*\* 모르는 단어는 같이 나누어 준 단어 리스트를 확인하시기 바랍니다. 이에 관한 질문은 받지 않습니다.

\*\*\*\* 시험 종료 후 본 시험지를 제출하기 바랍니다.

번호	문 장	점수						모르 겠음
		1	2	3	4	5	6	
		1	2	3	4	5	6	
P1	Look at that! <b>His teeth the man brush.</b>	(2)						(6)
P2	I have a lovely cat. <b>The cat drank the milk.</b>							
P3	The school opened. <b>The girl went to the school.</b>							
1	Jenny wanted to show her new apartment to her mother, who was visiting her city, so <b>Jenny walked her mother to her new apartment for 20 minutes.</b>							
2	Gary's mom promised to give him 100 dollars when he lost his final tooth, so <b>He wiggled loose.</b>							
3	Seri was golfing. When it was her turn to hit the ball, <b>she hit into the hole.</b>							
4	It is weird that my cat, Kitty, loves to eat flowers. One day, <b>the cat ate the flowerbed bare in five minutes.</b>							
5	A fire broke out in a tall building, so <b>the shopkeeper hollered the tourists out of the building.</b>							
6	Jake had a dog that he loved. <b>Jake cried when his dog died.</b>							

Version 1.1(2021.04.26.)

번호	문 장	점수						모름 응답
		1	2	3	4	5	6	
7	Tom was playing baseball. When he got the ball, <b>he threw the ball to the other side of the playground, and it got there just fine.</b>							
8	Lim's jacket was dirty, but she was supposed to give it to Ross. Lim wanted a clean jacket, so <b>Lim washed clean.</b>							
9	She knows that I have solved all the difficult math questions. <b>She always calls me a genius.</b>							
10	The suitcase was really important for the business. <b>Chen dragged the suitcase to the building, but stopped before he got there.</b>							
11	Tanaka found that the rope was too tight, so <b>he tugged the rope loose for ten seconds.</b>							
12	Jack wanted an ice flake with cookies. So, <b>he crumbled the cookies.</b>							
13	The dogs got up earlier than people. In the morning, <b>the dogs barked them awake, and they became awake.</b>							
14	It was a busy day, and Lee had a lot of hay to deliver to L.A. <b>She loaded the hay onto the truck in one minute.</b>							
15	My grandmother broke her leg in the car accident, so <b>I helped her the hospital yesterday.</b>							
16	My sister hates to tie her hair up during hot summer days, so <b>My mom cut her hair short.</b>							
17	She doesn't like sunbathing on the beach because, <b>it makes her to get a sunburn.</b>							
18	Chuck tried to tame a wild horse, but <b>the horse hopped him straight out of the saddle.</b>							
19	The soccer fans were mad at the player, so <b>they beat him unconscious.</b>							
20	Edward Kim, a chef, found that the eggs went bad. So, <b>he the eggs cracked.</b>							
21	Frank is a worker at a cushion factory. His boss told him to pile up the cushions. He piled them up and up and up, and finally, <b>he piled the cushions highly.</b>							
22	Grandfather asked Mary to get the cart to the room. So, <b>Mary pulled the cart to the door, but stopped before she got there.</b>							
23	Jennifer's dog got too fat, so <b>she ran the dog over the bridge in ten minutes.</b>							
번	문 장	점수						모름

호		1	2	3	4	5	6	점수
24	The pot is dirty. <b>Let's scrub the pot shiny.</b>	Ⓜ					Ⓜ	
25	Kwak is a nagging person who always complains about everything. <b>He will drive you mad for a few minutes.</b>							
26	Lisa wanted to look fancy. <b>She dyed her hair red, but it did not become red.</b>							
27	Matilda's dancing teacher believes that Matilda needs more exercise. <b>The dancing teacher waltzed Matilda around and around the room for hours.</b>							
28	It was difficult for me to build a sand castle on the beach <b>So, my father built a sand castle me.</b>							
29	Mary was half-dozing in English class. <b>The teacher shook her awake.</b>							
30	Grandmother asked Tom to get the cart to her house. So, <b>he pulled the cart to the door, but stopped before he got there.</b>							
31	My grandfather is a warm and friendly person. <b>He always speaks kindly to everyone.</b>							
32	The poor monkey was tied to the tree, so <b>the policeman tore the monkey free, so it became free.</b>							
33	Jake hates seeing his daughter, Yujin, lying on the sofa, idling her precious time away, so <b>Jake pushed Yujin off the sofa for fifteen minutes.</b>							
34	Dorothy did not want Sam to stay in her house. Dorothy knew that Sam hated spooky things. <b>Finally, Dorothy frightened out of the house!</b>							
35	Noah was floundering in the sea. <b>The lifeguard backstroked him to the shore.</b>							
36	The chef marinated the meat and pounded it thoroughly. <b>He made the meat tenderly.</b>							
37	It was time to start the class, but Bill was out of the classroom. <b>I urged him the room.</b>							
38	Piccolo made a pizza for lunch. <b>He rolled the dough thin, but it did not become thin.</b>							
39	The soldiers arrived at the stream, but the enemies were following them. <b>They went through the stream by splashing.</b>							
40	Irene is a waitress in that pizza house. <b>She wiped the table clean in two minutes.</b>							

\*\*\*\* 문법성 판단 평가가 모두 끝났습니다. 수고하셨습니다. \*\*\*\*



### 1.1.3. AJT in Delayed Posttest

IRB No. 2104/004-031      유효기간: 2022년 04월 23일

**\*\*문법성 판단 평가(Acceptability Judgment Test)**

CODE NUMBER: \_\_\_\_\_

<이름/학번을 절대로 기입하지 말고, 부여된 CODE NUMBER만을 기입하세요>

**굵은 글씨로 된 문장만의 문법적인 정확성을  
아래와 같이 6점 만점으로 판단하여  
해당 점수의 칸에 (동그라미) 하세요.**

1점                  2점                  3점                  4점                  5점                  6점  
 완전히 어색함    어색함            약간 어색함    약간 자연스러움    자연스러움    완전히 자연스러움

만약 판단하기 어렵다면, 맨 끝의 "모르겠음"에 표시하세요.

\* 본 시험은 총 **40문항**으로 구성됩니다. 풀고 싶은 시간 만큼 문제를 풀기 바랍니다.

\*\*\* 모르는 단어는 같이 나누어 준 단어 리스트를 확인하시기 바랍니다. 이에 관한 질문은 받지 않습니다.

\*\*\*\* 시험 종료 후 본 시험지를 제출하기 바랍니다.

번호	문 장	점수						모르 겠음
		1	2	3	4	5	6	
		← (1)						(6) →
P1	I have a lovely cat. <b>The cat drank the milk.</b>							
P2	Look at that! <b>His teeth the man brush.</b>							
P3	The school opened. <b>The girl went to the school.</b>							
1	Nakata wanted to show his new apartment to Mary who was visiting is city, so <b>Nakata walked Mary to his new apartment for 20 minutes.</b>							
2	David's mom promised to give him 100 dollars when he lost his first tooth, so <b>David wiggled loose.</b>							
3	Seri was golfing. When it was her turn to hit the ball, <b>she hit into the hole.</b>							
4	It is weird that the cow loves to eat flowers. One day, <b>the cow ate the flowerbed bare in seven minutes.</b>							
5	A fire broke out in a nursing home, so <b>John hollered the old people out of the building.</b>							
6	Zhang Shan had a cat that he loved. <b>Zhang shan cried when his cat died.</b>							

Version: 1.1(2021.04.26.)

번호	문 장	점수						모의 점수
		1	2	3	4	5	6	
7	Park was playing baseball. When he got the ball, <b>he threw the ball to the other side of the playground, and it got there just fine.</b>							
8	Jisoo's jacket was dirty, but he was supposed to give it to Minjeong. Minjeong wanted a clean jacket, so <b>Minjeong washed clean.</b>							
9	He could not sleep last night. <b>We found him tired.</b>							
10	The case was really important for the machine. <b>Tanaka dragged the box to the factory, but stopped before he got there.</b>							
11	We found that the rope was too tight, so <b>we tugged the rope loose for thirty seconds.</b>							
12	Jack's hobby is to play with the paperclip. <b>He the paperclip bent.</b>							
13	The dogs got up earlier than Jack. In the morning, <b>the dogs barked Jack awake, and he became awake.</b>							
14	It was a busy day, and Kimura had a lot of hay to deliver to Boston. <b>Kimura loaded the hay onto the truck in one minutes.</b>							
15	The beggar broke his leg in the car accident, so <b>People helped him the hospital yesterday.</b>							
16	My sisger hates to tie her hair up during hot summer days, so <b>My dad cut her hair short.</b>							
17	John does not like sunbathing on the beach because <b>it makes him to get a sunburn.</b>							
18	Laura tried to tame a wild horse, but <b>the Laura's horse hopped her straight out of the saddle.</b>							
19	The angry people were mad at the male criminal, so <b>they beat him unconscious.</b>							
20	People held balloons. To celebrate the anniversary, <b>They released the balloons.</b>							
21	Fu and Liao are workers at a cushion factory. Their boss told them to pile up the cushions. Fu and Liao piled the them up and up and up, and finally, <b>they piled the cushions highly.</b>							
22	Grandmother asked Sue to get the cart to the house. So, <b>Sue pulled the cart to the door, but stopped before he got there.</b>							
23	Shery's dog got too fat, so <b>she ran the dog over the bridge in forty minutes.</b>							

번 호	문 장	점수						모 든 것 을
		1	2	3	4	5	6	
24	Robert's pot looks dirty, so <b>let's scrub his pot shiny.</b>							
25	Jane is a nagging person who always complains about everything. <b>Jane will drive you mad for a few minutes.</b>							
26	I wanted to look fancy. So, <b>I dyed my hair red, but it did not become red.</b>							
27	Theresa's dancing teacher believes that she needs more practice. So, <b>Theresa's dancing teacher waltzed her around and around the room..</b>							
28	I didn't have time to go to a cake house to bring the cake for her. <b>So, my father brought me a cake.</b>							
29	I was half-dozing in English class. <b>John shook me awake.</b>							
30	The patient asked his daughter to get the cart to his room. So, <b>She pulled the cart to the door, but stopped before she got there.</b>							
31	My cousin, Ronaldo, is a warm and friendly person. <b>He always speaks kindly to everyone.</b>							
32	The poor slave was tied to the tree. Then, suddenly, <b>many soldiers tore the slave free, so the slave became free.</b>							
33	Jane hates seeing her son, Larry, lying on the sofa, idling his precious time away, so <b>Jane pushed her son off the sofa for hours.</b>							
34	Hyojin did not want Kate to stay in her house. Hyojin knew that Kate hated spooky things. <b>Finally, Hyojin frightened out of the house!</b>							
35	The old people were floundering in the sea. <b>The lifeguard backstroked the old people to the shore.</b>							
36	Marinate the pork and pound it thoroughly. The, <b>you can make the pork tenderly.</b>							
37	It was time to start the class, but Anne was out of the classroom. <b>My homeroom teacher urged Anne the room</b>							
38	Luigi made a pizza for lunch. <b>Luigi rolled the dough thin, but it did not become thin.</b>							
39	The towners are desperately waiting for a food truck. <b>Finally, a food truck rumbled into the town.</b>							
40	Betty is a waitress in that Spaghetti house. <b>She wiped the table clean in nine minutes</b>							

## 1.2. Elicited Writing Task (EWT)

### 1.2.1. EWT in Pretest

IRB No. 2104/004-031

유효기간: 2022년 04월 25일

#### Elicited Writing Task for Korean Learner Participants (40min)

코드번호: \_\_\_\_\_

\* 예시와 같이, 주어진 상황을 읽고 **굵은 글씨로 된 밑줄 친 문장만을** 괄호 안 단어를 활용하여 영어로 완성하세요.

예) 요즘은 Mary가 잘 웃지 않아 걱정이야. 그런데 오늘만은,

나는 Mary가 웃는 것을 보았어.

I **SAW MARY LAUGH** (laugh, see)

1. 영수가 나에게 책 한권을 더 줬어. 그래서  
나에게는 책이 두 권 있어.



I \_\_\_\_\_ (two books, have)

2. 한 소년이 물에 빠져있었어.  
마침 지나가던 여자가 물에 들어가  
그 소년을 물 밖으로 당겨 끌어냈어.



She \_\_\_\_\_ (pull, out)

Version 1.1(2021.04.26.)



3. 금속을 빨갱게 되도록 달구어서,

대장장이는 금속(metal)을 평평하게 두드려 썼어.



He \_\_\_\_\_ (flat, hammer)

4. 그는 손님들에게 갓 구운 따뜻한 빵을

대접하고 싶었어. 원래는 빵이 식은 후에

자르는게 좋은데, 이날만은

그는 빵이 따뜻할 때 잘랐지.



He \_\_\_\_\_ (cut, hot)

5. Tom 은 재채기가

심한 알러지가 있어. 그래서

그는 재채기 해서 티슈를 테이블에서 떨어뜨렸어.



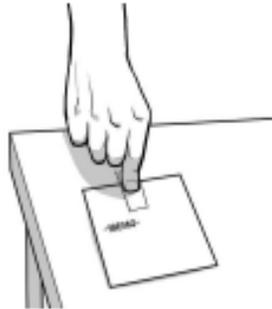
He \_\_\_\_\_ (off, sneeze)

6. Brandon 은 나무에서 노는 것을 좋아해.  
 그런데 어느날, 평소처럼 나무에서 놀다가,  
Brandon 은 나무 밖으로 떨어져 버렸어.



Brandon \_\_\_\_\_ (fall, out)

7. 김선생님을 만나러  
 갔는데 자리에 계시지 않더라고.  
 그래서 나는 메모를 테이블에 두었어.



I \_\_\_\_\_ (on, put)

8. 지난 금요일에 댄스 파티가 열렸는데,  
 그 파티에 가서는  
그녀는 어지럽도록 춤췄대.



She \_\_\_\_\_ (dance, dizzy)



9. 빵을 자르려고 빵칼을 찾았는데  
 빵칼이 고장이 났는지 전혀 잘리지 않더라고.  
 집에 칼은 날이 네모났고 커다란  
 중국식 주방 칼 밖에 없었는데, 다행히  
그 칼로 빵을 잘랐어.



The knife (the bread, cut)

10. Carey 는 길에 큰 돌이 길을  
 막고있는 것을 발견했어. 그래서  
그녀는 돌을 굴러 길 밖으로 치웠어.



She (roll, out)

11. 그는 이번 마라톤 대회가 마지막이었다.  
 그래서 각오를 단단히 하고  
그는 신발이 닳도록 달렸어.



He (threadbare, run)

12. 시험공부를 하면서 라디오를 듣는데,  
음악소리가 참 좋아.



The music \_\_\_\_\_ (sound, great)

13. 형이 아침식사로 오믈렛을 해 준다고 했어.  
먼저,  
그는 계란을 깨서 그릇(bowl)안에 넣었어.



He \_\_\_\_\_ (into, break)

14. 포테이토칩을 만들 때,  
감자조각이 얇지 않으면, 바삭바삭해지지 않아.  
그래서 어제  
나는 감자를 얇게 잘랐어.



I \_\_\_\_\_ (slice, thin)

15. 요즘 봄만 되면 미세먼지가 자꾸 눈에  
들어와. 맨손으로 눈을 비비면 눈병에 걸릴 것  
같아서,

나는 눈을 깜빡거리어서 먼지(dust)를 눈 밖으로 빼내.



I (out, blink)

16. 너 어제 밤 총소리 들었어?  
뒷산에 호랑이가 나타났는데,

경찰관 한 사람이 호랑이를 쏘아 죽였대.



A police officer (shoot, dead)

17. 수업시간에 못했던 질문이 생각나서,  
그는 선생님께 이메일을 보냈어.

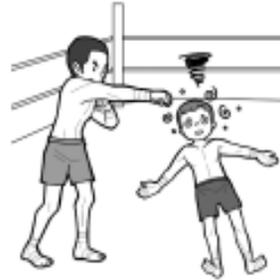


He (email, send)

18. Lee는 최고의 격투기 선수야.

지난 경기에서 펀치 한방으로,

Lee는 그를 의식을 잃도록 때려눕혔어.



Lee

(knock, unconscious)

19. John은 강남역에 갔어.

그런데 우연히

John은 Sophie를 보았어.



John

(Sophie, see)

20. 불의를 보면 참지 못하는 나는 갱패들에게

노인을 괴롭히지 말라고 말했다. 그러자

갱패들은 으르렁거리며 나를 집안으로 넣었어.



The gansters

(snarl, into)

21. 화장실에서 여자사이가 갖혀있었는데,  
열쇠를 잃어버려서  
아이 엄마가 문을 차서 열었다.



Her mom

(open, kick)

22. 고속도로를 달리고 있었는데,  
옆에서 차 한대가 엄청 빨리 달리더라고.  
그런데 갑자기 그 차 앞에 고라니가  
튀어나오는 거야. 고라니를 피하느라고,  
그 차는 끼익 소리를 내며 도로밖으로 나갔어.



The car

(screached, out)

23. 그녀는 남편이 골프공을 못보게 치워버리고  
싶었어. 그래서  
그녀는 공을 던져 지붕위에 두었어.



She

(on, throw)

24. 그 부부는 다음달에 아들을 낳을 예정이야.

그래서,

그들은 방을 파랗게 칠했어.



They

(paint, blue)

25. James가 요즘 무언가 숨기고 있는 것 같아.

넌 그를 정직하다고 생각해?



Do you

(honest, consider)

26. 축구 선수인 Jack 은 중요한 경기에서

자살공을 세 골이나 넣어버렸어. 락커룸에서

주장은 꺾려보아서 Jack 을 방 밖으로 내쫓아 버렸어.



The captain

(stare, out)

27. 난 밖으로 나갈 수가 없었어. 문도 잠겨 있고,

그가 창문을 못박아 달아 버렸어.



He

(shut, nail)

28. 지난 한 달간 너무 가물었어. 게다가 나는

지난 두 주 간 집을 비워 꽃에 물을

주지도 못했어. 다행히

비가 내려 꽃에 물을 주었어.



The rain

(the flower, water)

29. Millie는 모래놀이를 싫어했어.

왜냐하면 모래놀이를 할 때 마다

그녀는 손이 더러워졌거든.



She

(get, dirty)

30. 이사하는 날인데 엘리베이터가  
고장나 버렸지 뭐야? 그래서  
크리스는 계단(stairs) 위로 피아노를 밀어 올렸어.



Chris

(push, up)

- 쓰기 시험이 끝났습니다. 수고하였습니다. -

1.2.2. EWT in Immediate Posttest

IRB No. 2104/004-031

유효기간: 2022년 04월 25일

**Elicited Writing Task for Korean Learner Participants (40min)**

코드번호: \_\_\_\_\_

\* 예시와 같이, 주어진 상황을 읽고 굵은 글씨로 된 밑줄 친 문장 안 괄호 안 단어를 활용하여 영어로 완성하세요.

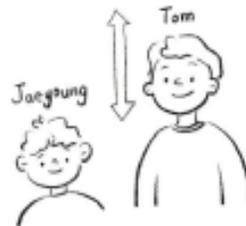
예) 요즘은 Mary가 잘 웃지 않아 걱정이야. 그런데 오늘만은,

나는 Mary가 웃는 것을 보았어.

I **SAW MARY LAUGH** (laugh, see)

1. 재영이는 키가 작아. 하지만,

툼은 키가 커.



I \_\_\_\_\_ (tall, Tom)

2. 한 소녀가 물에 빠져있었어.

마침 지나가던

남자아이가 그 소녀를 물 밖으로 당겨끌어냈어.



He \_\_\_\_\_ (out, pull)

Version 1.1(2021.04.26.)

SNURB

3. 금속을 빨갱게 되도록 달구어서,

그녀는 금속(metal)을 평평하게 두드려 뒀어.

She \_\_\_\_\_

(hammer, flat)



4. Tanaka 는 낚시를 좋아해. 특히 그가 좋아  
하는 것은, 배 위에서 물고기를 날로 잡아 먹는  
거야. 마침 큰 물고기가 잡혔어.

그래서 그는 물고기를 날로 먹었어.

He \_\_\_\_\_

(raw, eat)



5. Jane 은 재채기가

심한 감기에 걸렸어. 그래서

그녀는 재채기 해서 티슈를

테이블에서 떨어뜨렸어.

She \_\_\_\_\_

(sneeze, off)

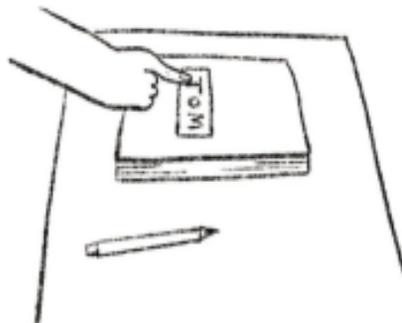


6. 과연 어떤 자동차가 이 험한 산 정상까지  
올라올 수 있을까 궁금했어. 그런데 놀랍게도,  
Jack의 차가 산의 정상으로 올라오고 있었어.



Jack's car \_\_\_\_\_ (top of the mountain, to)

7. Tom이 자꾸 책을 잃어버려서  
걱정이 많았어. 그래서  
나는 그의 책에 이름표를 붙여주었어



I \_\_\_\_\_ (on, put)

8. 지난 일요일에 댄스 파티가 열렸는데,  
그 파티에 가서는  
그는 어지럽도록 춤췄대.



He \_\_\_\_\_ (dance, dizzy)

9. 하도 칫솔이 오래되어서 새 칫솔을 샀어.  
그 칫솔로 이를 닦았어.



The toothbrush \_\_\_\_\_ (the teeth, clean)

10. Robert 는 큰 돌이 동굴의 입구를  
막고있는 것을 발견했어. 그래서  
그는 돌을 굴려 입구 밖으로 치웠어.



He \_\_\_\_\_ (roll, out)

11. 그녀는 이번 육상 대회가 마지막이었다.  
그래서 각오를 단단히 하고  
그녀는 신발이 닳도록 달렸어.

\* threadbare: 닳은



She \_\_\_\_\_ (run, threadbare)

12. 쉬는 시간에 라디오를 틀었는데  
그 남자 가수의 음악소리는 참 좋아.



His music \_\_\_\_\_ (great, sound)

13. 뷔페식당의 여자 요리사에게 계란 후라이를  
부탁했어. 먼저,

그녀는 계란을 깨서 그릇(bowl)안에 넣었어.



She \_\_\_\_\_ (into, break)

14. 엄마에게 감자전을 해 달라고 부탁드렸어.

엄마는 바삭한 감자전을  
만들어 주신다고 하셨어. 그리고

엄마는 감자를 얇게 잘랐어.



My mom \_\_\_\_\_ (thin, slice)

15. Judy는 눈이 엄청나게 많이 내리는  
날에 산을 걷고 있었어.

눈이 얼마나 내리던지 눈꺼풀  
안으로까지 눈이 들어갔더래. 그래서

그녀는 깜빡거리서 들어온 눈을 눈꺼풀 밖으로 빼내었어.



\*눈꺼풀: eyelashes

She \_\_\_\_\_ (out, blink)

16. 너 어제 밤 총소리 들었어?

뒷산에 호랑이가 나타났는데,

사냥꾼 한 사람이 호랑이를 쏘아 죽였대.



A hunter

(dead, shoot)

17. 내일은 내 생일이야. 그래서

아버지는 나에게 피아노를 사 줄거야.



My father

(piano, buy)

18. 그녀는 특전사 출신이야.

남자 도둑이 들었는데,

그녀는 그 남자를

의식을 잃도록 때려눕혔어.



She

(unconscious, knock)

19. Sophie는 여러 음식을 좋아해. 특히

그녀는 케익을 좋아해.



She

(cake, like)

20. 테러리스트들은 군인들에게

완전히 제압당했지만,

그래도 반항하려 했어.

그러자

군인들은 으르렁거리며

테러리스트들을 건물안으로 밀어넣었어.



The soldiers

(into, snarl)

21. 화장실에서 남자아이가 갇혀있었는데,

열쇠를 잃어버려서

경찰관이 문을 차서 열었다.



The policeman

(kick, open)

22. 자려고 하는데 어딘가에서 쿵쿵 소리가  
 들렸어. 밖을 내다 보니,  
다리에 나무 의족을 단 사람이  
쿵쿵거리며 방 안으로 들어왔어.



A wooden-legged man

(into, clump)

23. 그는 도둑들로부터 돈을 숨기려고 했어. 그래서  
그는 돈을 던져 지붕위에 두었어.



He

(throw, on)

24. Tanaka 는 방 색깔이 너무 칙칙하다고 생각했어.

그래서,

그는 방을 파랗게 칠했어.



He \_\_\_\_\_ (paint, blue)

25. Jenny가 요즘 무언가 숨기고 있는 것 같아.

넌 그녀를 정직하다고 생각해?



Do you \_\_\_\_\_ (honest, consider)

26. 내 동생은 이번시험에서도

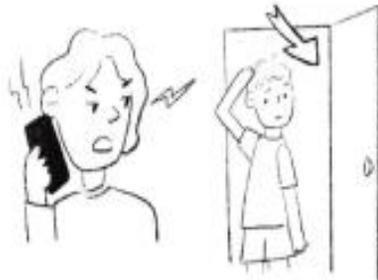
또 병점을 맞았어.

엄마는 마침 중요한 전화 통화를

하고 계시는 중이셔서,

엄마는 패러보아서

동생을 방 밖으로 내쫓아 버렸어.



My mom \_\_\_\_\_ (out, stare)

27. 무서운 녹대가 방에 들어가 있었어.  
 절대 녹대가 나가지 못하게 하려고  
그녀는 창문을 못박아 닫아 버렸어.



She \_\_\_\_\_ (nail, shut)

28. 겨울이 가고 봄이 왔어.  
햇빛덕분에 꽃이 따뜻하게 되었어.



The Sun \_\_\_\_\_ (the flower, warm)

29. Kihoon이는 모래놀이를 싫어했어.  
 왜냐하면 모래놀이를 할 때 마다  
그의 손이 더러워졌거든.



He \_\_\_\_\_ (dirty, get)

30. Jessica 사무실에  
새로운 가구가 들어왔어.  
그런데 이 건물에는  
엘리베이터가 없지 뭐야? 그래서  
그녀는 계단(stairs) 위로 가구를 밀어 올렸어.



She \_\_\_\_\_ (push, up)

- 쓰기 시험이 끝났습니다. 수고하셨습니다. -

1.2.3. EWT in Delayed Posttest

IRB No. 2104/004-031

유효기간: 2022년 04월 25일

**Elicited Writing Task for Korean Learner Participants (40min)**

코드번호: \_\_\_\_\_

\* 예시와 같이, 주어진 상황을 읽고 굵은 글씨로 된 밑줄 친 문장간을 필요한 단어를 활용하여 영어로 완성하세요.

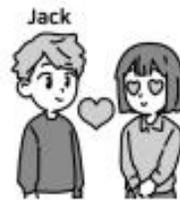
예) 요즘은 Mary가 잘 웃지 않아 걱정이야. 그런데 오늘만은,

나는 Mary가 웃는 것을 보았어.

I **SAW MARY LAUGH** (laugh, see)

1. 수영이는 Jack에게 호감을 느껴. 사실,

그녀는 Jack을 좋아해.



She

(Jack, love)

2. 토끼 한 마리(a rabbit)가 물에 빠져있었어.

마침 지나가던

할머니가 그 토끼를 물 밖으로 당겨끌어냈어.



An old lady

(out, pull)

Version 1.1(2021.04.26.)



3. 금속을 빨갱게 되도록 달구어서,

나는 금속(metal)을 평평하게 두드려 봤어.

I \_\_\_\_\_ (hammer, flat)



4. Jason 이 학교에 물병을 가지고 오지 않은 날이었어. Tom 은 Jason 에게 물을 엄청나게 연어먹었던 기억이 났어. 그래서

Tom 은 Jason 에게 물병이 가득차게 물을 주었어.

Tom \_\_\_\_\_ (full, give)



5. 김교수님은 재채기가

심한 감기에 걸렸어. 그래서

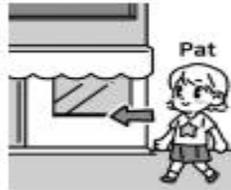
김교수님은 재채기 해서 티슈를

테이블에서 떨어뜨렸어.



Professor Kim \_\_\_\_\_ (sneeze, off)

6. Pat 은 코로나 시국에 운동이 부족해서  
걱정이 많았어. 그래서,  
Pat 은 그 상점까지 걸어서 갔어.



Pat (the shop, walk)

7. 귀여운 강아지 스티커를 나는  
참 좋아해. 그래서,  
나는 문에 스티커를 붙여두었어.



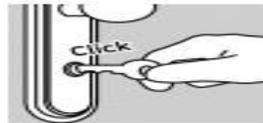
I (on, put)

8. 지난 토요일에 반 대향 장기자랑이 열렸는데,  
장기자랑에서  
영희는 어지럽도록 춤췄대.



Younghee (dance, dizzy)

9. 열쇠를 잃어버렸다가 새로 찾았지. 결국,  
열쇠로 문을 열었어.



The key (the door, open)

10. 산사태로 돌이 굴러내려와 길을  
막고 있더라고. 그래서  
우리는 돌을 굴러 길 밖으로 치웠어.



We \_\_\_\_\_ (roll, out)

11. 남아프리카에서 온 마라톤 선수는  
이번 마라톤 대회가 마지막이었다.  
그래서 각오를 단단히 하고  
그는 신발이 닳도록 달렸어.



\* threadbare: 닳은

He \_\_\_\_\_ (run, threadbare)

12. 아이유 콘서트에 갔는데,  
그녀의 음악소리는 참 좋아.



Her music \_\_\_\_\_ (great, sound)

13. 민영이는 아침식사로 계란요리를 참 좋아해.

그녀는 계란을 깨서 그릇(bowl)안에 넣었어.



She

(into, break)

14. 상명이는 요리를 배우는 학생이야.

프랑스식 감자요리를 만드는데,

요리실력이 워낙 좋아서,

상명이는 감자를 얇게 잘랐어.



Sangmyeong

(thin, slice)

15. Mrs. Sheen 은 봄, 가을만 되면

황사가 고생을 많이 해. 지난번에는

황사가 너무 심해서, 먼지가 눈 안으로

들어가더래. 그래서,

그녀는 깜빡거리서 먼지(dust)를 눈 밖으로 빼내었어.



She

(out, blink)

16. 너 어제 밤 총소리 들었어?  
 뒷산에 호랑이가 나타났는데,  
경찰들이 호랑이를 쏘아 죽였대.



Policemen (dead, shoot)

17. 어제는 Bill의 여자친구의 생일이었어.  
 그런데, 그녀는 저 멀리 아르헨티나를  
 여행중이야. 그래서  
Bill은 그녀에게 이메일을 보냈어.

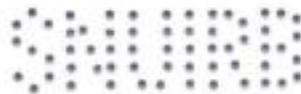


Bill (e-mail, send)

18. Mike Tyson은 전설적인  
 권투선수야. 20년 전, 링 위에서  
Mike는 상대선수(him)를  
의식을 잃도록 때려눕혔어.



Mike (unconscious, knock)



19. Peter는 건강을 지키기 위해  
여러 야외활동을 해. 지난주에  
그는 응봉산을 등산했어.



He (Eungbong mountain, climb)

20. 동네에서 행패를 부리던  
불량배들은 경찰에 의해  
완전히 진압되었어.



경찰아저씨는 으르렁거리며  
불량배(the gangsters)들을  
경찰서(the police station)안으로 들어가게 했어.

\* snarl: 으르렁거리다

The policeman (into, snarl)

21. 장난을 치다가 창고에서 수진이가  
간혀버렸는데,  
수진이 아버지가 문을 차서 열었다.



Sujin's father (kick, open)



22. 전쟁터는 정말로 무서운 곳이야...

적들을 피해 숨어든 집 안에서

총알이 쉬익 소리를 내며 창문을 뚫고 들어왔어.

\* whistle: 쉬익 소리를 내다



The bullet \_\_\_\_\_ (into, whistle)

23. 그녀의 고양이는 낮을

많이 가려. 그래서 지붕에서

내려오지 않아. 걱정이되어,

그녀는 고양이식품 깡통

(the catfood can)을

던져 지붕위에 두었어.

She \_\_\_\_\_ (throw, on)



24. Mrs. Parker 는 방 색깔이  
너무 칙칙하다고 생각했어.  
그래서,  
그는 방을 파랗게 칠했어.



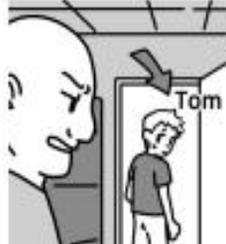
He \_\_\_\_\_ (paint, blue)

25. Sam이 가끔씩 보면 천재같고,  
또 가끔씩보면 바보같아.  
너는 Sam을 멍청하다고 생각해?



Do you \_\_\_\_\_ (stupid, consider)

26. Tom 은 프로젝트에서 큰 실수를  
저질렀어. 그것 때문에 사장님은  
크게 화가 났지.  
사장님은 꽤려보아서  
Tom 을 방 밖으로 내쫓아 버렸어.



\* stare: 꽤려보다

Boss \_\_\_\_\_ (out, stare)



27. 기사는 용을 방에 가두었어.  
절대 용이 나가지 못하게 하려고  
그 기사는 창문을 못박아 달아 버렸어.



The knight

(nail, shut)

28. 오늘은 유난히 바람이  
강하게 부는 날이야.  
바람이 문을 달아버렸어.



The wind

(the door, close)



29. Sujeong이는 청소를 열심히 했어.

Sujeong

검정 기름때를 없애는 청소를

너무 열심히 하다가,

그녀의 손이 더러워졌어.



She

(dirty, get)

30. Jack 사무실에

새로운 냉장고가 들어왔어.

그런데 트럭이 길가에

냉장고를 두고 가지 뭐야? 그의 사무실은

비탈길 위에 있는데!! 그래서,

그는 비탈길(slope) 위로 냉장고를 밀어 올렸어.



He

(push, up)

- 쓰기 시험이 끝났습니다. 수고하셨습니다. -



## Appendix 2. Chatbot Interaction Data Sample

### 2.1. A Conversation sample with AI chatbot for intransitive motion construction, *Changsoo biked to XX mountain to volunteer*

Get Started

[과업 설명]

Hi, # name of the participant #

1. 여러분은 26페이지의 그림에 대한 내용을 Elizabeth Ria와 영어로 대화할 것입니다.
2. Elizabeth Ria는 Changsoo가 어에 했던 선행에 관한 학교신문 기사를 작성하고자 합니다.
3. 26페이지 네오 안에 있는 열 문장은 Changsoo의 선행에 관한 우리말 이야기입니다.
4. Elizabeth Ria는 신문 기사를 쓰기 위하여 네오 안에 있는 내용에 관한 네 가지 질문을 할 것입니다. 여러분은 이 질문들에 대하여 올바른 영어 문장으로 답을 해야 합니다.
5. 이번 연습에서는 단어를 제공하지 않습니다. 다만 이전에 연습할때 모두 배웠던 것들입니다. 이전에 학습한 내용을 참고해 주세요.
6. 대화를 하며 Elizabeth Ria가 학교 신문 기사를 작성할 수 있도록 도와주세요!!

대화를 시작합니다.

대화를 시작합니다.



Hi, ##  
Good to see you!

김창휘!!  
I heard that Changsoo did something heroic last weekend!

[번역: ## 나는 창수가 지난 주말에 뭔가 영웅적인 일을 했다고 들어요]

Tell me what happened!! I'm going to write an article about it on school newspaper! 🤔

[번역: 무슨일이 있었는지 말해! 이것에 대해서 학교신문에 기사를 쓸 거야 🤔]

OK!

OK!

Thankyou!!! 🙏  
I have four questions!!  
[고맙워!!! 🙏  
질문이 네 개 있어]  
Next!

Next!

OK. First question.  
Where did Changsoo go yesterday? Why did he go there, and how?  
[좋아, 첫번째 질문이야.  
어제 창수는 어디를 갔었니? 거기에 왜 갔었고 어떻게 갔었니?]

\*\*\*\* 진행안내 \*\*\*\*  
1. 26페이지 내오 박스를 보고  
질문에 답을 하시기 바랍니다.  
2. 하나의 영어 문장으로 답을 하기 바랍니다.

yeterday changshu ride a bike for volunteer



yeterday changshu ride a bike for volunteer???  
Excuse me? Can you say that again?  
[번역: yeterday changshu ride a bike for volunteer라구??? 미안한데, 다시 말해볼래?]

뒤어를 잘 모르겠어요



Hm... How about starting with  
"Yesterday, Changsu ..."

[번역: 음... Yesterday, Changsu ... 라는 표현으로 시작해보면 어떨까?]



Ysterday changshu ride a bike volunteer



Oh!  
"Yesterday, Changsoo biked to XX mountain to volunteer"??

Please say it again.

[번역: 아! "Yesterday, Changsoo biked to XX mountain to volunteer"  
라구 한거야? 다시 한번 말해줄래?]



yesterdat cangshu biked to xx mountain to volunteer



yesterday cangshu biked to xx mountain to volunteer

OK. I got it!

Yesterday, Changsoo biked to XX mountain to volunteer.

(번역: 좋아! 알았어!

어제 창수는 자전거를 타고 자원봉사를 하러 XX 산에 갔다고.)

\*\*\*\*\* 진행멘트 \*\*\*\*\*

26페이지 박스안 문장 중

“어제, 창수는 자원봉사를 하기 위해 xx산에 자전거를 타고 갔다” 옆 빈칸에

정답문장: Yesterday, Changsoo biked to XX mountain to volunteer.

을 메모하세요.

메모완료하였음.



메모완료하였음.

Great! Then, tell me more!  
(좋아! 그러면 더 말해줘!)

OK!



2.2. A Conversation sample with AI chatbot for adjectival transitive resultative construction with a particle complement, *Changsoo stomped the fire out*



What did Changsoo do to protect the nature????  
[번역: 창수가 자연을 지키기 위해 무슨일을 했어??]  
\*\*\* 진행멘트 \*\*\*  
Changsoo ~ 로 답하세요.

changsoo guard for forest why are he doing



changsoo guard for forest why are he doing??  
Excuse me? Can you say that again?  
[changsoo guard for forest why are he doing??  
미안한데, 다시 말해줄래?]

모르겠습니다



Hm... How about starting with  
"Changsoo stomped the...?"

[음... "Changsoo stomped the..."로 시작하는것이 어떨까?]

\* stomp: 밟다



changsoo stomped the for forest the guard



Oh! "Changsoo stomped the fire out"? Please say it again.

[아! "Changsoo stomped the fire out." 이라구? 다시 말해줘.]



changsoo stomped the fire out

OK. I understood it. You said changsoo stomped the fire out.  
And ...

[번역: 좋아. 이해했어. 너는 changsoo stomped the fire out라고 말했어. 그리고...]



불이 일었는데,  
창수가 밟아서 불이 꺼졌어!  
(The fire became OUT!)

It's really heroic of him to stomp the fire out!

[번역: 불을 밟아 끄다니 정말로 영웅적인 일이구나!!!]

\*\*\*진행멘트\*\*\*

26페이지 박스안 문장 중

“창수는 밟아서 그 불을 꺼버렸다”

옆 빈칸에

정답문장: Changsoo stomped the fire out.  
를 메모하세요.

다 했음!



## 국 문 초 록

영어 형용사 타동결과구문(English Adjectival Transitive Resultative Construction)은 한국인 영어 학습자들에게 모국어의 대응 구문이 갖는 의미 통사론적 차이로 인해 학습하기 매우 어려운 것으로 알려져 있다. 따라서 영어 형용사 타동결과구문과 같은 복잡한 구문을 학습하기 위해서, 한국인 영어 학습자들에게는 목표 구조에 대한 명시적 교수와 교정적 피드백을 포함한 교수 처치가 요구된다.

수백 명의 학습자들에게 교정적 피드백을 제공하기 위해서는 과도한 시간과 노력이 요구되기 때문에, 인간 교사가 적절한 양의 교정적 피드백을 제공한다는 것은 사실상 불가능하다. 교정적 피드백을 제공할 때 직면하는 이러한 실용성 문제를 해결하기 위하여, 외국어 학습자들에게 인간 교사와 유사한 교정 피드백을 제공할 수 있는 수많은 인공 지능(AI) 챗봇이 개발되었다. 유감스럽게도, 현재 사용 가능한 많은 외국어 학습용 인공지능 챗봇은 아직 충분히 개발되지 않은 상태에 남아있으며, 인공지능 챗봇의 교정적 피드백이 갖는 교수효과를 비교 분석한 연구는 현재 이루어지지 않은 상태다.

이러한 선행연구의 한계에 초점을 두어, 본 연구에서는 인공지능 챗봇의 교정적 피드백이 한국 고등학생의 영어 형용사 타동결과구문의 이해와 생성에 미치는 교수 효과를 살펴보았다. 또한 본 연구에서는 이러한 교수 효과가 언어적으로 관련된 다른 영어 구문의 학습에도 영향을 끼치는지를 알아보기 위해 교실에서 직접 가르치지 않았던 구문인 영어 전치사 타동결과구문(English Prepositional Transitive Resultative Construction)의 학습 양상을 알아보았다. 이를 위해, 본 연구에서는 텍스트 메시지 기반의 페이스북 메신저에서 구동되는 인공지능 챗봇을 개발하였다.

인공지능 챗봇의 교수효과 검증을 위해 본 연구에 참여한 학생들은 네 개의 집단으로 구분되었다: 세 개의 교수 집단에는 교수처치가 적용되었고, 한 개의 통제 집단에서는 교수처치가 적용되지 않았다. 교수처치가 적용된 세 개의 집단은 챗봇그룹, 인간그룹, 영상그룹으로 분류되었으며, 이들은 모두 영어로 된 형용사 타동결과구문의 형태와 의미 쌍에 대한 5 분 길이의 학습 비디오를 시청함으로써 명시적 교수 처치를 받았다. 또한 비디오를 시청한 후 세 그룹의 학습자들은 교재를 통해 제공되는 언어연습자료를 해결하는 과업에 참여하였다.

다음으로 세 집단(챗봇그룹, 인간그룹, 영상그룹)은 다음과 같은 추가적 교수처치를 받았다: 챗봇그룹 학습자들은 교재 활동과 관련된 텍스트 기반 인공지능 챗봇과의 대화에 참여함으로써 오류에 대한 교정적 피드백을 받았다. 인간그룹 학습자들은 교재활동을 완수한 내용을 인간 교사에게 전송하고, 이에 대한 교정적 피드백을 받았다. 영상그룹 학습자들은 교재활동을 완수한 후 이에 대한 15 분의 추가적인 명시적 교수자료를 영상으로 시청하였다. 학습자의 교수효과는 사전시험, 사후시험 및 지연 사후시험으로 검증되었다. 한편 통제 집단 학생들은 교수처치 없이 세 번의 시험에만 참여하였다.

세 차례의 시험에서는 수용성판단과제(AJT)와 유도작문과제(EWT)의 두 가지 과제가 사용되었다. 수용성판단과제를 통하여, 교수된 영어 형용사 타동결과구문과 지시되지 않은 영어 전치사 타동결과구문 대한 참가자의 이해도를 측정하였다. 유도작문과제를 통하여 교수된 영어 형용사 타동결과구문과 지시되지 않은 영어 전치사 타동결과구문을 참여자가 정확하게 산출할 수 있는지를 측정하였다.

시험의 결과는 다음과 같았다. 수용성판단과제의 경우, 교수처치가 적용된 세 집단이 통제 집단보다 형용사 타동결과구문의 이해도 향상에 약간 더 효과적인 것으로 나타났다. 하지만 형용사 타동결과구문에 대한 교수적처치는 교수되지 않은 전치사 타동결과구문으로의 학습에 영향을 주지 못하였다. 유도작문과제의 경우, 인공지능 챗봇이나 인간 교사에 의해 제공되는 교정 피드백 그룹의 참가자가 영상그룹 및 통제집단의 참가자보다 형용사 타동결과구문의 올바른 생성에 더 유의미한 영향을 미치는 것으로 드러났다. 동일한 교수 효과가 전치사 타동결과구문의 학습에서도 관측되어, 형용사 타동결과구문의 학습이 전치사 타동결과구문의 학습에 일반화가 일어났다.

본 연구는 인간 교사가 직면해야 하는 실용성 문제를 극복하고, 인공지능 챗봇이 한국인 고등학교 L2 영어 학습자가 형용사 및 전치사 타동결과구문과 같은 복잡한 언어 구조를 이해하고 생성하는 데에 인간 교사와 비견될 정도로 교정적 피드백을 제공할 수 있을 것임을 시사한다. 또한, 본 연구는 인공지능 챗봇 기반 외국어 교육의 실제적 사례 및 효과를 선도적으로 보여주었다는 점에서 의미가 있다.

**주요어:** 영어 타동결과구문, 외국어 학습을 위한 인공지능 챗봇, 교정적 피드백, 외국어 교수, 영어 통사론, 한국어 통사론

**학 번:** 2013-30426