

TEACHING GRAMMAR AND VOCABULARY FOR THE TOEIC TEST WITH CORPORA: THE CASE OF LOWER INTERMEDIATE LEARNERS

Yen-Yu Lin

ABSTRACT

This study examined the effectiveness of guided data-driven learning (DDL) activities on helping technological university students with a lower-intermediate proficiency level to learn grammar and vocabulary topics for the TOEIC test. The question of whether inductive learners make more progress than deductive learners was also addressed. A total of fifty-one non-English majors in a freshman English class participated in the study. Five computer-based and four paper-based DDL lessons were developed to increase students' awareness of the usage of particular grammar items and the distinction between three sets of synonyms. In order to determine the effects of the treatments, all the participants took pre-tests and immediate posttests. The results showed that there was a statistically significant difference between pre- and post-tests on grammar and vocabulary. However, inductive learners did not outperform deductive learners. As for students' perception of the treatments, it is noteworthy that they tended to take a neutral to less than positive attitude towards this DDL experience even though such an experience helped them significantly improve. The results suggested the existence of a gap between students' perceived satisfaction and their actual achievement. Pedagogical implications of these findings were discussed to improve the efficacy of DDL in Taiwan's EFL context.

Key Words: corpus linguistics, data-driven learning, learning style, autonomous learning

INTRODUCITON

A number of studies find that the majority of students from Taiwan's vocational education system have lower English proficiency than students in liberal arts and sciences institutions. Lo (2011) and Tsao (2008), for instance, claimed that vocational school students acquired smaller vocabularies and had insufficient understanding of grammatical knowledge. The Score Data Summary for 2009 GEPT Elementary-Level shows that the technological university students incur a lower passing rate than student groups from general universities and other higher institutions. More specifically, Wu's (2009) study of teacher and student perceptions of the language learning environment in a Taiwanese technological university revealed a rather disappointing phenomenon: while the teachers reflected that they were actively presenting the teaching material through lecture and repetition, students commented that they were passive learners in class. It appears that technological university students, in a native Chinese-speaking English classroom, seldom get involved in the process of manipulating the target language and extending their knowledge. The present study argues that the learner-centered approach should further be explored in enhancing their acquisition of English vocabulary and grammar.

According to Johns (1988), the data-driven learning (DDL) approach has brought a major change in the learning process: "students now take on more responsibility for his or her learning, and the teacher acts as research director and research collaborator rather than transmitter of knowledge" (p. 14). In the extensive literature on DDL, relatively little research has focused on its effect on students with lower levels of English proficiency. The present study examines whether guided DDL activities help Taiwanese technological university students with lower-intermediate English proficiency levels make progress in the learning of particular grammar rules and sets of synonyms in the TOEIC test. Students' attitude towards corpus-based practices are also presented.

LITERATURE REVIEW

Recent years have seen increased attention being given to data-driven learning (DDL) in the literature of corpus linguistics for

English teaching and learning. The key principle of DDL is that learners should be put in the center of the learning process and be able to detect usage patterns among the data in corpora themselves. Through acting as researchers, learners become more independent and autonomous (Bloch, 2007; Boulton, 2010; Keck, 2004).

The DDL studies generally fall into two categories. Some studies attempt to explore how corpora consultation can benefit the L2 writing process and the development of EAP teaching materials (Cresswell, 2007; Coxhead, 2000; Coxhead & Byrd, 2007; Thurstun & Candlin, 1998; Tseng & Liou, 2006); others focus on the implementation of DDL on grammar learning (e.g., Bulton, 2008; Chujo & Oghigian, 2007; Vyatkina, 2006a, 2006b), collocation learning (e.g., Chan & Liou, 2005; Lindstromberg & Boers, 2008; Peters, 2014, 2016; Sun & Wang, 2003; Webb & Kagimoto, 2009, 2011), and lexical acquisition (e.g., Aston, 1998; Cobb, 1999; Murphy, 1996). The following review will focus on the studies of grammar and vocabulary learning with university students at intermediate or low L2 proficiency levels as participants.

In order to help Japanese university beginning-level students to improve their TOEIC scores, Chujo and Oghigian (2007) employed a Japanese-English parallel corpus and the *ParaConc* concordancing program to design sets of corpus-based activities to teach learners to identify recurring features of various words frequently appearing in the TOEIC. The learning outcomes showed this course design was useful for learning grammar basics. Similarly, in Chujo et al. (2013), remedial junior high school students used a parallel corpus tool, *AntPConc*, to learn grammar items through processing English target texts and their Japanese translations. By means of *AntPConc*, students can easily search and sort grammatical features in English and compare them with those in Japanese. Students gave favorable feedback on the lessons and the corpus tool.

Instead of offering lower level students L1 translation of English texts in the corpus, several researchers designed paper-based activities with sample concordance lines to see if they facilitate learning (e.g., Bulton, 2008, 2009, 2010, 2012; Frankenberg-Garcia, 2014; Koosha & Jafarpour, 2006; Vyatkina, 2016a; Yoon and Jo, 2014). The main benefit of paper-based DDL (or hands-off DDL) is that it does not require students to spend considerable time getting familiar with sophisticated corpora and their search techniques (e.g., Turnbull & Burston, 1998; Yoon & Hirvela, 2004); in addition, concordance print-outs reduce some of lower level students' cognitive burden and allow learners to focus on a single new element (Aston, 1997). Recently, a number of studies further explored the

difference in learning outcomes of the students at lower proficiency levels following paper-based and computer-based DDL (e.g., Boulton, 2012; Vyatkina, 2016b). Boulton (2012) evaluated the performance of the students whose TOEIC scores ranged between 500 and 555 points after experiencing DDL on computer and on paper. It is interesting that the results showed a slight but not significant advantage for paper-based materials. Boulton further suggests that hands-off DDL with highly controlled activities represents a feasible option and may be beneficial to learners at lower levels, but this does not mean that it is better than computer-based DDL. The variables, such as learner variation and different local conditions, should also be considered in deciding whether and how to implement hands-on or hands-off DDL in the teaching process. In a study on learning German verb-preposition collocations, Vyatkina (2016b) compared the effects of paper-based and computer-based activities for intermediate level English learners. The results showed that both DDL types are equally effective for all learners, regardless of students' proficiency and perceptions.

Observing that Taiwanese students are more used to the deductive teaching approach, some researchers were concerned about the feasibility of implementing inductive or deductive DDL in the English classroom. In 2003, Sun and Wang investigated whether or not any significant differences were exhibited by senior high school students learning collocations using a concordancer with an inductive approach or using a deductive teaching approach. Their findings show that the inductive approach to DDL helps students more than the deductive approach in learning the easier collocation patterns but not the more difficult ones. This implies that more structured guidance was needed for difficult sentence structures. In a more recent study, Tsai (2019) compared the inductive and deductive approaches in relation to Taiwanese college students' vocabulary acquisition. In her study, the inductive group searched for the target word in the corpus before consulting the dictionary, whereas the deductive group performed the same tasks in reverse order. The results show that the inductive group recalled more collocations but did not perform well in the retention of word definitions, whereas the deductive group performed better on definition recall.

Also working in the Taiwanese context, Lin has conducted a number of studies investigating the use of inductive DDL in the grammar classroom. In 2015, he investigated the perception of six early-career teachers after they taught grammar to Taiwanese college students by inductive DDL for the first time. Overall, these teachers regarded inductive DDL as an innovative approach to teaching

grammar; but at the same time, they also suggested several ways to improve DDL teaching, including cutting short the number of concordance lines for each grammar concept, using relatively more complete sentences, giving more guided questions and selecting sentences that contain only one grammatical item. In 2016, he further examined the effects of combined inductive DDL and TDA (traditional deductive approach) and pure TDA on teachers' and EFL students' attitudes to learning grammar. This study concluded that greater exposure to DDL activities may cultivate improved learning attitudes. In 2017, Lin conducted a follow-up study that not only examined the feedback of Taiwanese students on an inductive DDL treatment but also measured whether and to what extent the DDL-integrated treatments and pure TDA improved students' grammar proficiency in three EFL classes. Notably, no significant difference was observed between the classes in terms of their post-test grammar scores, implying that inductive DDL does not facilitate grammar learning any more than more traditional methods do.

Aside from the relationship between inductive/deductive DDL teaching and learning outcomes, another topic of recent discussion is the possible effects of DDL on Taiwanese inductive/deductive learners. Lee and Liou (2003), incorporating concordancing into the regular English curriculum for EFL senior high school, examined whether learners' preferred English learning styles influenced the effectiveness of concordancing on learning English vocabulary. The results show that inductive learners benefit more from concordancing and favored DDL more than deductive learners do. Yeh (2003) obtained a similar finding in a study on 23 college-level learners focusing on the effects of self-selected concordances and individual learning styles. The results indicate that concordancing-based vocabulary learning is particularly helpful for inductive learners. In contrast, Li's (2005) study on college students' DDL of synonymous adjectives shows that inductive learners do not outperform deductive learners. Quite distinct from the above-mentioned two studies, Mizumoto and Chujo's (2016) investigation explores the connection between DDL grammar learning and Japanese students' learning styles and suggests that DDL may be beneficial for both deductive and inductive learners. As this review shows, previous studies seem to have obtained inconsistent results concerning the relationship between DDL and learning styles, suggesting that further and deeper empirical investigation is required.

Tsai (2019) and Lin et al. (2017) claim that a possible reason for the mixed outcomes in previous studies may be the different ways by which teaching and learning approaches are implemented or the

different language points being examined. In the present study, paper-based and computer-based DDL lessons were integrated into the curriculum for the teaching of particular language points, most of which have not been covered in previous research of DDL. Students' preferred learning styles (inductive/deductive) and their connection with language achievement are also explored.

One other thing that is worth noting is that the majority of the DDL studies on Taiwanese students reviewed above have been conducted mainly in general colleges or high schools instead of technological universities. As Boulton (2010) remarks, DDL may be more suitable with particular groups of learners. It is more desirable to conduct empirical research to verify whether this exploratory corpus-based learning approach, which allows great autonomy for innovation, can benefit lower-intermediate-level technological university students.

RESEARCH QUESTIONS

For the purpose of this study, a guided induction approach was employed for DDL instruction. According to Smart (2014), guided induction, "provides a structured framework for inductive learning, places the learner at the center of the learning task" (p. 187). Considering the English proficiency of its participants, the current study takes an eclectic approach. That is, it combines the guided induction approach with a deductive teaching method for DDL. At the end of a discovery task, the teacher offers explanations to confirm or correct learners' hypotheses of the underlying pattern of the language feature being examined. Four research questions are addressed:

1. Can guided DDL improve lower-intermediate-level learners' learning of grammar?
2. Can guided DDL improve lower-intermediate-level learners' learning of vocabulary?
3. Which types of learner (inductive or deductive) benefit more from guided DDL instruction?
4. What are student perceptions of guided DDL learning?

METHODOLOGY

Participants and the Institutional Setting

The study was carried out in the spring semester of 2021. A total of 50 first-year college students at a technological university in central Taiwan participated in the study. They consisted of L1 Chinese speakers who had studied English for a minimum of six years in a classroom setting. They came from the College of Engineering and the College of Electrical Engineering and Computer Science. Most of them were 18 or 19 years of age. These students were placed in a lower-intermediate-level Freshman English class according to their English test scores in the Technological and Vocational Educational Examination. On average, they obtained 60 to 65 points of 100 in the test. The course was titled College English 1 and the researcher was the instructor for this compulsory freshman English course. The course aimed to help students acquire effective reading skills, enlarge their vocabulary size, and review important grammar rules taught in high school. The class met once a week for 100 minutes over 18 weeks.

The Development of the DDL Materials for the Study

In order to better prepare students in coping with global competition, the Taiwanese Ministry of Education has encouraged colleges and universities to implement an English graduation benchmark policy. In the school where the participants study, passing a standardized proficiency test (TOEIC) is a condition for graduation. Students should obtain at least 375 points in the TOEIC Listening and Reading Test to meet the requirement for graduation. The DDL grammar lessons in the present study, therefore, focus mainly on the grammar topics that frequently appear in the TOEIC. The target grammatical points include word classes, quantity phrases, conjunctions, subject-verb agreement, prepositions, and discourse markers. These six points were selected partly based on Chujo et al.'s (2013) study which lists the items that students tended to get wrong on the TOEIC test, and partly on the researcher's judgment of the difficulty level. The DDL vocabulary lessons were designed to help students distinguish between words of similar meanings by analyzing collocations and sentence structures. Three commonly confused sets

of words (*demand, require, and request; fare and fee; decline and decrease*) were selected from the TOEIC vocabulary list of 640 words provided by Chujo and Genung (2005); the efficacy of this list has been verified by calculating the text coverage of the vocabulary included in old TOEIC tests.

The treatment was implemented for about fifteen weeks. The details of the experimental arrangements are shown in Table 1.

Table 1

The Arrangement of the Treatments

Week	Spring semester
1	Introduction to DDL and corpus tools, background questionnaire, pre-test
2	Grammar Lesson 1: Quantity Phrases with Countable or Uncountable Nouns (<i>a large number of, a large amount of, a great deal of</i>)
3	Grammar Lesson 2: Word Classes (noun-forming morphemes, verb-forming morphemes, adjective-forming morphemes, adverb-forming morphemes)
4	Grammar Lesson 3: Time Prepositions (<i>until, from, during, by</i>)
5	Grammar Lesson 4: Subject-verb Agreement (<i>either...or..., neither...nor..., not only...but also..., ...as well as...</i>)
6	Grammar Lesson 5: Conjunctions (<i>despite, while, even though, even if</i>)
7	Grammar Lesson 6: Discourse Markers (<i>owing to, on account of, namely, consequently</i>)
8	Immediate post-test of grammar
9	Mid-term Exam
10	Mid-term Exam Review
11	Vocabulary Lesson 1: <i>Demand, Require & Request</i>
12	Vocabulary Lesson 2: <i>Fare & Fee</i>
13	Vocabulary Lesson 3: <i>Decline & Decrease</i>
14	Immediate post-test of vocabulary, evaluation questionnaire
15	Interviews with selected participants (after class)

For the grammar lessons, the corpus data came from the Corpus of Contemporary American English (COCA) (<http://corpus.byu.edu/coca>), which is a large, freely available online corpus of English, containing more than one billion words. The grammar lessons fell into two categories: computer-based and paper-

based. The presentation of the data through print materials or computers depended on the goals of the lessons. For example, in DDL Grammar Lesson 1 (see Appendix A), students were expected to expand their morphological knowledge and identify grammatically related forms of a word. Directly searching for the words of a particular part of speech in the corpus and observing their derivational endings was believed to enhance students' awareness of the connection among word structure, morphemes, and meaning. Given the requirements of the task, a computer-based exercise was favored. On the other hand, the grammar lesson of conjunctions (see Appendix B) aimed to encourage students to read a number of concordances and examine how different conjunctions fulfill various functions; modifying the texts from the corpus and presenting the simplified concordance lines on paper-based materials would be beneficial to learning since teachers can reduce lower-intermediate-level students' cognitive burden.

For the DDL vocabulary lessons, students were introduced to an online corpus query system, Sketch Engine (<http://www.sketchengine.co.uk/>). Sketch Engine was developed by Adam Kilgarriff, Pavel Smrz and David Tugwell. Sketch Engine contains a diverse set of functions, such as Word Sketch, Word Sketch Difference, Wordlist, Concordance and Thesaurus. The function 'Sketch Difference' shows a detailed summary page of the collocational differences and similarities between two similar words in various grammatical relations as shown below (see Figure 1).

decrease 617,362x			decline 459,367x		
objects of "decrease/decline"			subjects of "decrease/decline"		
likelihood	2424	0	concentration	704	214
risk	11104	55	percent	2070	1092
cost	10510	807	level	2339	1199
incidence	2321	110	consumption	649	453
consumption	3128	233	mortality	474	424
rate	9872	3620	rate	4212	5041
percent	2535	3958	revenue	781	1521
revenue	1217	2204	price	1420	2807
population	1496	3730	sale	962	2046
enrollment	259	1538	share	415	1051
offer	23	4318	population	1907	4954
invitation	0	2553	spokesman	0	800

Figure 1. A sample page of Sketch Difference of *decrease* and *decline*

A sample of the vocabulary lesson can be found in Appendix C. A series of guiding questions were provided to call students' attention to specific vocabulary. These questions guided students to note the data, form their hypothesis of the language, and become active participants in thinking and analysis.

Instruments

The background questionnaire consisted of three parts. The questions in the first two parts were adapted from Crosthwaite (2017) and Quinn (2015). The first section contained nine items designed to elicit information about the students' computer use habits and the experience of being taught a language through computerized means. In the second section, there were 10 items, which focused on students' use of the online tool/resources for self-directed language learning. The subjects were asked to rate how often they performed the described behavior on a Likert scale (with anchors of 'never; seldom; occasionally; frequently; almost always'), ranging from 1 (never) to 5 (almost always). The third part was sourced from Cohen, Oxford, and Chi's (2001) Language Style Survey. This survey has been adapted for use in a number of previous studies to measure how frequently learners use different strategies (e.g., Cohen et al., 2001; Mizumoto & Chujo, 2016; Oxford & Lee, 2007). In this study, seven items concerning how learners dealt with language rules were adapted in minor ways from the survey to measure the subjects' preferred learning styles by using a 5 point Likert scale. The first four items concerned deductive learning methods and the latter three items dealt with inductive learning methods. This background questionnaire is reliable since it is derived from existing questionnaires and many of the questions have been used before and lead to correct answers.

The main test instruments consisted of pre- and post-tests (see Appendices D and E). They were developed by the researcher. The reason the researcher chose to create tests instead of using the TOEIC test as pre- and post-tests was that the self-developed measures could focus on specific topics of grammar and vocabulary covered in the study, suiting the needs of the research. The pre-test contained 20 questions: 14 for grammar and six for vocabulary; each was worth five points, resulting in a total of 100 points. The number of questions for grammar and vocabulary in the pre-test was the same as that in the posttest. It should be noted that there were three items for Grammar Lessons 3 and 5. The other fourteen items were equally distributed to four grammar lessons and three vocabulary lessons. The

pre- and post-tests for grammar took the form of multiple-choice. The test questions were primarily based on sample sentences collected from COCA and adapted to meet the needs of the participants whose academic English proficiency is lower. To ensure the validity of the tests, two other experienced EFL instructors examined whether all the questions focused on the concepts learned through the lessons.

In order to gather students' feedback on their DDL experience, they were asked to fill out an evaluation questionnaire at the end of the study. The questionnaire was developed by examination of the literature, e.g., Yoon and Hirvela (2004), along with consideration of the specific use of the corpus in computer-based and paper-based activities for the study. The questionnaire sought to elicit students' responses to the following four dimensions: (1) their satisfaction with the user interface of the corpus (seven items), (2) the design of the learning materials and in-class activities (seven items), (3) the lessons' learning effects (11 items), and (4) the willingness to use the corpus for language learning in the future (four items). It consisted of both closed questions on a 5-point Likert scale (ranging from 1 being "don't agree at all" to 5 "strongly agree") and open questions that allowed students to freely express their thoughts. Lastly, one-on-one interviews were conducted. To enhance the ease of expression, the interviews were conducted in the participants' native language, Mandarin Chinese. Each interview, digitally recorded, lasted at least 20 minutes. The recording was then transcribed for the purposes of analysis.

Procedure

As mentioned above, the class met once a week for 100 minutes throughout the entire semester. For the purpose of this study, the second 50-minute period of every class was mainly devoted to DDL lessons. In the first week, the participants spent around 30 minutes taking the pre-test and filling out the background questionnaire and learning to search for the corpus data they needed. Next, in the nine-week treatment stage, students were required to learn one lesson in class for 50 minutes each week. The instructional approach adopted in the treatment to facilitate guided DDL followed the 'three I's (illustration, interaction, induction) proposed by Carter and McCarthy (1995). As Carter and McCarthy put it, "illustration" means looking at real data, "interaction" means discussing and sharing opinions and observations, and "induction" means making one's own rule for a particular feature, which "will be refined and honed as more and more

data is encountered” (p. 155). It is worth noting that a teacher-led whole class discussion was added as the final step of this inductive approach. The teacher can scaffold students’ understanding of the target learning points with providing hints and clearer guides and keep the discussion focused. Two flowcharts summarizing the steps of learning grammar and vocabulary are given in Figures 2 and 3.

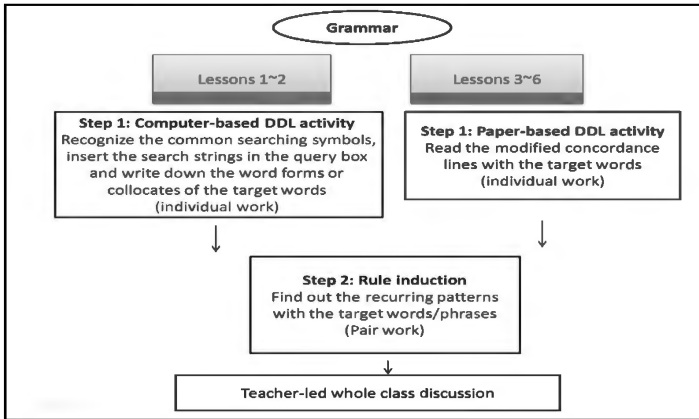


Figure 2. Steps of DDL grammar lessons

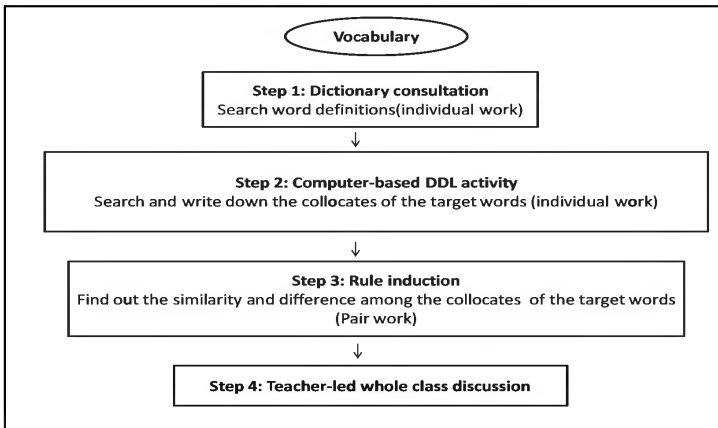


Figure 3. Steps of DDL vocabulary lessons

In computer-based grammar lessons, recognizing the common symbols used in the corpus query system to extract data was a crucial step. More specifically, for DDL Grammar Lesson 1, students had to learn how to use the wildcards or search symbols in COCA to retrieve the words with particular suffixes that determine word classes. After typing in different commands, the lists of words attached by various derivational suffixes could be obtained. Through noting and writing down the sample words on the lists, students' awareness that suffixes indicate the part of speech was stimulated. Finally, the teacher led a discussion to increase students' familiarity with the functions of morphemes. In the paper-based lessons, students individually observed modified concordance lines of target words/phrases and then analyzed the pattern of usage. To foster idea sharing, prior to the discussion, the students were required to post the results of their pair discussion onto Padlet (<https://padlet.com/dashboard>), a live online bulletin that allows people to express their thoughts on a common topic (see Figure 4). The teacher then took note of the mistakes or misconceptions in students' posts of data analysis. In the whole-class discussion, the teacher evaluated the appropriateness of the students' interpretations and corrected errors. Also, she offered the result of her analysis to students and explained the thinking and reasoning process (see Figure 5). It was believed that, by doing so, students gained a better understanding of what linguistic feature in the data they could pay attention to and how it was associated with the surrounding semantic and syntactic context. At the end of the class, students wrote a learning log to record observations and briefly summarize what they learned in the lesson. The purpose of this step was to help students review and synthesize the important information which was discussed in class.

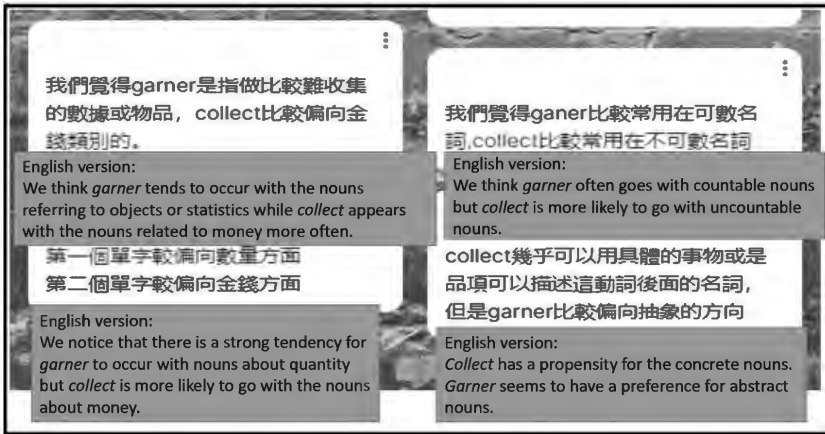


Figure 4. A snapshot of the Padlet wall with students’ analysis of the differences between the collocates of “garner” and “collect”

garner	collect
ATTENTION	DATA
SUPPORT	INFORMATION
PERCENT	MONEY Related to facts or details:
VOTES	SAMPLES data, information, evidence
MILLION	TAXES
AWARDS	EVIDENCE
PRAISE	YEARS
LOT	WATER
MEDIA	SIGNATURES Related to the money
ACCLAIM	THINGS people pay or get:
RESPECT	TAX tax, fee, benefits
REVIEWS	PEOPLE
INTEREST	FEES
SIGNATURES	DUST
NOMINATIONS	FOOD Related to substances or
	BENEFITS elements:
	ART water, dust
	THOUGHTS
	SALES

Figure 5. The semantic classification of the collocates of “garner” and “collect” suggested by the teacher

After the treatments, a posttest was immediately administered to evaluate the students’ knowledge of the target learning points. It should be noted that the posttests for both grammar and vocabulary lessons were conducted separately on different days. Next, an

evaluation questionnaire was then distributed for students to report on their opinions about their DDL learning experience. Finally, follow-up interviews were conducted for further clarification of students' responses to the feedback questionnaire. Interview questions can be seen in Appendix F.

Data Analysis

To investigate students' computer use habits, learning styles, and feedback on the DDL lessons, the data from the background questionnaire, perception survey and interview were coded and analyzed. For the analysis of this Likert-type data, the mean scores of the students' responses were computed relative to the five-point scale employed.

To examine whether the participants made significant progress in grammar and vocabulary after the experiment, paired-sample *t*-tests were run to compare participants' pre-test and post-test scores. The mean scores of the pre-test and post-test and the gain scores were calculated as well. In addition, we further explored which target learning point was more effectively acquired by the participants. Since the test questions of each learning point were equally distributed in the pre- and post- tests, the mean scores and gain scores were obtained and compared. Moreover, the Mann-Whitney Test (for small numbers of subjects) was performed to see whether there is a significant difference between students' gain score with respect to their preferred learning styles (inductive/deductive).

RESULTS AND DISCUSSION

A few students missed several lessons because of personal concerns, such as illness and death in the family. At the end of the study, a total of 40 students received all the treatments. Based on the results of the background questionnaire, around half of the students (51%) had experiences in learning English in class with CALL. Regarding the use of online resources, 65% of the students reported that they searched for information online upon encountering difficulties in grammar or vocabulary. Interestingly, about half of the class (45%) indicated frequent use of translation apps while only 11% of the students often took advantage of the resources in online

dictionaries. A great majority of the students (85%) never used an online thesaurus.

To answer the aforementioned research questions, the results of the tests, students' learning style preference, and their feedback after the treatments are presented below.

Learners' Performance in Pre- and Post-tests

Table 2 shows comprehensive comparisons of the performances before and after the interventions. The paired-sample t-test revealed significant differences between the pre- and post-tests ($p < 0.001$), with median effect sizes (Cohen's $d = .597$). The result suggests that the treatments in the study had positive efficacy. As far as students' performance in the grammar tests is concerned, a significant difference was found between pre- and post- tests ($p = 0.024$) (see Table 3). Similarly, as displayed in Table 4, students showed significant improvement on the posttest of vocabulary ($p = 0.014$).

Table 2

Paired-sample T-test Result of the Pre- and Post- test

Test	<i>n</i>	Mean	<i>SD</i>	<i>df</i>	<i>T</i>	<i>p</i>	Cohen's <i>d</i>
Pre-test	40	37.12	11.13	39	-3.777	< 0.001	.597
Post-test	40	45.50	13.43				

Table 3

Paired-sample T-tests for the Grammar Tests Scores

Grammar Test	<i>n</i>	Mean	<i>SD</i>	<i>df</i>	<i>T</i>	<i>p</i>	Cohen's <i>d</i>
Pre-test	40	24.13	9.53	39	-2.353	.024	.372
Post-test	40	28.25	11.46				

Table 4

Paired-sample T-tests for the Vocabulary Tests Scores

Vocabulary Test	<i>n</i>	Mean	<i>SD</i>	<i>df</i>	<i>T</i>	<i>p</i>	Cohen's <i>d</i>
Pre-test	40	13	7.40	39	-2.567	.014	.406
Post-test	40	17.25	8.00				

Further investigation was conducted to examine how well the students performed in each lesson and which lesson was more effectively learned by the participants with guided DDL. Table 5 shows the mean gain of the students across the nine lessons. For the lessons of discourse markers and time prepositions, the mean scores and gain scores were converted to a score from 0 to 10 as shown in the parentheses so that they were comparable with those of other lessons. Regarding the grammar lessons, the result revealed that the mean score of quantity phrases was the lowest in the pre-test; however, students obtained the highest gain score for it. It can be inferred that the students got the most benefits from this lesson. The learning unit for discourse markers, on the other hand, seems to have been difficult and much less effective for students. The mean scores in the pre- and post-test were low. It appears that the participants made almost no progress.

With regard to the results of the vocabulary test, the gain score of *demand*, *require*, and *request* was the highest (2.125), suggesting that students were more able to differentiate them from each other after the treatment. The mean score for the synonym set of *fare* and *fee* appeared to be the lowest (2.75) in the pre-test. The difference between it and the post-test mean score showed that students originally had far less understanding of the distinction between this pair of synonyms, but their awareness was raised after receiving instruction. Even though the gain score for the synonym set of *decrease* and *decline* was quite low, the highest mean scores in pre- and post-test reflected the fact that the participants had already had a good understanding of these two words in the beginning of the treatment.

Table 5

Changes in the Mean Scores for the Nine Lessons across Pre- and Post-tests

	Pre-test	Post-test	Gain score
Grammar Lessons			
1. Conjunctions	3.375	4.5	1.125
2. Word classes	3.75	4	0.25
3. Subject-verb agreement	3	4.125	1.125
4. Quantity phrases	2.75	4.25	1.5
5. Discourse markers	5.25 (3.5)	5.375 (3.58)	0.125 (0.08)
6. Time prepositions	5.875 (3.91)	6 (4)	0.125 (0.09)
Vocabulary Lessons			
7. Synonyms (<i>fare & fee</i>)	2.75	4.375	1.625
8. Synonyms (<i>demand, require, & request</i>)	4.125	6.25	2.125
9. Synonyms (<i>decrease & decline</i>)	6.25	6.625	0.375

Generally speaking, students seemed to be more familiar with the usage and features of the target items in grammar and vocabulary through DDL. There was a statistically significant difference between pre-test and post-test scores. Such results echo those of previous studies (Boulton, 2009; Lin, 2017; Liu, 2007; Smart, 2014) in which the DDL groups improved their grammar or vocabulary competency; nevertheless, it is worth noting that the post-test mean score was still less than 60 (out of 100). This suggests that the DDL lessons were challenging for students. More specifically, although students appeared to know more about the concepts covered in the lessons, it was not easy for them to choose correct answers in the post-test. It is likely that the level of difficulty of the lessons is higher than students' current proficiency level. Therefore, their ability to determine a correct answer was limited. Particularly, the mean and gain scores for discourse markers were low. This result is in line with Smart's (2012)

finding that not every grammar item showed a statistically significant improvement under the instruction of DDL. Lin's (2016) argument that DDL may be more suitable for teaching some items than others (p. 13) was also verified. Future research could further clarify whether modifying the lessons with fewer target items and more contextual clues and extending treatment duration improves the effectiveness of DDL for lower-intermediate English learners.

Comparison of Inductive and Deductive Learners' Performance in Pre- and Post-tests

The students' preferences for different learning styles were surveyed using the background questionnaire. Table 6 shows the means and standard deviations for the items for inductive and deductive learning styles.

Table 6

Descriptive Statistics for the Questionnaire Items of Inductive and Deductive Learning Styles

Learning styles	No.	Items	Mean	SD
Deductive	1	I like to go from general patterns to the specific examples when learning the target language.	3.63	0.73
	2	I like to start with rules and theories rather than specific examples.	3.76	0.78
	3	I like to begin with generalizations and then find examples that relate to them.	3.65	0.89
	4	I like to learn the rules of language indirectly by being exposed to many examples of grammatical structures.	3.46	0.90
Inductive	1	I like to discover underlying patterns by seeing many examples.	3.76	0.84
	2	I like to learn concrete examples first and then generalizable rules later.	3.46	0.85
	3	I like to figure out rules based on the way I see language forms being used over time.	3.68	0.72

The students who scored 16 to 20 points from the first four items were identified as inductive learners, and those who obtained 12 to 15 points in the last three items were labeled as deductive learners. It was found that nine students preferred learning inductively and 10 students strongly liked learning deductively. The Wilcoxon signed-rank test, a nonparametric statistic method suitable for the small size of the subjects, was employed to determine if significant differences existed between the total scores of the inductive and deductive learning groups on the pre- and post-tests. As Tables 7 and 8 show, the inductive learner category exhibited no significant difference between their pre- and post-test scores ($p = 0.283$), whereas the deductive learner category had a statistically significant difference between their pre- and post-test scores ($p = 0.020$).

Table 7

Comparison of the T Pre- and Post-test Scores of the Inductive Learners

		<i>N</i>	Mean rank	Sum of ranks	Z score	Asmp. (two-tailed)	Sig.
Post-Pre	Negative ranks	3 ^a	4.50	13.50	-	1.073 _b	.283
	Positive ranks	6 ^b	5.25	31.50			
	Ties	0 ^c					
	Total	9					

^a post-test < pre-test. ^b post-test > pre-test. ^c post-test = pre-test.

* $p < 0.05$.

Table 8

Comparison of the T Pre- and Post-test Scores of the Deductive Learners

		N	Mean rank	Sum of ranks	Z score	Asmp. Sig. (two-tailed)
Post-Pre	Negative ranks	1 ^a	1.50	1.50	-2.325 ^b	.020
	Positive ranks	7 ^b	4.93	34.50		
	Ties	2 ^c				
	Total	10				

^a post-test < pre-test. ^b post-test > pre-test. ^c post-test = pre-test.

* $p < 0.05$.

The Mann-Whitney test was then applied to compare the pre- and posttests of the two groups of learners (see Table 9). The results revealed no significant difference in the performance of inductive and deductive learners ($p = 0.604$ and 0.905 , respectively, > 0.05 , ns) in both tests. Inductive learners did not outperform deductive learners.

Table 9

Results of the Mann-Whitney Test for the Differences between the Pre- and Post-test Performance of Inductive and Deductive Learners

	Pre-test Total Score	Post-test Total Score
Mann-Whitney U	38.000	43.500
Z	-.583	-.125
Exact Sig [2*(one-tailed Sig.)	.604 ^b	.905 ^b

It appears that the results of learning styles seem to run counter-

intuitive to our tacit knowledge and are in contrast with some earlier studies reporting that inductive learners benefited more from DDL than deductive learners (Chan & Liou, 2005; Lee & Liou, 2003; Lewis, 2006). The inconsistency with previous research and the small number of the subjects identified as either type of learner is worth noting. Social desirability, which can be understood as research participants' tendency to bias their responses in a survey in order to appear in a more favorable light (Crowne and Marlowe, 1960) may cause the students to choose the responses which help achieve a better social impression of themselves. According to Table 6, each of the mean scores of the seven items concerning learning styles is above the neutral position on the Likert scale. It is likely that many students were motivated to 'show' that they usually learned English with BOTH inductive and deductive approaches in a flexible way. This factor can possibly explain why only a few subjects displayed a stronger tendency towards inductive or deductive learning and those labeled as inductive learners did not outperform those classified as deductive learners. In addition, the variables such as the differences in learning proficiency, aptitude, and the design of DDL tasks and instructions may also contribute to the unusual findings. On the other hand, the research result is consistent with Bulton (2009) and Mizumoto (2016), both of whom suggest that DDL is just not suitable for inductive learners. There is still a likelihood that guided inductive DDL may be beneficial for learners despite their learning styles. It is recommended that future research reduce the influence of biases and investigate how the above-mentioned variables may interact with each other, further illustrating the value of DDL to different types of learners.

Students' assessments of the treatments

After the posttest, all the students were asked to complete an evaluation questionnaire survey. Afterwards, nine students from the high-, middle-, and low-scoring groups on the post-test were invited to participate in a one-on-one interview.

The feedback from the evaluation questionnaire

Table 10 shows the statistics for the four dimensions in the questionnaire. The mean scores on the four investigated areas ranged

from 3.36 to 3.5, suggesting that, on average, students held neutral to less than positive attitudes toward a guided DDL experience.

Table 10

Descriptive Statistics for the Four Dimensions in the Evaluation Questionnaire

Dimensions	Mean	SD
1. The use of corpora	3.36	0.857
2. The design of the learning materials and activities	3.70	0.758
3. Learning effects	3.65	0.764
4. The future use of the corpora	3.38	0.730

For the first dimension, the usability of corpora, merely 41% of the participants agreed that COCA and Sketch Engine were easy to use. This rather neutral opinion may be attributed to the fact that students lacked the experience of using corpora before and did not feel they had enough hands-on experience to develop any clear or strong preference.

Regarding the design of the learning materials and activities, over half of the students agreed that the questions on the worksheets helped guide them in observing the corpus data and exploring the tendencies step by step. Of the respondents, 51% indicated that reading their classmates' analysis on Padlet and writing learning logs strengthened their learning impression.

Concerning learning effects, the students agreed that they learned more from the lessons on time prepositions (73%), discourse markers (68%), conjunctions (67%), and quantity phrases (65%); however, they were less certain about the effectiveness of lessons of word classes (45% expressing "agree" or "strongly agree") and synonyms (53% expressing "agree" or "strongly agree"). Such results seem to indicate a gap between the students' perceived usefulness of the lessons and their achievements. As seen before, the participants made more progress in the lessons of word classes and synonyms. It appears that the preference was not clearly reflected in improved performance. Future research should be conducted to investigate the relationship between perceived effectiveness of DDL lessons and learning output.

Concerning the ways of exploring data, while 56.1% of the students said that they liked to analyze the sample concordance lines on the worksheets, only 9.8% showed stronger intention to search and observe the collocates or sentences in corpora by themselves. It appears that the participants were more used to learning through teacher-directed materials.

In terms of the students' future use of corpora, it is interesting to see that even though about 50% of the students said they would recommend the application of corpora in a language class, only 39% expressed willingness to make actual use of corpora for their own study of grammar or vocabulary.

The feedback from the interview

For a hands-on experience with corpus tools, all of the interviewees indicated that they did not find it difficult to formulate the corpus queries in searching for collocates of synonyms and retrieving words with particular suffixes. Nevertheless, those from the low- and middle-scoring groups mentioned that they felt overwhelmed by the length and complexity of the result page of collocate lists. Feeling intimidated by the corpus data may explain why in the survey the respondents took a rather neutral attitude towards the usability of corpora.

For the learning activities, even though the majority of the interviewees agreed that idea sharing on Padlet and during pair discussion facilitated their learning process, some concerns were encountered. Two students from the low scoring group said that they felt anxious when they observed findings on the posts that differed from theirs. For the pair work, a student suggested that it would be better if the teacher could place one high achiever and one low achiever on the same team. Such heterogeneous grouping provides the high achiever an opportunity to offer a useful model in observing and interpreting the data and the lower achiever may feel more motivated to try. Regarding the activity of writing learning logs, some students reflected that it was helpful, whereas some did not. The following extract from one student's response shows that the activity did not get him thinking more actively:

The teacher usually shared with us her results of analysis at the end of the whole-class discussion. I thought the information she offered should be accurate. What I did was just copied the information from her results into my learning log.

Concerning learning effects, several respondents from the low- and middle-scoring groups reflected that the sample sentences provided in the paper-based lessons could be shortened and reduced. They also suggested that narrowing down the number of the target items in each lesson would be beneficial. Other students reported that more follow-up exercises or homework might improve long-term retention.

On the future use of corpora, all students recognized the value of consulting corpora for language learning and mentioned that implementing corpora in the classroom was desirable. Nevertheless, some of the interviewees expressed reservations about the use of corpora for self-learning. One respondent reflected that, in preparing for the TOEIC test, practicing mock tests and resorting to vocabulary books might be more helpful. Two other students expressed that they were more used to infer the meaning of unknown words through context clues for better comprehension and were not motivated to do further “research” for self-learning.

Judging from the above, students’ attitude towards guided inductive DDL appears neutral. The reason this finding does not resonate with those of previous DDL studies which obtained more positive responses from Taiwanese students can also be ascribed to the difference in participants’ academic major background and the course they are in. A number of studies with more successful results involved English major students in composition classes (e.g., Lin & Lee, 2017; Liou & Li, 2007). It is expected that English major students have higher motivation in exploring grammar and vocabulary in depth using approaches that require more time and effort. Engaging in more productive activities, like writing, may lead to better recognition and retention of the usage of the target items. Moreover, Taiwan’s educational settings should be taken into consideration when discussing student satisfaction with their learning experience. Kilickaya’s (2015) warning that the efficacy of DDL depends on different cultural contexts is noteworthy. As reported in previous research, many Taiwanese students are accustomed to learning through deductive and teacher-centered approaches (Lee, 2013; Lee & Kennedy, 2017; Lin & Lee, 2015; Smith, 2011), which may cause them to develop a less than positive attitude towards a more inductively oriented learning method. The washback effect, another crucial variable influencing students’ perception, was also identified in the present study. One interviewee reflected that he preferred practice tests in preparing for standardized English tests. The social pressure of having to pass the test to receive a college diploma may cause learners to focus only on subjects or activities that

directly contribute to their obtaining higher scores on the exam (Chen, 2002; Cheng, 1999; Chou, 2015; Shih, 2010), thereby diminishing their acceptance rate of DDL.

PEDAGOGICAL IMPLICATIONS AND CONCLUSION

The present study advances our understanding of the effect of guided inductive DDL on lower-intermediate-level technological university students' learning of particular vocabulary and grammatical points frequently occurring in TOEIC tests. It also gives us insight into how they perceive corpus-based DDL. The possible challenges encountered and students' admiration of this learning approach are revealed.

From this study's findings, some pedagogical implications can be drawn for EFL teachers. First, before introducing corpora into the class, teachers should raise students' awareness of the importance of being an independent language learner for long-term EFL learning. It is crucial that teachers convince students that, in contrast to other online resources, the data in corpora always tell a story – that is, they are always found in the living context of their usage rather than in a vacuum of meaning – and we can take good advantage of them for authentic language use. Teachers can arouse students' curiosity of word behavior in context by comparing the usages of the synonym sets they are familiar with. Showing the evolution of particular linguistic properties over time may also interest students because this reflects the trends of change registered within social history.

Second, teachers can demonstrate the thinking aloud of the steps of analysis to facilitate students' processing and classifying a huge amount of data. When students practice analyzing the data, teachers should provide more time and guidance in inferring patterns of word behavior and observing the concordance lines where the grammar items are used. Truncating the concordance lines may lessen learners' cognitive burden.

Third, to prevent students from becoming passive receivers of information, teachers are recommended to ask learners to complete more open out-of-class learning tasks before presenting the model data analysis to them. For lower-intermediate-level students, assigning homework such as constructing or collecting sentences with target words/structures, is feasible. It is important that teachers offer opportunities for students to engage in active thinking processes.

Finally, it is suggested that heterogeneous grouping be used when assigning students to a group of two or three for discussion. Students with higher proficiency may serve as facilitators to assist peers. Students with lower proficiency tend to lack confidence and be unsure of what they should think about when inferring rules of grammar and vocabulary so that the support they can derive from higher achievers may help them reduce their level of anxiety.

Although this study offers some insights into the application of DDL in a lower-intermediate-level class, a number of limitations should be noted. First of all, due to limited access to enough participants, there was no control group. Therefore, the intervention (guided DDL) was not compared to any other form of traditional teaching approach. Secondly, multiple-choice items are somewhat susceptible to guessing. The guessing factor may put some limits on the accuracy of the test results. Moreover, a delayed interval posttest was not incorporated into this study. It would be worthwhile to measure knowledge acquisition and retention over a sustained period of time. Lastly, students' production of the target grammar/vocabulary items can be collected and investigated to determine how well they apply the inferred rules into practice.

REFERENCES

- Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 51–64). Addison Wesley Longman.
- Aston, G. (1998, May 28-30). *Learning English with the British National Corpus* [Paper presentation]. 6th Jornada de Corpus, UFF, Barcelona. <https://godzilla.sslmit.unibo.it/~guy/barc.htm>
- Bloch, J. (2007). *Technologies in the second language composition classroom*. University of Michigan Press.
- Boulton, A. (2008) Looking for empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk (Ed.), *Corpus linguistics, computer tools, and applications: State of the art* (pp. 581–598). Peter Lang.
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37–54.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572.
- Boulton, A. (2012). Hands-on/hands-off: Alternative approaches to data-driven learning. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 152–168). Masaryk University Press
- Carter, R., & McCarthy, M. (1995). Grammar and spoken language. *Applied Linguistics*, 16 (2), 141–158.
- Chan, T., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18, 231–250.
- Chen, L.-M. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English* [Unpublished doctoral dissertation]. Ohio University.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15, 253–271.
- Chou, M.-H. (2015). Impacts of the test of English listening comprehension on students' English learning expectations in Taiwan. *Language, Culture and Curriculum*, 28, 191–208.
- Chujo, K., & Genung, M. (2005). *Utilizing the British National Corpus to analyze TOEIC tests: The quantification of vocabulary-usage levels and the extraction of characteristically used words*. The Institute for International Business Communication. http://www.iibc-global.org/library/default/english/lr/about/data/research/pdf/3_E.pdf
- Chujo, K., & Oghigian, K. (2007). Discovering grammar basics with parallel concordancing in the beginner-level EFL classroom. In *Proceedings of the 2007 Asia TEFL International Conference* (pp. 1–14).

- http://hanamizuki2010.sakura.ne.jp/public_html/data/2007%20Discovering%20grammar%20basic.pdf
- Chujo, K., Yokota, K., Hasegawa, S., & Nishigaki, C. (2013). Identifying the general English proficiency and distinct grammar proficiency of remedial learners. *Journal of the College of Industrial Technology, Nihon University*, 45, 43–54.
- Cobb, T. (1999). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research and Development*, 47(3), 15–31.
- Cohen, A. D., Oxford, R. L., & Chi, J. C. (2001). *Learning style survey: Assessing your own learning styles*. Center for Advanced Research on Language Acquisition.
http://carla.umn.edu/maxsa/documents/LearningStyleSurvey_MAXSA_IG.pdf
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129–147.
- Cresswell, A. (2007). Getting to ‘know’ connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267–287). Rodopi.
- Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning*, 30(6), 447–473.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349–354.
- Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, 26(2), 128–146.
- Johns, T. F. (1988). Whence and whither classroom concordancing? In T. Bongaerts, T. van Els, & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9–27). Foris.
- Keck, C. M. (2004). Book review: Corpus linguistics and language teaching research: Bridging the gap. *Language Teaching Research*, 8(1), 83–109.
- Kilickaya, F. (2015). Computer-based grammar instruction in an EFL context: Improving the effectiveness of teaching adverbial clauses. *Computer Assisted Language Learning*, 28, 325–340.
- Koosha, M., & Jafarpour, A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal Quarterly*, 8(4), 192–209.
- Lee, C.-Y., & Liou, H.-C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching and Learning*, 27, 35–56.
- Lee, H.-C. (2013). Investigating the effects of student learning of English using COL approach based on situational theories. *Computers in Human Behavior*,

29, 2211–2217.

- Lee, J. C.-K., & Kennedy, K. J. (Eds.). (2017). *Theorizing teaching and learning in Asia and Europe: A conversation between Chinese curriculum and European didactics*. Routledge.
- Lewis, J. (2006). *Connecting corpora to learner style: To what extent is the effectiveness of an online corpus-based approach to grammar learning dependent on whether students prefer to learn grammar deductively or inductively?* [Unpublished master's thesis]. Universidade do Porto.
- Lin, M. H., & Lee, J.-Y. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69, 264–279.
- Lin, M. H. (2016). Effects of corpus-aided language learning in the EFL grammar classroom: A case study of students' learning attitudes and teachers' perceptions in Taiwan. *TESOL Quarterly*, 50, 871–893.
- Lin, M. H. & Lee, J.-Y. (2017). Pedagogical suitability of data-driven learning in EFL grammar classes: A case study of Taiwanese students. *Language Teaching Research*, 23(5), 541–561.
- Lindstromberg, S., & Boers, F. (2008). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, 29, 200–222.
- Lo, G. Y.-H. (2011). ESP versus EGP: A case study of an ESP program for vocational high school students of tourism. *Taiwan International ESP Journal*, 3(2), 71–100.
- Mizumoto, A., & Chujo, K. (2016). Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System*, 61, 55–64.
- Murphy, B. (1996). Computer corpora and vocabulary study. *Language Learning Journal*, 13, 53–57.
- Oxford, R. L., & Lee, K. R. (2007). L2 grammar strategies: The second Cinderella and beyond. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: 30 years of research and practice* (pp. 117–139). Oxford University Press.
- Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18, 75–94.
- Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20, 113–138.
- Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, 69(2), 165–177
- Shih, C.-M. (2010). The washback of the general English proficiency test on university policies: A Taiwan case study. *Language Teaching Research*, 20(7), 113–138.
- Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, 26, 184–201.
- Smith, S. (2011). Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning*, 24(4), 291–316.
- Sun, Y.-C., & Wang, L.-Y. (2003). Concordancers in the EFL classroom: Cognitive

- approaches and collocation difficulty. *Computer Assisted Language Learning*, 16, 83–94.
- Thurstun, J., & Candlin, C. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, 17(3), 267–280.
- Tsai, K.-J. (2019). Corpora and dictionaries as learning aids: Inductive versus deductive approaches to constructing vocabulary knowledge. *Computer Assisted Language Learning*, 32(3), 1–22.
- Tsao, C.-H. (2008, May 6). *English-learning motivation and needs analysis: A case study of technological university students in Taiwan*. [Paper presentation]. The Basic Research Conference of Chinese Military Academy on its 84th school anniversary. Kaohsiung, Taiwan.
- Tseng, Y., & Liou, H. (2006) The effects of online conjunction materials on college EFL students' writing. *System*, 34, 270–283.
- Turnbull, J., & Burston, J. (1998). Towards independent concordance work for students: Lessons from a case study. *ON-CALL*, 12(2), 10–21.
- Vyatkina, N. (2016a). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, 28(2), 207–226.
- Vyatkina, N. (2016b). Data-driven learning of collocations: learner performance, proficiency and perceptions. *Language Learning & Technology*, 20(3), 159–179.
- Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43(1), 55–77.
- Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, 32, 259–276.
- Wu, W.-C. (2009). Criteria for establishing an authentic EFL learning environment in Taiwan. *Asian EFL Journal*, 11(3), 159–189.
- Yeh, Y.-L. (2003). Vocabulary learning with a concordancer for EFL college students. In English Teachers' Association (Ed.), *Proceedings of the 2003 International Conference on English Teaching and Learning in the Republic of China* (pp. 467–476). The Crane Publishing Company.
- Yeh, Y., Liou, H.-C., & Li, H.-Y. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20(2), 131–152.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing*, 13(4), 257–283.
- Yoon, H., & Jo, J. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, 18(1), 96–117.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers of the *Taiwan Journal of TESOL* for their valuable suggestions on the previous versions of this paper. This paper was sponsored by the Ministry of Science and Technology, Taipei, Taiwan [Grant 109-2410-H-167 -015].

CORRESPONDENCE

Yen-Yu Lin, Language Center, National Chin-Yi University of Technology, Taichung, Taiwan
Email: yylin@ncut.edu.tw

PUBLISHING RECORD

Manuscript received: August 5, 2021; Revision received: April 12, 2022; Manuscript accepted: April 16, 2022.

APPENDIX

Appendix A. Data-driven Learning Material for Word Classes

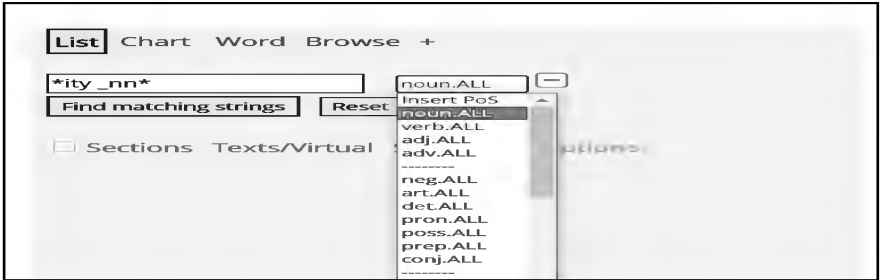
✧ Derivational morphemes are used to change the grammatical categories of words. For example, the derivational morpheme *-ly* transforms the adjective *fluent* into the adverb *fluently*. The morpheme *-is* can be used to change the verb *hypothesize* into the noun *hypothesis*.

- Below are some common derivational morphemes which can change a class of words to another class. Please identify the categories of derivational morphemes (noun-forming, verb-forming, adjective-forming, and adverb-forming morphemes) they belong to.

1. <i>-en</i>	11. <i>-ist</i>	Noun-forming morpheme(s):
2. <i>-ity</i>	12. <i>-al</i>	
3. <i>-ous</i>	13. <i>-sion</i>	
4. <i>-ify</i>	14. <i>-ive</i>	Verb-forming morpheme(s):
5. <i>-ful</i>	15. <i>-ence</i>	
6. <i>-ment</i>	16. <i>-ic</i>	
7. <i>-tion</i>	17. <i>-ish</i>	Adjective-forming morpheme(s):
8. <i>-ize</i>	18. <i>-er</i>	
9. <i>-ary</i>	19. <i>-able</i>	Adverb-forming morpheme(s):
10. <i>-ness</i>	20. <i>-ly</i>	

- Please search each morpheme in COCA to find **THREE** examples with it. The wildcard search is useful. For example, you can enter “**ity*” into the word query box to find the list of

words ending with *-ity*. To be more specific about your search morpheme, you can specify part of speech categories with POS tags. The search page of *-ity* in COCA using the List Function is shown below for your reference.



Derivational morphemes	Examples	Derivational morphemes	Examples
1. <i>-en</i>		11. <i>-ist</i>	
2. <i>-ity</i>		12. <i>-al</i>	
3. <i>-ous</i>		13. <i>-sion</i>	
4. <i>-ify</i>		14. <i>-ive</i>	
5. <i>-ful</i>		15. <i>-ence</i>	
6. <i>-ment</i>		16. <i>-ic</i>	
7. <i>-tion</i>		17. <i>-ish</i>	
8. <i>-ize</i>		18. <i>-er</i>	
9. <i>-ary</i>		19. <i>-able</i>	
10. <i>-ness</i>		20. <i>-ly</i>	

- **Exercise:** Please read the following sentences and identify the part of speech of the missing word which should be filled in the blank. Choose the best answer to complete the sentences.

<p>(1). The catalog for the gallery contains an accurate _____ of each piece of artwork.</p> <p>(A) described (B) description (C) descriptive (D) describes</p>	<p>(2). The participants will be judged _____ on _____ performance, and the winners will be announced later in the awards ceremony.</p> <p>(A) athletically (B) athletic (C) athletes (D) athlete</p>
<p>(3). Marketers believed that if the packaging were more _____ colored, consumers might pay more attention to the product.</p> <p>(A) Variously (B) Vary (C) Various (D) Variety</p>	<p>(4). On next week's radio program, our host will interview Kristen Dabney about her time working as an _____ for a UN official.</p> <p>(A) interpretation (B) interpret (C) interpreting (D) interpreter</p>

Appendix B. Data-driven Learning Material for Conjunctions

- Conjunctions are used to show the relationship between the ideas in the clauses.

1. **Even if** an applicant was not selected on stage (在台上), his/her name might be drawn at random at a later date for an audition (試鏡會).
2. **Even if** the caller has your full name and more, don't confirm any of it if you don't know for sure who the caller is.
3. **Even if** the student does poorly on AP exam, that score will not affect his/her high school course grade.
4. **Even though** I always lose, I love gambling.
5. **Even though** this book makes me cry, it's not sad at all.
6. Buckley said that **even though** she has added various contemporary (當代的) songs to her new production, she does not leave out classics from her albums.
7. **While** Toyota scored four wins in the 10 vehicle categories, Japanese automaker Subaru was named as the best brand in the industry.
8. **While** I feel I am benefiting from this experience, I miss my old job and the more comfortable lifestyle.
9. **While** other kids played with dolls, Vicky had strong love for doll.
10. **Despite** his controversial (爭議性的) nature (特質), he is widely considered to be a musical genius.
11. **Despite** his cancer diagnosis, Jeff hasn't let it change his lifestyle.
12. **Despite** many obstacles, India's young generation displays an

extraordinary amount of talent and energy.

13. A foreign accent is generally not being frowned (皺眉) upon by the locals in the United States, **provided that** you speak good English and your accent is not too strong.
14. Sending aged parents to nursing homes (療養院) is not an immoral thing to do, **provided that** the children take the time to visit them frequently.
15. Persons with reserved accessible seating will be permitted to enter the studio before general audience members **provided that** they arrive 30 minutes before the taping is scheduled to start.

1. Please take a look at the fifteen example sentences above and write down the meaning of the five conjunctions (*even though, even if, while, despite, provided that*).

even though: _____; *even if*: _____; *while*: _____; *despite*: _____; *provided that*: _____

2. Please identify the functions of the conjunctions.

(1) Expressing cause and effect	(2) Expressing condition (條件假設)
(3) Expressing contrast	

even though: _____; *even if*: _____; *while*: _____; *despite*: _____; *provided that*: _____

3. In what condition are each of the conjunctions used? What are the differences between them?

- **Exercise:** Please read the following sentences and choose the best answer to complete the sentences.

<p>(1) _____ anything you've done and what we've been through, I will always love you.</p> <p>(A) Even though (B) Even if (C) Despite (D) While</p>	<p>(2) _____ God took away the person we love, we still praise God.</p> <p>(A) Even though (B) Provided that (C) Despite (D) While</p>
<p>(3). Some plants may be used for antifertility purposes in some places _____ in other places they are used for fertility purposes.</p> <p>(A) provided that (B) even if (C) despite (A) while</p>	<p>(4). Such technology can transform learning in positive ways, _____ important classroom changes occur to support these new models of teaching.</p> <p>(A) even though (B) even if (C) provided that (D) while</p>

Appendix C. Data-driven Learning Material for *Fare* and *Fee*

- **Word Definitions**: Please consult *Longman Dictionary of Contemporary English* online for the word definitions of *fare* and *fee*.

1. *fare* (n.): _____

2. *fee* (n.): _____

- **Collocations**: Please make use of Sketch Diff to explore the significant collocates of *fare* and *fee*.

1-1. Please write down the modifiers of *fare* (n.).

1-2. Please write down the modifiers of *fee* (n.).

1-3. Please write down any shared linguistic feature of the modifiers of *fare* (n.) and *fee* (n.) based on your observation.

1-4. Please write down the nouns modified by *fare* (n.).

1-5. Please write down the nouns modified by *fee* (n.).

1-6. Please write down any shared linguistic feature of the nouns modified by *fare* (n.) and *fee* (n.) based on your observation.

- **Exercise:** Decide whether *fare* or *fee* fills in each blank the best.
1. Late Tuesday, Southwest Airlines raised all of its round-trip _____s by \$10. (fares)
 2. The research looks at the cheapest train _____ for a family of four from cities and county towns to their nearest seaside resort. (fare)
 3. You need to pay your membership _____ and submit your membership form before Sunday 14th July 2013. (fees)
 4. Beginning Jan. 19, the registration _____ will be \$995 per person. (fee)
 5. Elevators and _____ gates designated for persons with disabilities may also be used by airline passengers with luggage. (fare)
 6. In order to apply for a student _____ waiver, please send a copy of your student card and a short motivation to clin26org@googlegroups.com. (fee)

Appendix D. Pretest

1. _____ an ankle (膝蓋) injury, the baseball player participated in the last game of the season.
 - A. Despite
 - B. Even if
 - C. Even though
 - D. While
2. _____ some of us didn't agree with it, he made his decision.
 - A. Despite
 - B. Even though
 - C. While
 - D. Even if
3. She is a certified (認證的) _____ in infectious disease.
 - A. especially
 - B. specialty
 - C. specialist
 - D. special
4. We have a strategy that could help us _____ a suspect (嫌疑犯).
 - A. identification
 - B. identify
 - C. identifies
 - D. identified
5. Either the manager or the artist _____ the right to end the agreement.
 - A. has
 - B. have
 - C. is
 - D. are
6. Tom as well as his parents _____ fond of going swimming.
 - A. is
 - B. are
 - C. be
 - D. will be
7. Maryland (馬里蘭州) has _____ universities and colleges that offer postgraduate (研究所的) courses (課程).
 - A. a large number of
 - B. a large amount of
 - C. a great deal of

8. That movie fostered (促進) excitement and _____ interest about the exploration of space (太空).
- A. a large number of
 - B. a large amount of
 - C. a great deal of
9. They decided to cancel the flight _____ the heavy rain.
- A. consequently
 - B. namely
 - C. owing to
 - D. because
10. There are two rooms to paint, _____, the kitchen and the living room.
- A. consequently
 - B. namely
 - C. owing to
 - D. because
11. One of the cause of global warming is the emission (釋放) of greenhouse (溫室) gasses (氣體) such as carbon dioxide. _____, in recent years, many scientists have been trying to develop low-carbon technologies to meet our energy needs.
- A. Consequently
 - B. Namely
 - C. Owing to
 - D. Because
12. I promised your father that I would protect both of you _____ you didn't need me.
- A. by
 - B. from
 - C. until
 - D. during
13. A research report this month forecast (預言) China's Internet population could reach (達到) 190 million _____ 2010.
- A. by
 - B. from
 - C. until
 - D. during
14. There were 3,000 journalists (記者) in Pakistan _____ the war in Afghanistan (阿富汗).
- A. by
 - B. from
 - C. until
 - D. during

15. We serve low-income families who _____ assistance (協助) with immigration (移民) legal services.
A. demand
B. require
C. request
16. The protesters (示威抗議者) _____ the resignation (辭職下台) of the president and that he is put into court (審判).
A. demanded
B. required
C. requested
17. The registration (註冊) _____ is \$125, payable by cash or check (支票) upon arrival of the meeting.
A. fee
B. fare
18. This rate (費率) includes all train tax, hotel tax, round-trip train _____, and Grand Canyon shuttle bus transportation.
A. fee
B. fare
19. The recent stock (股票) market _____ from its high point has caused concerns (擔憂) for many investors (投資者).
A. decline
B. decrease
20. This important information will help parents to prevent or _____ anxiety in young children. *單選
A. decline
B. decrease

Appendix E. Posttest

Grammar:

1. The overall situation is good, _____ a few minor (輕微的) problems.
A. despite
B. even though
C. while
D. even if
2. _____ the test was pretty easy, he still made several careless mistakes.
A. Despite
B. Even though
C. While
D. Even if
3. At the age of thirty-two, Peter decided to become a _____.
A. psychology
B. psychological
C. psychologist
D. psychologically
4. Girls should _____ their parents before getting an abortion (墮胎).
A. notify
B. notification
C. notified
D. notifies
5. Either Monika or her friend _____ responsible for the accident.
A. have been
B. has been
C. has
D. have
6. The manager as well as his associates (合夥人) _____ going to prison.
A. is
B. are
C. be
D. will

7. Teachers should check whether students have understood the subject matter by asking _____ questions.
 - A. a large number of
 - B. a large amount of
 - C. a great deal of
8. Although I have _____ respect for my colleagues (同事), I believe my ideas are better and smarter for our company.
 - A. a large number of
 - B. a large amount of
 - C. a great deal of
9. _____ his careless driving, he had a terrible accident.
 - A. Consequently
 - B. Namely
 - C. Owing to
 - D. Because
10. Producer prices in April were higher due to increased costs (成本) for food—_____, meat, eggs and dairy (奶類) products.
 - A. consequently
 - B. namely
 - C. owing to
 - D. because
11. Basketball teams discovered that bigger players had an easier time scoring (得分). _____, they began adding more bigger players.
 - A. Consequently
 - B. Namely
 - C. Owing to
 - D. Because
12. She lived at home _____ she turned 18 and then bought her own place in the Hollywood Hills.
 - A. by
 - B. from
 - C. until
 - D. during
13. Application (申請) materials must be e-mailed _____ the deadline (截止日期) date of November 16, 2012.
 - A. by
 - B. from
 - C. until
 - D. during

14. He spent six years in the Air Force (空軍) stationed (駐紮) in Korea _____ the Vietnam (越南) War. He was proud of his service to his country.
- A. by
 - B. from
 - C. until
 - D. during

Vocabulary:

15. Greece was the first Eurozone country to _____ official financial assistance (金融援助) in May 2010.
- A. demand
 - B. require
 - C. request
16. February 20 the protesters (示威抗議者) _____ a political change and the end of the corruption(貪腐).
- A. demanded
 - B. required
 - C. requested
17. The registration (註冊) _____ of 1000 USD include attendance, lunches, and coffee breaks.
- A. fee
 - B. fare
18. The cost to students includes: train _____ and museum entrance.
- A. fee
 - B. fare
19. The company's stock (股票) has _____ 70 percent over the last few months, partly due to a steep (劇烈的) drop (下跌) in the stock prices of its two YieldCos (太陽能營運發電商).
- A. declined
 - B. decreased
20. A number of studies have shown that regular exercise can _____ anxiety and unhappiness.
- A. declined
 - B. decreased

Appendix F. Interview Questions

1. What did you think when you heard that you would learn grammar concepts and word usage through data-driven learning?
2. What did you think about the introduction to corpora? Were the instructions clear? Were the introductory exercises comprehensible?
3. About the corpus work in class:
 - (a) Did you encounter any technical problems?
 - (b) Did you usually understand the instructions well enough to carry out the exercises?
 - (c) How well did you manage to formulate your own corpus queries?
4. Did you manage to draw conclusions about grammatical rules and the usage patterns of vocabulary based on the corpus results?
5. What did you think about the effectiveness of pair work? Did you feel that the workload was fairly divided between you?
6. Did you think reading other classmates' posts on Padlet facilitated the process of generating linguistic rules and conclusion with your partner/on your own?
7. Did you think writing a learning log helped your review and synthesize what you learned in each lesson?
8. What do you think about learning grammar/vocabulary in this way (by drawing conclusions based on corpus results) compared to learning in the more traditional way (from a grammar book)?
9. Do you think the corpus will be of use to you in the future?
10. Any other comments?