Yingying Yang*, Lin Chen and Xumin Tian

Student perceived effectiveness of task-based instructional design of data-driven synonym learning featuring “mini-lecture”

https://doi.org/10.1515/jccall-2023-0024
Received November 5, 2023; accepted December 29, 2023; published online March 1, 2024

Abstract: This study evaluates student perceived effectiveness of a guided inductive approach for synonym learning through a 36-week, two-round data-driven learning (DDL) and task-based instruction design. Presented as a student-centered “mini-lecture” task, it integrates pre-task training, during-task guidance, and post-task feedback. Data were collected from 23 business English university students via questionnaires and interviews. Results show that learners generally held consistent positive attitudes toward this long-term data-driven learning. Enhanced instruction in the second round underscored the importance of scaffolding, reducing learners’ anxiety, and promoting cognitive development. Task complexity and language proficiency seemed unrelated to learner perceptions. Although task motivations influenced learners’ perceptions, learners with lower motivation were not overwhelmed by technical difficulties involved in DDL, supporting the effectiveness of the long-term practice in improving learners’ corpus query skills. Learners often used their native language to explain key concepts in the “mini-lecture” presentation. The results suggest the task-based “mini-lecture” design is effective and that guided inductive DDL is a promising alternative in language pedagogy. Based on the findings, a task-based data-driven synonym learning model is proposed.

Keywords: guided inductive approach; task-based; DDL; perceptions; synonym learning

***Corresponding author: Yingying Yang**, School of Education, Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China, E-mail: yyyzhuye@163.com. https://orcid.org/0000-0001-7343-4596
Lin Chen, School of Education, Shanghai International Studies University, Shanghai, China, E-mail: chenlin36@foxmail.com
Xumin Tian, Bijie Institute of Education Science, Bijie, China, E-mail: 2357165965@qq.com

1 Introduction

Synonym/near-synonym acquisition has a great influence on vocabulary learning (Alanazi, 2017). Mastering the nuances among synonyms contributes to accurate expressions and effective communication (Hatch & Brown, 1995; Liu & Zhong, 2016; Sun et al., 2011; Yeh et al., 2007). However, synonymy poses a great challenge to language learners (Hatch & Brown, 1995; Nasser, 2021), especially L2 learners (Martin, 1984; Yevchuk, 2022); even advanced learners encounter difficulties using synonyms (Dushku & Paek, 2021; Khazaal, 2019; Laufer, 1991; Liu & Zhong, 2016).

Regarding L2 learners, the inefficiency of synonym acquisition may be caused by a lack of rich authentic language input. Data-driven learning (DDL; Johns, 1991), featuring the use of corpora, enables learners to interact with large amounts of authentic language data, observe concordance results, and explore the collocations and patterns of language use, and helps them to identify the semantic meaning and usage patterns of synonyms (Liu & Zhong, 2016). Also, it helps improve learners' skills in discovery learning and self-regulated learning (Alanazi, 2022; Divjak & Gries, 2006; Gries & Otani, 2010; Liu, 2010, 2013; Liu & Espino, 2012; Yevchuk, 2022). Only recently has synonymy received its due attention (e.g., Divjak & Gries, 2006; Gu, 2017; Liu & Zhong, 2016; Song, 2021; Yeh et al., 2007). However, knowledge about how data-driven synonym instruction can be implemented in the EFL classroom contexts is limited (Boontam, 2022; Soruç & Tekin, 2017).

Technology and task-based language teaching (TBLT) bears a reciprocal relationship: technology provides an authentic venue facilitating TBLT implementation, and TBLT provides a pedagogical framework for the selection and research of technology (Doughty & Long, 2003; Skehan, 2003). Research exploring new technologies in task-based contexts has the potential to further understanding of TBLT and insights into Corpus-aided Language Learning (CALL) (Ziegler, 2016). Therefore, as a processed-oriented approach, TBLT and DDL can well be integrated into teaching practice.

A "mini-lecture" instructional design was implemented in this study. It is a learner-centered, task-based learning model in which learners demonstrate the outcomes of their corpus exploration to their peers. This approach emphasizes the role of guidance and scaffolding in the DDL process. The current study aims to investigate learners' perceptions of this model to provide further empirical evidence for DDL research and more implications for technology-empowered foreign language teaching.

2 Literature review

2.1 Traditional approaches to synonym learning

In the EFL context, learners primarily gain synonym knowledge from teacher instruction and/or dictionary consultation. The former, mainly based on teachers' experience and intuition, is form-focused instruction and may fall short in providing learners with accurate and abundant information on word recognition and collocation (Tsai, 2019), while the latter seldom represents lexical regularity in extended texts (Quinn, 2014), and may even lead to incorrect substitution between synonyms in production (Murphy, 2003; Thienthong, 2020; Yeh et al., 2007). Both methods lack sufficient nuance information and may present confusing, overlapping interpretations of synonyms (Song, 2021), resulting in learners' shortage of collocational knowledge, which makes synonyms differentiation difficult.

2.2 Data-driven synonym learning

Synonym learning requires far more than just providing the implied equivalence; the knowledge of collocational patterns, semantic features, and stylistic guidance is essential (Harvey & Yuill, 1997; Phoocharoensil, 2021). With the development of cognitive linguistics and corpus linguistics, synonym research has gained attention and progress (Liu & Zhong, 2016). Corpus linguistics proposes lexical semantic theory (Firth, 1957; Halliday, 1966; Sinclair, 1966, 1991), arguing that the meaning of words is determined by their collocates and other contextual features that accompany them, and thus collocation is a powerful indicator for synonym discrimination (Kilgarriff et al., 2014; Liu & Zhong, 2016). In this vein, many studies have investigated specific sets of synonyms with online tools such as COCA (Corpus of Contemporary American English), Sketch Engine, BNC Web, and so forth., by examining distribution across genre, collocations, and semantic preference (e.g., Divjak & Gries, 2006; Gu, 2017; Liu, 2010, 2013; Phoocharoensil, 2020, 2021, 2022). These studies show that synonyms differ in terms of collocational patterns despite their similar conceptual meaning (Murphy, 2009), and the collocational interchangeability between synonyms is rare (Xiao & McEnery, 2006). Research in this area yields abundant information about corpus-based synonym differentiation. However, studies have conventionally focused on certain sets of synonyms, and this seems to be too narrow in terms of synonym learning. What matters most is to develop learners' independent synonym query skills and consultation literacy, yet how they can achieve this is still not clear.

Some corpus-based studies focused on learners' synonym acquisition. Ahmadian and Darabi (2012) investigated the relationship between EFL learners' knowledge of near synonyms and their performance on a corpus-driven test of collocational behavior. They found students were unaware of the subtle distinctions among synonyms, and their knowledge of synonymy was related to their performance on the test of collocational behavior. Liu and Zhong (2016) used a forced-choice question instrument and conducted a comparative analysis of synonym usage of native speakers and ESL/EFL learners, and they found that salience of target pattern and learners' construal of the communication context/goal were two major factors influencing synonym acquisition. Wongkhan and Thienthong (2021) employed a forced-choice collocation test and examined the relationships between learners' academic experience at university and their knowledge of academic collocation and synonymy. The results revealed that students with more experience significantly outperformed those with less experience in most collocation questions. These studies provide evidence for learners' difficulty in synonym acquisition and the importance of corpus-based synonym information (i.e., collocational patterns, distribution, and language formality), but they focused on learners' synonym knowledge and used corpus data only as a reference. How corpus can be applied in L2 synonym teaching remains unknown.

Yeh et al. (2007) investigated the effectiveness of the use of corpora in enhancing university students' synonym learning. They designed five online units for increasing students' awareness of underused specific adjectives for EFL college writing. The results showed that students' synonym knowledge not only improved in their immediate post-test but was retained in the two-month delayed post-test. Students also reported positive attitudes toward inductive learning despite the difficulty in verbalizing the semantic differences. Jafarpour et al. (2013) compared the effect of the corpus-based approach and traditional approach in teaching 90 university students the collocation of near synonyms. The results revealed that the group with the corpus-based approach performed significantly better than the traditional group in the comprehension and production of collocations of synonyms. Similarly, Soruç and Tekin (2017) conducted a study with 72 secondary school students to examine the effect of the DDL approach on vocabulary learning. They found that the DDL group achieved significantly higher scores on both the immediate and the delayed post-test, and they displayed favorable attitudes toward the DDL practice. Boontam (2022) examined the effectiveness of 30 Thai second-year EFL university students' paper-based data-driven learning of three synonymous English adjectives, "naughty," "disobedient," and "rebellious." The instructional materials were designed according to the meanings, collocations, formality, and grammatical patterns from BNC and Sketch Engine. The results showed that learners' vocabulary knowledge developed after this four-week learning through DDL activities.

The participants held positive attitudes toward this DDL practice. Notably, DDL research has been conducted with languages than other English. For example, Yao (2019) implemented a three-week quasi-experimental design to examine the difference between the DDL approach and the dictionary-based approach to vocabulary learning with 32 Chinese learners of Spanish. The study provided further evidence for students' positive attitudes toward DDL practice. The DDL group significantly outperformed the control group. This is "meaningful for the popularization and acceptance of DDL" in foreign language learning other than ELT (Yao, 2019, p. 38).

While these studies offer examples of how to design concordance-based materials with electronic referencing tools, the treatment span was relatively short (i.e., about four weeks in Yeh et al.'s and Boontam's studies, and only four classroom hours for each instructional group in Soruç & Tekin's study). Therefore, the conclusions have to be considered carefully, and the long-term effect should be tested further. Besides, these studies are fundamentally empirical, and the teaching materials design was purely instructional but without "theoretical positioning" (Pérez-Paredes, 2019), which underscores one of the problems in the 30 years of DDL research (Boulton & Vyatkina, 2021).

2.3 Learners' perceptions of DDL

Existing studies are rich in investigating learners' perceptions of DDL implementation (e.g., Bernardini, 2004; Kennedy & Miceli, 2010; Lee et al., 2020; O' Sullivan & Chambers, 2006; Yoon, 2008; Yoon & Hirvela, 2004), but no conclusive results have been identified. Some studies found learners' positive attitudes toward DDL (e.g., Mizumoto & Chujo, 2015) while others found negative attitudes (e.g., Quan, 2016) or mixed attitudes within the same cohort of learners (e.g., Kennedy & Miceli, 2001). Given the differences in conceptualization, operationalization, and research paradigm, most studies were small-scale and short-term with different research questions and teaching contexts, thus failing to ensure generalizable conclusions (Geluso & Yamaguchi, 2014; Tsai, 2019). In addition, DDL has little impact concerning the real problem of "how to implement big data in language teaching" (Boulton, 2009) for it has not been normalized in language teaching (Conrad, 2005; Flowerdew, 2012; Römer, 2006). Therefore, more empirical evidence, especially studies generated from cutting-edge teaching practices, is needed to support DDL in the teaching context (Chambers, 2019). Longitudinal research or long-term DDL treatment is also needed (Boulton, 2009; Boulton & Vyatkina, 2021).

Against this backdrop, the present study examines the effectiveness of a tentative data-driven synonym learning model to investigate learners' perceptions

(changes) in the two rounds of DDL practice. It addresses the following research questions (RQs):

RQ1: How do learners perceive the data-driven synonym learning model featuring a task-based "mini-lecture"?

RQ2: Are there any changes in learners' perceptions after experiencing a long duration of data-driven synonym learning?

RQ3: What are the potential factors influencing learners' perceptions (changes) of task-based data-driven synonym learning?

3 Research design

According to Flowerdew (2015), DDL may draw on the theories of noticing hypothesis, constructivist learning, and scaffolding in sociocultural theories; the present study takes these theories as the underpinnings of the corpus-based TBLT instructional design. This design adopts a guided inductive approach (Johansson, 2009) to conduct the corpus-based TBLT instructional design (Ellis, 2003; Skehan, 1998).

3.1 Theoretical framework

Schmidt (2001) argued that the noticing hypothesis is intentional guidance for noticing some features of words in context, which promotes language acquisition, thus supporting the use of the inductive method in language learning (Kavaliauskiene, 2003; Shaffer, 1989). As learners read concordances in KWIC (keyword in context) or other input-enhanced forms presented by corpus tools, the salient input helps the noticing necessary for meaningful input processing to occur (VanPatten & Benati, 2010, as cited in Crosthwaite & Boulton, 2022).

By "noticing" the corpus data, learners can be stimulated to use inductive learning strategies, especially those of "perceiving similarities and differences and of hypothesis formation and testing" (Johns, 1994, p. 297). These high-order cognitive skills can be practiced with process-oriented constructivist pedagogy. Within this framework, DDL highlights exploratory language learning that requires learners to employ cognitive mechanisms and background knowledge to process new language data and discover salient new language input.

Nonetheless, learners with varying learning styles react differently to the constructivist approach; many learners are reluctant to use it, and some may even get lost in language data because of ineffective independent exploration (O'Keeffe, 2021), which calls for scaffolding teaching (Cobb & Boulton, 2015). As stated by Yoon and Hirvela (2004), "it may be pedagogically unwise to simply let students browse through a tremendous amount of sample texts or corpora without proper guidelines" (p. 278). Similarly, Smart (2014) argued that DDL "is not necessarily characterized by direct interaction with language corpora or by total learner autonomy" (p. 186). Instead, the implementation of DDL relies on "carefully designed and scaffolded activities" to help learners explore language data, and these activities "place the learner at the center of the learning task" (p. 187). Johansson (2009) proposed "a guided inductive approach or a combination of an inductive and deductive approach where the elements of explanation and corpus use are tailored to the needs of the student" (p. 42).

The present study adopts this guided inductive approach (Flowerdew, 2009; Mizumoto & Chujo, 2016; Smart, 2014) and conducts an instructional design, which takes the form of student-centered "mini-lecture" tasks, with teacher training and feedback before, during, and after the tasks. Through these scaffolding activities, learners would engage more as they can gain support and interaction in the process of discovering language.

3.2 Participants and settings

Participants in this study were 23 students (business English majors, 19 women and 4 men, aged 20–21) from a university in China. Three participants did not fully engage in the whole process; ultimately data from 20 students were collected for analysis. Using convenience sampling, this research was conducted in an "*Advanced English (I/II)*" course, with the first author serving as both researcher and instructor. This course was a required course for junior English majors. Altogether, learners underwent two semesters (each 18 weeks) with four sessions (180 min) per week. Within each unit text, many synonymous words posed great challenges. As such, a corpus-based student-centered "mini-lecture" task was designed to enhance learners' vocabulary learning vision. Arguably, junior English majors may have the potential to work more independently with cognitively load corpus query tasks, which benefits them in both language improvement and cognitive development.

Each "mini-lecture" presentation of the synonyms was synchronous with one unit of learning. Participants were divided into seven groups (five groups of three and two groups of four). To fully promote learners' self-regulated learning, DDL should avoid conformity when possible, and it was better to start in the most

comfortable way for learners (Gilquin & Granger, 2022; Hunston, 2002). The "mini-lecture" task allowed students to independently choose 3 to 5 new words from the text and the pairing synonyms. Thus, the task catered to their learning needs, and the selection of the follow-up quiz items also aligned with their language proficiency. By increasing the task relevance, this freedom of choice will motivate the students in DDL practice.

The research was conducted with two rounds (each one semester) of DDL. The first round was the running-in stage where learners were familiarized with the basic query skills, and the second round was the enhanced stage in which we adjusted the teaching design based on the learners' first perceptions of DDL. The emphasis was placed on optimizing the task procedure and providing the scaffolding worksheet to streamline the process.

3.3 "Mini-lecture" instructional design

We used COCA as the platform for data-driven synonym learning. COCA provides a web-based corpus-query interface. Specifically, the "Compare" function makes discriminating synonyms more accessible. Figure 1 displays the flowchart for the task-based instruction design.

The design involved three main stages for learners:

Stage 1: Before the task, the teacher introduced the purpose of integrating corpus into language learning and conducted a 90-min training about how to use COCA for synonym differentiation, especially on the function "COMPARE" and "COLLOCATES."

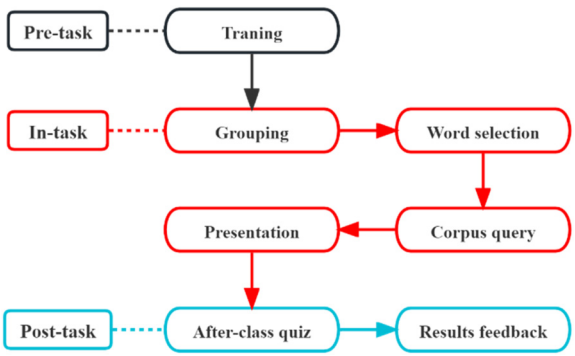

Figure 1: The flowchart for the data-driven "mini-lecture" instructional design.

Stage 2: During the task, groups selected 3–5 pairs of target synonyms; then, they conducted a corpus query and made an in-class report of their findings and conclusions. In this process the students fully engaged to learn how to use different query skills for information gathering and participated in more high-order cognitive activities by observing the data, forming hypotheses, and verifying the conclusions. Scaffolding and immediate feedback were provided to help students better acquire query skills and high-order cognitive skills.

Stage 3: After the task, the group distributed a self-designed synonym quiz based on their presentation and collected their peers' answers. Feedback was provided in the next class. This step is very crucial: the quiz helped them verify the effectiveness of their "mini-lecture"; it also helped to consolidate all the students' understanding of target synonyms.

3.4 Research procedures and instruments

This study was well-grounded in the classroom setting and the research procedures are displayed in Figure 2; data were collected through three questionnaires and two interviews (see Figure 2).

Questionnaire A (see Appendix A) was distributed before the first-round DDL practice to gain learners' basic information, such as English learning experiences, their attitudes toward English learning, habits of using online resources, tools of vocabulary learning, the evaluation of their learning effect, and background knowledge of corpus use.

Questionnaire B (see Appendix B) was distributed at the end of the first round of the DDL practice. It was adapted from Yoon and Hirvela (2004) and the Cronbach Alpha coefficient was 0.799. Questionnaire C (see Appendix C) was distributed it at the end of the second-round DDL practice. The Cronbach Alpha coefficient was 0.743. Questionnaire C maintained the consistency in most of the items but was

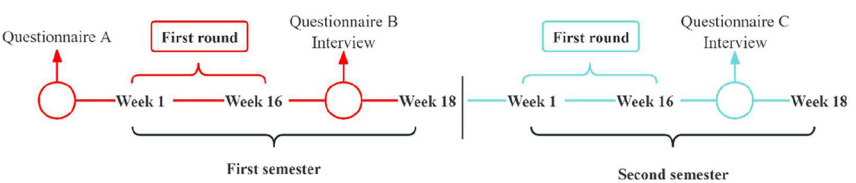

Figure 2: The research procedure.

Table 1: The constructs of Questionnaire B & C.

Questionnaire B			Questionnaire C	
Item number	**Sub-categories**	**Dimensions**	**Sub-categories**	**Item number**
21, 22, 25	Effectiveness	Perceptions of the	Effectiveness	20, 24, 25, 48
26, 27	Interaction	"mini-lecture"	Interaction	7, 8
20, 23, 24	Task motivation		Task motivation	21, 23, 38, 39
	Task complexity		*Task complexity	26
	Working language		*Working language	44, 45, 46, 47
	Cognitive development		*Cognitive development	22, 28, 49, 51
	Knowledge construction		*Knowledge construction	10, 36, 37
29, 35, 36	Purpose	Perceptions of DDL	Purpose	
1–3, 5–11	Effectiveness		Effectiveness	2, 6, 18, 29, 30, 32, 35
4, 28, 32–34, 37, 39	Continuous use		Continuous use	3, 14, 15, 27, 31, 33, 34
12–19		Perceptions of difficulties		5, 40-43
30, 38, 40		Suggestions for improvement		
		*Feedback of the optimized procedures		16, 17,19
				9, 11, 12, 13
		*Learning style		1, 4

modified in line with the optimized measures in Round 2.[1] Table 1 presents the constructs of the two questionnaires.

In Questionnaire C, to counterbalance the testing effect, many items were displayed in random order (see Table 1). Both questionnaires aimed to collect the students' perceptions of the DDL practice in three aspects: (1) the student-centered "mini-lecture"; (2) the practice of DDL; (3) the difficulties in the process. However, since they served different purposes, differences existed as well: Questionnaire B attempted to tentatively investigate learners' first perceptions of DDL in a more

1 Specifically, we removed the subcategory "Purpose" in DDL perceptions because we came to notice that students were fully aware of the purpose of DDL practice by the second round, and we reduced the number of the items concerning "perceptions of difficulty" and only kept the key items in Questionnaire C as the responses in Questionnaire B were highly consistent with previous findings. We added items concerning "knowledge construction, cognitive development, task complexity, working language, feedback of the optimized procedures, and learning style" (as shown with a "*" sign in Table 1). The purpose of doing so was to ask learners to reflect on their DDL experience, which allowed us to explore their cognitive development and perception changes in depth via the responses.

tentative way, while Questionnaire C focused more on learners sustained efforts for DDL practice and possible perception changes.[2]

Both Questionnaires B and C investigated learners' perceptions from three dimensions, namely, how they responded to the "mini-lecture" task (the Cronbach Alpha coefficient were 0.610 and 0.640), corpus-aided synonymous learning in general (0.813 and 0.802), and the difficulties in corpus consultation (0.648 and 0.551[3]). Additionally, we maintain that this study focuses on designing the learning model of DDL classroom operationalizations and examining its effectiveness, so the learners' perceptions of the "mini-lecture" task itself would be different from their overall perceptions of the DDL approach, namely, the use of the corpus, and the difficulty perception was helpful to reveal the potential factors influencing DDL in practice.

Semi-structured interviews were conducted at the end of each semester. To follow and investigate the perceptions (changes), we selected the same three focus groups (10 students) for the two interviews. The topics covered learners' use of corpus, the experience of preparing for the "mini-lecture", their comments on their presentation, and also the evaluation of their learning habits and difficulties.

4 Results

4.1 Learners' perceptions of DDL in the first round

A cohort of 20 learners participated in this study (17 females, 3 males). The results of Questionnaire A show that participants were positive toward English learning, and most frequently used electronic resources (88 %), including online websites (25 %), dictionary APPs (49 %), and electronic dictionaries (12 %). However, 94 % of them were not satisfied with their learning effects. The majority (91 %) used COCA mainly to query collocations (41 %), word meaning (16 %), and stylistic features (16 %), but none used it to discriminate synonyms.

To calculate and display the results of Questionnaire B and C more explicitly, we merged the responses of 1 and 2, 4 and 5 in the Likert scale respectively as "disagree" and "agree" in general. Results of Questionnaire B show that learners seemed to be

2 This is also the reason for the subtle difference in the wording of some shared items in the two questionnaires.

3 Since learners may have found it difficult to accurately recall their feelings, some deviations in their responses could have reduced the reliability of certain dimensions, particularly in the case of the difficulty dimension which had fewer items compared to Questionnaire B.

more positive toward the "mini-lecture" task (M = 4.33) than data-driven synonym learning (M = 3.76); the difficulties encountered were also at a moderate level (M = 3.24).

Moreover, no significant correlation between difficulty perceptions and the other two perception dimensions was found, indicating that the corpus consultation was not beyond learners' ability. Therefore, the positive responses paved the way for the second-round DDL implementation.

To further investigate the possible sources of learners' difficulties, Questionnaire B included items (B-20, 23, 24) about task motivation. Table 2 shows that both active and passive learners were negatively correlated with the difficulty perceptions, but the correlations were more significant for passive learners at the individual task level; this indicates some relevance between learners' task motivation and their perceptions of DDL.

Questionnaire B investigated learners' needs and willingness to continue this learning model (see Figure 3). Overall, learners were enthusiastic about DDL while proposing suggestions for improvements, especially on the need to increase corpus

Table 2: Correlations between difficulty perceptions and task motivation in the 1st round.

Difficulty perceptions	Motivation		
	Active: positive exploration (B-23)	Passive: use corpus only for completing individual assigned tasks (B-24)	Passive: use corpus only because the task was part of course activities (B-20)
Network instability (B-12)	−0.077	0.526*	−0.152
Time/energy cost (B-14)	−0.213	−0.101	−0.024
Need more training (B-19)	−0.082	−0.288	0.405

*$p < 0.05$.

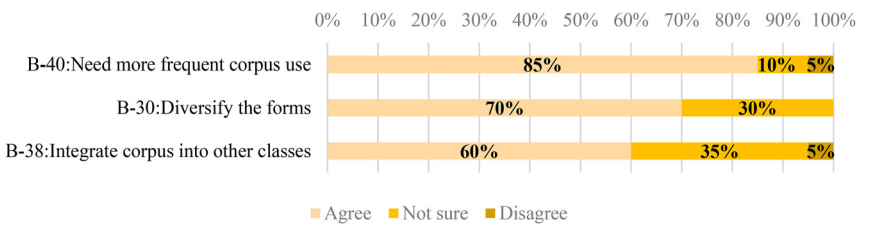

Figure 3: Suggestions for improving the teaching design.

use. DDL in the first round was a unit-based learning process, and provided only one opportunity for students to present their query outcomes. Limited corpus use hampered learners' understanding of DDL and its effectiveness. In addition, in the open questions, some learners responded that "the presenting forms and query skills used in each group were different, and this made some conclusions confusing and unreliable." The collected responses revealed learners' need for continuing this model and optimizing the operationalizations in the next round.

In the second round, we standardized and condensed the procedures into a worksheet with clear steps for the students to prepare and present the "mini-lecture" task, which provided scaffolding for learners to better understand and grasp the skills. The worksheet highlighted the process of observing concordance, formulating hypotheses, and verifying hypotheses, to promote cognitive development. In addition to breaking up tasks into smaller segments, the optimized instruction encouraged the teacher and peers to provide immediate comments and feedback as learners presented "mini-lectures." The activity of after-task vocabulary quiz design remained while corpus-aided peer-review writing practice[4] was added to increase the corpus use frequency. In short, the optimized steps and enhanced scaffolding further promoted learners' synonym learning by developing their cognitive ability and query skills in DDL.

4.2 Exploring learners' perceptions (changes) in the longer term practice

4.2.1 Comparison of learners' perceptions during two rounds of DDL

As aforementioned, Questionnaire C also investigated learners' perceptions from the three dimensions. Table 3 shows the comparison results between the two rounds DDL practice.

In the second round, it seems that learners still had positive responses to the "mini-lecture" task, and there was no significant difference between the two rounds; the standard deviation in the second round was smaller, indicating that learners' attitudes seemed to be consistent after the reinforcement. In the first round, 80 % of the students acknowledged their improvement in vocabulary knowledge; in the

4 There is an individual writing task with each unit; the prompt was in line with the unit topic. After completing the writing assignment, each student should have two peers review their writing with COCA to check their language use.

Table 3: Comparison between two-round perceptions of three dimensions.

Dimension	Round	Case number	Mean	Std.	Sig.
"Mini-lecture" perceptions	First	*N* = 20	4.33	0.617	0.899
	Second	*N* = 20	4.11	0.330	
DDL approach perceptions	First	*N* = 20	4.13	0.448	0.756
	Second	*N* = 20	4.12	0.471	
Difficulty perceptions	First	*N* = 20	4.08	0.518	0.428
	Second	*N* = 20	4.16	0.600	

second round, 75 % of students still held this perception (B-21, C-25). With the after-task vocabulary quiz, positive supporters increased from 70 % to 90 % (B-22, C-48); this supports the effectiveness of the optimized learning model.

As for the DDL approach, learners also had positive responses. Similar to "mini-lecture" task perception, no significant difference was found between the two rounds.

Nevertheless, it is worth noting that learners' attitudes toward future corpus use present a mixed and even contradictory picture. They acknowledged the effectiveness of corpus (see Figure 4) while still prioritizing dictionaries in vocabulary learning; this trend increased by 20 % in the two rounds. Also, the answers ("would not take DDL as a main learning method") to the open question (C-53) echoed this result although they tended to describe COCA as "accurate," "professional," or "good" (C-54).

As for difficulty perceptions (see Figure 5), the results show that learners encountered network instability, high time and energy cost, and inadequate query skills, but there was no significant difference between the two rounds.

The perceptions of network instability did not change, but those of time and energy costs increased (from 90 % to 100 %); the relative complexity brought by the

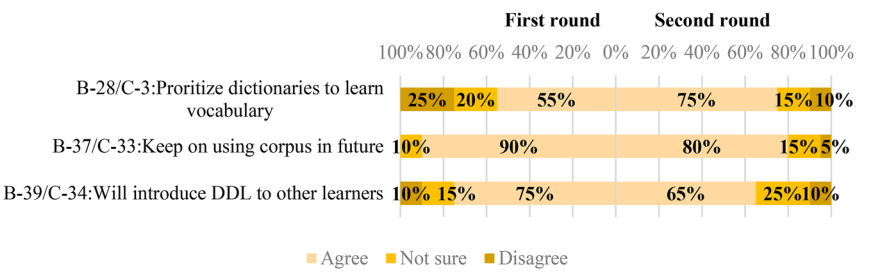

Figure 4: Learners' willingness of future corpus use.

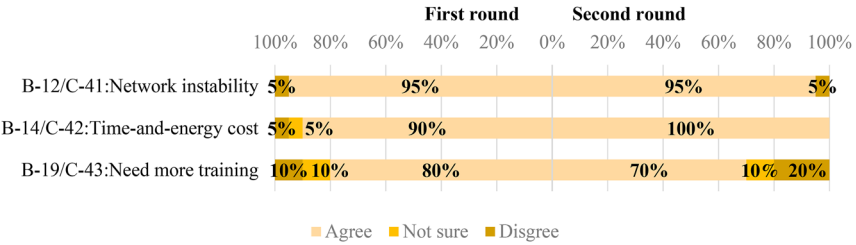

Figure 5: Learners' difficulty perceptions.

optimized and standardized procedures in the second round may contribute to the results. Meanwhile, the majority (90 %) perceived increased query skills development (C-6), and 80 % reported decreased anxiety (C-17). Accordingly, their demand for technical training also decreased, indicating that the optimized instruction provided efficient scaffolding for learners' DDL practice. This inference finds further evidence from the answers to the question (C-52) concerning difficulties in Questionnaire C, namely, the decreased perceptions of "network operation" (19 %), "time- and energy-consuming task" (18 %), "selecting proper synonyms" (15 %), and "generalization of results" (13 %). Generally, after two rounds of DDL, learners' skills of corpus consultation greatly improved. In the first round, 85 % admitted to having technical problems (C-5) while in the second round, 90 % realized that their query skills had improved (C-6).

Similar to the first round, no significant correlations were found between difficulty perceptions and the other two dimensions. The correlational analysis covers comparable dimensions in the two-rounds, while the new dimensions that emerged in the second round will be elaborated.

Also, we investigated the correlation between learners' motivation and difficulty perceptions.

In the first round, active learners did not display significant correlations with their difficulty perceptions at all levels, while passive learners did in terms of network stability (see Table 2). Therefore, in the second round, efforts were focused on further verifying this result. Table 4 indicates that learners who considered the "mini-lecture" task complex tended to be more sensitive about the time and energy cost, but not about the need for training, whereas learners who "only used corpus to complete tasks" showed significant negative correlations with "need for technical training." The results imply that the training before the DDL practice and the standardized procedures seem to have guaranteed learners' completing the "mini-lecture" tasks, which evidenced the effectiveness of the learning model.

Table 4: Correlations between difficulty perceptions and negative task motivation in the 2nd round.

Difficulty perceptions	Motivation	
	Passive: tasks are complex (C-26)	Passive: use corpus only for task (C-38)
Network instability (C-41)	−0.309	0.142
Time-consuming (C-42)	0.492*	0.371
Need more training (C-43)	−0.031	−0.511*

*$p < 0.05$.

4.2.2 The effectiveness of optimized instruction in the second round

Figure 6 shows learners' acknowledgment of the effectiveness of the standardized procedure. The majority believed that they became more skillful in using corpus (90 %, C-6), and the "mini-lecture" was better presented (95 %, C-19), and they highly appreciated the additional steps added in the second round.

Figure 7 displays how learners responded to knowledge construction and cognitive ability development. In the second round, 60 % of the learners valued peer collaboration, and 95 % valued teacher-student interaction and scaffolding. The majority were aware of cognitive development in themselves (90 %) and their peers (75 %), especially in language awareness (90 %). Meanwhile, they believed that the cognitive skills acquired in the second round could be transferred to other courses (85 %).

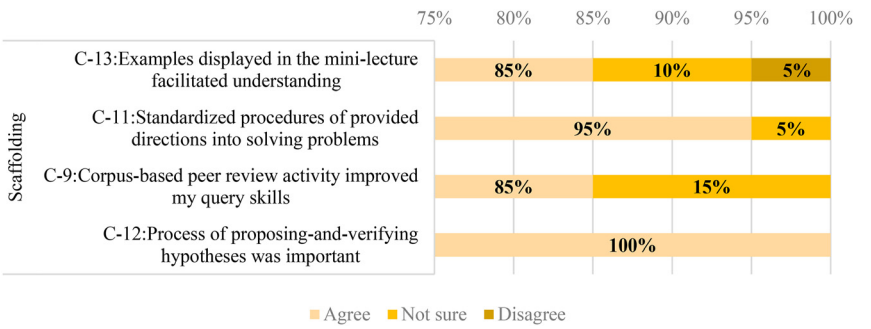

Figure 6: Perceptions of scaffolding provided.

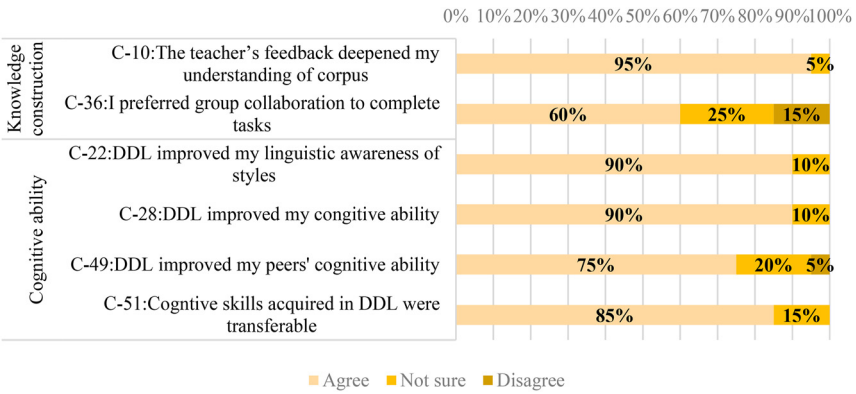

Figure 7: Perceptions of knowledge construction and cognitive development.

4.3 Potential factors influencing learners' perceptions

4.3.1 Task complexity

Since more steps were added to the "mini-lecture" design in the second round, it would be reasonable to consider the task complexity when analyzing the possible factors affecting learners' perceptions.

Overall, 15 students in the second round perceived increasing task complexity (C-26), which may be due to enhanced requirements and added steps in task completion. Despite this, task complexity displayed only weak correlation with the "mini-lecture" perceptions and no correlation with the dimensions of DDL or difficulty (see Table 2). Therefore, task complexity seemed to have little influence on learners' perceptions, indicating the practicality of the teaching design for learners' synonym learning.

4.3.2 Task motivation

It is worth noting that learners' motivation during two rounds experienced a significant change (p = 0.000). In the first round, only 25 % learners disagreed that "I used corpus only for in-class tasks" (B-20), but the proportion soared to 65 % in the second round (C-38). Though 35 % were still uncertain, the wider application of corpus in language learning could not be denied.

Tables 5 and 6 reveal the relationships between task motivation with learners' perceptions of "mini-lecture" and the DDL approach in the two rounds.

Table 5: Correlations of task motivation with learners' perceptions of "mini-lecture" and the DDL approach in the 1st round.

Motivation	Perceptions				
	DDL approach perceptions			"Mini-lecture" perceptions	
	Purpose (B-29, 35, 36)	Effectiveness (B-1, 2, 3, 5–11)	Continuous use (B-4, 28, 32–34, 37, 39)	Effectiveness (B-21, 22, 25)	Interaction (B-26, 27)
Active: positive exploration (B-23)	0.711**	0.121	0.254	−0.214	0.051
Passive: use corpus only for completing individual assigned tasks (B-24)	0.228	−0.231	−0.305	−0.467*	−0.198
Passive: use corpus only because the task was part of course activities (B-20)	−0.107	−0.044	−0.366	0.130	−0.095

*$p < 0.05$, **$p < 0.01$.

Table 6: Correlations of task motivation with learners' perceptions of "mini-lecture" and the DDL approach in the 2nd round.

Motivation	Perceptions				
	DDL approach perceptions		"Mini-lecture" perceptions		
	Effectiveness (C-2, 6, 18, 29, 30, 32, 35)	Continuous use (C-3, 14, 15, 27, 31, 33, 34)	Effectiveness (C-20, 24, 25, 48)	Interaction (C-7, 8, 10, 36, 37)	Cognition (C-22, 28, 49, 51)
Active: interesting task (C-39)	0.645**	0.419	0.213	0.454*	0.432
Passive: use corpus only for the task (C-38)	−0.493*	−0.453*	−0.267	−0.141	−0.653**

*$p < 0.05$, **$p < 0.01$.

In the first round (see Table 5), learners' motivation seemed relevant to their "mini-lecture" and DDL perceptions. The results show that those who explored the corpus actively could better understand and perceive the benefits of the task design.

Table 7: Correlation between task motivation and continuous-corpus-use perception in the 2nd round.

Motivation	Continuous-use perceptions			
	A dictionary is more effective for synonym learning (C-15)	A corpus is more effective for synonym learning (C-14)	Keep on using corpus (C-33)	Recommend corpus to others (C-34)
Active learners	−0.368	0.061	0.664**	544*
Passive learners	0.700**	−0.539*	−0.272	−0.698**

*$p < 0.05$, **$p < 0.01$.

The results in Table 6 not only support the findings from Table 5 but also display more details about their perceptions of cognitive development: positive motivation was significantly correlated with learners' DDL perceptions and peer interaction in the "mini-lecture" task. In contrast, a significant negative correlation was revealed between negative motivation with learners' perceptions of DDL, their choice of further corpus use, and their cognitive development (see Table 7).

Learners with positive motivation displayed a strong willingness to keep using corpus ($p < 0.01$) and introducing it to others ($p < 0.05$). By contrast, learners with negative motivation displayed a strong preference for a dictionary ($p < 0.01$) and reluctance to use corpus ($p < 0.05$). Moreover, the negative correlations show that they would not use corpus in the future and would not introduce it to others ($p < 0.01$). Task motivation seemed to be a factor affecting learners' perceptions in this context.

4.3.3 Language proficiency

Learners may be cognitively burdened if the learning activities are beyond their current L2 proficiency level (Allan, 2009; Lee et al., 2017, 2019). However, the existing literature has not come to an agreement regarding the suitability of DDL for learners of different language levels (Boulton, 2009; Mizumoto & Chujo, 2016). To further examine the role of L2 proficiency in the effectiveness of the DDL practice, this study divided the learners into groups according to their final scores of intensive reading courses[5] in the second semester of their sophomore year. Those who scored higher than the average were grouped in the higher-intermediate group ($n = 10$) while the others were in the lower-intermediate group ($n = 10$).

Table 8 shows that for different groups, in each round independent *t*-tests show no significant differences in each dimension; paired sample *t*-tests indicate no

5 This is the same type of course as the "Advanced English (I/II)" in this study.

Table 8: Comparison of perceptions of learners at different language levels.

Dimension	Language level	Round	Case number	Mean	Std.	Sig.
Learners' perceptions of "mini-lecture" tasks	Higher-intermediate	First	*N* = 10	4.45	0.550	0.343
		Second	*N* = 10	4.15	0.669	
	Lower-intermediate	First	*N* = 10	4.2	0.752	0.790
		Second	*N* = 10	4.1	0.699	
Learners' perceptions of the DDL approach	Higher-intermediate	First	*N* = 10	4.17	0.435	0.284
		Second	*N* = 10	3.98	0.540	
	Lower-intermediate	First	*N* = 10	4.08	0.481	0.358
		Second	*N* = 10	4.22	0.614	
Learners' perceptions of difficulties	Higher-intermediate	First	*N* = 10	4.33	0.631	0.749
		Second	*N* = 10	4.42	0.665	
	Lower-intermediate	First	*N* = 10	4.37	457	0.208
		Second	*N* = 10	4.21	0.471	

significant differences between the two rounds either. Generally, their perceptions of both "mini-lecture" tasks and the DDL approach slightly decreased, while the difficulty perceptions increased. In contrast, the lower-intermediate's difficulty perceptions decreased and their perceptions of the DDL approach increased.

To further analyze this phenomenon, the study examined the learners' difficulties. All the learners perceived higher time/energy costs, but the changes in perceptions were more significant in the high-intermediate group; this group displayed a significant negative correlation between the perception of time/energy cost and their DDL perceptions ($r = -0.679^{**}$), indicating that this group was more sensitive to the time/energy required for DDL ($p = 0.045^{*}$). Given these results, it was understandable that the high-intermediate group showed a slight increase in difficulty perceptions and a decrease in both the "mini-lecture" and DDL perceptions.

Further, we found that the demand for training in the second round decreased with both groups, and was more obvious for the lower-intermediate group ($M = 4 \rightarrow 3.22$) despite no significant difference. This slight change further confirmed that the optimized "mini-lecture" teaching design provided learners with adequate and effective scaffolding, especially for lower-intermediate learners.

The majority were aware of cognitive development, but no significant difference was found between the two groups ($p = 0.633$). This finding reflects the effectiveness of the "mini-lecture" instructional design in enhancing learners' cognitive development regardless of learners' language proficiency.

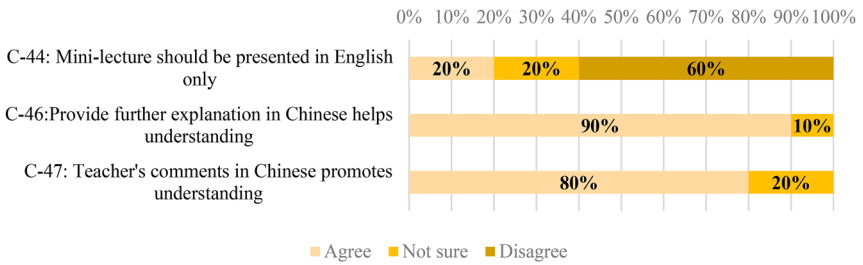

Figure 8: Learners' responses to the working language in DDL.

4.3.4 The working language in DDL

The cognitive load of data-driven synonym learning is fairly high, and may be greater if learners are required to use English for expressing technical terms in their "mini-lecture." Therefore, the working language for DDL is likely to be a potential factor affecting DDL effectiveness.

During the two-round DDL practice, learners predominantly used English to present the "mini-lecture," and the mother tongue (Chinese) was occasionally interspersed in peer feedback or teacher comments. However, Figure 8 shows that Chinese was far more favorable for explicit process explanations.

5 Results and discussions

5.1 Learners' perceptions of the DDL featuring task-based "mini-lecture"

This study conducted a task-based "mini-lecture" learning model with a guided approach and examined learners' perceptions from the dimensions of the "mini-lecture" tasks, the DDL approach, and the difficulties. In the first round, learners generally held a more positive attitude toward the "mini-lecture" tasks and the DDL approach, but they also encountered some difficulties. These findings echo the existing literature (see Boulton & Vyatkina, 2021). However, no significant correlations between learners' difficulty perceptions and the other two dimensions were found. Thus, we can infer that the learners acknowledged the "mini-lecture" teaching design and the DDL approach was within their reach in this context. However, learners suggested more practice with query skills to help improve their analysis of the query results.

5.2 Learners' perceptions (changes) of the long duration DDL practice

Accordingly, the second round of DDL sought to optimize the procedures and provide learners with more adequate scaffolding. Learners maintained positive responses to the "mini-lecture" tasks and the DDL approach. Again, no significant correlations were found between difficulty perceptions and the other two dimensions. Importantly, learners' perceptions of the three dimensions did not show significant differences between the two rounds.

There were some slight fluctuations in learners' perceptions during the two rounds. In the first round, many learners intuitively took DDL only as a compulsory task (only one-time practice for each student); in the second round learners viewed DDL with a wider perspective and the longer duration allowed them to reflect on its compatibility with their own learning needs. This change was evidenced in their lower willingness of future use. More explanations can be found with the task itself. Synonym discrimination poses a great challenge to language learners (Liu & Zhong, 2016), and the threshold for the data-driven approach would be much higher. In the interviews after the second round, some learners mentioned that "the whole process was time/energy-consuming, and they were even not sure whether the conclusions were convincing." They also commented "the synonym discrimination task was more suitable for learners of higher language proficiency because of their richer vocabulary knowledge to observe, hypothesize and testify language patterns from the concordances." Indeed, learners had different needs for vocabulary learning; this was confirmed with interviews and the answers to the open questions in Questionnaire C. Some studies have also shown that learners no longer use corpus for data-driven learning outside classroom training or activities (e.g., Crosthwaite & Cheung, 2019; Meunier, 2019).

5.3 Potential factors influencing learners' perceptions of the task-based DDL

This study also attempted to explore the potential factors (i.e., task complexity, task motivation, learners' language proficiency, and the working language) that may influence learners' perceptions. No significant correlation was found between task complexity and learners' perceptions of the three dimensions, implying the practicality of the teaching design for synonym learning. But learners with passive motivations were more likely to perceive difficulties in DDL (e.g., time/energy cost), and were less willing to use it in the future (see Table 7) despite their acknowledgment of

query improvements (see Figure 5). This supported the scaffolding effect of the instructional design. Besides, learners' perceptions did not seem to be influenced by L2 proficiency. This differs from Quan's (2016) findings, which showed advanced learners were the only suitable learners of DDL. However, we maintain that the main difference may not lie in learners' L2 proficiency but in the setting. To be specific, in Quan's study, the learners employed DDL in a self-directed manner, without support from peers or teachers, which constituted a big challenge for intermediate learners, who tended to view themselves as language learners, "thereby creating more interest in this new technique for improving their language skills" (Yoon & Hirvela, 2004, p. 277). Generally, if learners are equipped with the necessary and sufficient scaffolding, lower-intermediate learners could also integrate corpus into effective language learning and hold positive attitudes toward DDL. Additionally, most learners would welcome the use of their native language during the presentation of "mini-lecture," which proves the need for translanguaging (Li, 2018) in EFL context, for translingual practices in language classrooms help students to support one another, build rapport, assert their cultural identity, and enable meaningful access to in-depth knowledge construction (Cenoz & Gorter, 2015; Rajendram, 2021; Stavrou, 2020).

5.4 Further discussions

First, learners need to be trained to practice DDL and this takes more time. This finding is further supported in the current study. First, before implementing DDL, learners need training in reading concordances and induction skills (Yeh et al., 2007), which helps provide learners with directions for knowledge construction and avoids large deviations in the learning process. But these skills cannot be acquired at once and learners need more time to get familiar with the corpus tools (Ädel, 2010; Alsolami & Assrar, 2020). In the current study, learners' first response to DDL also showed the needs and the effectiveness was well illustrated by the learners' decreased training needs in the second round even though some were passive in completing the tasks. The significance was echoed in other studies; for example, Ädel (2010) commented that an isolated experiment with no follow-up work constituted a serious limitation of her research; Götz and Mukherjee (2006) also noted that their learners would have preferred a longer introduction. This is consistent with the literature (Cobb, 1997; Kennedy & Miceli, 2001; Yoon & Hirvela, 2004), and also supports the conclusion that "DDL can take many forms to meet the needs and proficiency of learners" (Mizumoto & Chujo, 2016, p. 62).

Second, sustainable DDL calls for the constructive alignment of the instruction design and language assessment. The results of this study reveal the need for

sustainability of DDL practice. Learners showed mixed and even contradictory attitudes toward future corpus use. This echoed previous studies that found the auxiliary role of a corpus in language learning (Boulton, 2009; Chambers, 2019); the interviewees in Meunier's study (2019) said they did not need corpora except for the class. This can be explained by the educational system and cultural traditions (Chan & Liou, 2005), namely, "the more authoritative roles of teachers in Chinese culture… (and learners) are more accustomed to deductive learning in which teachers presented rules in order to save time" (Yeh et al., 2007, p. 136). Lee and Lin (2019) also found that students raised in deductive and teacher-led reasoning educational settings may struggle with and even personally reject them. Similar findings in the studies indicate the need to increase the sustainability of the DDL practice. Revising some language assessment practices to prompt learners' motivation would be a solution (Meunier, 2019).

Third, the theoretical framework provides solid underpinnings for the DDL instruction design. Theoretical positioning is lacking in the DDL research (O'Keeffe, 2021; Pérez-Paredes, 2019). Using noticing hypothesis in SLA, constructivism, and sociocultural theories (Flowerdew, 2015) as the theoretical framework, this study offers an innovative example of how guidance can be fully employed in DDL, filling a gap "between researchers' claims about the importance of and the need for more guidance and the small number of attempts to empirically support these claims" (Boulton & Vyatkina, 2021, p. 79). The key features are: this instruction design well integrates DDL and TBLT, emphasizing more of teacher's guidance and learners' interactions in the process. This also highlights learners' problem-based learning and promotes their autonomy (Templeton & Timmis, 2023). Besides, learners' needs were taken into consideration, namely, the universal difficulty in learning synonyms and the freedom of choosing targeted synonymous words, which greatly trigger learners' motivation, creating conditions for cultivating self-regulated learning. It remains consistent with the zone of proximal development from sociocultural theory and represents differentiated teaching (Tomlinson, 2001). After the second round, learners indicated that the "mini-lecture" task expanded their understanding of target words from various perspectives, and scaffolding provided clues for reading concordances, which alleviated anxiety and improved the effect of the class presentation. This finding is consistent with previous studies (Jansen et al., 2019; Nelson, 1990). Further, this instructional design allowed for learner differences (i.e., L2 proficiency, and task motivation, among others), and displays the feasibility and generalizability in the EFL context. According to Templeton and Timmis (2023), these considerations make DDL more accessible to learners; they are also in line with Tomlinson's (2011) three basic criteria for designing language learning materials: be motivating and relevant to the learners; present an achievable level of challenge; draw attention to selected features of the input.

5.5 Task-based data-driven synonym learning model

Accordingly, we present the flowchart in Figure 9 to put forward the task-based data-driven synonym learning model with meticulous pedagogical details, which is uncommon in studies and can create replication challenges (Vyatkina, 2020). The instruction process consists of pre-task, in-task, and post-task stages. The dotted part represents the standardized procedures added in the second round. Overall, the three-stage procedure takes a guided inductive approach, which is "a hybrid of an inductive DDL approach and a deductive teaching method" (Mizumoto & Chujo, 2016, p. 58), with students' major role in discovering language and the scaffolding from the teacher and peers throughout the process.

Pre-task was the teacher-led training activities. Appropriate training helps learners understand the advantages of corpus so their attitudes can be transformed from using corpus only to complete in-class tasks to using corpus to promote language learning (Chambers, 2019; Lee et al., 2019; Poole, 2022). The in-task activity was based on the four-step strategy proposed by Kennedy and Miceli (2010): (1) formulating the question; (2) devising a search strategy; (3) observing the data and selecting examples; (4) drawing conclusions. Consultation with COCA provided learners with the KWIC display and drew learners' attention to the target synonyms, thus encouraging learners' noticing (Hyland & Milton 1997, as cited in Boulton, 2009).

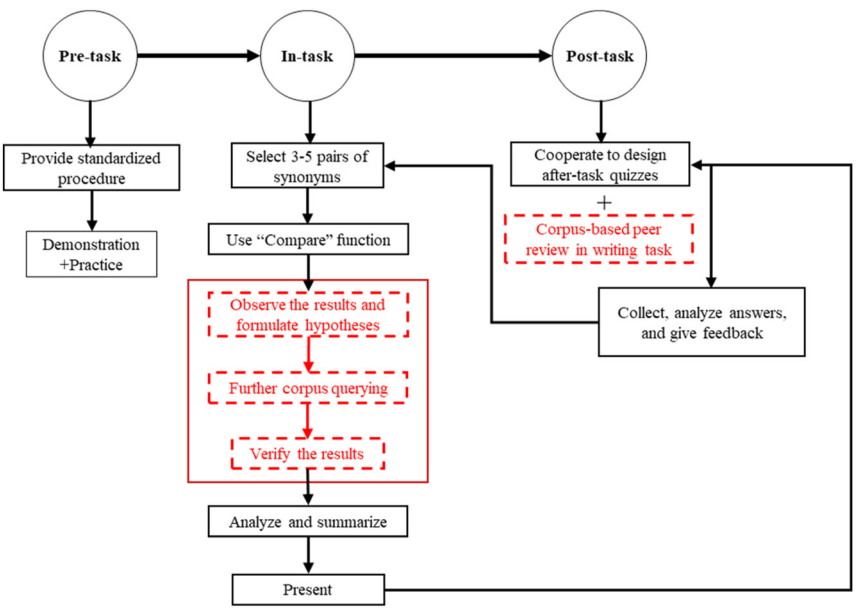

Figure 9: The student-centered data-driven "mini-lecture" instructional design.

Post-task was an after-class quiz and follow-up feedback. It helped verify learners' query results and construct synonym knowledge.

In this instructional design, some key elements such as task-based learning, teacher guidance, students' autonomy, sufficient exposure duration, and scaffolding from both teachers and peers, and so on were considered. It creates an immersive learning environment. Therefore, compared to traditional synonym learning, "guided induction could be seen as a promising methodological alternative" (Mizumoto & Chujo, 2016, p. 62).

6 Conclusions

This two-round action research examines the effectiveness of the data-driven instructional design with the guided inductive approach, addressing "how to make DDL play a role in language learning" (Boulton & Vyatkina, 2021, pp. 78–79). Learners displayed consistency in their perceptions of this model and became increasingly aware of the advantages of corpus in synonym learning. The effectiveness of the design was recognized and amplified with the optimized operationalizations. The teaching design generally ensured learners completed the "mini-lecture" tasks and embraced DDL with an open mind; it also emphasized the significant role of scaffolding in promoting knowledge construction and cognitive development in the EFL context.

This study presents an feasible model for implementing data-driven language learning. It underscores the significance of theoretical underpinning, and is a worthwhile attempt to bridge the gap between corpus research and teaching practice; foreign language teaching should encourage teachers to integrate modern technology to promote language pedagogy. Teacher training on corpus literacy, especially on corpus query skills, is also crucial to making concrete progresses in this aspect (Ma et al., 2023).

Nevertheless, there were some limitations. Synonymous knowledge was not examined during this long-term DDL practice. Learners' target synonymous words were based on their own needs, so it would be difficult to set up any pre-test. Data collected from learners' performance can be different from subjective self-report with questionnaires. Future studies can examine the synonymous knowledge through writing to check their use of the words in sentence production. Additionally, the sample size was small although we argue it could be a trade-off between a convenience sample and the long-term DDL operationalizations. Despite such limitations, the present study provides some teaching implications not only at the local EFL classroom level but also contributes to promoting the use of corpus consultation in language education.

Acknowledgments: We acknowledge the feedback and comments from the editors and reviewers, also Dr. Benjamin H. Nam's help with the proofreading.
Research funding: This research was funded by Shanghai International Studies University [grant number: 2022114015, 2022113043].

Appendix A

1. 学号 Student ID_____ 专业 Major_____

2. 姓名 Name_____ 微信号 WeChat ID_____

3. 性别 Gender_____ A. 男性 Male B. 女性 Female

4. 你喜欢学英语吗 Do you like learning English in general?_____

A. 非常喜欢 Extremely like B. 比较喜欢 Like C. 不清楚 Not sure

D. 不太喜欢 Dislike E. 非常不喜欢 Extremely dislike

5. 在出于个人目的上网的时候, 你是用英语还是母语? _____

When surfing the Internet for personal purposes, do you use English or Chinese? _____

A. 英语 English B. 中文 Chinese C. 二者使用差不多 Both

6. 你用中文上网的时间占比是多少? _____

How much of your total computer time is in Chinese? _____

A. 几乎没有 Almost none B. 大约 25 % About 25 % C. 大约 50 % About 50 %

D. 大约 75 % About 75 % E. 几乎全部 Almost all

7. 你在学习单词的时候, 是更侧重扩大词汇量还是词汇应用? _____

When learning vocabulary, you focus more on_____.

A. 扩大词汇量 Vocabulary expansion B. 词汇应用 Vocabulary application

C. 二者都有但偏重词汇量 Both but prefer vocabulary expansion D. 二者都有但偏重词汇应用 Both but prefer vocabulary application

E. 不知道 Not sure

8. 在接触 COCA 实际操作之前, 我的词汇学习方式主要是 (单选):_____。

Before using COCA , I learn English vocabulary mainly through (single choice) _____.

A. 查纸质版词典 Paper dictionaries B. 使用百度等在线搜索工具 Online websites

C. 查阅电子词典设备 Electronic dictionaries D. 使用词典 APP Dictionary APP

E. 我不怎么主动背单词, 主要通过上下文猜测词义。

I don't recite vocabulary but rely on guessing from context.

F. 其他 Others_____

(continued)

9. 你最常使用以下那种词典? _____ (可多选)

What kind of dictionary do you often use? Please check all that apply. _____

A. 英汉词典 Bilingual (e.g., English-Chinese) B. 英英词典 Monolingual (English-English)

C. 纸质词典 Paper dictionary D. 网络词典 Web dictionary E. 电子词典 E-dictionary

10. 你如何评价你目前的词汇学习方式?

How do you evaluate your current vocabulary learning method?

A. 我觉得很有效 I found it effective.

B. 我认识了一定词汇, 但是在写作或是口语表达的时候想不起来用
I know certain words, but I can't remember to apply them in writing or speaking.

C. 我认识了一定词汇, 但我不确定词语的具体搭配
I know certain amounts of words, but I'm not sure about the collocations.

D. 我认识了一定词汇, 但我不知道这些词语应该使用在何种语境下
I know certain words, but I don't know how to use them in different contexts.

11. 在这门课之前, 你有听过语料库吗? _____
Had you heard about corpus before you took this class? _____

A. 有 Yes B. 没有 No

12. 如果有, 你用过什么语料库? _____

If yes, which corpus did you use? _____

13. 当时使用语料库的目的是为了查询_____ (可多选)

At that time, you used corpus for information about _____ (Multi-choice).

A. 词义 Meaning B. 语义搭配 Collocation C. 词义辨析 Word discrimination

D. 语体的适用性 Style E. 都有 Both F. 其他 Others_____

Appendix B

以下问题是你对使用 COCA 的看法, 请根据如下量表圈出你认为最能反映你观点的数字。

The following questions are regarding your opinions on using the Corpus of Contemporary American English (COCA). Please use the scale below to circle the response that most closely resembles your perspectives.

1. 很不赞同 Strongly disagree
2. 有点不赞同 Somewhat disagree
3. 不清楚 Not sure
4. 有点赞同 Somewhat agree
5. 很赞同 Strongly agree

1. 通过词汇索引的输出更容易构建词汇的语义原型和搭配倾向。 It is easy to construct prototype strings by use of concordance output.	1 2 3 4 5
2. 构建词汇语义原型和搭配倾向丰富了我对词汇学习的认知。 Constructing prototype enhances my cognitive of vocabulary learning.	1 2 3 4 5
3. 我对构建词汇语义原型和搭配倾向有信心。 I feel confident in constructing prototype strings.	1 2 3 4 5
4. 对于我的英语词汇学习, 语料库比字典更有帮助。 The corpus is more helpful than a dictionary for my English vocabulary learning.	1 2 3 4 5
5. 语料库对学习词汇的含义很有帮助。 Using the corpus is helpful for learning the meaning of vocabulary.	1 2 3 4 5
6. 语料库对学习词汇的用法很有帮助。 Using the corpus is helpful for learning grammar.	1 2 3 4 5
7. 语料库对学习短语的用法很有帮助。 Using the corpus is helpful for learning the usage of phrases.	1 2 3 4 5
8. 语料库对学习同义词辨析有帮助。 Using the corpus is helpful for synonym learning.	1 2 3 4 5
9. 语料库增强了我的语言意识 (如词与词之间的搭配、单词在不同语体下的使用差异)。 Using the corpus enhances my language awareness (e.g., Word collocations, the usage of vocabulary in different styles).	1 2 3 4 5
10. 使用语料库提高了我的英语阅读能。 Using the corpus improved my English reading skill.	1 2 3 4 5
11. 使用语料库提高了我的英语写作能力。 Using the corpus improved my English writing skill.	1 2 3 4 5
12. 网站的不稳定性导致我在使用语料库时遇到了困难。 I have some difficulty in using the corpus due to network instability.	1 2 3 4 5
13. 未付费导致的检索限制导致我使用语料库进行词汇检索分析遇到了困难。 I have some difficulty in using the corpus due to the restricted times for search caused by unpaid accounts.	1 2 3 4 5
14. 分析数据需花费大量时间和精力, 导致我在使用语料库时遇到了困难。 I have some difficulty in using the corpus due to time and energy spent on analyzing the data.	1 2 3 4 5
15. 索引行句子不完整, 导致我在使用语料库时遇到了一些困难。 I have some difficulty in using the corpus due to cut-off sentences in concordance output.	1 2 3 4 5
16. 分析索引行结果对我来说有点困难。 I have some difficulty in analyzing concordance output.	1 2 3 4 5
17. 我在使用某种特定检索引擎功能的时候有点困难。 I have some difficulty in performing certain search techniques/functions.	1 2 3 4 5

(continued)

18. 语料库提供的原始语言文本很难理解。 The real texts in the corpus are too difficult to understand.	1 2 3 4 5
19. 我觉得老师还应该再持续加强对我们使用 COCA 的培训。 I recommend the teacher to continue to strengthen training on COCA practice.	1 2 3 4 5
20. 我使用语料库仅仅是为了完成与课程作业相关的任务。 I use corpus only when the teacher assigns in-class tasks.	1 2 3 4 5
21. 基于语料库的微型讲座有助于提升我的词汇学习效果。 The corpus-based "mini-lecture" activity improves my vocabulary learning.	1 2 3 4 5
22. 微型讲座后由同学设计的练习提升了我的词汇学习效果。 The self-designed exercises based on the "mini-lecture" improves my vocabulary learning.	1 2 3 4 5
23. 我 (或小组) 在准备微型讲座时会尽可能充分检索有效信息。 I (or my group) will search as much information as possible when preparing for the "mini-lecture".	1 2 3 4 5
24. 我在准备微型讲座时只负责自己分配到的部分。 I use corpus only for completing individually assigned tasks.	1 2 3 4 5
25. 准备微型讲座的过程提高了我语言学习的自主能力。 Preparing the "mini-lecture" enhances my self-regulated language learning	1 2 3 4 5
26. 在准备微型讲座时, 组员给我提供了一定技术操作方面的帮助。 My groupmates offer me technical helps when preparing for the "mini-lecture".	1 2 3 4 5
27. 在准备微型讲座时, 组员给我提供了一定分析总结方面的帮助。 My groupmates offer me analytical helps when preparing for the "mini-lecture".	1 2 3 4 5
28. 相比之下, 我更愿意查字典、百度检索或者直接听老师对词汇的讲解。 I still prioritize learning vocabulary through dictionaries, websites, or the teachers' instruction.	1 2 3 4 5
29. 我能理解老师在这门课上介绍并培训我们使用 COCA 的目的。 I understand the purpose of using the corpus in this course.	1 2 3 4 5
30. 我认为语料库辅助的词汇学习课堂活动形式还可以更加多样化。 I recommend more diversified corpus-based activities in future class.	1 2 3 4 5
31. 我更喜欢语料库检索后归纳总结的词汇学习方法。 I prefer inductive method of vocabulary learning that requires corpus query.	1 2 3 4 5
32. 目前, 我在高级英语课程以外的词汇学习遇到困难或没有把握时也会倾向于求助语料库进行检索验证。 I use the corpus when encountering difficulties or uncertainties outside the *Advanced English* course.	1 2 3 4 5
33. 我在语料库总能找到自己需要的信息。 When I search for information in the corpus, I usually get the information that I need.	1 2 3 4 5
34. 我对语料库的了解越多, 我就越喜欢它。 As I have learned more about the corpus, I have come to like them more	1 2 3 4 5
35. 对 COCA 的培训和应用使我意识到使用语料库进行词汇学习的重要性。Training and practice with COCA make me realized the importance of using corpus for vocabulary learning.	1 2 3 4 5

(continued)

36. 总体来说, 语料库对于我的英语词汇学习而言是一个很有用的资源。 Overall, the corpus is a very useful resource for my English vocabulary learning.	1 2 3 4 5
37. 我将会在以后的英语词汇学习中继续运用语料库。 I will use the corpus for my English writing in the future.	1 2 3 4 5
38. 我认为如果有更多课程引入语料运用会对我的语言学习有更大的提升。 I think there will be more improvement in my language learning if more courses introduce the use of corpus.	1 2 3 4 5
39. 我会将语料库推荐给周围其他的学生或者是其他学校的学生使用。 I will recommend the corpus to my schoolmates or students from other universities.	1 2 3 4 5
40. 我认为频次可以加大, 更多的操经验有助于我们提升利用语料库学习词汇的效率。 I think the use frequency can be increased to improve the efficiency of using corpus in vocabulary learning.	1 2 3 4 5
41. 学习使用 COCA 之后, 我在需要深度学习某些词汇的时候, 我会: After learning to use COCA, when I need in-depth knowledge of some words, I will: A. 先查词典再用语料库 look up into a dictionary then a corpus B. 先用语料库再用词典 look up into a corpus then a dictionary C. 查词典, 必要时使用语料库 look up into a dictionary and a corpus if necessary D. 只查词典 look up into a dictionary only E. 只查用语料库 look up into a corpus only	
42. 请问你对于语料库词汇学习还有什么其他意见或想法吗? Do you have any other suggestions or ideas about data-driven vocabulary learning? Please write it down.	

Appendix C

以下问题是你对使用 COCA 的看法, 请根据如下量表圈出你认为最能反映你观点的数字。

The following questions are regarding your opinions on using the Corpus of Contemporary American English (COCA). Please use the scale below to circle the response that most closely resembles your perspectives.

1. 很不赞同 Strongly disagree
2. 有点不赞同 Somewhat disagree
3. 不清楚 Not sure
4. 有点赞同 Somewhat agree
5. 很赞同 Strongly agree

1. 我喜欢自下而上的探索式学习方式。 I prefer inductive and exploratory learning method that requires searching before analyzing.	1 2 3 4 5
2. 通过语料库检索方式我能获得更多的词汇知识。 Using corpus is helpful for acquiring more vocabulary knowledge.	1 2 3 4 5

(continued)

3. 虽然现在我会借助语料库检索词汇, 但我还是更喜欢老师讲解词汇或者查字典学习词汇。 Although I use corpus for vocabulary learning, I still prioritize dictionaries or the teachers' instruction .	1 2 3 4 5
4. 我愿意学习并尝试新技能以促进学习。 I am willing to learn and try new techniques to promote learning.	1 2 3 4 5
5. 在上学期微型讲座准备过程中, 我在语料库的操作方面存在一些技术困难。 In the last semester, I had some technical difficulty in using corpus when preparing for the "mini-lecture".	1 2 3 4 5
6. 经过这学期的练习, 我在技术上明显更熟练, 能够进行更充分的词汇检索。 After practicing in this semester, I am able to do a fuller research of target words in corpus with more proficient skills.	1 2 3 4 5
7. 在准备微型讲座时, 组员给我提供了很多技术操作方面的帮助。 My groupmates offer me technical helps when preparing for the "mini-lecture".	1 2 3 4 5
8. 在准备微型讲座时, 组员给我提供了很多归纳分析方面的帮助。 My groupmates offer me analytical helps when preparing for the "mini-lecture".	1 2 3 4 5
9. 在每单元学习所要求的写作任务中融入基于语料库的同伴互改活动提升了我使用语料库的技能。 The corpus-based peer-corrective activity integrated in the writing task for each unit of study enhances my skills in using corpus.	1 2 3 4 5
10. 老师在小组展示微型讲座的过程中给予的点评和建议加深了我对使用语料库的理解。 The teacher's comments and suggestions for the "mini-lecture" presented by each group deepen my understanding of using corpus.	1 2 3 4 5
11. 这学期老师所展示的关于微型讲座准备与呈现的规定性要求 (例如提出—验证假设、运用例句补充说明、用表格呈现一组同义词的异同点等) 为我提供了使用语料库解决问题的指引。 The standardized procedures for "mini-lecture" preparation and presentation (e.g., proposing-verifying hypotheses, using examples for further explanation, using charts to present similarities and differences between target synonyms) offer me guidance of using corpus to solve the problems.	1 2 3 4 5
12. 在语料库中搜索辨析同义词的证据时, 我认为"提出—验证假设"这一步骤十分重要。 When searching for evidence for synonym discrimination in corpus, "proposing and verifying hypotheses" is a very important step.	1 2 3 4 5
13. 运用例句对检索和归纳结果进行补充说明加深了我对目标同义词的理解。 Using examples to further explain the concordance and induced results deepens my understanding of target synonyms.	1 2 3 4 5

(continued)

14. 我认为在辨析英语同义词方面, 使用语料库更加有效。 Corpus is more effective in English synonyms discrimination.	1 2 3 4 5
15. 我认为在辨析英语同义词方面, 使用词典(包括专门的同义词词典)或老师直接讲解更有效。 Dictionaries or the teacher's instruction is more effective in English synonyms discrimination.	1 2 3 4 5
16. 在上学期, 我对使用语料库进行英语词汇检索的学习方式感到焦虑。 Last semester, I felt anxious about using corpus for English vocabulary learning.	1 2 3 4 5
17. 经过这学期的强化练习, 我对于使用语料库检索和辨析词汇感到更轻松。 After practicing in this semester, I feel more relaxed about using corpus for concordance and word discrimination.	1 2 3 4 5
18. 经过这学期的强化练习, 我更加明确了使用语料库进行词汇学习的必要性和实用性。 After practicing in this semester, I am clearer about the necessity and practicality of using corpus for vocabulary learning.	1 2 3 4 5
19. 在这学期, 我(或小组)的微讲座的效果比上学期更好。 This semester, my (or my group's) "mini-lecture" is better than that in the last semester.	1 2 3 4 5
20. 我从使用语料库准备微型讲座的过程中收获很多。 I have gained a lot from using corpus for "mini-lecture" preparation.	1 2 3 4 5
21. 就基于语料库的微型讲座这种教学设计形式本身而言, 我觉得这是一种有意义的尝试。 The corpus-based "mini-lecture" teaching design itself is a meaningful try.	1 2 3 4 5
22. 数据驱动的同义词学习方法帮助我内化语言的语体适用性意识。 The data-driven synonym learning method helps me internalize the language awareness of stylistic applicability.	1 2 3 4 5
23. 我对自己所检索、验证过的目标同义词相关知识点的印象更深刻、持久。 I have a deeper and lasting impression on relevant information of target synonyms that I searched for and verified.	1 2 3 4 5
24. 我需要更多时间理解、消化其他同学在微型讲座中展示和总结的目标同义词知识。 I need more time to comprehend and digest relevant information of target synonyms presented in other groups' "mini-lectures".	1 2 3 4 5
25. 我能从其他小组的微型讲座展示中学习到很多词汇知识。 I acquired a lot of vocabulary knowledge from other groups' "mini-lectures".	1 2 3 4 5
26. 我觉得微型讲座任务比较复杂、麻烦。 I found the "mini-lecture" task annoying and troublesome.	1 2 3 4 5
27. 遇到词汇辨析或搭配问题时, 我会很自然地想到用语料库进行验证。 When encounter problems of word discrimination or collocation, I will instantly think of corpus for verification.	1 2 3 4 5

(continued)

28. 经过这学期的强化练习, 我认为自己在认知能力方面(观察、分析、对比、假设、验证等)有较大提升。 After practicing in this semester, my cognitive ability (e.g., observing, analyzing, comparing, hypothesizing, verifying, etc.) has enhanced.	1 2 3 4 5
29. 语料库有助于聚焦我的词汇学习范围。 Using corpus helps me spotlights certain scopes for vocabulary learning.	1 2 3 4 5
30. 在进行同义词替换时, 现在我会考虑到语体适用性的问题。 I will now consider stylistic applicability when interchanging synonyms.	1 2 3 4 5
31. 现在语料库已经成为我英语学习的常用工具之一。 Corpus has now become one of my normalized tools for English learning.	1 2 3 4 5
32. 数据驱动的同义词学习方法增强了我的自主学习能力。 The data-driven synonym learning method enhances my self-regulated learning.	1 2 3 4 5
33. 我将会在以后的英语词汇学习中继续运用语料库。 I will use the corpus for my English writing in the future.	1 2 3 4 5
34. 我会将语料库推荐给周围其他的学生或者是其他学校的学生使用。 I will recommend the corpus to my schoolmates or students from other universities.	1 2 3 4 5
35. 经过一个学年的探索和尝试, 我认为语料库在词汇学习中有更好、更持久的效果。 After exploration and practice in a school year, I think corpus has a better and more lasting effect in vocabulary learning.	1 2 3 4 5
36. 在进行语料库探索学习时, 我喜欢和同伴一起进行检索、分析、验证和归纳。 In corpus-based exploratory learning, I prefer searching, analyzing, verifying, and inducing results with partners.	1 2 3 4 5
37. 在进行语料库探索学习时, 我更相信自己的独立分析和判断 In corpus-based exploratory learning, I prefer searching, analyzing, verifying, and inducing results on my own.	1 2 3 4 5
38. 我使用语料库仅仅是为了完成与课程作业相关的任务 I use corpus only when the teacher assigns in-class tasks.	1 2 3 4 5
39. 我觉得使用语料库检索目标词汇很有趣。 I found using corpus for information of target words is interesting.	1 2 3 4 5
40. 我在使用某种特定检索引擎功能的时候有点困难。 I have some difficulty in performing certain search techniques/functions	1 2 3 4 5
41. 网站不稳定是我使用该技术的最大障碍。 I have some difficulty in using the corpus due to network instability.	1 2 3 4 5
42. 花费的时间和经历太多是我使用语料库的最大障碍。 I have some difficulty in using the corpus due to time and energy spent on analyzing the data.	1 2 3 4 5
43. 我还需要更多语料库使用技术方面的强化训练和指导。 I still need more guidance and training on technical skills of corpus use	1 2 3 4 5

(continued)

44. 我认为微型讲座的展示过程应该使用全英语进行讲解。 I think the “mini-lecture” should be presented in English only.	1 2 3 4 5
45. 若展示小组在微型讲座开始前先共享PPT, 我会更容易理解讲座中的内容。 I will comprehend the content more easily if the presenters share the PPT before the “mini-lecture”.	1 2 3 4 5
46. 若展示小组用中文补充讲解微型讲座中的重点内容, 我会对该内容产生更深刻的印象。 I will have a deeper impression on key points in the “mini-lecture” if the presenters illustrate them in Chinese.	1 2 3 4 5
47. 若老师用中文对微型讲座提供点评和建议, 我会更容易关注到并理解老师的反馈。 I will notice and comprehend the teacher’s feedback more easily if the teacher provide comments and suggestions in Chinese.	1 2 3 4 5
48. 微型讲座后由同学设计的练习体现了微型讲座中展示的词汇要点, 提升了我的词汇学习效果。 The self-designed exercises embody key points presented in the “mini-lecture”, which improves my vocabulary learning.	1 2 3 4 5
49. 通过我对不同微型讲座的观察, 我认为同学们在认知能力方面 (观察、分析、对比、假设、验证等) 有很大提升。 According to my observation, my classmates’ cognitive ability (e.g., observing, analyzing, comparing, hypothesizing, verifying, etc.) has enhanced.	1 2 3 4 5
50. 准备微型讲座的操作流程 (观察检索结果—提出假设—验证假设—提供更多例句) 拓展了我学习英语词汇的空间。 The procedures of the “mini-lecture” preparation (observing concordance results—proposing hypotheses—verifying hypotheses—providing more examples) offer me more perspectives of English vocabulary learning.	1 2 3 4 5
51. 上述认知技能会潜移默化地影响或迁移到我对其他事物的认知和学习中去。 The above cognitive skills will imperceptibly influence or be transferred to my cognition and learning of other things.	1 2 3 4 5
52. 对我而言, 目前在准备微型讲座的流程中仍存在的困难有 (可多选): _____ By far, when preparing for the "mini-lecture", I still have difficulty in: _____ (Multi-choice).	
A. 选合适的词进行对比 Selecting proper synonyms for comparison	
B. 对比的切入点选择(即检索的方法或技术问题) Choosing perspectives for comparison (i.e., concordance methods and query skills)	
C. 对检索结果的解读分析(即形成假设的过程) Analyzing the concordance results (i.e., the process of proposing hypotheses)	

(continued)

D. 对结果或假设的验证
Verifying the concordance results and hypotheses

E. 对比结果的最终总结和提炼(即对表格内容的呈现)
Generalizing and refining of results (i.e., the formation of charts for content exhibition)

F. 对PPT的展示 (即课上对微型讲座的展示)
Presenting the PPT (i.e., the in-class "mini-lecture" presentation)

G. 根据自己小组词汇知识点总结再设计练习
Designing exercises based on key points presented in my groups' "mini-lecture"

H. 耗费时间久
Dealing with the high time-and-energy cost

I. 网站自身问题(如服务器不稳定、未付费导致的检索次数受限等)
Dealing with website problems (e.g., network instability, restricted times for research)

J. 没有困难
None

53. 对我而言, 利用语料库技术手段进行词汇学习会是主要的英语词汇学习方法吗?
For me, will corpus-aided learning be a main English vocabulary learning method?

A. 是 Yes B. 否 No C. 不知道 Not sure

请给出所做选项的主要原因 Please explain the reason for the choice

54. 看到"语料库技术应用到词汇学习", 我所能想到的第一个词是: _____
When I see "the integration of corpus into vocabulary learning", the first word comes to my mind is: _____.

References

Ädel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. In M. C. Campoy-Cubillo, B. Bellés-Fortuño & M. L. Gea-Valor (Eds.), *Corpus-based approaches to ELT* (pp. 39–55). Continuum.

Ahmadian, M., & Darabi, A. (2012). A study of the relationship between EFL learners' knowledge of near synonyms and their collocational behaviour. *Journal of Educational and Social Research*, *2*(3), 179–187.

Alanazi, M. (2017). The comprehension of synonyms by Saudi EFL learners: Acquisition and pedagogical implications. *International Journal of Applied Linguistics and English Literature*, *6*(3), 58.

Alanazi, Z. (2022). Corpus-based analysis of near-synonymous verbs. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 1–25.

Allan, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal*, *63*(1), 23–32.

Alsolami, T., & Assrar, A. (2020). Saudi EFL learners' perceptions of the use of corpora in academic writing teaching. *Studies in English Language Teaching*, *8*(4), 94–111.

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* (pp. 15–36). John Benjamins.

Boontam, P. (2022). The effect of teaching English synonyms through data-driven learning (ddl) on Thai EFL students' vocabulary learning. *Shanlax International Journal of Education*, *10*(2), 80–91.

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

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, *25*(3), 66–89.

Cenoz, J., & Gorter, D. (2015). Translanguaging as a pedagogical tool in multilingual education. In J. Cenoz, D. Gorter & S. May (Eds.), *Language awareness and multilingualism* (pp. 309–321). Springer International.

Chambers, A. (2019). Towards the corpus revolution Bridging the research-practice gap. *Language Teaching*, *52*(4), 460–475.

Chan, T.-P., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, *18*(3), 231–250.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, *25*, 301–315.

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *Cambridge handbook of English corpus linguistics* (pp. 478–497). Cambridge University Press.

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393–410). Lawrence Erlbaum.

Crosthwaite, P., & Boulton, A. (2022). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In H. Tyne (Ed.), *Discovering language: Learning and affordance*. In press.

Crosthwaite, P., & Cheung, L. (2019). *Learning the language of dentistry. Disciplinary Corpora in the teaching of English for specific academic purposes*. Benjamins.

Divjak, D., & Gries, S. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, *2*(1), 23–60.

Doughty, C. J., & Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology*, *7*(3), 50–80.

Dushku, S., & Paek, Y. (2021). Investigating ESL learners' awareness of semantic prosody across proficiency levels. *Language Awareness*, *30*(3), 234–256.

Ellis, R. (2003). *Task-based language learning and teaching*. OUP.

Firth, J. R. (1957). *Papers in linguistics 1934–51*. Oxford University Press.

Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, *14*, 393–417.

Flowerdew, L. (2015). Data-driven learning and language learning theories. In A. Leńko- Szymańska & A. Boulton (Eds.), *Multiple affordance of language corpora for data- driven learning* (pp. 15–360). John Benjamins. Frankenberg-Garcia.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, *26*(2), 225–242.

Gilquin, G., & Granger, S. (2022). Using data-driven learning in language teaching. In *The Routledge handbook of corpus linguistics* (pp. 430–442). Routledge.

Götz, S., & Mukherjee, J. (2006). Evaluation of data-driven learning in university teaching: A project report. In S. Braun, K. Kohn & J. Mukherjee (Eds.), *Corpus technology and language Pedagogy: New resources, new tools, new methods* (pp. 49–67). Peter Lang.

Gries, S. T., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, *34*, 121–150.

Gu, B. J. (2017). Corpus-based study of two synonyms: Obtain and gain. *Sino-US English Teaching*, *14*(8), 511–522.

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In Memory of J. R. Firth* (pp. 148–162). Longman.

Harvey, K., & Yuill, D. (1997). A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing. *Applied Linguistics*, *18*(3), 253–278.

Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language acquisition*. Cambridge University Press.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, *6*, 183–205.

Jafarpour, A. A., Hashemian, M., & Alipour, S. (2013). A corpus-based approach toward teaching collocation of synonyms. *Theory and Practice in Language Studies*, *3*(1), 51–60.

Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review*, *28*(3), 100292.

Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 33–44). John Benjamins

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

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (Cambridge Applied Linguistics) (pp. 293–313). Cambridge University Press.

Kavaliauskiene, G. (2003). Correction and self-correction of written assignment at tertiary level. *Journal of Language and Learning*, *1*(2), 51–58.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus Investigation. *Language Learning & Technology*, *5*(3), 77–90.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, *14*(1), 28–44.

Khazaal, E. N. (2019). Investigating and analyzing ESP college students' errors in using synonyms. *International Journal of English Linguistics*, *9*(5), 328–339.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, *1*(1), 7–36.

Laufer, B. (1991). The development of lexis in the production of advanced L2 learners. *Modern Language Journal*, *75*, 440–448.

Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System*, *81*, 14–25.

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, *21*, 32–51.

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753.

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *Modern Language Journal*, *104*(2), 345–362.

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, *39*(1), 9–30.

Liu, D. (2010). Going beyond patterns: Involving cognitive analysis in the learning of collocations. *TESOL Quarterly*, *44*(1), 4–30.

Liu, D. (2013). Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics*, *24*(1), 67–113.
Liu, D., & Espino, M. (2012). Actually, genuinely, really, and truly: A corpus-based behavioral profile study of the near-synonymous adverbs. *International Journal of Corpus Linguistics*, *17*(2), 198–228.
Liu, D., & Zhong, S. (2016). L2 vs. L1 use of synonymy: An empirical study of synonym use/acquisition. *Applied Linguistics*, *37*(2), 239–261.
Ma, Q., Chiu, M. M., Lin, S., & Mendoza, N. B. (2023). Teachers' perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL*, *35*(1), 19–39.
Martin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal*, *68*(2), 130–137.
Meunier, F. (2019). Data-driven learning: From classroom scaffolding to sustainable practices. *EL.LE: Educazione Linguistica/Language Education*, *8*(2), 423–434.
Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, *22*, 1–18.
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, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge University Press.
Murphy, M. L. (2009). *Semantic relations and the lexicon*. Cambridge: Cambridge University Press.
Nasser, S. M. (2021). A cognitive linguistics diagnosing of Iraqi EFL students' difficulties in using synonyms. *International Journal of Language and Literary Studies*, *2*(3), 135–145.
Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125–173.
O'Keeffe, A. (2021). Data-driven learning, theories of learning and second language acquisition, in search of intersections. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 35–55). John Benjamins.
O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, *15*(1), 49–68.
Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, *35*, 36–61.
Phoocharoensil, S. (2020). A genre and collocational analysis of consequence, result, and outcome. *3L: Southeast Asian Journal of English Language Studies*, *26*(3), 1–16.
Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms *persist* and *persevere*. *Eurasian Journal of Applied Linguistics*, *7*(1), 240–258.
Phoocharoensil, S. (2022). Primary, main, and major: Learning the synonyms through corpus data. *GEMA Online Journal of Language Studies*, *22*(4), 76–89.
Poole, R. (2022). "Corpus can be tricky": Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, *35*(7), 1620–1641.
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.
Quinn, C. (2014). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, *69*(2), 165–177.
Rajendram, S. (2021). Translanguaging as an agentive pedagogy for multilingual learners: Affordances and constraints. *International Journal of Multilingualism*, *20*(3), 1–28.
Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, *54*(2), 121–134.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Shaffer, C. (1989). A comparison of inductive and deductive approaches to teaching foreign languages. *The Modern Language Journal*, *73*(4), 395–403.

Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (Eds.), *In Memory of J.R. Firth*. Longman.

Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language*. Oxford University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2003). Focus on form, tasks, and technology. *Computer Assisted Language Learning*, *16*, 391–411.

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, *26*, 184–201.

Song, Q. (2021). Effectiveness of corpus in distinguishing two near-synonymous verbs: "Damage" and "destroy". *English Language Teaching*, *14*(7), 8–20.

Soruç, A., & Tekin, B. (2017). Vocabulary learning through data-driven learning in an English as a second language setting. *Educational Sciences: Theory & Practice*, *17*(6), 1811–1832.

Stavrou, S. (2020). Introducing translanguaging in the educational context of cyprus: A dynamic view of spaces of translanguaging in a primary classroom. *Anthropology & Education Quarterly*, *51*(4), 440–458.

Sun, K. T., Huang, Y. M., & Liu, M. C. (2011). A WordNet-based near-synonyms and similar-looking word learning system. *Educational Technology and Society*, *14*(1), 121–134.

Templeton, J., & Timmis, I. (2023). A flexible framework for integrating data-driven learning. In K. Harrington & P. Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 39–58). Palgrave Macmillan.

Thienthong, A. (2020). Bilingual dictionary synonyms for paraphrasing: A solution or a problem? *International Journal of Lexicography*, *33*(1), 17.

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Association for Supervision and Curriculum Development.

Tomlinson, B. (Ed.). (2011). *Materials development in language teaching*. Cambridge University Press.

Tsai, K. J. (2019). Corpora and dictionaries as learning aids: Inductive versus deductive approaches to constructing vocabulary knowledge. *Computer Assisted Language Learning*, *32*(8), 805–826.

VanPatten, B., & Benati, A. (2010). *Key terms in second language acquisition*. Continuum.

Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals*, *53*(2), 359–370.

Wongkhan, P., & Thienthong, A. (2021). EFL learners' acquisition of academic collocation and synonymy: Does their academic experience matter? *RELC Journal*, *52*(3), 523–538.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103–129.

Yao, G. (2019). Vocabulary learning through data-driven learning in the context of Spanish as a foreign language. *Research in Corpus Linguistics*, *7*, 18–46.

Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, *20*(2), 131–152.

Yevchuk, A. (2022). An empirical study of near-synonym choice: A comparison of advanced EFL learners to L1 English speakers. *Taikomoji Kalbotyra*, *17*, 79–94.

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

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257–283.

Ziegler, N. (2016). Taking technology to task: Technology-mediated TBLT, performance, and production. *Annual Review of Applied Linguistics*, *36*, 136–163.

Bionotes

Yingying Yang
School of Education, Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China
yyyzhuye@163.com
https://orcid.org/0000-0001-7343-4596

Yingying Yang received her PhD in Applied Linguistics from Shanghai International Studies University and is currently an Associate Professor at the School of Education, and Institute of Corpus Studies and Applications, Shanghai International Studies University, China. Her research interests include corpus linguistics, foreign language teaching, and second language writing.

Lin Chen
School of Education, Shanghai International Studies University, Shanghai, China
chenlin36@foxmail.com

Lin Chen received her Bachelor's degree in English Education from Shanghai International Studies University and is currently a graduate student majoring Curriculum and Instruction at Shanghai International Studies University. Her research interests include corpus linguistics and foreign language teaching.

Xumin Tian
Bijie Institute of Education Science, Bijie, China
2357165965@qq.com

Xumin Tian received her Bachelor's degree in Pre-school Education from Guizhou Education University and is currently a graduate student majoring Curriculum and Instruction at Shanghai International Studies University. Her research interests include corpus linguistics and foreign language teaching.