



## Increasing Lexical Awareness through Data-Driven Learning: Polysemy in EFL Pedagogy

Inseul Hwang · Minyoung Cho (Korea University)



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

Received: August 12, 2022

Revised: October 16, 2022

Accepted: October 31, 2022

Inseul Hwang (1<sup>st</sup> author)  
Graduate Student, M.A. Dept.  
of English Language and  
Literature, Korea University  
Tel: 02-3290-1980  
E-mail: [inseul122@korea.ac.kr](mailto:inseul122@korea.ac.kr)

Minyoung Cho (corresponding  
author)  
Associate Professor, Dept. of  
English Language and  
Literature, Korea University  
Tel: 02-3290-1980  
E-mail: [mycho27@korea.ac.kr](mailto:mycho27@korea.ac.kr)

### ABSTRACT

Hwang, Inseul and Minyoung Cho 2022. Increasing lexical awareness through data-driven learning: Polysemy in EFL pedagogy. *Korean Journal of English Language and Linguistics* 22, 1116-1132.

Many L2 students struggle to use vocabulary effectively and flexibly in a variety of situations. For learners to recognize words' different meanings in diverse circumstances, they need to be exposed to diverse contexts that manifest various meanings of the word. Traditional vocabulary learning (TL) tools such as dictionaries and online translators, however, are limited in raising the learner's awareness of meanings beyond the first two or three of polysemy, and data-driven learning (DDL) has been recommended as an alternative teaching method. This study, thus, explores the efficacy of data-driven learning (DDL) using concordance in acquiring polysemous word knowledge. Fifty middle school students were assigned to either DDL or TL condition and had vocabulary learning sessions for the respective condition. The participants' acquisition of polysemous word knowledge was examined in terms of their explicit and implicit knowledge of the words in the immediate posttest and the delayed posttest. The results indicated that the DDL group outperformed the TL condition in only the implicit knowledge test in the immediate post-test. However, the delayed post-test showed no significant difference in retention of the knowledge. The findings are discussed in light of the use of DDL in vocabulary learning and the use of corpus as a pedagogical tool.

### KEYWORDS

polysemy, data-driven learning, concordances

## 1. Introduction

The meaning of a word can be transformed, extended, or derived depending on its context. Polysemous words possess diverse meanings that are used in different contexts. However, it is often a challenge for L2 learners to utilize vocabulary effectively and flexibly for social, academic, and professional goals (Boulton and De Cock 2017) since they only know one of the representative meanings of a word. It seems important for language pedagogy to help L2 learners recognize multiple meanings of a word and develop their cognitive and metacognitive knowledge of a word (Stahl and Nagy 2006).

When students are faced with unknown words, one of the common ways to solve this problem is to search for words in dictionaries or use online translators, which do not seem as effective for their learning. For a dictionary, students tend to look at the primary meaning and rarely refer to the other entries (Jin and Deifell 2013), and online translators only give users a translation of one of the target word's meanings. According to Abou-Khalil (2019), when language learners use a translation tool to search for a single word, they lack access to the context needed to understand the word, which limits learners' understanding of the use of the words outside of the particular context. This approach is particularly problematic for polysemous words and learning polysemy necessitates copious contextualized input to recognize the different usage of a certain word.

Using concordance lines is one of the ways to expose learners to diverse authentic language input. Concordance lists a number of instances of a word in a KWIC (keywords in context) format, which helps learners notice the target word within the neighboring contexts. According to Schmidt's (1990) noticing hypothesis, noticing is the essential condition for converting input into intake. Therefore, the tool used in DDL appears to create an input flood that triggers notice, thereby initiating the transformation of input into the intake and thus learning.

Despite the further potential advantages of DDL, however, previous DDL studies have mostly concentrated on the learning of lexico-grammatical components of vocabulary (Boulton and Cobb 2017) and were only concerned with knowing the one conventional meaning of the word, focusing on the breadth of learning rather than the depth of vocabulary knowledge. Therefore, polysemy is an unexplored field that has not been applied in DDL yet, whose knowledge, nonetheless, is critical for language use. Furthermore, previous studies on DDL focused on whether DDL assists language learning, rather than how DDL affects the acquisition of different types of knowledge (e.g., implicit vs. explicit). Addressing this research gap, the present study investigates whether data-driven learning (DDL) with concordance lines aids in the acquisition of polysemous words in the measurement of implicit and explicit knowledge tests.

## 2. Literature Review

### 2.1 Vocabulary Acquisition in SLA

Knowing a word means more than understanding its meaning. The knowledge includes not only knowing its meaning but understanding its use, synonyms, compositions, and etymologies. Thus, vocabulary knowledge refers to both breadth and depth dimensions. Vocabulary breadth represents how many words are learned, or the number of terms for which a learner has at least a cursory comprehension. Depth of vocabulary knowledge, on the other hand, indicates a student's level of understanding of various aspects of a given word – for example, how well the words are understood (Li and Kirbt 2015) as well as their semantic associations, use in collocations, or idioms, and multiple meanings (Nation 2001).

As vocabulary knowledge is extensive, how to assist language learners in developing vocabulary knowledge has been an issue in language teaching, and the effectiveness of explicit and implicit approaches has been discussed. Schmitt (2010) asserted that explicit vocabulary learning is advantageous in robust and faster learning, and better retention, focusing on selective vocabulary learning. By contrast, implicit learning is beneficial for the acquisition of difficult words, for filling in contextual word knowledge that is challenging to learn explicitly, and for recycling words already known, while simultaneously improving other language skills (e.g., reading) (p. 40). Similarly, vocabulary can be acquired intentionally and incidentally. Intentional vocabulary learning makes “students engage in activities that focus attention on vocabulary” (DeCarrico 2001, p. 286), whereas incidental vocabulary learning is used to describe “learning that occurs when the mind is focused elsewhere, such as on understanding a text or using language for communicative purposes” (Celce-Murcia 2001, p. 289).

While both intentional and incidental learning are complementary in terms of quantity and quality of vocabulary acquisition (Schmitt, 2000), many studies highlight the significance of inferring the meanings of words from context (e.g., Itzes 1991, Nagy 1997, Schouten-van Parreren 1985), which allows room for “incidental” learning. The ability to comprehend a word in its metalinguistic context increases awareness of it in its comprehensive sense (Nation 2001), which implies becoming more flexible with the languages one knows and developing a deep understanding of vocabulary rather than a superficial understanding of terminology, both of which are beneficial through incidental learning.

As such, the effectiveness of different types of vocabulary learning has been discussed (e.g., Baddeley 1997, Craik and Lockhart 1972), existing studies focused on vocabulary knowledge as the ability to map a single lexical unit to a single meaning (Booton et al. 2022) and paid little attention to acquiring the depth of vocabulary. According to Fang (2014), remembering and reciting words by rote will not help students master extended meanings as it lacks context. Instead, increasing lexical awareness can help learners develop a positive attitude toward vocabulary learning, improve their learning skills, develop a long-term interest in the analysis of vocabulary, and understand how vocabulary is used for a wide range of purposes (Nation 2008).

## 2.2 Data-Driven Learning for Vocabulary Acquisition

Corpus linguistics has provided new tools, approaches, and resources for language instruction (Quan 2016). Concordance generators, concordancers, and concordancing software, to mention a few words, are used to analyze natural language data that are saved electronically (Conrad 2005, Tribble and Jones 1990). Word concordance is one of the useful tools for language learning as concordance lines show co-occurrences of keywords and their context. The target words are always presented in a KWIC (keywords in context) format (see Figure 1), which puts all the instances of the same word together and emphasizes the target word, saving time and assisting learners in focusing their attention. This enables the target words to be intentionally taught to the learners. Concordance lines and other corpus features can reveal a wealth of information about a lexical item, including its syntax, collocational information, different meanings, frequent settings in which it appears, forms, and how it is applied in real life (Nation and Chung 2009).

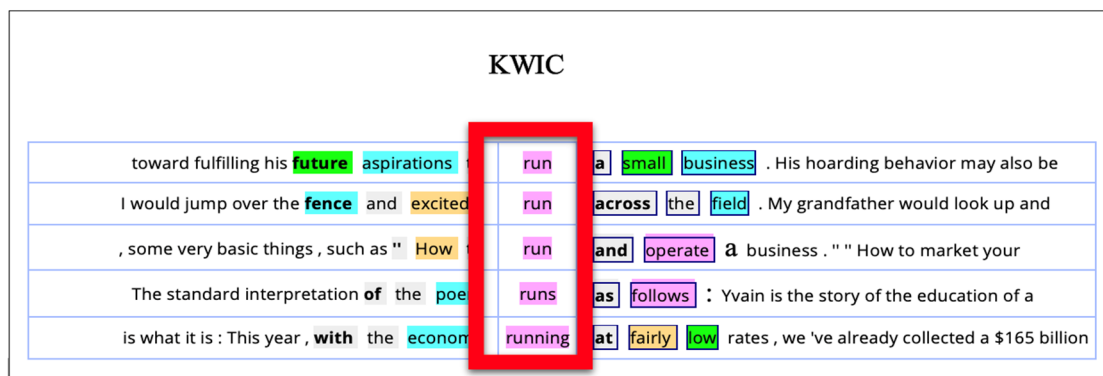


Figure 1. KWIC Format in Concordance Lines

As a model of corpus-assisted learning in corpus linguistics, Johns (1991) devised DDL in which students explore “real and authentic language data” to find linguistic patterns and regularities in concordance lines inductively (Gabrielatos 2005). Learners uncover language regularities and generalize about linguistic phenomena based on observation, analysis, induction, and conclusion. This method of education is also known as discovery learning (Bernardini 2000, 2004) or characterized by “autonomic learning”, “authentic language input”, “self-discovery”, and “bottom-up inductive learning”.

Previous studies examined the effectiveness of DDL on vocabulary learning by comparing concordance-based vocabulary instruction with traditional instructions such as memorizing dictionary definitions, synonyms, and practicing fill-in-the-blank exercises. Cobb (1999) examined students’ vocabulary learning outcomes when viewing multiple concordance lines in comparison to using a word list and dictionary. The findings showed that while using a word list and dictionary resulted in more gains in definitional knowledge in the short term, students were not able to retain the acquired knowledge and apply it to new contexts. However, viewing concordance lines helped acquire both definitional knowledge and the ability to transfer comprehension in novel situations. These results are consistent with Balunda’s (2009) findings that students’ vocabulary knowledge, retention, and transferable word acquisition can be enhanced when learning occurs through concordance lines compared to dictionary-based vocabulary teaching.

Furthermore, the benefits of contextualized vocabulary learning compared to rote memorization appeared in other domains. Presenting contexts enhances autonomous learning and teaches proper language usage including collocation, colligation, and semantic prosody (Jiao 2012, Liya 2021). Learners have also shown positive attitudes toward DDL activities (Aşık et al. 2016, Liya 2021, Yılmaz and Soruç 2014). For example, Aşık et al. (2016) explored 126 Turkish EFL learners’ perceptions of DDL in terms of lexical awareness and development. According to their survey and interview data, participants reported a noticeable improvement in their awareness of synonyms and collocations, though there was a little improvement regarding frequency, idioms, and learning strategies. Similarly, in a study of vocabulary teaching using corpora in comparison to online dictionary use by Jiao (2012), the group who received DDL remembered new words significantly more than a group of online dictionary users. Additionally, students appreciated corpus work to help autonomous learning and proper language usage.

### 2.3. Polysemy

Polysemy is a lexical phenomenon where a single word has several meanings (Cruse 2000, Ravin 2000), and often the meanings are related. (Kieffer and Logan 2017). Lamb, for example, can refer to both the animal and the meat of that animal, and the two senses have a lot in common semantically (Lyons 1995). Due to the interrelatedness of meanings, polysemy has been proposed to be more difficult to cope with than homonymy, which has multiple unrelated meanings (e.g., the pen for writing and the pen for pigs) (Kang 1992).

Polysemy consists of its prototype and extended definitions. The prototype refers to its literal or primary meanings from which metaphorical or figurative meanings have been created over time. Fillmore and Atkins (2000) described the characteristics of a polysemous word: (a) multiple interpretations share a common origin, (b) the connections between these senses form a network, and (c) understanding the “inner” one contributes to comprehending the “outer” one.

According to McCarthy, “the power of the central meaning and its transferability across languages may be important features in how words are learned and how different senses are felt to relate to the center or periphery of a word’s meaning potential” (1990, p. 25). This suggests that one might retrieve the core meaning and get a sense of what the speaker is talking about when meeting a polysemous word with no biasing context. Later information would add to this core by adding other features, culminating in the precise, intended sense. Verspoor and Lowie (2003) also contend that the core meaning of the polysemous words would help acquire the peripheral meaning of the words. Learners retain abstract, figurative senses of polysemous terms better when given core senses as cues because supplying a core sense helps learners construct a “precise elaboration.” As a result, students should be able to infer any following senses based on the defining aspects, starting with the basic meaning.

This is implicative for vocabulary acquisition because previous findings suggest an easier and more economical way of acquiring additional meanings through a semantic network of polysemous words (Verspoor and Lowie 2003). Considering the connections of word senses in polysemy, it seems necessary to provide contexts for diverse use of polysemous words, as contextualized vocabulary use can help learners guess and develop word sense connections with the knowledge of core meaning. In this sense, concordance can provide contextualized examples for learners to derive extended meanings of polysemous words. Also, learners can increase their lexical awareness beyond the surface understanding of a word’s definition and develop a sense of transferrable knowledge of the word by engaging in DDL activities. Despite the potential benefits of DDL for developing polysemous word knowledge, there is a lack of empirical research on this topic. To address this gap, the present study asked the following questions:

- RQ 1. How do different types of instruction (DDL, Traditional Learning) affect L2 students’ acquisition of implicit and explicit knowledge of polysemy?
- RQ 2. How do different types of instruction (DDL, Traditional Learning) affect L2 students’ retention of acquired polysemy knowledge?

### 3. Method

#### 3.1 Participants

This study was conducted with middle school students at a private English institution in Korea, ranging in age from fourteen to fifteen. Fifty students participated in the study, and they were assigned to one of the two learning conditions: one group learned polysemy words with concordance-based DDL materials while the other group learned them through dictionary-based materials. 58% were female and 42 % were males (male = 12, female = 13 for the DDL group; male =16, female = 9 for the TL group).

#### 3.2 Target Words

Five polysemy words ‘*run, set, stand, take, see*’ were selected on the basis of three criteria: frequency, familiarity, and multiplicity of meanings. First, among polysemous words, higher-frequency words were chosen so that students could encounter them relatively more prevalently. Second, the primary meaning of the word had to be known by students so that they can derive other uses of the word from the primary meaning. Third, polysemes with as many different meanings were selected so that some meanings are unknown to L2 learners. In this way, L2 learners are familiar with the primary definition of the words, yet they rarely know derived meanings or extended usage of the words.

#### 3.3 Learning Materials (DDL vs. TL)

Two different learning procedures were provided for the two groups to compare the effectiveness of data-driven learning (DDL) and traditional learning (TL) in polysemy learning. For the DDL condition, concordance lines in KWIC format were used, making it easy for students to detect the target word. To identify the most understandable and applicable sample sentences, about 20 concordance lines for each word were selected from an electronic concordance application called Corpus of Contemporary American English (COCA) and given in paper material.

In contrast, students in the TL condition were provided with a list of the five representative definitions for each word as they appear in a dictionary and were asked to memorize these definitions. As it is important that both groups spend the same amount of time learning target words, the TL condition was given a follow-up activity in which they match the memorized definitions to each word.

#### 3.4 Measurements

Students’ knowledge of polysemous words was measured with two different types of tests. One type of test measures implicit knowledge of polysemous words, and this included 30 questions. Implicit knowledge was measured through the judgment test. As seen in Figure 2, each sentence contains the target polysemous word in different contexts and students should judge if the word is used correctly in the given context. Questions were based on example sentences from the online Oxford English Dictionary (OED), and incorrect sentences were created by substituting target words with one of the other learned words that did not fit the context of the given sentence. The naturalness of the questions was reviewed by an English native speaker.

<b>O / X Questions (N= 30)</b>	
1. Holmes <b>ran</b> an impressive race to take the gold medal.	<u>O</u> / X
2. Today the jury began to <b>run</b> the evidence.	O / <u>X</u>

**Figure 2. Example of Implicit Knowledge Test**

The second type of test measures learners' explicit knowledge of polysemous words which contained 20 questions. In this test, each item is presented with synonyms for various dictionary definitions of the target words, and students are expected to match the given definition with the words listed in the box. The synonyms of the definitions corresponding to each item were extracted from the online thesaurus dictionary.

**Matching definition (N= 20)**

A. (Fast moving on foot); race, rush, pace

It corresponds to the meaning of 'run' and is listed next to the word 'run' in the below box.

Run: A
Set:
Stand:
Take:
See:

**Figure 3. Example of Explicit Knowledge Test**

Both types of tests were measured at three time points: (a) pretest before the instruction, (b) immediate posttest after the instruction, and (c) delayed posttest which occurred one week after the posttest. The test materials were the same, but the tests were presented in different orders.

### 3.5 Procedures

The experiment took over three weeks. In the first week, the participants received orientation on what polysemy is and took a pre-test on their implicit and explicit knowledge of polysemy. The test took about 15 minutes, followed by the instruction session. The participants received instructions on target words according to their instructional conditions (DDL or TL) which lasted about 30 minutes. In the DDL condition, students were asked to read the sentence in turn and infer the meaning and speech pattern of the target word by themselves. After then, under the teacher's guidance, classmates shared their opinions to find a suitable interpretation together. In contrast, students in the TL condition reviewed each definition of every target word together with their instructor and were given some time to memorize them. Afterward, they completed an exercise which asked them to match the target words with their definitions.

In the following week, the participants received the same type of instruction (DDL, TL) and took a posttest under the same condition. Lastly, one week later, they had a delayed posttest.

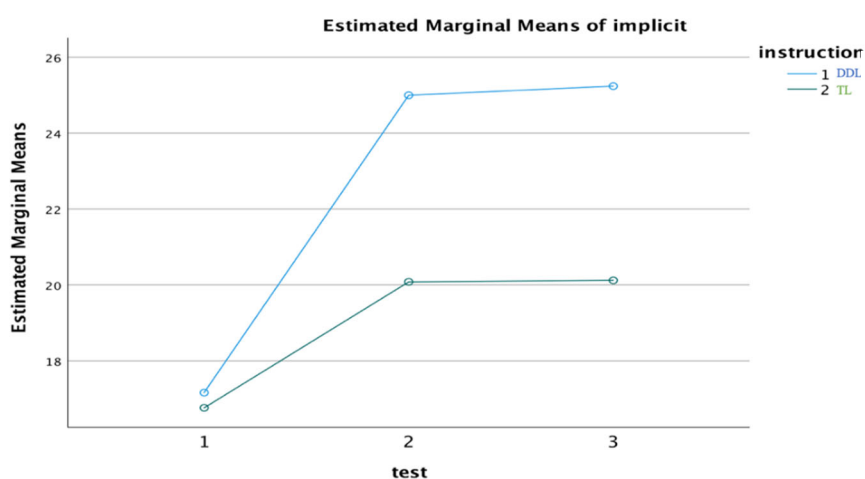
## 4. Result

As the present study compared the effectiveness of DDL- and TL-based learning of polysemy in terms of implicit and explicit knowledge over three weeks. Table 1 and Table 2 summarize descriptive statistics for implicit and explicit test scores, and profile plots are provided in Figures 4 and 5.

**Table 1. Descriptive Statistics for Implicit Test-scores**

Instruction	Implicit pretest		Implicit posttest		Implicit delayed posttest	
	DDL	TL	DDL	TL	DDL	TL
Mean	17.16	16.76	25.00	20.08	25.24	20.12
Std. Deviation	3.236	3.527	2.121	3.161	2.166	3.206

The mean score on the implicit knowledge test increased from the pretest to the posttest by 7.84 for the DDL group and by 3.32 for the TL group. Similar to this, the mean score from the posttest to the delayed test in the DDL group increased by 0.21 and in the TL group by 0.04.



**Figure 4. Profile Plot for Implicit Knowledge**

The lines of the graph show an overall increase for the posttest and the delayed posttest for both groups. The pretest scores of the two groups are almost close. When compared to the TL group, however, the DDL group grows to a greater degree in the post-test than the TL group. Furthermore, while both groups increase significantly from the pretest to posttest, the degree of improvement in the delayed test from the posttest score is rather small.

Next, changes in explicit knowledge scores over time are shown below.

**Table 2. Descriptive Statistics for Explicit Test Scores**

Instruction	Explicit pretest		Explicit posttest		Explicit delayed posttest	
	DDL	TL	DDL	TL	DDL	TL
Mean	8.32	9.12	13.16	13.40	13.64	13.72
Std. Deviation	2.719	2.538	3.460	2.483	4.081	3.221



In the explicit knowledge test, the mean score from the pre-test to the posttest in the DDL group increased by 4.84, whereas it increased by 4.28 for the TL group. Similarly, for the DDL group and the TL group, the mean scores increased from the posttest to the delayed test by 0.48 and 0.32, respectively. Figure 5 presents the changes of the explicit knowledge test scores.

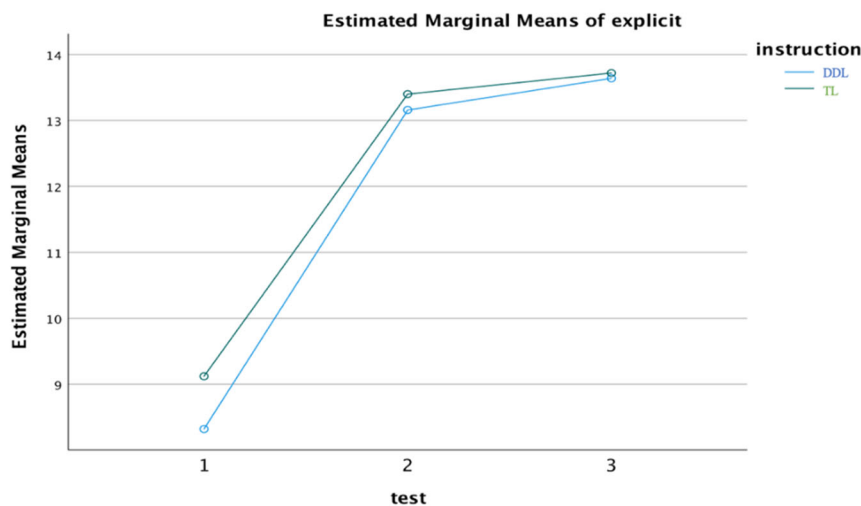


Figure 5. Profile Plot for Explicit Knowledge

In both groups, the line of the graph increases for the posttest and the delayed posttest. No differences existed in the pretest scores for both groups, but after the treatment, the posttest scores increased for both groups, and this acquired knowledge seemed to be retained in the delayed posttest scores. Furthermore, the posttest and delayed posttest results for both groups seemed similar.

Based on the descriptive statistics, to examine whether the observed differences are statistically significant, a repeated measures MANOVA was performed. The independent variable was instruction conditions at two levels (DDL vs. TL), and the dependent variables were scores of implicit and explicit knowledge tests at three points (pretest, posttest, and delayed posttest). Table 3 reveals the results.

Table 1. Multivariate Tests with the Effects of Between-Subjects and Within Subjects

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Between Subjects	Intercept	.015	1517.384 <sup>b</sup>	2.000	.985	<.001*	.985
	instruction	.650	12.652 <sup>b</sup>	2.000	.350	<.001*	.350
Within Subjects	week	.074	141.350 <sup>b</sup>	4.000	.926	<.001*	.926
	week * instruction	.484	11.985 <sup>b</sup>	4.000	.516	<.001*	.516

\* *p* value < .05.

Table 3 shows both between-subjects and within-subjects effects. The between-subjects result demonstrates how the implicit and explicit knowledge scores differ depending on the instructions (DDL or TL). The results revealed that different instructions resulted in statistically significant differences between the groups (*p* < .001). Moreover, the within-subject results showed that there are differences in learners' knowledge of polysemy across

three time points ( $p < .001$ ). In addition, the interaction between week and instruction was significant, suggesting that instructional effects differ across different points of measurement. Based on the results, a post hoc analysis was conducted, and the results are summarized in Table 4.

**Table 2. Tests of Within-Subjects Contrasts Over Weeks**

Source	Measure	week	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	
k	implicit	Week 1 vs. Week 2	1556.820	1	1556.820	290.994	<.001*	.858	
		Week 2 vs. Week 3	.980	1	.980	.950	.335	.019	
		Week 1 vs. Week 2	1039.680	1	1039.680	276.633	<.001*	.852	
	explicit	Week 2 vs. Week 3	8.000	1	8.000	2.211	.144	.044	
		implicit	Week 1 vs. Week 2	255.380	1	255.380	47.735	<.001*	.499
			Week 2 vs. Week 3	.500	1	.500	.485	.490	.010
explicit	Week 1 vs. Week 2		3.920	1	3.920	1.043	.312	.021	
	Week 2 vs. Week 3	.320	1	.320	.088	.767	.002		

\*  $p$  value < .05.

There was a significant difference in implicit and explicit knowledge scores from pretest to posttest for both implicit and explicit knowledge. However, there was no significant difference from the week 2 to the week 3, suggesting that learners' knowledge has been retained over one week of their learning (i.e., delayed posttest). The week\*instruction interaction effects existed only in the implicit knowledge from the pretest to posttest, suggesting that more learning occurred in the DDL condition compared to the TL condition in the immediate posttest, yet no differences existed in the delayed posttest.

## 5. Discussion

The goal of this study was to investigate whether DDL is more effective in helping students gain polysemous word knowledge in a second language compared to a dictionary-based traditional learning style. The results of the present study revealed that although both groups improved their performance from the pretest to the posttest only in an implicit knowledge test, the DDL group outperformed the traditional learning group. These effects, however, did not last in the delayed posttest. In other words, both groups' performance was not different in the delayed posttest.

Focusing on the immediate learning effects, these results are consistent with findings of previous studies that showed the positive effects of DDL over traditional learning strategies (Cobb 1999, Balunda 2009, Frankenberg-Garcia 2012, Jiao 2012, Liya 2021). For example, in a study of vocabulary teaching using corpora in comparison to online dictionary use by Jiao (2012), the group who received DDL remembered significantly more than a group of online dictionary users in the posttest. In the present study, the interaction effects between week and instruction

observed in the implicit knowledge test suggest that the benefits of DDL over TL existed in the implicit knowledge test, but not in the explicit knowledge test.

The findings are explained in light of the benefits of corpus-based learning or exposure to implicit language learning. First, concordance output exposes learners to linguistic phenomena in authentic contexts (Cobb 1999). This provides learners with a variety of inductive and deductive language learning opportunities which were not previously available in traditional vocabulary learning with little exposure to instances (Liu and Jiang 2009). Also, students can expand their metalinguistic knowledge by inferring the meaning and speech patterns, and it expands their sociocultural perspectives. Furthermore, DDL generates procedural knowledge as well as essential learning skills (Aston 1996, Cohen 2003). In other words, working with corpus data allows students to develop a process-oriented approach to language learning as learners are actively engaged in the learning process (O'Sullivan 2007). The knowledge and learning techniques that students acquire during DDL may be just as important since they lay the groundwork for future learning. Lastly, DDL enables students to notice certain characteristics of a word that could not be achieved by traditional vocabulary learning strategies.

However, though many studies supported the effectiveness of DDL in vocabulary learning, there are studies that revealed the benefit of traditional learning styles in understanding the meaning of a word in the short term (Cobb 1999, Frankenberg-Garcia 2012). Frankenberg-Garcia (2012) compared vocabulary learning through dictionary definitions and corpus instances for EFL students in Portugal and found that a dictionary helped learners understand the meaning of the terms better while a corpus helped them correctly write the words syntactically.

Regarding retention, the acquired knowledge lasted at least another week. However, while the DDL condition was more beneficial than the TL condition for immediate posttest, there were no condition effects for the delayed posttest, suggesting that the beneficial effects of DDL seem only immediate and after a week, the benefits disappear. Put differently, both groups retained the acquired knowledge similarly well after one week of practice. Furthermore, as the current study examined the retention of knowledge only after a week, it is unclear how the acquired knowledge will last afterward. A delayed posttest that occurs after a longer period of time is needed to better understand the longer-term effects of DDL-based learning.

## 6. Conclusion

Like a jigsaw puzzle piece, vocabulary fits into the set when combined with surrounding pieces. Vocabulary learning, in that light, is something that should be internalized in context. However, in the EFL environment, there is a lack of contextualized context to expose learners to the natural context in which the word is used. For that reason, L2 learners' knowledge of the word and their ability to produce and comprehend it is limited. Moreover, it is more noticeable in polysemy learning since the most frequent and important words are polysemous and learners usually lack knowledge about various meanings in polysemous words besides the most frequent ones.

According to Cook (2016), "we don't know a word properly until we have learned its forms, its different types of meaning, and how it is used in sentences" (p. 80). As a result, lexical polysemy is an essential feature of vocabulary depth and emerges as a key aspect of language proficiency, which means learners must be able to employ vocabulary, particularly high-frequency terms, receptively and productively to achieve proficiency in L2 learning.

However, the sheer number of meanings of polysemy might frustrate students and it might be inefficient to memorize all meanings that appear to be similar at the same time. Instead, polysemy knowledge should be accumulated by detecting multiple meanings employed in novel circumstances through their experience so that

the knowledge can extend along with its primary meaning. Thus, concordance or DDL-based learning seems to assist learners to be exposed to diverse contexts and language uses.

DDL highlights three major benefits: “text authenticity”, “students' automatic attitude toward learning”, and “enhanced interest among learners”. The DDL approach to learning vocabulary seems to facilitate a deeper understanding of a word that extends beyond dictionary definitions of terms to include metalinguistic awareness about that word. Steven (1991) claims that concordance-based learning is effective, “if the purpose of the exercise is to reinforce the vocabulary, rather than testing, and if the teacher's proclivity is to instill a sense of confidence and well-being in the students” (p. 55).

The present study's results show that corpus-assisted learning in polysemous words can be an effective pedagogical method in developing, at least, in the short term and the implicit knowledge of polysemous words. Teachers and language practitioners can design diverse DDL-based teaching materials and corpus resources to use in the classroom. This would aid students in understanding the multiple meanings of polysemous words by increasing their awareness of language patterns in natural language contexts. Despite this, the current study is limited in terms of the short duration of instruction time. As there are limited vocabulary items to learn in the present study, the total of an hour of instruction over two weeks is considered to be appropriate, but the length of instruction could have been extended to include more target items. Additionally, the study did not incorporate learner interviews or surveys on how students have experienced DDL or on why they have chosen incorrect answers, which would have given better insights into DDL. Future research should address these issues. Nonetheless, it is hoped that this study will provide additional insight into DDL-based research as it attempted to tap into the depth of vocabulary in the context of DDL, by examining polysemous words rather than focusing on the vocabulary meaning of a single unit.

## References

- Abou-Khalil, V., S. Helou, B. Flanagan, M. R. A. Chen and H. Ogata. 2019. Learning isolated polysemous words: identifying the intended meaning of language learners in informal ubiquitous language learning environments. *Smart Learning Environments* 6(1), 1-18.
- Asik, A., A. S. Vural and K. D. Akpınar. 2016. Lexical awareness and development through data driven learning: Attitudes and beliefs of EFL learners. *Journal of education and training studies* 4(3), 67-96.
- Aston, G. 1996. What corpora for ESP? Paper presented at a conference on ESP at the University of Pavia, November 1996.
- Baddeley, A. D. 1997. *Human Memory: Theory and Practice*. Hove: Psychology Press.
- Balunda, S. 2009. *Teaching Academic Vocabulary with Corpora: Student Perceptions of Data-Driven Learning*. Master's thesis, Indianan University.
- Bernardini, Silvia. 2000. Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard and T. McEnery, eds., *Rethinking Language Pedagogy from a Corpus Perspective*, 225-234. Hamburg: Peter Lang.
- Bernardini, S. 2004. Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair, ed., *How to Use Corpora in Language Teaching*, 5-36. Amsterdam: John Benjamins.
- Booton, S., A. Hodgkiss, S. Mathers and V. Murphy. 2022. Measuring knowledge of multiple word meanings in children with English as a first and an additional language and the relationship to reading comprehension. *Journal of Child Language* 49(1), 164-196.

- Boulton, A. and T. Cobb. 2017. Corpus uses in language learning: A meta-analysis. *Language Learning* 67(2), 348-393.
- Boulton, S. De Cock. 2017. Dictionaries as aids for language learning. *International Handbook of Modern Lexis and Lexicography*, 1–17.
- Brown, H. D. 2007. *Principles of Language Learning and Teaching*. Pearson Longman.
- Celce-Murcia, M. 2001 *Teaching English as a Second or Foreign Language*, 3rd ed. Heinle & Heinle Publisher, Boston.
- Conrad, S. 2005. *Corpus Linguistics and L2 Teaching*. Routledge.
- Cobb, T. 1999. Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning* 12, 345-360.
- Cohen, A. 2003. *Strategy Training for Second Language Learners*. ERIC Digest.
- Craik, F. I. and R. S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.
- Cruse, A. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford, England: Oxford University Press.
- DeCarrico, J. S. 2001. Vocabulary learning and teaching. *Teaching English as a Second or Foreign Language* 3, 285-299.
- Fang, X. 2014. Conceptual metaphor and vocabulary teaching in the EFL context. *Open Journal of Modern Linguistics*, IV(2).
- Frankenberg-Garcia, A. 2012. Learners' use of corpus examples. *International Journal of Lexicography* 25(3), 273-296.
- Gabrielatos, C. 2005. Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ* 8(4), 1-37.
- Ittze's, I. 1991. Lexical guessing in isolation and context. *Journal of Reading* 34, 360-366.
- Jiao, B. 2012. An empirical study on corpus-driven English vocabulary learning in China. *English Language Teaching* 5/4, 131-137.
- Jin, L. and E. Deifell. 2013. *Foreign language learners' use and perception of online dictionaries: A survey study*. *MERLOT Journal of Online Learning and Teaching* 9(4), 515-533.
- Johns, T. 1991. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal* 4, 27-45.
- Kang, H.-W. 1992. How can a mess be fine? polysemy and reading in a foreign language. *Mid-Atlantic Journal of Foreign Language Pedagogy* 1, 35-49.
- Kuiper, K., R. Fromont and D. Gerhard. 2017. Polysemy and word frequency: A replication. *Journal of Research Design and Statistics in Linguistics and Communication Science* 4, 144-155.
- Li, M. and J. R. Kirby. 2015. The effects of vocabulary breadth and depth on English Reading. *Applied Linguistics* 36(5), 611-634
- Liu, D. and P. Jiang. 2009. Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal* 93(1), 61-78.
- Liya X. 2021 Using data-driven learning activities to improve lexical awareness in intermediate EFL learners. <https://www.tandfonline.com/doi/full/10.1080/2331186X.2021.1996867>
- Lyons, J. 1995. *Linguistic Semantics: An Introduction*. Cambridge, England: Cambridge University Press.
- McCarthy, M. 1990. *Vocabulary*. Oxford: Oxford University Press.

- Nagy, W. 1997. On the role of context in first- and second-language vocabulary learning. In N. Schmitt and M. McCarthy, eds., *Vocabulary: Description, Acquisition and Pedagogy*, 64-83. Cambridge, England: Cambridge University Press.
- Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P. 2008. Lexical awareness in second language learning. In N. H. Hornberger, eds., *Encyclopedia of Language and Education*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30424-3\\_14](https://doi.org/10.1007/978-0-387-30424-3_14)
- Nation, P. and T. Chung. 2009. Teaching and testing vocabulary. In M. H. Long and C. J. Doughty, eds., *The Handbook of Language Teaching*, 543-559. Malden, Ma: Wiley-Blackwell.
- O'Sullivan, I. 2007. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL* 19, 269-286. DOI: <https://doi.org/10.1017/S095834400700033X>
- Quan, Z. 2016. Introducing “mobile DDL (data-driven learning)” for vocabulary learning: an experiment for academic English. *Journal of Computers in Education* 3(3), 273-287.
- Ravin, Y. and C. Leacock. 2000. *Polysemy: Theoretical and Computational Approaches*. New York: Oxford University Press.
- Schmidt, R. 2001. Attention. In P. Robinson, ed., *Cognition and Second Language Instruction*, 3-32. Cambridge, MA: Cambridge University Press.
- Schouten-van Parreren, C. 1985. *Woorden leren in het vreemde-taalonderwijs [Learning Words in a Foreign Language Setting]*. Apeldoorn, the Netherlands: Van Walraven.
- Stahl, S. A. and W. E. Nagy. 2006. *Teaching Word Meanings*. Mahwah, NJ: Erlbaum.
- Stevens, V. 1991. Classroom concordancing: Vocabulary materials derived from relevant, authentic text. *English for Specific Purposes* 10, 35-46.
- Tribble, C. and G. Jones. 1990. *Concordances in the Classroom: A Resource Book for Teachers*. Essex, U.K.: Longman.
- Yilmaz, E. and A. Soruç. 2015. The use of concordance for teaching vocabulary: A data-driven learning approach. *Procedia - Social and Behavioral Sciences* 191, 2626-2630.

Examples in: English  
Applicable Languages: English  
Applicable Level: Tertiary

## APPENDIX A

### A Sample of the DDL Instruction Material

*Run*

1	them as essential toward fulfilling his <b>future</b> aspirations to	run	a <b>small</b> business . His hoarding behavior may also be important
2	a stop , I would jump over the <b>fence</b> and <b>excitedly</b>	run	<b>across</b> the <b>field</b> . My grandfather would look up and smile as
3	, some very basic things , such as " <b>How</b> to	run	<b>and</b> <b>operate</b> a business . " " How to market your business
4	# (9 .) The standard interpretation <b>of</b> the <b>poem</b>	runs	<b>as</b> <b>follows</b> : Yvain is the story of the education of a
5	with completely transected spinal cord . Now , <b>we</b> asked to	run	<b>few</b> <b>experiments</b> to find out whether the same treatment will be
6	employees of a 501(c) (3) organization <b>campaign</b> for <b>candidates</b>	running	<b>for</b> <b>election</b> as directors of the organization ? All members of
7	" " Then I 'm here to help . " He	ran	<b>his</b> <b>hand</b> <b>through</b> slicked hair . There was something wrong about
8	Sur , the underlying problem is simply that <b>the</b> <b>otters</b> <b>are</b>	running	<b>out</b> of <b>food</b> . # While they are not starving to death
9	, no they `re not . " And <b>the</b> <b>media</b> <b>will</b>	run	<b>stories</b> and <b>we</b> had it during the funeral period of Diana ,
10	his three young children . # By October , <b>the</b> <b>schedule</b>	ran	<b>to</b> <b>14</b> <b>hours</b> , seven days a week . Techs blew off
11	when he got to the warehouse . A <b>motor</b> did n't	run	<b>without</b> <b>fuel</b> . # A tall , skinny , gray-haired man waited
12	a reason ... # In most cases , <b>just</b> a <b>quick</b>	run	<b>through</b> the <b>machine</b> and the Pokemon would be good as new .

## APPENDIX B

### A Sample of the TL Instruction Materials

#### *Definition*

##### Run

1. 달리다
2. 운영하다
3. 출마
4. 열리다
5. 계속하다

##### Set

1. 세우다
2. 견디다
3. 서 있다
4. 입장
5. 위치하다

#### *Matching Game*

- 운영하다
- 이해하다
- 세트
- 달리다
- 서다
- 발견하다
- 데려가다

- Run
- Set



## APPENDIX C

### Example of Implicit Knowledge Test

*The meaning of a word often changes slightly depending on the way it is used in a sentence. In the following activity, you will be asked to read sentences that contain polysemous words in different contexts. Choose O / X based on your judgment if the word is used correctly in the given sentences.*

For example:

1. Holmes **ran** an impressive race to take the gold medal.  / X
2. Today the jury began to **run** the evidence. O /

#### Questions

1. Her last musical **ran** for six months on Broadway. O / X
2. The wedding is **run** to take place at the end of November. O / X
3. He made an unsuccessful **run** for governor in 2008. O / X

### Example of Explicit Knowledge Test

*Polysemous words contain different but related meanings depending on the way it is used. In the following activity, you will be given various definitions of the words; 'run, set, stand, take, see'. Find the alphabetical symbols that correspond to the meanings of each word and list them next to the word in the box below.*

For example:

- A. (Fast moving on foot); race, rush, pace

It corresponds to the meaning of 'run' and is listed next to the word 'run' in the below box.

#### Question

Run: A Set: Stand: Take: See:
---

- A. (Fast moving on foot); race, rush, pace
- B. (Attempt to be elected to public office); compete, race, challenge, contest
- C. (Accept, adopt, use); bring, enjoy have, include, perform