ReCALL **30**(1): 48–67. 2017 © European Association for Computer Assisted Language Learning
doi:10.1017/S0958344017000246
First published online 20 September 2017

Unlearning overgenerated be *through data-driven learning in the secondary EFL classroom*

SOYEON MOON

Seoul National University, Republic of Korea
(email: soyeonmoon@snu.ac.kr)

SUN-YOUNG OH

Seoul National University, Republic of Korea
(email: sunoh@snu.ac.kr)

Abstract

This paper reports on the cognitive and affective benefits of data-driven learning (DDL), in which Korean EFL learners at the secondary level notice and unlearn their "overgenerated *be*" by comparing native English-speaker and learner corpora with guided induction. To select the target language item and compile learner-corpus-based materials, writing samples of 285 learners were collected. The participants were randomly divided into traditional grammar learning and DDL groups. After providing instruction for each group, one immediate and one delayed writing sessions were implemented. Revealing a lower ratio of overgenerated *be* after the instruction than the control group, the DDL group showed statistically significant retention as well as immediate effects in terms of the raw counts of the target item. Based on this improvement in grammar learning and retention, DDL is considered helpful for these learners as it facilitated their efforts to discover and apply rules. In addition, their positive attitudes toward DDL including both native speaker and learner data provide useful pedagogical implications. Learning from the negative evidence produced in their own classroom helped learners, especially at lower levels, to raise their grammar consciousness and boost their motivation to learn.

Keywords: data-driven learning, learner corpus, overgenerated *be*, secondary education, guided induction

1 Introduction

For the past couple of decades, great attention has been drawn to corpus linguistics in the field of language learning and teaching (Daskalovska, 2015; Huang, 2014; Lin & Lee, 2015; Römer, 2011; Smart, 2014; Vyatkina, 2016a, 2016b). By virtue of the development of computer technology (Hunston, 2002), learners and instructors can now access "an enormous amount of authentic language input" (Liu & Jiang, 2009: 62) at the click of a mouse. The range of corpus applications in the field is wide, including developing curricula and materials, choosing syllabi, producing dictionaries, publishing references, and testing

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

language (Oh, 2004). Since the late 1990s, researchers have also begun to pay attention to learner corpora, which refer to "electronic collections of language data produced by L2 learners, that is, second or foreign-language learners" (Granger, 2013: 3235). Research based on an analysis of learner corpora has provided valuable information on language used by particular groups of learners, including their patterns of overuse, underuse, and misuse of target linguistic features compared to a reference native speaker (NS) corpus (Braun, 2007; Cotos, 2014; Smart, 2014).

On the other hand, corpora can be applied to language teaching and learning in a more direct way, as a teaching methodology. This method is known as data-driven learning (DDL), in which learners explore authentic language data and discover useful information by identifying, classifying, and generalizing the data (Johns, 1991). In their recent meta-analysis of 64 studies representing 88 unique samples, Boulton and Cobb (2017) found that DDL results in large overall effects in second language learning. Indeed, advantages of DDL as a teaching methodology have been highlighted by a number of studies conducted in various instructional contexts at tertiary (Boulton, 2009, 2010; Smith, 2011) and secondary levels (Braun, 2007; Hong & Oh, 2008). Empirical implementations of DDL, however, have almost exclusively been based on NS, not learner, data (Gilquin, Granger & Paquot, 2007). It has been suggested that exploring and comparing learner corpora with NS data in DDL greatly helps learners to notice their own errors or patterns of language use that differ from NS uses (Nesselhauf, 2004). Yet one obstacle to DDL is that some learners apparently find it overwhelming to navigate and analyze a large amount of L2 data using concordancing software by themselves. In such cases, instructors' guided induction along with preselected paper-based DDL materials can scaffold learners to interact with corpora, reducing the cognitive burden, especially at low levels (Boulton, 2010; Mizumoto & Chujo, 2016; Smart, 2014). Despite the need for such research, studies of paper-based DDL with learner corpora have thus far been rare at secondary levels (Braun, 2007), which has motivated the present study.

This study aims to investigate the effects of guided DDL grammar induction based on a comparison of NS and learner corpora of Korean middle school English learners. In addition, the study examines learners' perceptions toward learning grammar by comparing NS data with their own. The learners' affective aspects are included in the scope of the study because they are known to have an important influence on the effectiveness of general learning including grammar instruction (Green, 1993; Mantle-Bromley, 1995; Schulz, 2001). With "overgenerated *be*" as the target grammar feature (see Section 3.1 for details), this study provides two learner groups with guided DDL and traditional instruction and collects their immediate and delayed post-writing samples. Comparison of these, alongside a feedback survey, demonstrates that EFL learners at the secondary level can not only benefit from this particular type of DDL, but also enjoy learning grammar in this way.

2 Data-driven learning comparing NS and learner corpora

DDL, a term first introduced by Johns (1991), allows learners to undertake the role of researcher or language detective and engage in autonomous and active learning while investigating and analyzing authentic data. In recognition of its benefits of fostering learner autonomy and raising language awareness in examining naturally occurring language and discovering patterns (Boulton, 2009), efforts have been made to implement DDL in actual instructional contexts. A considerable number of DDL studies targeting various linguistic

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

features have reported the effectiveness of inductive instruction with corpus data in comparison with traditional teaching methods (Boulton & Cobb, 2017). Not only were new vocabulary items learned more efficiently (Frankenberg-Garcia, 2014; Stevens, 1991), but problematic linguistic items were better remedied as well (Boulton, 2010). In Huang's (2014) usage-based learning, DDL showed positive effects upon learners' understanding regarding the lexical collocations and prepositional colligations of five abstract nouns. Other linguistic items that have been dealt with in teaching practice applying DDL include linking adverbials (Boulton, 2009), passive voice (Smart, 2014), and verb–preposition collocations (Vyatkina, 2016a).

Several studies, on the other hand, have focused primarily upon learners' attitudes toward concordancing or experiences during DDL in writing. Most of these studies have dealt with undergraduate or graduate students, who overall displayed positive attitudes toward corpus-based learning or DDL. College ESL students at or above the intermediate level, for instance, evaluated the corpus use in their L2 writing positively (Yoon & Hirvela, 2004). Similarly, college students in Vyatkina's (2016b) study enjoyed corpus-based activities that provided useful information not given in dictionaries or other reference materials. Graduate learners' positive perception of corpus consultation is reported in Sun (2007) and Yoon (2008). In terms of the participants, Braun's (2007) study is an exception in the sense that it zoomed in on the secondary level of education. Still, learners' responses were not substantially different from those at the tertiary level, revealing satisfaction with the integration of diverse corpus-based activities, mostly related to grammar and lexical points, into their learning. One thing that needs to be pointed out, though, is that the studies discussed thus far have utilized NS corpora only in their implementation of corpus-based learning.

The current study pays special attention to the view that greater benefit can accrue from DDL utilizing a learner corpus alongside an NS corpus. Seidlhofer (2002) called this application of the learner corpus to language teaching the "learning-driven data approach." Comparing parallel NS and learner concordance lines of error-prone items may encourage learners to notice otherwise fossilized negative evidence (Granger & Tribble, 1998) by making their problematic language uses more salient (Meunier, 2002). Such uses, in fact, are considered as features not only to be corrected, but also to be *discovered* by learners by analyzing and comparing them with the NS corpus (Fan, Greaves & Warren, 1999), which accelerates acquisition (Seidlhofer, 2000). With appropriate follow-up communicative tasks and teachers' guidance to consolidate proper uses, the potential of DDL based on NS and learner corpora may be maximized.

The NS-learner comparison model in the foreign language classroom was initially recommended by Tribble (1990). Later, Granger and Tribble (1998) presented form-focused instruction through classroom concordancing with the learner corpus, suggesting that providing negative feedback on learners' fossilized patterns or forms (Granger, 1996) acts as effective remedial work on their underuses, overuses, or misuses (e.g. the misuse of the infinitive after *accept* or *possibility*, or the overuse of the word *important*). In the same vein, Horváth (2001) collected evidence of advanced Hungarian learners' language uses from the Janus Pannonius University corpus and used the data for individual and classroom study guides in writing pedagogy. For instance, when learners' overuse of the noun *thing* was detected, they were encouraged to replace it with more specific terms. The NS-learner corpus comparison can also be used as a valuable resource for classroom discussion

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

(Ragan, 2001) or presentation (Lee & Swales, 2006). Approaching learner-corpus-based DDL from a developmental perspective, Belz and Vyatkina (2005, 2008) tracked learners' developmental changes in using German modal particles. They showed that corpus-driven pedagogical interventions with an ongoing dynamic assessment of a learner corpus positively affect the accuracy as well as frequency of use of the target items. In Hsieh and Liou's (2008) study, EFL graduate students perceived the comparison between the expert and novice corpora of research articles (along with online peer editing) to be helpful for reducing errors in academic writing. More recently, Cotos (2014) brought together research and pedagogy of the learner corpus, first investigating learners' uses of linking adverbials and then exploring DDL pedagogy combining NS and learner data. Although both groups of learners improved in the use of linkers, students with access to both corpus types revealed more frequent, diverse, and appropriate uses than those with just the NS corpus. It appears that particularly favorable reactions toward working on the texts of their own and those of their peers contributed to better cognitive processing.

Although the vast majority of learner-corpus-based DDL studies tend to consider tertiary education as their target level (Braun, 2007), this heuristic approach has strong potential to extend its positive effects to secondary levels. These lower-level learners, however, may be unaccustomed to interacting directly with corpora (Smart, 2014); the cognitive burden of using new technology (the concordancing software tool) during DDL may inhibit learning, thereby demotivating learners (Gavioli, 2005). Thus, for learners of lower proficiency and age, a guided induction approach can be more beneficial than a pure inductive hands-on method, especially in terms of learning grammar (Haight, Herron & Cole, 2007; Herron & Tomasello, 1992), on the condition that Smart's (2014: 186) two particular characteristics of DDL are shared:

1. Real language data are used as sources of language learning materials or reference resources.
2. Learning activities are student-centered and focus on language discovery.

Here, teachers' co-constructing roles encouraging learners to achieve their aims by observing data, identifying patterns, solving problems, and completing follow-up activities are regarded as crucial (Mizumoto & Chujo, 2016; Smart, 2014).

In the light of this, the present study attempts to conduct learner-corpus-based guided DDL induction using paper-based materials with Korean middle school students, who tend to produce errors by overgeneralizing the use of *be* (see Section 3.1). The study will examine whether, compared to being taught using the traditional approach, learners can increase their grammar consciousness and extend learning more effectively while analyzing and comparing corpora of NS data and their own writing. It also aims to understand whether learners develop positive perceptions toward this particular type of DDL, which may have a strong influence on the efficacy of grammar learning with the new method (Green, 1993; Mantle-Bromley, 1995; Schulz, 2001). The following two specific research questions are addressed in the study:

1. Which is more effective in unlearning overgenerated *be*: traditional grammar learning or DDL based on NS and learner corpora?
2. How do learners perceive grammar learning based upon DDL comparing NS and learner corpora?

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

3 The study

3.1 The target language feature

The target language feature of the current study was determined based on the analysis of learner data and the relevant literature. It has been widely reported that L2 learners of English (Ionin & Wexler, 2002; Sasaki, 1990; Starren, 2006), including Korean learners (Hahn, 2000; Shin, 2001; Yang, 2002), often produce an additional *be* occurring before thematic verbs. Indeed, Korean middle school students, who constitute the participants of the current study, showed frequent overgeneralization of *be*, generating sentences such as *He* ***is*** *dance very well* and *She* ***is*** *like swimming*. As over-generated *be* occurs predominantly in such a pattern in the current database – as well as in Shin's (2001) investigation – the present study focuses on this particular structure, excluding less evident forms of overgenerated *be* in sentences such as *He* ***is*** *test is good* or *I* ***am*** *not like milk*.

Ongoing controversies exist regarding the syntactic properties of overgenerated *be* (Ionin & Wexler, 2002), between a topic marker (Hahn, 2000; Sasaki, 1990; Shin, 2001) and a functional category of agreement, tense, or aspect (Ionin & Wexler, 2002; Starren, 2006; Yang, 2002) – or both (Choi, 2013; Kim, 2011). Without subscribing to any particular view, this study focuses instead on the cognitive and affective effects of DDL on unlearning overgenerated *be*.

3.2 Participants

All 285 ninth-grade students (151 boys and 134 girls, aged 14) enrolled in a middle school in Seoul, Korea, in 2011 voluntarily participated in the compilation of the learner corpus. Most had completed four years of formal EFL instruction, for one hour per week, and their English proficiency was at a relatively elementary level. Of these, 191 students (101 boys and 90 girls) from six of the nine classes were randomly divided into a traditional learning (TL) or a DDL group, and the study was conducted during their regular school hours. Based upon the analyses of the learner corpus, approximately one quarter of the students from each group were selected as target participants. Excluded were those whose writing showed no thematic verb before which the overgenerated *be* could occur. The detailed information of the target participants is shown in Table 1, with the mean and standard deviation calculated on the final-term English test scores. A *t*-test was performed, and the result indicated no statistically significant difference of means ($p = .990$). Furthermore, the effect size of this non-significant relationship ($d = 0.06$) was found to be smaller than Plonsky and Oswald's (2014) L2 research-specific criterion for a small effect size ($d = 0.40$), confirming the comparability of the two groups in their English ability (Larson-Hall, 2016).

Table 1. *Descriptive statistics of participants*

Group	*N*	Mean	*SD*
DDL	21	72.98	16.96
TL	24	72.08	15.58

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

3.3 The construction of NS and learner corpora

3.3.1 NS corpus. Despite the many benefits of providing authentic data in English language learning (Liu & Jiang, 2009), using an NS corpus in classrooms often results in a number of concerns, such as the difficulty and unfamiliarity of vocabulary and texts (Hunston, 2002). A common solution to this problem is to edit concordance lines, replacing difficult words with easier ones, reducing the length of the original sentences, deleting words not critical to understanding, and simplifying complex syntactic structures (Braun, 2005). Editing concordance lines, however, runs counter to "the objective of exposure to naturally occurring language in use" (Adolphs, 2006: 108). Another way of leveling language to the proficiency of learners is to limit the corpus to "limited and manageable" text sources (Gavioli & Aston, 2001: 244), such as graded readers. Offering an acceptable approximation to natural language, graded reader texts seem sufficient to give input reflecting authentic language to learners, who will thus be less overwhelmed by the language and encouraged to draw conclusions from the data (Allan, 2009).

The latter approach was adopted in the present study, and a small tailor-made corpus (Basanta & Martín, 2007), designed to be easier to manage and more useful for beginners, was compiled from online issues of *Time for Kids* (TFK) published from 2008 to 2011.[1] TFK was considered an ideal source for an NS corpus to be used by the participants of this study in terms of both content and language. Issued by *TIME* magazine, one of the best-selling weekly magazines, TFK covers a wide range of interesting topics on world and national events; its language is geared toward the proficiency of native English readers from grades K1 to K6. The NS corpus consisted of 269,649 words from 973 articles and is referred to as the Time for Kids' Corpus (TFKC).

3.3.2 Learner corpus. Learner data collection began in the first week of the school year. After signing participation consent forms, the ninth graders from all nine classes were asked to write a one-page letter of self-introduction, the first writing trial (W1), in English during class time (45 minutes). A variety of sub-topics were covered, including family members, future dreams, favorite food or subjects, and mottos. The instructor did not allow learners to use dictionaries but helped them with any unknown words. Ultimately, 285 learner writings were collected (29,972 words), comprising the middle schoolers' corpus (MSC). Subsequently, based on the analysis of the learner corpus, overgenerated *be* was selected as the target feature and learner-corpus-based DDL materials were developed. Table 2 displays the basic information about the NS and learner corpora used in the study.

3.4 Materials

3.4.1 Data-driven learning materials. Materials for DDL consisted of concordance lines extracted from both TFKC and MSC. *Is*, the most frequent form of *be* in both TFKC and MSC, was used as the keyword to generate concordances.[2] The concordance lines were carefully selected according to students' levels and interests, and none of the sentences was cut off in the middle in order to enhance familiarity and comprehensibility. The materials

[1] http://www.timeforkids.com/TFK

[2] The other forms of *be* (i.e. *am*, *are*, *was*, and *were*) were also dealt with briefly in the instruction.

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Table 2. *Description of NS and learner corpora*

Corpora	Words	Number of texts	Words per text
TFKC (NS corpus)	269,649	973	277
MSC (learner corpus)	29,972	285	105

were designed so that the learners could be prompted to unlearn the overgenerated *be* through guided induction while reviewing two main categories of the *be* verb, namely, copular and auxiliary functions (Greenbaum & Quirk, 1990).[3] In the case of copular *be*, the three most common sub-categories (i.e. *is* + NOUN, *is* + ADJECTIVE, and *is* + GERUND) were included; for the function of auxiliary *be*, two sub-types (i.e. *is* + PRESENT PARTICIPLE and *is* + PAST PARTICIPLE) were dealt with. Part of the actual DDL materials is provided in Appendix A. The concordance lines were arranged in a way to help learners compare the two sets of language data. Learners were helped to explore and analyze the data interactively, filling out guide questions in their worksheet ranging from "What types of words are located before/after *is*?", "Can we see any patterns?", "How does *is* function in group #?", "Shall we match groups in the NS data to ours?" to "Is there any group that does not have the corresponding group in the other data?" Noticing the additional category (i.e. overgenerated *be*) only in the learner concordances (f-1 to f-5), learners were scaffolded to correct their own and/or peers' misuses of the *be* verb. For instance, learners correctly revised the original sentence *He* ***is*** *dance very well* to *He dances very well*. At the bottom of the materials, some space was provided for learners to write down their feelings or attitudes toward DDL based on the comparison of NS and learner data.

3.4.2 Traditional learning materials. Consulting the relevant grammar sections of the learners' textbooks, the instructor developed conventional learning materials, in which the forms, meanings, and uses of *be* verbs were explained (see Appendix B). As in the DDL materials, five categories of copular and auxiliary *be* were included. In the case of overgenerated *be*, learners were explicitly informed that it is incorrect to put an additional *be* in front of the main verb. Practice activities included gap-filling, multiple-choice, and rewriting tasks. The critical difference from the DDL materials was the absence of a direct comparison about the occurrences of *be* verbs between NSs and learners, which was the key to the guided DDL induction.

3.5 Instructional treatments

With the separate learning materials, the instructional sessions were conducted for both DDL and TL groups (see Section 3.3). As the students were not acquainted with a "corpus," getting them to experience what a corpus was and to understand why and how to use a

[3] According to the school curricula, learners were introduced to copula and progressive auxiliary *be* in the seventh grade and gerund and passive auxiliary *be* in the eighth grade. However, the participants of the present study, who were in the ninth grade, frequently made errors in the use of *be* and needed additional attention to, and/or instruction on, this feature.

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

corpus in learning was a prerequisite for the DDL instruction. One of the researchers, who was also the teacher, showed the DDL group sample concordance lines by visiting several websites, such as the British National Corpus and the Corpus of Contemporary American English.[4] In addition, the DDL group observed how their own texts could be analyzed using concordance tools like *WordSmith* (Scott, 2008). This was essential to help the learners get acquainted with a "learner corpus" and realize that their own writings could be used as learning materials. Next, the DDL and TL groups participated in information-gap activities with two articles from the TFK, from which it was confirmed that the level of the language fitted the learners' English proficiency.

Before the main instructional sessions, the DDL group had an introductory session of analyzing corpus data comparing the TFK concordances of *interesting* and *interested*, one of the typical errors in the MSC. Meanwhile, those in the TL group were given traditional instruction about the same linguistic features. Two class periods of the main sessions, approximately 45 minutes per session, were then conducted for each group. The instruction for the DDL group was based on Flowerdew's (2009) framework for corpus-based activities, namely the "4 *I*s": illustration, interaction, intervention, and induction:[5]

1. Step 1 (Illustration). Learners first looked at the teacher's preselected paper-based data, focusing on the patterns of concordance lines including *is*.
2. Step 2 (Interaction). In pairs and groups, learners shared their observations and discussed opinions following guide questions in their worksheet.
3. Step 3 (Intervention). Moving away from presenting explicit rules, the teacher provided broader hints for induction where necessary. For the lower-level learners especially, the teacher helped them with any unknown words and narrowed their focus down to the target patterns, thereby offering clearer guidance on finding the gap between the NS and learner data.
4. Step 4 (Induction). Learners individually formed their hypotheses, shared their findings with the whole class and, when necessary, revised them according to the feedback from the class. They then applied the rules that they had discovered in productive exercises by rewriting incorrect sentences including the overgenerated *be*.

In addition, at the end of the instructional treatments, learners were asked to write down their perceptions about the particular type of learner-corpus-based instruction they had received.

In the TL group, the learners were taught in the traditional way, following the paradigm of the "three *P*s" (presentation, practice, and production). By giving instructions about the form, meaning, and use of *be* verbs, the instructor clearly explained that the overgenerated *be* that is often used by the learners is an incorrect usage. Learners then practiced what they

[4] The two corpora are available online at http://corpus.byu.edu/bnc/ and http://corpus.byu.edu/coca/.

[5] Flowerdew's (2009) guided approach to DDL has also been adopted by Smart (2014). Note that the type of DDL implemented in the current study may be seen as sharing features of both "soft" and "hard" DDL (Gabrielatos, 2005; Mizumoto & Chujo, 2016). Although it displays many characteristics of soft DDL (e.g. using simplified data, paper-based concordances, and convergent tasks in classroom), it is closer to hard DDL in that the focus of the instruction was put on (guided) induction as opposed to deduction.

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

were taught with gap-filling and multiple-choice activities. Finally, they completed some productive activities, rewriting the sentences containing the overgenerated *be*.

After the main sessions, the students in both groups wrote a one-page letter of self-introduction (W2) to their native English-speaking teacher (NEST). In order to examine the retention of learning, the students were asked to produce the third writing (W3) after seven weeks. During the interim between W2 and W3, the NEST replied to each learner asking for more detailed information based on W2, which composed the topic of W3. Here students had to respond to the reply from the NEST, often answering unexpected or unfamiliar questions. The three writing tasks, which showed subtle differences, are summarized as follows:

W1: a self-introduction letter to the Korean English teacher
W2: a self-introduction letter to the NEST
W3: a reply to the NEST's letter

Finally, the data were analyzed to examine the effects of each type of instruction on unlearning overgenerated *be* and the learners' attitudes toward the learner-corpus-based grammar learning.

3.6 Data analysis

The first step in the data analysis was to identify all the instances of overgenerated *be* from the learners' writings (W1, W2, and W3). In order to examine the immediate and delayed effects of DDL compared to TL, those who did not show any token of overgenerated *be* on W1 were excluded from further analysis, which yielded 45 target participants: 21 in the DDL and 24 in the TL group. Given the finding that overgenerated *be* typically occurs before thematic verbs bearing obligatory inflections (Shin, 2001), such potential environments for overgenerated *be* were identified and counted in the writings of the two groups. The ratio of overgenerated *be* was then computed for each writing by dividing the number of overgenerated *be* by that of thematic verbs with obligatory inflections. Next, the means of those ratios were calculated and compared by groups and writing trials. A chi-square test was subsequently run (with the significance level set at .05) with the raw counts of overgenerated *be* in order to examine the immediate and delayed learning effects between groups.

The present study also examined how learners affectively perceived DDL based on a comparison of NS and learner corpora. To answer this research question, as survey results were open-ended, learners' top five common responses were categorized and discussed.

4 Results and discussion

4.1 Effects of DDL on unlearning overgenerated be

The first research question was concerned with determining which instruction is more effective in unlearning overgenerated *be* between traditional grammar learning and guided DDL induction comparing NS and learner corpora. Table 3 presents the mean ratios and the numbers of overgenerated *be* for each group during the three writing trials (W1, W2, and W3). The mean ratios for W1 were similar (DDL = 24.76, TL = 24.37), meaning that they initially made comparable proportions of overgenerated *be* errors. For W2, however, the difference in mean ratios increased by 7.73 (DDL = 4.91, TL = 12.64), which indicates that the DDL group was more successful in unlearning overgenerated *be*. The gap narrowed to 5.86 for W3 (DDL = 7.44, TL = 13.30) but was still greater than for W1 (0.39).

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Table 3. *Mean ratios and numbers of overgenerated* be

	W1	W2	W3
DDL	24.76% (40)	4.91% (7)	7.44% (14)
TL	24.37% (40)	12.64% (17)	13.30% (28)

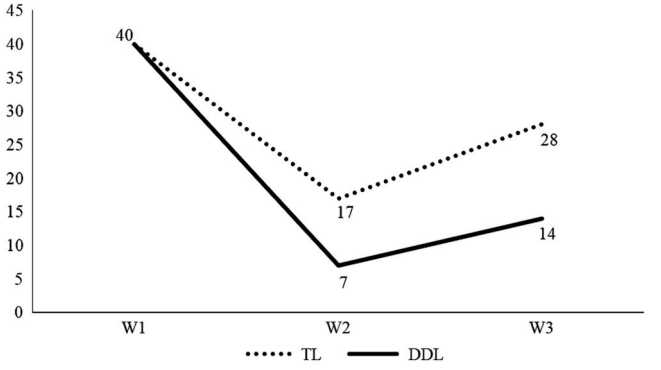


Figure 1. Changes in raw counts of overgenerated *be*

This finding suggests that, although both groups used overgenerated *be* more often during W3 than W2, the DDL group made many fewer errors on this feature than the TL group during both of the writing tasks.

However, it needs to be pointed out that the ratio of overgenerated *be* is greatly influenced by the number of inflected thematic verbs included in the writing. For example, if student A produced one overgenerated *be* out of four obligatory inflections, and student B produced one out of two, the same raw frequency (i.e. 1) would double up in ratio for student B (50%) compared to student A (25%). Taking this possibility into consideration, a chi-square test was run with the raw frequencies of overgenerated *be*. The results revealed significant differences between the two groups for both W2, X^2 $(1, N = 24) = 4.167$, $p < .05$, and W3, X^2 $(1, N = 42) = 4.667$, $p < .05$. The effects of the analyses were $r = 0.42$ on W2 and $r = 0.33$ on W3, which are near the L2 research-specific benchmark of 0.40 for a medium effect (Plonsky & Oswald, 2014).

Figure 1 illustrates that the raw counts of overgenerated *be* prior to the instruction (i.e. 40) were significantly reduced in the two groups, but at different rates. The number of overgenerated *be* produced immediately after the instruction was 2.4 times greater in the TL group than in the DDL group. In terms of retention, the TL group produced twice as many overgenerated *be* during W3 than the DDL group. The raw count of 28 overgenerated *be* for the TL group during W3 was 70% of the initial count for W1 (40).

From the results of the current study, the positive effects of the guided DDL induction on unlearning overgenerated *be* and its retention seem quite clear. Learners in the DDL group appeared to increase their linguistic awareness through discovery by identifying, comparing, and learning linguistic patterns regarding the use of *be* in the NS and learner corpora and coming up with their own hypotheses (Aston, 2001). These learners seemed to

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

internalize the target feature by noticing "the necessary and sufficient condition for the conversion of input to intake for learning" (Schmidt, 1994: 17), which prompted them to pay more attention and, accordingly, learn more. In particular, learning materials developed not only from NS but also from learner production enabled the participants to concentrate on negative evidence regarding the difference in use between NSs and themselves (Oh, 2004). The positive pedagogical effects of negative evidence obtained through the investigation of learner corpora have been reported in several studies (Cotos, 2014; O'Sullivan & Chambers, 2006). Through this negative input enhancement (Sharwood Smith, 1993), learners could themselves notice that the given form (overgenerated *be*) is incorrect. Such negative evidence could not be as easily and/or effectively noticed or obtained by learners from the traditional resources of teachers' deductive instruction or course books (O'Sullivan & Chambers, 2006). In contrast to the DDL group, learners in the TL group simply received the explicit information about overgenerated *be* through the teacher-centered conventional approach, where the explicit knowledge was weakly established (Takimoto, 2008).

4.2 Learners' perceptions toward DDL

It is generally believed that learners' perceptions, such as attitudes, beliefs, or enjoyableness, significantly affect the efficacy of grammar learning in their classroom achievement (Green, 1993; Mantle-Bromley, 1995; Schulz, 2001). In this regard, many studies have examined learners' attitudes during their corpus uses, and most have reported positive learner responses (Braun, 2007; Hsieh & Liou 2008; Sun, 2007; Vyatkina, 2016b; Yoon, 2008; Yoon & Hirvela, 2004). Yet there has been little exploration regarding secondary-level learners' perceptions of learner-corpus-based activities, especially in grammar learning, one of the most challenging areas for EFL learners. Accordingly, the second research question focused on the affective responses of learners as they encounter DDL based on the comparison of NS and learner corpora.

Students in the DDL group responded to an open-ended survey regarding what they felt about the specific type of instruction that they received.[6] They were strongly encouraged to provide as free and honest responses as possible. The survey result indicates that the vast majority of the learners – 92% ($n = 87$) – showed positive perceptions, with 5% ($n = 5$) and 3% ($n = 3$) answering negatively and "I don't know," respectively.

The learners' positive reactions toward DDL may be grouped into five main categories. First, 48 learners (55%) answered that they were able to increase grammar consciousness and would actively try to avoid grammar errors in the future. Second, 47 students (54%) appreciated both the fun and pedagogical benefits of the DDL activities. Next, 35 (40%) were satisfied with the use of both NS and learner corpora in DDL; some wrote about the advantages of "authentic" data from the NS corpus, TFKC, while others mentioned that it was very interesting and amusing to see not only their own writing but also sentences written by their peers. Furthermore, 31 learners (36%) commented that grammar learning, which had previously seemed complicated and frustrating, became, through this new method, something that they eagerly anticipated. They gained confidence in grammar learning and became motivated to study grammar.

[6] Note that all the students who received the DDL instruction ($n = 95$) were surveyed, including those who were excluded from the target participants of the study (see Section 3.2).

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

What is truly encouraging is the finding that 16 participants (18%), whose English proficiency was lower than the rest (as judged by the final-term test scores), responded that they tried to overcome difficulties in grammar learning through DDL instead of merely "giving up." Indeed, these lower-level learners often got stuck trying to figure out the meanings of words and phrases included in the concordance lines. Nevertheless, they seemed to lower their affective filter (Krashen, 1981) and enjoyed learning while analyzing their own and/or their peers' writing. Realizing that not only they but also their friends make similar errors may have exposed them to a positive affective atmosphere, indirectly facilitating grammar learning, which had previously been their least favorite learning subject. In a similar vein, Vyatkina (2016b) reported that the lowest proficiency level learner in her study could enjoy the NS corpus-based collocation learning activities despite the difficulties arising from the limited L2 knowledge. Given the current finding, it is possible that making use of not only NS but also learner corpora in DDL may be more beneficial and effective for low-level learners. The clear positive influence that learner-corpus-based learning has on motivation, a key to learning (Dörnyei, 1998), offers important implications.

On the other hand, the learners who felt negatively about the DDL, no matter how few, should not be ignored. These learners tended to struggle with comprehending basic structures and meanings of sentences included in the concordance data. When such is the case, identifying the disparities between the NS language and their own would be a rather demanding task. Some of the learners also expressed their reluctance to use the new type of instruction, which is distinctly different from the typical and familiar class and resulted in general demotivation on their part. To maximize the positive effects of DDL, the scaffolding role of the teacher, previously undervalued, becomes essential as the teacher assists students in bridging the gap between their language and the NS model. Teachers, for instance, may try to grasp the nature of difficulties for each learner and provide personalized help while encouraging all learners to approach the novel task of data analysis with willingness and less anxiety. Learners' differences, including diverse learning styles and strategies, should also be given special consideration (Mizumoto & Chujo, 2016).

5 Conclusion

With a focus on the cognitive and affective effects of DDL based on learner as well as NS corpora, this study compared guided DDL induction with traditional grammar learning for the purpose of unlearning overgenerated *be* in secondary-level EFL classrooms. The results pointed to positive immediate and retention effects of DDL with regard to the target feature, confirming Boulton and Cobb's (2017: 386) conclusion that "DDL works pretty well in almost any context where it has been extensively tried." First, the mean ratios of overgenerated *be* dropped to a greater degree in the DDL group than with traditional teaching. Next, the number of overgenerated *be* decreased after instruction in both groups, but to a significantly greater extent with DDL for both subsequent writing tasks. It seems that – compared with the learners who received traditional instruction through the teacher-centered three *P*s – learners in the DDL group actively observed, classified, and generalized (Johns, 1994) the use of *be* in the NS and learner corpora, assisted by the teacher's guidance. Noticing the differences between NS and learner corpora in this way may have stimulated deeper processing and, ultimately, improved subsequent retention

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

(Craik & Lockhart, 1972). Regarding learners' perceptions of grammar learning based on NS and their own data, the findings showed that the vast majority of learners in the DDL group, especially at lower levels, were satisfied with this new instructional method. The benefits that they recognized include increased grammar consciousness, heightened motivation, and interest in learning. DDL based on a learner corpus encouraged them not to give up learning by helping them realize that their peers also make the same errors, thereby lowering their affective filters (Krashen, 1981).

From the findings of the current study, some pedagogical implications can be drawn. Above all, there is a call for data-driven grammar learning, especially based on both NS and learner corpora. A substantial amount of literature mentions positive effects of corpus-based research and its implications (Conrad, 2000; Yeh, Liou & Li, 2007). Teachers, however, often face obstacles in the actual implementation of DDL for a variety of reasons, including its time-consuming nature (Yoon & Hirvela, 2004) and the lack of both teacher and student training (O'Keeffe & Farr, 2003). DDL based not only on NS but also on learner corpora at the secondary level is especially rare. The present study demonstrates that secondary-level learners may not only learn better but also enjoy learning grammar once they are placed in the role of active language detectives (Johns, 1991), rather than passive recipients of a detached set of rules. There appear to be innumerable benefits of using learner corpora in the EFL classrooms and, for this reason, it is important for researchers and educators to work together to overcome impediments in the classroom application. One thing to always remember is the significance of constructing and/or using both NS and learner corpora that are most appropriate and relevant in terms of language and content for local students. The target language features in DDL instruction may be successfully extended from lexis, the typical items in DDL, to other items, such as grammar, when they are determined on the basis of the instructor's careful analysis of his or her students' needs, as was the case in this study. In this regard, it is important to note the value of a corpus-based instruction on the part of not only learners but also educators. Building up and analyzing a local learner corpus itself helps teachers better diagnose and address the linguistic problems of their own students. Instruction based on such data may also help build a strong bond between the teacher and the students, as the material and the method of learning become more personalized.

The findings of the present study are based on a specific group of learners in Korea and need to be corroborated and/or challenged by studies conducted in various other contexts. Although the present learners examined paper-based DDL materials that the teacher had prepared, taking the form of guided induction, the range of application can be broadened toward direct corpus use with less guidance adapted to the local context. Given the paucity of empirical research despite the potential benefits and usefulness of applying DDL with both NS and learner corpora, future studies should focus more on how to enhance students' motivation and autonomy through proper assistance and training according to learners' learning styles and levels as well as target linguistic items. The current research will hopefully serve as a sound foundation for such future studies.

References

Adolphs, S. (2006) *Introducing electronic text analysis: A practical guide for language and literary studies*. London: Routledge. https://doi.org/10.4324/9780203087701

Allan, R. (2009) Can a graded reader corpus provide 'authentic' input? *ELT Journal*, **63**(1): 23–32. https://doi.org/10.1093/elt/ccn011

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Aston, G. (2001) Learning with corpora: An overview. In Aston, G. (ed.), *Learning with corpora*. Houston, TX: Athelstan, 7–45.

Basanta, C. P. and Martín, M. E. R. (2007) The application of data-driven learning to a small-scale corpus: Using film transcripts for teaching conversational skills. In Hidalgo, E., Quereda, L. & Santana, J. (eds.), *Corpora in the foreign language classroom: Selected papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6), University of Granada, Spain, 4–7 July, 2004*. Amsterdam: Rodopi, 141–158. https://doi.org/10.1163/9789401203906

Belz, J. A. and Vyatkina, N. (2005) Learner corpus analysis and the development of L2 pragmatic competence in networked inter-cultural language study: The case of German modal particles. *Canadian Modern Language Review*, **62**(1): 17–48. https://doi.org/10.1353/cml.2005.0038

Belz, J. A. and Vyatkina, N. (2008) The pedagogical mediation of a developmental learner corpus for classroom-based language instruction. *Language Learning & Technology*, **12**(3): 33–52.

Boulton, A. (2009) Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, **21**(1): 37–54. https://doi.org/10.1017/s0958344009000068

Boulton, A. (2010) Data-driven learning: Taking the computer out of the equation. *Language Learning*, **60**(3): 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A., and Cobb, T. (2017) Corpus use in language learning: A meta-analysis. *Language Learning*, **67**(2): 348–393. https://doi.org/10.1111/lang.12224

Braun, S. (2005) From pedagogically relevant corpora to authentic language contents. *ReCALL*, **17**(1): 47–64. https://doi.org/10.1017/s0958344005000510

Braun, S. (2007) Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, **19**(3): 307–328. https://doi.org/10.1017/s0958344007000535

Choi, I. (2013) Why is 'be' appear there?: Topic marker vs. underdeveloped functional category. *Linguistic Research*, **30**(3): 567–581. https://doi.org/10.17250/khisli.30.3.201312.008

Conrad, S. (2000) Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, **34**(3): 548–560. https://doi.org/10.2307/3587743

Cotos, E. (2014) Enhancing writing pedagogy with learner corpus data. *ReCALL*, **26**(2): 202–224. http://doi.org/10.1017/S0958344014000019

Craik, F. I. M. and Lockhart, R. S. (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**(6): 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x

Daskalovska, N. (2015) Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, **28**(2): 130–144. http://doi.org/10.1080/09588221.2013.803982

Dörnyei, Z. (1998) Motivation in second and foreign language learning. *Language Teaching*, **31**(3): 117–135. https://doi.org/10.1017/s026144480001315x

Fan, M., Greaves, C. and Warren, M. (1999) Identifying characteristic patterns in students' writing using a corpus of learner data. In Berry, R., Asker, B., Hyland, K. and Lam, M. (eds.), *Language analysis, description and pedagogy*. Hong Kong: Language Centre, HKUST, 147–161.

Flowerdew, L. (2009) Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, **14**(3): 393–417. https://doi.org/10.1075/ijcl.14.3.05flo

Frankenberg-Garcia, A. (2014) The use of corpus examples for language comprehension and production. *ReCALL*, **26**(2): 128–146. https://doi.org/10.1017/s0958344014000093

Gabrielatos, C. (2005) Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ*, **8**(4). http://files.eric.ed.gov/fulltext/EJ1068106.pdf

Gavioli, L. (2005) *Exploring corpora for ESP learning*. Amsterdam/Philadelphia: John Benjamins. https://dx.doi.org/10.1075/scl.21

Gavioli, L. and Aston, G. (2001) Enriching reality: Language corpora in language pedagogy. *ELT Journal*, **55**(3): 238–246. https://doi.org/10.1093/elt/55.3.238

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Gilquin, G., Granger, S. and Paquot, M. (2007) Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, **6**(4): 319–335. https://doi.org/10.1016/j.jeap.2007.09.007

Granger, S. (1996) Exploiting learner corpus data in the classroom: Form-focused instruction and data-driven learning. *TALC 1996*. Lancaster, 9–12 August.

Granger, S. (2013) Learner corpora. In Chapelle, C. A. (ed.), *The encyclopedia of applied linguistics*. Malden, MA: Wiley-Blackwell, 3235–3242.

Granger, S. and Tribble, C. (1998) Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In Granger, S. (ed.), *Learner English on computer*. London: Longman, 199–209. https://doi.org/10.4324/9781315841342

Green, J. M. (1993) Student attitudes toward communicative and non-communicative activities: Do enjoyment and effectiveness go together? *The Modern Language Journal*, **77**(1): 1–10. https://dx.doi.org/10.2307/329552

Greenbaum, S. and Quirk, R. (1990) *A student's grammar of the English language*. London: Longman.

Hahn, H. (2000) *UG availability to Korean EFL learners: A longitudinal study of different age groups*. Seoul National University, unpublished PhD.

Haight, C. E., Herron, C. and Cole, S. P. (2007) The effects of deductive and guided inductive instructional approaches on the learning of grammar in the elementary foreign language college classroom. *Foreign Language Annals*, **40**(2): 288–310. https://doi.org/10.1111/j.1944-9720.2007.tb03202.x

Herron, C. and Tomasello, M. (1992) Acquiring grammatical structures by guided induction. *The French Review*, **65**(5): 708–718.

Hong, S.-Y. and Oh, S.-Y. (2008) The effects of corpus-based learning of vocabulary and grammar on Korean high-school students. *English Language Teaching*, **20**(1): 261–283. https://doi.org/10.17936/pkelt.2008.20.1.013

Horváth, J. (2001) *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Pécs: Lingua Franca Csoport.

Hsieh, W.-M. and Liou, H.-C. (2008) A case study of corpus-informed online academic writing for EFL graduate students. *CALICO Journal*, **26**(1): 28–47.

Huang, Z. (2014) The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, **26**(2): 163–183. https://doi.org/10.1017/s0958344014000020

Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524773

Ionin, T. and Wexler, K. (2002) Why is 'is' easier than '-s'?: Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research*, **18**(2): 95–136. https://doi.org/10.1191/0267658302sr195oa

Johns, T. (1991) Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal*, **4**: 1–16.

Johns, T. (1994) From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In Odlin, T. (ed.), *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press, 293–313. https://doi.org/10.1017/cbo9781139524605.014

Kim, K. (2011) Overgenerated *be* from topic marker to verbal inflection. *Foreign Language Education Research*, **14**: 1–22.

Krashen, S. (1981) *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Larson-Hall, J. (ed.) (2016) *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). New York, NY: Routledge.

Lee, D. and Swales, J. (2006) A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, **25**(1): 56–75. https://doi.org/10.1016/j.esp.2005.02.010

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Lin, M. H. and Lee, J.-Y. (2015) Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, **69**(3): 264–274. https://doi.org/10.1093/elt/ccv010

Liu, D. and Jiang, P. (2009) Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, **93**(1): 61–78. https://doi.org/10.1111/j.1540-4781.2009.00828.x

Mantle-Bromley, C. (1995) Positive attitudes and realistic beliefs: Links to proficiency. *The Modern Language Journal*, **79**(3): 372–386. https://doi.org/10.2307/329352

Meunier, F. (2002) The pedagogical value of native and learner corpora in EFL grammar teaching. In Granger, S. Hung, J. and Petch-Tyson, S. (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam/Philadelphia: John Benjamins, 119–141. https://doi.org/10.1075/lllt.6.10meu

Mizumoto, A. and Chujo, K. (2016) Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System*, **61**: 55–64. https://doi.org/10.1016/j.system.2016.07.010

Nesselhauf, N. (2004) How learner corpus analysis can contribute to language teaching: A study of support verb constructions. In Aston, G., Bernardini, S. and Stewart, D. (eds.), *Corpora and language learners*. Amsterdam/Philadelphia: John Benjamins, 109–124. https://doi.org/10.1075/scl.17.08nes

Oh, S.-Y. (2004) Corpus and English language education. *Foreign Language Education Research*, **7**: 1–38.

O'Keeffe, A. and Farr, F. (2003) Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, **37**(3): 389–418. https://doi.org/10.2307/3588397

O'Sullivan, Í. and Chambers, A. (2006) Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, **15**(1): 49–68. https://doi.org/10.1016/j.jslw.2006.01.002

Plonsky, L. and Oswald, F. L. (2014) How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, **64**(4): 878–912. https://doi.org/10.1111/lang.12079

Ragan, P. H. (2001) Classroom use of a systemic functional small learner corpus. In Ghadessy, M., Henry, A. and Roseberry, R. L. (eds.), *Small corpus studies and ELT: Theory and practice*. Amsterdam/Philadelphia: John Benjamins, 207–236. https://dx.doi.org/10.1075/scl.5.14rag

Römer, U. (2011) Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, **31**: 205–225. https://doi.org/10.1017/s0267190511000055

Sasaki, M. (1990) Topic prominence in Japanese EFL students' existential constructions. *Language Learning*, **40**(3): 337–368. https://doi.org/10.1111/j.1467-1770.1990.tb00667.x

Schmidt, R. (1994) Deconstructing consciousness in search for useful definitions for applied linguistics. *AILA Review*, **11**: 11–26.

Schulz, R. A. (2001) Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *The Modern Language Journal*, **85**(2): 244–258. https://doi.org/10.1111/0026-7902.00107

Scott, M. (2008) *WordSmith Tools* (Version 5). Liverpool: Lexical Analysis Software.

Seidlhofer, B. (2000) Operationalizing intertextuality: Using learner corpora for learning. In Burnard, L. and McEnery, T. (eds.), *Rethinking language pedagogy from a corpus perspective*. Berlin: Peter Lang, 207–223.

Seidlhofer, B. (2002) Pedagogy and local learner corpora: Working with learning-driven data. In Granger, S., Hung, J. and Petch-Tyson, S. (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam/Philadelphia: John Benjamins, 213–234. https://doi.org/10.1075/lllt.6.14sei

Sharwood Smith, M. (1993) Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, **15**(2): 165–179. https://doi.org/10.1017/s0272263100011943

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Shin, J.-S. (2001) L1 influence in foreign language learning: Topic-prominence in Korean EFL learners' interlanguage grammar. *Foreign Language Education*, **8**(1): 1–21.

Smart, J. (2014) The role of guided induction in paper-based data-driven learning. *ReCALL*, **26**(2): 184–201. https://dx.doi.org/10.1017/s0958344014000081

Smith, S. (2011) Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning*, **24**(4): 291–316. https://doi.org/10.1080/09588221.2011.557024

Starren, M. (2006) Temporal adverbials and early tense and aspect markers in the acquisition of Dutch. In van Geenhoven, V. (ed.), *Semantics in acquisition*. Dordrecht: Springer, 219–244. https://doi.org/10.1007/1-4020-4485-2_9

Stevens, V. (1991) Concordance-based vocabulary exercises: A viable alternative to gap-fillers. In Johns, T. and King, P. (eds.), *Classroom concordancing: English Language Research Journal*, **4**: 47–63.

Sun, Y.-C. (2007) Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, **20**(4): 323–343. https://doi.org/10.1080/09588220701745791

Takimoto, M. (2008) The effects of deductive and inductive instruction on the development of language learners' pragmatic competence. *The Modern Language Journal*, **92**(3): 369–386. https://doi.org/10.1111/j.1540-4781.2008.00752.x

Tribble, C. (1990) Concordancing and an EAP writing program. *CAELL Journal*, **1**(2): 10–15.

Vyatkina, N. (2016a) Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, **28**(2): 207–226. http://doi.org/10.1017/S0958344015000269

Vyatkina, N. (2016b) Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, **20**(3): 159–179.

Yang, H.-K. (2002) Korean EFL learner's acquisition of English inflectional features. *Korean Journal of English Language and Linguistics*, **2**(2): 227–248.

Yeh, Y., Liou, H.-C. and Li, Y.-H. (2007) Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, **20**(2): 131–152. https://doi.org/10.1080/09588220701331451

Yoon, H. (2008) More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, **12**(2): 31–48.

Yoon, H. and Hirvela, A. (2004) ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, **13**(4): 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

Appendix A

Example of DDL learning materials

Note: In the original materials, instructions were provided in the learners' mother tongue (i.e. Korean) in consideration of their English proficiency levels.

♦Let's examine native speakers' actual data for the uses of *be* verbs.

A-1	One solution to the child-labor problem in poor countries	**is**	**education.**
A-2	Empathy	**is**	an **understanding** of other people's feelings.
A-3	The United States	**is**	a **country** of immigrants.
A-4	Election Day, November 2,	**is**	a big **day** for voters.
A-5	Barack Obama	**is**	the **man** for this job.
A-6	Merapi	**is**	the **volcano** in the northwestern Pacific Ocean.
B-1	"Now the mummy	**is**	**safe**." Zahi Hawass, Egypt's top archaeologist, told TFK.
B-2	Some birds ate the oil, which	**is**	**poisonous**.
B-3	"Despite our hardships," he said, "our union	**is**	**strong**."
B-4	But studies show that free playtime	**is**	**good** for kids' minds.
B-5	He	**is**	very **big**. His wings are nine feet across.
C-1	"Logic? Why, dear, logic	**is**	**knowing** what things are true and not true."
C-2	Poetry is the art of saying a whole lot in as little space as possible. It	**is**	**writing** "concentrate." (*c.f.* I like orange juice *concentrate*.)
C-3	One of the best parts of my job	**is**	**meeting** kids like you across the country.
C-4	Part of knowing the company	**is**	**understanding** its people and their roles within the organization.
D-1	Obama will report on how well the country	**is**	**doing**.
D-2	The company	**is**	**destroying** their forests.
D-3	Scientists say the Earth is getting warmer. That	**is**	**making** mountain glaciers near Bangladesh melt.

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press

E-1	That's how a mummy	**is**	**made**.
E-2	No hunting or habitat destruction	**is**	**allowed** there.
E-3	And Atlantis's last trip	**is**	**planned** for the end of June.

♦Let's examine our actual data for the uses of *be* verbs.

a-1	My favorite food	**is**	**pizza** because it is very delicious.
a-2	He is very tall and cute. He	**is**	an **actor**.
a-3	I saw my new school teacher. He	**is**	a kind **man**.
a-4	Seoul	**is**	the **capital** of my country.
b-1	I really hate singing because my song	**is**	**terrible**.
b-2	I have a brother. He	**is**	**smart**.
b-3	My room	**is**	very **dirty** because I don't clean my room once a week.
b-4	Harry Potter	**is**	very i**nteresting** and **fun**.
c-1	My hobby	**is**	**playing** soccer and reading books.
c-2	My topic	**is**	**using** real names on the Internet.
c-3	My habit	**is**	**playing** computer games.
d-1	Also I don't like English, too. However, your class	**is**	**changing** me.
d-2	The clock	**is**	**pointing** to 11:29 a.m.
d-3	The iceberg in North Pole	**is**	**melting** because of global warming.
e-1	My favorite food	**is**	**made** by my mom. She cooks well.
e-2	If "ozon"	**is**	**destroyed**, the earth will be hotter than before.
e-3	After this diary	**is**	**checked**, I'll probably be better at English.
f-1	My favorite sport is soccer. Soccer	**is**	**make** me excited.
f-2	Seung-ri is the youngest of "Bigbang." He	**is**	**dance** very well.
f-3	My older brother is very bad. He	**is**	**hit** me.
f-4	Youngwoong-Jaejung	**is**	**sing** well and he is very pretty.
f-5	My sister is smart. She	**is**	**go** to university.

Appendix B

Example of traditional learning materials

2. **The Uses of *Be* Verbs**

2.1 **Nouns**, **adjectives**, or **gerunds** come **after *be*** verbs to indicate state of being.

- My father **is** a **businessman** and my mother **is** a **teacher**.
- In Korea we have four seasons. The weather **is beautiful**.
- Spring begins in March. It **is warm** and pleasant.
- My hobby **is drawing** cartoons.

2.2 ***Be*** verbs are followed by a **present participle** or a **past participle** to express **progressive** or **passive** forms, respectively.

- Look at this picture. A koala **is carrying** its baby on its back.
- Lately a lot of work **is done** by volunteer workers.

★ **Note** that it is **incorrect** to use *be* verbs before the infinitive (or tensed) form of a verb.

- My sister is smart. She ***is*** **go** to university. [incorrect]

→ My sister is smart. She **goes** to university.

About the authors

Soyeon Moon teaches English as a foreign language at a public high school in Korea. She is also a PhD candidate in the Department of English Education at Seoul National University. Her main research interests include corpus-based language learning and teaching.

Sun-Young Oh received her doctoral degree in Applied Linguistics at the University of California, Los Angeles. Currently, she is a professor in the Department of English Education at Seoul National University, where she teaches courses in corpus linguistics, discourse analysis, and second/foreign language pedagogy.

https://doi.org/10.1017/S0958344017000246 Published online by Cambridge University Press