Effects of learner uptake following automatic corrective recast from Artificial Intelligence chatbots on the learning of English caused-motion construction

Rakhun Kim, Hankuk University of Foreign Studies

Abstract

This study investigated the instructional effects of learner uptake following automatic corrective recast from artificial intelligence (AI) chatbots on the learning of the English caused-motion construction. 69 novice-level EFL learners in a Korean high school were recruited to investigate the instructional effects of corrective recast from AI chatbots on the learning of the English caused-motion construction. Results from the elicited writing tasks (EWT) revealed that statistically significant gains were observed in both immediate and delayed posttests for the production of the English caused-motion construction by experimental group participants. Also, the relationship between learner uptake from AI chatbots’ corrective recast and the learning of the English caused-motion construction were analyzed. The results demonstrated that learners’ successful repair from AI chatbots’ corrective recast was positively correlated with the learning gains in the two EWT posttests. The study concludes by highlighting the significance of noticeability in AI chatbots’ corrective feedback for foreign language learning.

Keywords: Artificial Intelligence Chatbot, Caused-motion Construction, Corrective Feedback, Crosslinguistic Interference

Language(s) Learned in This Study: English

APA Citation: Kim, R. (2024). Effects of learner uptake following automatic corrective recast from artificial intelligence chatbots on the learning of English caused-motion construction. Language Learning & Technology, 28(2), 109–133. https://hdl.handle.net/10125/73574

Introduction

Extensive research on linguistic typology reveals significant syntactic and semantic differences between Korean and English, particularly with respect to English caused-motion construction (e.g., The lifeguard swam the children to the shore.) (e.g., Croft, 1998; Folli & Harley, 2006; Matsumoto, 1996; Talmy, 1985). While English employs a satellite-framed language structure, the same construction lacks an equivalent form in Korean, a verb-framed language, leading to persistent difficulties for Korean learners of English (Choi, 2020; Choi & Bowerman, 1991; Folli & Harley, 2006; Lee & Kim, 2011; Shibatani & Pardeshi, 2002; Sung, 2019; Talmy, 1985; Zubizarreta & Oh, 2007).

One solution to this issue is to provide learners with consistent corrective feedback (CF). However, the tremendous amount of manual labor required makes it practically impossible to provide Korean English as a Foreign Language (EFL) learners with individualized, consistent, explicit CF. To address this practicality issue, artificial intelligence (AI) chatbots have been introduced in foreign language learning contexts (Chen et al., 2017; Wang & Petrina, 2013). Artificial intelligence (AI) chatbots can offer more optimal and naturalistic CF under intelligent computer-assisted language learning (ICALL) frameworks (Chen et al., 2017; Choi, 2019; Chowdhary, 2020; Deng & Liu, 2018; Hirschberg & Manning, 2015; Wang & Petrina, 2013).

Despite the progress in AI chatbot-based foreign language instruction, two research gaps exist. Firstly,
previous studies (Bibauw et al., 2019; Heift, 2010; Petersen, 2010; Vlugter et al., 2009; Wilske, 2015) have overlooked typological variations when selecting target constructions for AI chatbots, limiting their effectiveness. Secondly, current AI chatbots are unable to consistently provide noticeable CF for EFL learners’ erroneous utterances, falling short compared to human instructors (Doughty & Varela, 1998; Kartchava & Ammar, 2014; Lyster & Ranta, 1997; Schmidt, 1990, 2001). This highlights the need for further advancements of the AI chatbot, providing noticeable and consistent CF in foreign language learning contexts. In fact, one of the ways to make CF more noticeable and consistent is by employing corrective recast: Corrective recast is operationally defined as providing multiple CFs (e.g., prompt, recast) in dealing with a learner’s persistent production of error within an interactional context (Doughty & Varela, 1998).

To address these gaps, the present study encompasses interdisciplinary studies over linguistics, SLA research, and ICALL framework (Ellis & Bogart, 2007), and introduces a novel AI-based chatbot program for foreign language learning, providing Korean EFL learners with corrective recast for English caused-motion construction. Drawing on insights from theoretical approaches to linguistic typology, the study aims to bridge the gap between the two languages’ differences in construction formulation.

By adopting corrective recast as a pedagogical technique and leveraging the interactionist approach in second language acquisition (SLA) research (Long, 2015), the study provides experimental group learners with noticeable CF from the AI chatbot designed explicitly for English caused-motion construction over a four-week period. The participants’ English caused-motion construction production abilities were assessed through an Elicited Writing Task (EWT) across pretest, immediate posttest, and delayed posttest sessions. Additionally, the correlations between Korean EFL learners’ uptake from the AI chatbot and the acquisition of English caused-motion construction were evaluated.

To the best of the author’s knowledge, this study represents the first attempt to incorporate AI chatbot corrective recast into the instruction of English caused-motion construction, highlighting its distinctiveness from learners’ first language typology. This sheds new light on promoting Korean EFL learners’ production of English caused-motion construction for communicative purposes under the ICALL framework (Ellis & Bogart, 2007).

**Literature Review**

**Theoretical Framework**

This study is grounded in a theoretical framework encompassing construction grammar, linguistic typology, and corrective feedback (CF) research within the interactionist perspective of the ICALL research. Ellis and Bogart (2007) highlighted the importance of “construction” in language, particularly in the design of ICALL programs for foreign language learners. They argued that the typological properties of the learners’ first language (e.g., Korean) could negatively influence their cognitive mechanisms, potentially hindering the processing of the target language (e.g., English). Thus, Korean English language learners need to adjust their interlanguage representation to align with the second language’s constructional representation, namely, English Caused-motion construction, based upon CF in interactive contexts.

**Linguistic Challenges Posed by Korean EFL Learners: English and Korean Caused-motion Construction**

According to cognitive linguistics, formal properties of the argument structure construction are directly associated with semantic structures, which reflect visual scenes basic to human experience (Goldberg, 1995). Caused-motion constructions include a form and meaning pairing, reflecting specific visual scenes. For example, the expression of a caused-motion construction, “Susan sneezed the tissue off the table,” combines formal properties (e.g., subject, verb, object, preposition phrase) with semantic features (e.g., causer argument, action, cause argument, and oblique expression). The constructional schema of the caused-motion expression corresponding to its visual scene is illustrated in Figure 1.
English caused-motion construction semantically packages a manner-denoting element (i.e., Verb) and a result-denoting component (i.e., Object and Oblique) into a mono-clausal formulation. As Figure 1 demonstrates, even an intransitive verb in English (e.g., sneeze) can take result-denoting phrases (e.g., off the table) as its complement (Folli & Harley, 2006; Goldberg & Jackendoff, 2004).

However, substantial research on linguistic typology has confirmed that the Korean caused-motion construction does not allow the single-clause packaging strategy (Croft, 1998; Matsumoto, 1996; Talmy, 1985). As Figure 2 demonstrates, the Korean caused-motion construction prefers a bi-clausal formulation strategy (Choi, 2020; Choi & Bowerman, 1991; Zubizarreta & Oh, 2007). In this formulation, the first clause denotes the causal element while the other clause denotes the result state; each clause is mediated by a conjunctive marker, -ese (Shibatani & Pardeshi, 2002).
Research in L2 acquisition consistently reports challenges faced by Korean EFL students in mastering the English caused-motion construction, mainly attributed to cross-linguistic differences between Korean and English (Kim & Rah, 2021; Sung & Yang, 2016). In fact, Korean EFL learners often encounter difficulties in accurately formulating English-like result-denoting elements even in the later developmental stage (Lee & Kim, 2011; Sung, 2019). One of the ways to solve such problems is to provide learners with consistent and noticeable CFs (Schmidt, 1990, 2001).

**Noticeability in Corrective Recast and its Effect on Learner Uptake**

While earlier investigations into the acquisition of caused-motion construction by Korean EFL learners have primarily centered on the effects of explicit instruction (Kim & Rah, 2021; Sung & Yang, 2016), little attention has been devoted to exploring the pedagogical efficacy of CF within interactional settings. CF from human instructors has generally been classified as either reformulations or prompts (Lyster & Ranta, 1997): Reformulations include recasts and explicit correction, where teachers correct erroneous linguistic forms (e.g., input-providing). Meanwhile, prompts, including metalinguistic feedback, elicitation, repetition, and clarification requests, drive language learners to recognize the corrective intent of the feedback and lead them to modify or to self-correct their errors (e.g., output-promoting; Egi, 2007; Ellis et al., 2006; Lyster & Ranta, 1997).

The modified responses of L2 learners following CF play a crucial role in language learning. Swain (1985) highlighted the significance of learners’ production of modified output in response to teachers’ CF for L2 language development. In the teacher-learner interaction, this modified output is termed “uptake,” which refers to a student’s utterance that immediately follows the teacher’s feedback and represents a reaction to the teacher’s corrective intention to address an erroneous aspect of the student’s utterance (Lyster & Ranta, 1997). Uptake can be categorized as either successful repair (SR) or needs repair (NR): SR denotes uptake that leads to a successful correction of the error, while NR involves uptake where the language learner does not fully or only partially correct the error (Sheen, 2006).

The noticeability of CF is positively correlated with learner uptake and foreign language learning (Long, 1996; Long & Robinson, 1998; Schmidt, 1990, 2001). However, there is ongoing debate regarding which type of CF is more noticeable and instructionally effective in promoting learner uptake and L2 learning. Some studies suggest that reformulations (i.e., recasts) are more noticeable than prompts (Long & Robinson, 1998; Mackey & Philp, 1998), while others demonstrate that prompts are more salient than reformulations (Ammar & Spada, 2006; Lyster, 2004; Yang & Lyster, 2010). In this context, Doughty and Varela (1998) propose that incorporating both prompts and reformulations in a CF loop, or corrective recast, may be more advantageous in enhancing noticeability compared to the repetitive use of a single type.
Corrective recast constitutes a hybrid corrective attempt that draws learners’ attention to the erroneous part in their utterance to implicitly encourage self-correction (i.e., prompt) and explicitly provides a correct linguistic exemplar to model the input (i.e., reformulation; Doughty & Varela, 1998). Doughty and Varela (1998) experimented with an approach in which the recast was preceded by repetition, with intonational stress for emphasis. This approach successfully drew the learners’ attention to the target form, as seen by their increasing self-repairs. An episode of such corrective recast is exemplified in Excerpt 1 (Han & Kim, 2008, p. 41).

**Excerpt 1**

1. S: I quarreled with my brother.
2. T: *Uh?*  
   **Clarification Request**
3. S: I quarreled with my brother when I was young. I just came out…  
   I’ve just memorized …  
   **Uptake (Needs Repair)**
4. T: *Memorized?*  
   **Confirmation Check**
5. S: I’ve just memorized …  
   **Uptake (Needs Repair)**
6. T: *OK, I’ve just remembered …*  
   **Recast**
7. S: Yeah, I’ve just remembered I quarreled with my brother. I hit him when I was young.  
   **Uptake (Successful Repair)**
8. T: Did he hit you back?

*Note. S: Student, T: Teacher*

Excerpt 1 demonstrates that the corrective recast incorporates the benefits of both prompts and reformulation by promoting the noticeability in the CF (Kartchava & Ammar, 2014). In lines 2 and 4, the teacher (T) delivered a clarification request and a confirmation check, respectively, which offered the student (S) opportunities for self-repair. However, as demonstrated in lines 3 and 5, the student failed to provide the correct formulation, constituting NR of uptake. When the recast was provided in line 6, S finally used the CF as the basis for the grammatical formulation, constituting a SR of uptake (Han & Kim, 2008). This process, packaged as a whole, promotes the saliency of the corrective function in the recasts so that they become more noticeable and instructionally effective (Doughty & Varela, 1998; Han & Kim, 2008; Li et al., 2016; Lyster & Saito, 2010; Quinn, 2014; Zhao & Ellis, 2022).

Although providing corrective recast may promote learners’ L2 learning processes, it is practically impossible for human instructors to provide persistent CFs for erroneous utterances within interactional contexts (Semke, 1984; Truscott, 1999; Valero et al., 2008). To account for such practicality issue, AI chatbots from the ICALL framework has been introduced into foreign language contexts (Heift, 2004).

**Instructional Effects of AI Chatbot for Foreign Language Learning**

AI chatbots have received attention in the field of language learning, as a supplement—not a replacement—to human language instructors with respect to the practicality issue (Shadiev & Feng, 2023). Researches demonstrate that AI chatbots can provide ceaseless pedagogical interventions to learners’ utterances, leading to improved learner uptake and L2 development (Ai, 2017; Heift, 2004, 2010). In fact, AI chatbots can automatically identify learners’ errors and provide the most appropriate CF in an immediate and consistent manner within interactional contexts as much as and even more than human instructors (Basiron, 2008; Golonka et al., 2014; Nagata, 1995, 1996; Petersen, 2010; Van Deusen-Scholl, 2008; Wilske, 2015).
Numerous conversational AI chatbots offer CF in foreign language learning, beyond English: E-tutor, an innovative ICALL predecessor, provided explicit corrections and hints to German L2 learners (Heift, 2004). Te Kaitito, a bilingual communicative ICALL system for New Zealand English speakers learning Maori, offered explicit pronoun usage corrections (Vlugter et al., 2009). Additionally, Wilske (2015) developed AI-based chatbot systems teaching German dative prepositional phrases and causal subordinate clauses to diverse second language learners, comparing recast and metalinguistic feedback effects. Further, Ai (2017) explored how an AI chatbot delivering a series of CF whose noticeability graduated from implicit to explicit facilitated the learning of Chinese sentence-level construction for American university students.

Moreover, significant efforts have been devoted to developing AI chatbots providing CF for English as a Second Language (ESL) learners. Petersen (2010) created an AI chatbot named Sasha to investigate the instructional effects of recast on the learning of English question constructions and morphosyntactic properties among learners with mixed L1 backgrounds. The results indicated that learners achieved significant learning gains from the chatbot’s CF, which was as effective as that provided by human instructors.

Research Questions

Despite the potential advantages of AI chatbots providing CF, few studies have investigated how Korean EFL learners acquire English caused-motion construction through the use of corrective recast from AI chatbots to address crosslinguistic variation (McManus & Marsden, 2019; Schenck, 2020). Additionally, previous studies generally overlooked the correlation among corrective recast from AI chatbots, learner uptake, and the acquisition of the specific constructions (Long, 1996; Long & Robinson, 1998; Schmidt, 1990, 2001). To account for the research gaps, the following research questions were addressed in the present study:

1. Does learner uptake from AI chatbots’ corrective recast lead to Korean high school EFL learners’ production of the English caused-motion construction?
2. What is the relationship among AI chatbots’ corrective recast, learner uptake, and the acquisition of the English caused-motion construction?

Based on the two research questions, the following research hypotheses are formulated:

Hypothesis 1: It is hypothesized that learner uptake from AI chatbots’ corrective recast will positively impact Korean high school EFL learners’ production of the English caused-motion construction (Heift, 2004; Petersen, 2010; Vlugter et al., 2009; Wilske, 2015). Specifically, learners who receive consistent and noticeable corrective recast by AI chatbot, whose CF is graduated from implicit to explicit, is predicted to demonstrate improved performance in producing English caused-motion constructions by overcoming L1 interference (Ai, 2017; Aljaafreh & Lantolf, 1994; Doughty & Varela, 1998; Poehner & Lantolf, 2013).

Hypothesis 2: It is hypothesized that L1-Korean EFL learners’ uptake from AI chatbots’ corrective recast may have a positive correlational relation with the production of English caused-motion construction. Specifically, it is predicted that the more experimental group participants successfully responded to the corrective recast, the better they will learn the target construction (Doughty & Varela, 1998; Li et al., 2016; Lyster & Saito, 2010; Quinn, 2014; Zhao & Ellis, 2022).

Method

Participants

The study involved 69 volunteers from two high schools in Seoul, South Korea, who were all 11th-grade students with an average age of 18 in Korean age and no prior experience living in an English-speaking country. These participants were currently enrolled in regular English language instruction at school, attending classes three times a week. However, the regular curriculum did not include any specific
instruction on English caused-motion construction. Instead, the primary objective of the regular English classes was to prepare students for English reading comprehension tasks on the Korean College Scholastic Aptitude Test (K-CSAT). In addition to their school classes, many participants reported attending cram schools for the preparation for school exam or K-CSAT.

The participants were divided into three groups: the chatbot group \((n = 23)\), the grammar group \((n = 23)\), and the no-instruction group \((n = 23)\). No-instruction group only participated in testing sessions, while the chatbot and grammar groups were engaged in instructional activities. Both group participants took part in the instructional activities during afterschool hours. The grammar group participants were assigned the same number and type of activities as the chatbot group participants. The crucial distinction between the chatbot and grammar groups was how incorrect utterances were treated. When errors were produced, the grammar group learners practiced target constructions through paper-and-pencil activities, receiving only correct formulations without any corrective feedback (CF) (Nagata, 1996). On the other hand, the chatbot group learners were provided with corrective recast from the AI chatbot, as a way to enhance the opportunities to learn the English caused-motion construction through trial and error.

Participants’ knowledge of this construction prior to the instructional treatment was measured by a pretest consisting of the Elicited Writing Task (EWT). The results showed that the three groups were homogeneous in terms of their knowledge of caused-motion construction, with no significant differences across the three groups \((M = 4.35, SD = 1.80, F(2, 66) = .15, p = .85)\).

**Target Structure**

Korean L2 English learners were insensitive to the syntactic and semantic properties of English caused-motion construction. Thus, they made persistent errors in formulating the relevant construction, requiring automatic CF in dyadic interactional contexts.

The corrective recast from the AI chatbot employed in the present study was focused on caused-motion constructions. As Table 1 demonstrates, other interrelated structures (i.e., intransitive motion constructions, and intransitive resultative, simple transitive, and transitive resultative) were also provided as instructional material, considering the network effect (Kim & Rah, 2021).

**Table 1**

*List of Target Constructions*

<table>
<thead>
<tr>
<th>Construction</th>
<th>Example</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>[Constructional Formulation]</strong></td>
<td></td>
</tr>
<tr>
<td>Intransitive Motion</td>
<td>The bird flew into the bird house.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>[Subject + Verb + Prepositional Phrase]</td>
<td></td>
</tr>
<tr>
<td>Intransitive Resultative</td>
<td>His eyes became itchy.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>[Subject +Verb +Adjectival Phrase]</td>
<td></td>
</tr>
<tr>
<td>Simple Transitive</td>
<td>He has a lot of candies.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>[Subject +Verb +Object]</td>
<td></td>
</tr>
<tr>
<td>Transitive Resultative</td>
<td>Kenny worried himself sick.</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>[Subject + Verb + Object + Adjectival Phrase]</td>
<td></td>
</tr>
<tr>
<td>Caused-motion</td>
<td>In the audition, Jack danced her off.</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>[Subject + Verb + Object + Prepositional Phrase]</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>45</td>
</tr>
</tbody>
</table>
Procedure

From March 3 to May 2, 2021, the chatbot and grammar group participants completed ten interactional tasks on the English caused-motion construction and its interrelated structures during afterschool periods. The total number of target task was 50, and one instructional session provided learners with four to five language tasks. The overall procedure is illustrated in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Week</th>
<th>Instructional Session</th>
<th>Experimental Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>Pretest (EWT)</td>
<td></td>
</tr>
<tr>
<td>2–4</td>
<td>1–10</td>
<td>Corrective recast from Chatbot</td>
<td>No instructional treatment</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>Immediate posttest (EWT)</td>
<td></td>
</tr>
<tr>
<td>6–9</td>
<td>-</td>
<td>Interval</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>Delayed posttest (EWT)</td>
<td></td>
</tr>
</tbody>
</table>

After completing the instructional sessions, all students took an immediate posttest and a delayed posttest with the interval of four weeks.

Instructional Material: AI Chatbot

The AI chatbot used in this study aimed to provide corrective recast for the English caused-motion construction and related structures during learners’ interactive sessions. As existing AI chatbots did not meet the requirements, the researcher hand-coded the chatbot on Chatfuel, a popular Facebook messenger chatbot development platform.

Chatfuel, a code-free chatbot building platform, was utilized for developing the AI chatbot in this study. The “Automation” of the Chatfuel dashboard consisted of three subfields: “Blocks,” “Flows,” and “Keywords.” Special emphasis was placed on the “Blocks” and “Keywords” sections during the construction of the chatbot. The “Blocks” tab (Figure 3-(a)) enabled the chatbot designer to manage the dialogue flow by sequencing blocks, while the “Keywords” tab automatically categorized learners’ utterances as correct or incorrect using natural language processing (NLP) engines (Figure 3-(b)). Learners’ utterances were compared to predefined correct answer sets, and if they matched, they were categorized as correct; if not, incorrect. Synonymous expressions were not considered in the predefined correct answers, as the focus of the present study lies in the instruction of a specific English construction.

Figure 3

Dashboard of Chatfuel: Blocks and Keywords

![Dashboard of Chatfuel: Blocks and Keywords](image-url)
Once the learners’ utterances were categorized as default answer or incorrect formulation, the NLP technology behind Chatfuel provided a corrective recast. The corrective recast provided in this study followed a graduated approach from implicit to explicit, as suggested by previous research (Aljaafreh & Lantolf, 1994; Doughty & Varela, 1998; Ellis, 2009, p. 13). As illustrated in Figure 4, the corrective recast commenced with less noticeable feedback, such as (1) repetition and clarification request, progressed to more noticeable feedback like (2) elicitation, and finally, the most noticeable feedback was given as (3) recast.

**Figure 4**

*Technological realization of corrective recast on Chatfuel: Configuration of Default answer Block*

---

**Figure 5** illustrates the general mechanism of corrective recast from AI chatbot, which was uniformly provided to all experimental group learners. Upon detecting the first error (e.g., “Amy’s mom take a rub.”), the chatbot offered implicit CF in the form of repetition and clarification request (e.g., “[repetition] Amy’s mom take a rub??? [clarification request] Excuse me? Can you say that again?”). Despite this, the learner made another error and responded with “I don’t know.” Subsequently, the AI chatbot provided a more explicit form of CF, elicitation, by implicating only the subject and verb relevant for the correct formulation (e.g., “Hm... How about starting with, “Amy’s mom rubbed ...””). Based on the elicitation, the learner improved the formulation compared to the initial utterance, but some ungrammatical aspects remained (e.g., “Amy’s mom rubbed lotion.”), with the omission of an obligatory prepositional phrase. To deal with such erroneous expressions, the AI chatbot generated an even more explicit form of CF, recast (e.g., “Oh! Amy’s mom rubbed the lotion on Amy’s back!!!”). Finally, based on the final CF in the corrective recast, the learner self-repaired the initial erroneous utterance completely (e.g., “Amy’s mom rubbed lotio on Amy’s back.”). Although it contains a misspelling (e.g., “lotio”), the NLP engine parsed this expression as the correct one.
Figure 5
Corrective Recast from AI Chatbot and Learner Uptake of Successful Repair (Participant C)

Tell me what Amy's mom does!

Amy's mom take a rub

Amy's mom take a rub???

Excuse me? Can you say that again?

i don't know

Hm.... How about starting with ....

"Amy's mom rubbed ..."

Amy's mom rubbed lotion

Oh! "Amy's mom rubbed the lotion on Amy's back"?

Please say it again.

Amy's mom rubbed lotion on Amy's back

OK! Amy's mom rubbed lotion on Amy's back.

Question from AI Chatbot

Pictorial Cue

1) Learner's First Error:

Repetition & Clarification Request

2) Learner's Second Error:

Elicitation

3) Learner's Third Error:

Recast

Corrected Response

I. Block name:
Step03-3

II. Block name:
Repeat CR3

III. Block name:
Elicitation 3

IV. Block name:
Recast 3
Testing Materials: Elicited Writing Task (EWT)

The EWT was employed to examine how Korean English learners produce transitive resultative constructions. A number of studies have employed EWT as one of the ways to evaluate learners’ ability to produce the relevant English sentence construction, considering potential L1 transfer effects (Ionin & Zizyk, 2014; Kim et al., 2020). The EWT consists of 30 items: 10 experimental items for the caused-motion construction and 20 distracters. The distribution of experimental items in the EWT is demonstrated in Appendix A.

Figure 6 demonstrates an example of the EWT. The depicted task included a context written in Korean, a picture cue, and a partial English sentence with two English cue words. Based upon the proposal of Ionin and Zizyk (2014), Korean translation was employed to provide more detailed contextual information as a way to trigger their use of the English caused-motion construction. In addition, the visual cues and linguistic stimuli provided in the pretest items were minimally adjusted for the immediate and delayed posttests. Finally, English cue words (a verb and a preposition) were provided inside parentheses to elicit the use of the target construction (Kim et al., 2020).

Data Collection and Coding Scheme

To address the first research question, participants’ responses to the Elicited Writing Task (EWT) were electronically transcribed and coded. Following the approach of Kim and Rah (2021), the correctness of sentences was evaluated based on the presence of all four semantic components of English caused-motion construction (i.e., Causer, Action, Causee, and Oblique) in the designated order. Sentences meeting these criteria were considered correct, while those lacking any component or having incorrect word order were categorized as incorrect. Grammatical morpheme misuses and spelling errors were excluded from the analysis, since the primary focus of this study was to investigate whether the participants utilized the constructional information to construct a sentence.

For the second research question, interactional data between AI chatbot and Chatbot group participants (n = 23) were collected. Chatbot group participants’ uptake responses during interactions with the AI chatbots during instructional session contained 1,050 tokens. Following Lyster and Ranta’s (1997) framework, the corrective recast employed in the resent study included three types of CF: repetition and clarification request (CF1), elicitation (CF2), and recast (CF3). Learners’ uptake was coded into two categories; needs repair (NR) and successful repair (SR) (Lyster & Lanta, 1997). Depending on the corrective feedback move that led to successful correction, SR was further divided into three subtypes; SR on CF1, SR on CF2, and SR on CF3 (Sheen, 2006).
Analysis

The data were analyzed using R, a free language and environment for statistical computing (R Development Core Team, 2008). Descriptive statistics were calculated to report on general patterns across different groups and test sessions.

The first research question was answered by comparing the instructional effects between the two groups and across the three tests via a 3*3 repeated measure (RM) ANOVA for the EWT tasks. It included group as the between-group variable (experimental and control groups) and test as the within-group variable (pretest, immediate posttest, delayed posttest). For any interaction between group and test, post-hoc Tukey HSD comparisons were applied to compare group performances by each test period. The statistical significance level was set to .05 in all statistical analyses. The model was created in R (version 4.05), using the car package.

The second research question was addressed through a correlational analysis, which aimed to examine the relationship between variables. Descriptive statistics, including posttest scores on English caused-motion construction (IP, DP), and rates of SR, NR, SR on CF1, CF2, and CF3 were analyzed. A two-tailed Pearson’s product-moment correlation coefficient (r) determined the strength and direction of the relationships, with values ranging from +1 to -1 (Guilford, 1956). Results were categorized as negligible (less than .2), low (.2 to .4), moderate (.4 to .7), high (.7 to .9), or very high (.9 and above) based on Guilford’s guidelines. The statistical alpha level was set at .05, and ggplot2 package in R was used for visualization.

Results

Analysis of Mean Score of EWT

This subsection is dedicated to addressing the first research question, which examines the impact of learner uptake from the AI chatbot’s corrective recast on the production of the English caused-motion construction by Korean high school EFL learners. The mean scores of the Elicited Written Task (EWT) for the caused-motion construction in Table 3 demonstrate that learners in the experimental group outperformed those in the control group across all test sessions.

Table 3

Descriptive Statistics of Elicited Writing Test Score Results

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Immediate Posttest Mean</th>
<th>Immediate Posttest SD</th>
<th>Delayed Posttest Mean</th>
<th>Delayed Posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chatbot</td>
<td>4.43</td>
<td>2.31</td>
<td>8.09</td>
<td>2.23</td>
<td>8.04</td>
<td>2.70</td>
</tr>
<tr>
<td>Grammar</td>
<td>4.5</td>
<td>1.14</td>
<td>5.08</td>
<td>1.06</td>
<td>4.83</td>
<td>1.63</td>
</tr>
<tr>
<td>No Instruction</td>
<td>4.17</td>
<td>1.87</td>
<td>4.08</td>
<td>1.78</td>
<td>3.95</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Note. The maximum mean score is 10.

Figure 7 plots the EWT results across the pretest, immediate posttest, and delayed posttest. Overall, the AI chatbot-based instruction group outperformed the control group on both the immediate posttest and delayed posttest.
To compare group performance in detail, a $3 \times 3$ repeated-measures ANOVA was conducted on participants’ scores for each task. The main effect of group ($F(2, 66) = 35.46, p < .001$) indicates distinct performance among groups. The main effect of test ($F(2, 132) = 11.51, p < .001$) was induced by the overall improvements on the immediate and delayed posttests compared to the pretest across all groups. A statistically significant interaction existed between group and test ($F(4, 132) = 7.71, p < .001$), suggesting that the degree of score improvements varied depending on the availability of chatbot instruction.

The pairwise Tukey comparisons demonstrated that the three groups significantly differed from one another on the immediate posttest ($F(2, 66) = 30.74, p < .001$), with the chatbot group scoring higher than the grammar and no instruction groups (all $p$s < .001). However, no statistical difference was found between the grammar and no instruction groups ($p = .81$). On the delayed posttest, a significant difference emerged among the three groups ($F(2, 66) = 22.98, p < .001$). Subsequent post hoc tests revealed that the chatbot group had significantly higher scores than the grammar group ($p < .001$) and the no instruction group ($p < .001$), yet the scores between the grammar and no instruction groups did not differ ($p = .96$).

In summary, the EWT results indicated that the corrective recast from the AI chatbot had a facilitative effect on Korean high school EFL learners’ production of the English caused-motion construction.

**Correlational Analysis**

This subsection reports the general findings of learners’ responses following AI chatbots’ corrective recast by analyzing the descriptive statistics of learner corpus data made up of 1,050 CF tokens. Table 4 summarizes the distribution of overall frequency of learner responses, feedback types, and repair moves. As can be seen, the participants supplied fewer correct responses (39.14%) than wrong formulations (60.86%) in the initial phase. Targeting the initial erroneous utterances, the AI chatbots provided learners with corrective recast, thereby constituting learner uptake episodes.
Table 4

Frequency and Rate of Learners’ Production of Correct Responses and Wrong Responses from the AI Chatbots

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial_Correct</td>
<td>411</td>
<td>39.14</td>
</tr>
<tr>
<td>Initial_Erroneous</td>
<td>639</td>
<td>60.86</td>
</tr>
<tr>
<td>Total</td>
<td>1050</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5 summarizes the frequency and rate of needs repair and successful repair in learner uptake episodes for initial erroneous utterances. It depicts the general tendency of learners’ responses to the CF provided for wrong responses (n = 639). The NR and SR rates were calculated based upon the total number of NR and SR divided by the total number of wrong responses, respectively. In the learner uptake episodes, more learners successfully self-corrected (SR: 76.06%) the initial ungrammatical formulation than left them unfixed (NR: 23.94%).

Table 5

Frequency and Rate of Needs Repair and Successful Repair in Learner Uptake Episodes from the AI Chatbots

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needs Repair (NR)/Uptake</td>
<td>153/639</td>
<td>23.94</td>
</tr>
<tr>
<td>Successful Repair (SR)/Uptake</td>
<td>486/639</td>
<td>76.06</td>
</tr>
<tr>
<td>Total</td>
<td>639/639</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 6 demonstrates the quantitative aspects of successful learner repair (SR) following the corrective recast of the three different CF sequences (i.e., SR on CF1, SR on CF2, and SR on CF3): Most of the SRs were obtained on CF3, which was provided as the last sequence in the recast (SR on CF3: 68.31%), followed by SR on CF2 (N= 94, 19.34%) and SR on CF1 (N = 60, 12.35%).

Table 6

Frequency and Rate of CF Types by Successful Repair (SR)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR on CF1</td>
<td>60/486</td>
<td>12.35</td>
</tr>
<tr>
<td>SR on CF2</td>
<td>94/486</td>
<td>19.34</td>
</tr>
<tr>
<td>SR by CF3</td>
<td>332/486</td>
<td>68.31</td>
</tr>
<tr>
<td>Total</td>
<td>486/486</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 8 presents the Pearson correlation matrix for the relevant factors in the study. Following Zhao and Ellis’s (2022) approach, a Pearson correlation analysis was conducted to examine the relationships among the variables, including successful uptake (e.g., SR), successful uptake by CFs (e.g., SR on CF1, SR on CF2, and SR on CF3), and L2 development of the caused-motion construction indicated by the scores of the two posttests (i.e., IP and DP).
A correlational relationship was found between the total SR and SR on CF1, CF2, or CF3. The Pearson correlation was moderate between SR and SR on CF3 (r = .63, p < .001), SR on CF1 (r = .55, p < .01), and SR on CF2 (r = .41, p = .05). The highest correlation relationship was found between SR and SR on CF3. See Appendix B for the scatter plots between SR and SR on CF1, CF2, CF3.

The correlation between IP and DP for the caused-motion construction was .61 (p < .001), indicating consistent instructional effects for the experimental group learners in the two posttest sessions. The data also revealed that SR was associated with L2 development in the caused-motion construction. Specifically, a moderate correlation was observed between IP and SR (r = .58, p < .001) and between DP and SR (r = .42, p = .05). Scatter plots depicting the relationship between the two posttest scores (IP and DP) and SR for L2 development in the English caused-motion construction are provided in Appendix C.

To conclude, as summarized in Figure 9, the results of the current study demonstrate that there is a correlational relation among corrective recast, successful repair of learner uptake, and the acquisition of a target grammatical formulation.

Figure 8
Pearson Correlation Matrix among Successfully Repaired Corrective Feedback, Total Successful Repair Rate from Uptake, and Acquisition of Caused-motion Construction

A correlational relationship was found between the total SR and SR on CF1, CF2, or CF3. The Pearson correlation was moderate between SR and SR on CF3 (r = .63, p < .001), SR on CF1 (r = .55, p < .01), and SR on CF2 (r = .41, p = .05). The highest correlation relationship was found between SR and SR on CF3. See Appendix B for the scatter plots between SR and SR on CF1, CF2, CF3.

The correlation between IP and DP for the caused-motion construction was .61 (p < .001), indicating consistent instructional effects for the experimental group learners in the two posttest sessions. The data also revealed that SR was associated with L2 development in the caused-motion construction. Specifically, a moderate correlation was observed between IP and SR (r = .58, p < .001) and between DP and SR (r = .42, p = .05). Scatter plots depicting the relationship between the two posttest scores (IP and DP) and SR for L2 development in the English caused-motion construction are provided in Appendix C.

To conclude, as summarized in Figure 9, the results of the current study demonstrate that there is a correlational relation among corrective recast, successful repair of learner uptake, and the acquisition of a target grammatical formulation.
Conclusion

Summary of the Findings

The present study found that the corrective recast from an AI chatbot has a positive impact on the learning of the English caused-motion construction by Korean EFL learners. Additionally, this study revealed significant correlations among this corrective recast, learner uptake, and the acquisition of the target construction.

Regarding the first research question, the EWT results showed that the chatbot group participants who received corrective recasts from an AI chatbot outperformed the other groups on the immediate posttest, indicating improved productive skills. Moreover, the experimental group participants maintained significantly higher scores on the delayed posttest, suggesting that the instructional effects of the AI chatbot’s corrective recast was persistent. This finding aligns with prior research indicating that AI chatbots have pedagogical potential in facilitating the acquisition of the target construction in foreign language learning contexts (Ai, 2017; Heift, 2004; Petersen, 2010; Vlugter et al., 2009; Wilske, 2015). Moreover, like explicit instruction by human instructors (Doughty & Varela, 1998), corrective recasts from AI chatbots have the pedagogical potential to consistently assist Korean EFL learners in acquiring English caused-motion construction, mitigating difficulties arising from typologically distinct L1 transfer effects (Kim & Rah, 2021; Sung & Yang, 2016; Talmy, 1985; Zubizarreta & Oh, 2007).

The second research question explored the relationship between learner uptake and L2 development. Supporting the previous studies (Ai, 2017; Heift, 2004, 2010; Petersen, 2010; Vlugter et al., 2009; Wilske, 2015), the present study demonstrated that pedagogical intervention by AI chatbots can contribute to learner uptake. The present research also revealed that AI chatbots’ corrective feedback (CF) effectively facilitated learner uptake, with 75% of erroneous utterances being self-repaired. Notably, the CF3 in the corrective recast appeared to contribute to successful repair and learner uptake, which resonates with the earlier findings that the noticeability of CF is positively related to the effectiveness of learner uptake (Long, 1996; Long & Robinson, 1998; Schmidt, 1990, 2001). To sum up, these findings corroborate the previous research that AI chatbots’ CF provided in a graduated fashion—namely, from more implicit (repetition and clarification request) to more explicit (recast or metalinguistic feedback)—can promote learner uptake and L2 learning (Ai, 2017; Heift, 2004).

Limitations and Suggestions for Future Research

The present study has several methodological limitations. Firstly, the small number of experimental group participants (n = 23) restricts the generalizability of the results and limits the statistical foundation for exploring causal relationships. Future research should conduct longitudinal studies with a larger sample size. Secondly, technological restrictions in the AI chatbot resulted in inaccurate feedback delivery, reflected in the high needs repair rate (23.94%): Among them, approximately 6% of the inaccurate feedback delivery was attributed to the chatbot’s miscategorization of learners’ correct formulations as incorrect ones. Technological improvements to promote AI chatbot accuracy are necessary for more effective feedback. Lastly, the study’s focus on the caused-motion construction may have implications for learning other interrelated constructions (Goldberg, 1995). Future research should consider the potential linguistic generalizability of AI chatbots’ corrective recast (Kim & Rah, 2021).

Despite these limitations, the study provides a promising framework for the practical application of AI chatbots in foreign language learning, emphasizing the importance of integrating computer science, linguistic findings, and SLA research to enhance L2 development through optimal feedback. (Ai, 2017).

Acknowledgements

The author wishes to express deepest gratitude to Professor Emeritus Hyun-Kwon Yang for offering invaluable perspectives that have significantly shaped the current research framework. Also, heartfelt
thanks are extended to Professor Kitaek Kim and Professor Hyunkee Ahn for their constructive guidance regarding this study. Profound appreciation is directed towards the editorial board and three anonymous reviewers whose insightful feedback and recommendations have substantially enhanced the quality of this manuscript.

Notes

1. More recent studies have suggested that providing a series of corrective feedbacks graduated from implicit to explicit promotes foreign language learning. These studies theoretically drew upon Aljaafreh & Lantolf (1994) which indicated that providing a multiple CF whose noticeability is graduated from implicit to explicit may promote L2 language learning. Many of the previously cited studies focused on either reading, listening, or writing (e.g. Kamrood et al., 2021; Poehner et al., 2015; Poehner & Lantolf, 2013; Yang & Qian, 2020; Vakili & Ebadi, 2019; Udeshinee et al., 2022), or enhancement of vocabulary (e.g. Ebadi et al., 2018; Jeon, 2021).

2. This study follows ethical guidelines to ensure the protection of the participants under the age of 19. Participation was completely voluntary, and participants were guaranteed that their responses would remain confidential and anonymous. Before the study started, informed consent was obtained, both from the participants and their parents or legal guardians. The participants could withdraw from the experiment at any time. Additionally, each participant received the equivalent of 30 U.S. dollars in Korean won as financial reimbursement for their research participation.

3. Chatfuel granted permission for the use of the chatbot for research purposes.

4. A number of corrective feedback studies have demonstrated that recast is a form of implicit corrective feedback. However, when embedded in a conversational sequence, recast becomes more explicit (Doughty & Varela, 1998); elicitation is somewhat implicit and explicit (Wilske, 2015, p. 90).

5. A reviewer has raised a valid concern regarding the classification of learners’ response, “I don’t know” as an explicit confirmation of their lack of understanding of the target construction rather than an erroneous formulation. Since the present study categorized an “error” as any expression not predefined as correct formulations in Chatfuel AI system, it was challenging to provide additional treatments for learners’ explicit expressions indicating a lack of understanding for the target construction. This limitation highlights a potential area for future research exploring how AI chatbots should handle learners’ explicit expressions of uncertainty or lack of understanding of the target construction. I express my gratitude to the reviewer for highlighting this important issue.

6. A reviewer suggests that this may be seen as a “providing clue” rather than elicitation. However, elicitation stands apart by being embedded in a communicational context (Norris & Ortega, 2000). In this study, this corrective feedback is regarded as elicitation, as it occurs within interactive sequences. I appreciate the reviewer for emphasizing this crucial distinction.

7. A reviewer pointed out that the corrective recast used in this study lacked an explicit explanation of the underlying rules for the correct form. Although none of the experimental group participants explicitly sought an explanation from the chatbot, offering learners rule-explanations could greatly enhance their acquisition of the target construction. Future studies should explore the use of an AI chatbot that provides metalinguistic feedback with rule-explanations at the end of the corrective recast. Future research should investigate the effect of corrective recast with metalinguistic explanation on participants’ independent performance, which was not the primary focus of the present study.

References


**Appendix A. Distribution of Experimental Items in EWT**

Caused-motion Construction \((N = 10)\)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>Subj <strong>blinked</strong> Obj off LOC</td>
</tr>
<tr>
<td>2)</td>
<td>Subj <strong>stared</strong> Obj out of LOC</td>
</tr>
<tr>
<td>3)</td>
<td>Subj <strong>sneezed</strong> Obj off LOC</td>
</tr>
<tr>
<td>4)</td>
<td>Subj <strong>snarled</strong> Obj into LOC</td>
</tr>
<tr>
<td>5)</td>
<td>Subj <strong>pulled</strong> Obj out of LOC</td>
</tr>
<tr>
<td>6)</td>
<td>Subj <strong>broke</strong> Obj into LOC</td>
</tr>
<tr>
<td>7)</td>
<td>Subj <strong>put</strong> Obj on LOC</td>
</tr>
<tr>
<td>8)</td>
<td>Subj <strong>threw</strong> Obj on LOC</td>
</tr>
<tr>
<td>9)</td>
<td>Subj <strong>rolled</strong> Obj out of LOC</td>
</tr>
<tr>
<td>10)</td>
<td>Subj <strong>pushed</strong> Obj up LOC</td>
</tr>
</tbody>
</table>
Appendix B. Scatter Plot: Correlation between Successful Repair Rate on CF1, CF2, CF3 and Total Successful Repair Rate
Appendix C. Scatter Plot: Correlation between Total Successful Repair Rate and the Two Posttests

About the Author

Rakhun Kim, an Adjunct Professor at Hankuk University of Foreign Studies in South Korea, explores the convergence of theoretical linguistics, Second Language Acquisition (SLA) theories, and their practical application through artificial intelligence (AI) within Korean English as a Foreign Language (EFL) contexts.

E-mail: kimrhee22@hufs.ac.kr
ORCiD: https://orcid.org/0009-0004-1312-0076