CURRICULUM & TEACHING STUDIES | RESEARCH ARTICLE

Using data-driven learning activities to improve lexical awareness in intermediate EFL learners

Liya Xue

Abstract: A large number of empirical studies have demonstrated the impact of lexical chunks on foreign language learning. However, reliable and efficient use of the lexical approach remains a problem. In this paper, a revised chunk teaching approach, based on Data-Driven Learning (DDL) activities and language awareness theory, is proposed and tested. Taking an action research approach, the effects of DDL activities on increasing the lexical awareness of intermediate EFL learners were investigated. A 16-week course, using the teaching pattern of EIODI (Exposure-Identify-Observe-Discover-Internalize) along with recognition and production tasks, was used to teach 158 intermediate EFL first-year university students. Quantitative questionnaire data from an in-class survey and a university course assessment survey revealed that the learners increased their awareness of lexical chunks and developed positive attitudes toward this approach. In addition, they reported a positive impact on their English proficiency; claims were supported by a statistical comparison of Finals for College English Course test scores for the two succeeding academic years. This research highlights the importance of increasing learners’ language awareness and using effective DDL activities.

Subjects: Databases; Bilingualism / ESL; Teaching & Learning

Keywords: lexical chunks; lexical awareness; DDL activities; intermediate EFL learners

ABOUT THE AUTHOR

Dr. Xue Liya woks as a language lecturer at the School of International Studies of Zhejiang University in China and as a researcher in the Institute of Linguistics and Applied Linguistics in Foreign Languages of Zhejiang University. Her research interests are second language acquisition and speech prosody. Previously she had done some research on prosodic disambiguation in second language at the University of Pittsburgh in USA as a visiting scholar. In recent years, she has been working on a teaching reform project on the promotion of language awareness in EFL classroom. It’s found that lexical awareness can be effectively raised through Data-Driven Learning (DDL) activities, which is reported in this paper. Future work will be more focused on prosodic awareness in language education.

PUBLIC INTEREST STATEMENT

In foreign language teaching, intermediate students pose a big challenge to language teachers, partly because the students continue to have gaps in their language knowledge that need to be filled, and partly because their progress is less noticeable than that of beginners; thus, their motivation to learn tends to drop. In this paper, a revised chunk teaching approach, based on Data-Driven Learning (DDL) activities and language awareness theory, is proposed and tested. The results revealed that the learners increased their awareness of lexical chunks and developed positive attitudes toward this approach. This study highlights the importance of increasing learners’ language awareness and using effective DDL activities.
1. Introduction
According to CEFR (Common European Framework of Reference for Languages) language proficiency level descriptors (as cited in Figueras, 2012), the majority of Chinese first-year university students are B1 (intermediate) or B2 (upper-intermediate) EFL learners. With ten or so years of English education at school, these students are familiar with the basic grammatical structures of English and with moderate English vocabulary and can usually “produce simple connected text on topics which are familiar or of personal interest” (B1-Intermediate), but have difficulty in “using language flexibly and effectively for social, academic and professional purposes” (C1-Advanced). To be specific, Chinese first-year university students often find themselves plateaued at a certain level of English competence, producing simple speech and using Chinglish expressions when writing. Many learners have difficulty breaking through this “plateau stage” (Chen & Cheng, 2018).

Intermediate students pose a big challenge to language teachers, partly because the students continue to have gaps in their language knowledge that need to be filled, and partly because their progress is less noticeable than that of beginners; thus, their motivation to learn tends to drop. Lewis (2000, p. 14) argues, “The reason so many students are not making any perceived progress is simply because they have not been trained to notice which words go with which. They may know quite a lot of individual words which they struggle to use, along with their grammatical knowledge, but they lack the ability to use those words in a range of collocations which pack more meaning into what they say or write”. Here “collocations” also refers to other terms, such as lexical phrases, formulaic sequences, and lexical chunks. Lewis’s statement explicitly explains why intermediate EFL learners fail to flexibly and idiomatically use the English language. Although in recent years, collocation learning has been the focus of research on vocabulary learning and teaching for second language learners (Wang & Yang, 2020), studies have repeatedly demonstrated the difficulties encountered by learners in collocation learning and use (Hsu & Chiu, 2008).

Traditional English language teaching in China approaches vocabulary and grammar separately and pays little attention to lexical items, which leads to the fact that “new word acquisition is not based on lexical grammar, so meaning is separated from colligation and collocation” (Li, 2018b, p. 498). As a result, learners’ active vocabulary is often composed of a limited number of lexical chunks, as learners have not even been “trained to notice” these lexico-grammatical units. In other words, they have not developed an authentic awareness of lexical chunks. Therefore, despite years of hard work in English classes, their lack of vocabulary knowledge is often cited as one of the obstacles they struggle to overcome (Zheng, 2009).

To break the plateau of English learning of intermediate EFL learners caused by the lack of lexical chunks, this study experiments on a teaching approach based on data-driven learning activities and attempts to raise learners’ lexical awareness.

2. Theoretical frameworks

2.1. Lexical chunks
Wray (2002, p. 9) defines lexical chunks as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated; that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”. Originating in the field of psychology, chunking is considered the all-important principle of human cognition, implying “the ability to build up structures recursively; thus leading to a hierarchical organization of memory” (Newell, 1990, p. 7). Some applied linguists argue that this concept can be applied to language acquisition and, rather than taking the traditional view that language is composed of grammar and vocabulary, it consists of lexical chunks, which are the essential mediator between conceptualization and grammatical and phonological encoding (Lewis, 1993, 1997; Nattinger & DeCarrico, 1992; Leech, 1983; Levelt, 1989). Thus, the lexical approach proposed by Lewis (1993) focuses on teaching lexico-grammatical units and encouraging learners to perceive these units as a whole and store
them in the mind as such. These units fall into four types: 1) polywords, such as by the way, in a nutshell, to sum up; 2) collocations, such as meet one's needs, striking increase; 3) institutional expressions, such as how do you do, nice meeting you; and 4) sentence frames, such as A new study suggests that… (Lewis, 1997). Since these chunks serve the basic units of human memory, they play an essential role in the learning process. An average native speaker knows hundreds of thousands of chunks, thus achieving a taken-for-granted degree of real-time fluency (Pawley & Syder, 1983).

In recent decades, there has been a growing recognition of the importance of lexical chunks in foreign language acquisition. Research has found a strong relationship between lexical chunks and second language skills. Ding and Qi (2005), for example, have found that the ability of second language learners to use formulaic language is a better predictor of oral and written English performance than grammatical accuracy. In other words, good language learners are good users of formulaic sequences. Boers (2006) has suggested that it is helpful for language learners to establish a set of formulaic sequences to improve their oral proficiency, while Mohammadi and Enayati (2018) have found that the input of lexical chunks helps to avoid errors caused by vocabulary selection and cultural differences, thus improving the accuracy of spoken English. Kazemi et al. (2014) have demonstrated that lexical chunks help students to improve their writing ability and argue that all writing courses should pay attention to the teaching of lexical chunks. Ilaria (2015) argues that “The most talented reader is the one who is able to identify the set of patterns which the words belong to and continually guesses the following item, in a steady dialogue between the so-called ‘Encyclopedia’, and the new information.” Research by both Tang (2013) and Li (2018a) has revealed that the number of lexical chunks that a learner possesses is closely related to listening performance, suggesting that lexical chunks can improve the efficiency with which a learner processes language information. Although a number of empirical studies demonstrate the impact of chunk input on EFL learning, there has been limited research into how to reliably and effectively teach lexical chunks.

2.2. Language awareness

Language awareness refers to “knowledge about language”, which encourages “the development in learners of an enhanced consciousness of and sensitivity to the forms and functions of language” (Carter, 2003, p. 64). A language awareness approach has been developed in contexts of both mother-tongue language education and second language acquisition. A key element of this approach is that learners “discover language for themselves” (Bolitho et al., 2003, p. 251). As a concept that has long played an important role in language education, language awareness is more often associated with more prescriptive approaches, characterized by, for example, the 1980s method of analyzing language form. It has since evolved beyond “knowledge about language”, or a focus on language itself, to emphasizing the cognitive advantages of reflecting on language, and to the belief that learners’ attitudes towards language and language learning can be changed by methods which highlight specific language features through learners’ active involvement (Bolito & Tomlinson, 1995). Therefore, language awareness emphasizes both language structure and language use, and considers both formalistic and communicative language teaching methodologies. More specifically, language awareness involves awareness of both comprehension and use of grammar and awareness of the use of language in practical communication (Yu, 2016). Traditionally, language awareness research has been confined to the study of grammar (Ellis, 1998; Valeo, 2013). This has also traditionally been the case in China (Pang, 1996; Peng, 1999). Meanwhile, awareness of the use of language in practical communication has been neither extensively researched nor developed. As “fixed or semi-fixed frequently-used syntactical structures with discourse functions” (He, 2016, p. 143), lexical chunks are units that can help learners connect language structure with language use. As Lewis (1993, pp vi–vii) notes, “A central element of language teaching is raising students’ awareness of, and developing their ability to, ‘chunk’ language successfully.” Therefore, the value of enhanced “noticing” and of “consciousness raising” of lexical chunks in second language learners should be emphasized.
However, although the traditional present-practice-produce (3P) mode of teaching may work for grammar, it is less successful in raising learners' awareness of the pragmatic function of lexical chunks in authentic contexts, leaving learners “void of sensitivity of Those chunks which, though grammatically correct, do not conform to the idiomaticity of native use” (Yu, 2016, p. 14). Based on a corpus study (Wei, 2007), lack of pragmatic quality, including interaction, cooperation, courtesy, and appropriateness of discourse, is a major problem in the spoken English of Chinese university students. Pragmatic failure is an important cause of cross-cultural communication breakdown. Misshapen, hybrid language might be described as “English with Chinese characteristics” is commonly found in EFL writing and translation, even among some who are highly trained and experienced (Pinkham, 2000). The reason for this is an inadequate store of pragmatic lexical chunks. Therefore, a revised approach to chunk teaching based on Data-Driven Learning (DDL) activities is proposed.

2.3. Data-driven learning

Data-Driven Learning is a term coined by Tim Johns, who stated that DDL is “the attempt to cut out the middleman as far as possible and give the learner direct access to the data” (Johns, 1991, p. 30). Through direct exposure to real-world contexts in the form of data, learners “gradually get aware [sic] of observing linguistic phenomena, analyzing and summarizing language features, and in this way, strengthen their language abilities” (Zhang, 2018, p. 107). In this paper, “data” include both attested uses of language recorded in real communicative contexts and authentic uses of language stored in language corpora and accessed predominantly through concordancers. Hassan (2018, p. 24) describes a corpus as a “collection of naturally occurring language stored on a computer and used to tell how language is used”. Compared to the 3P mode of teaching, DDL focuses on authentic language exposure and student-centered exploratory learning.

Over the past three decades, a growing body of empirical DDL research demonstrates its positive effects on promoting learner autonomy, increasing language awareness, enhancing noticing skills, and extending learners’ cognitive abilities (Boulton, 2009, 2010; Lin & Lee, 2015; Luo, 2016). In recent years, more studies have been conducted on the role of DDL in the EFL classroom. For example, DDL activities have been proved effective in English preposition learning and increasing learners' grammatical knowledge (Boontam & Phocharoenasil, 2018), in the acquisition of Lexico-grammatical Patterns in EFL Writing (Yılmaz, 2017), in improving collocational competence (Hua & Azmi, 2021). However, to date corpus tools still have not “fully ‘arrived’ on the pedagogical landscape” and the practice of foreign language teaching seems to have been largely unaffected by advances in corpus research, with relatively few teachers and learners aware of the availability of useful resources and having personal access to corpora (Römer, 2010, p. 18). Besides the low awareness and inaccessibility to corpora, reasons may also lie in teachers’ lack of detailed practice guidance of DDL activities, the difficulty in integrating DDL into the training of a certain language skill, such as lexical ability, the uncertainty of the effectiveness of DDL and students’ attitudes towards DDL approach.

Therefore, in this study a teaching pattern—EIODI (Expose-Identify-Observe-Discover-Internalize) with DDL activities is designed and practised to answer the following three questions:

1. Can the EIODI teaching pattern with DDL activities increase the learners’ awareness of lexical chunks?
2. Does the improvement of lexical awareness influence the learners’ English proficiency?
3. What are the learners’ attitudes toward this DDL approach?

3. Methodology

3.1. Context

Based on their College Entrance Examination English test scores, all first-year Zhejiang University students are classified into English level Band 3 (intermediate) and Band 4 (upper-intermediate),
among which the top 30% are Band 4 and the remaining 70% are Band 3. Research was conducted at an optional semester-long College English Course open to first-year students from Zhejiang University with English level Band 3. The course consisted of twice-weekly 90-minute sessions for 16 weeks. The classroom was equipped with computers and Wi-Fi internet connection.

3.2. Population
Four parallel classes consisted of a total of 158 non-English majors, ranging in age from 18 to 20 years, of whom 58 were female (36.7%) and 100 were male (63.3%). The vast majority of students had studied English at school for ten years, while a small minority had only six years of school-level English. All students were Chinese, and none had reported learning disabilities. All students agreed to participate in this study.

3.3. Data used in DDL
Two types of data were selected for DDL in this study. The first type of data was iWeb corpus (https://corpus.byu.edu/iweb) released in May 2018. A number of features made this an ideal choice. This corpus is approximately 14 billion words in size, making it 25 times as big as COCA (560 million words). It allows the user to create a “virtual corpus” for any topic, and to access a wide range of searches, including words, phrases, substrings, lemmas, part of speech, synonyms, and customized wordlists. A unique feature of iWeb, which makes it very useful for language learners and teachers, is the ability to browse through a list of the top 60,000 words (lemmas) in the corpus, together with a wide range of information on each of those words. Appendix 1 contains an example of a word home page. Zhejiang University has officially purchased the iWeb corpus, so that students have free access. This affordance enables students to have direct interaction with the real uses of words, collocations, and clusters. The second type of data was the authentic materials in real communicative contexts, including TED talks (www.ted.com), news articles, The World of English magazine, original English language books, and New College English textbooks. Of these, TED talks, BBC news, and New College English VLS (Viewing, Listening and Speaking) were used for listening input, and The World of English, original English language books, and the New College English textbook were used for reading input. Given that interest may increase motivation to learn, students were given the freedom to choose some materials, while some were chosen by the teacher and related to textbook topics so that the learning process could be easily monitored. The students had easy access to all the above resources via the internet and the university library.

3.4. Procedure
During the first two weeks, the course aims were explained and key concepts, such as lexical chunks, language awareness, and DDL, were introduced. The students were then provided with basic information about corpus-based learning and were guided in the use of iWeb corpus.

For the remaining 14 weeks, an EIODI (Expose–Identify–Observe–Discover–Internalize) pattern of teaching was adopted. First, students were exposed to real language data. In addition to the use of textbooks, each week the students were required to listen to one TED talk chosen by the teacher relating to the textbook topic and one news report chosen by the students relating to their own interests. In addition, they were required to read one magazine article chosen by the teacher relating to the textbook topic and one chapter of an original book chosen by the students and relating to their own interests. Next, in the “identify” stage of learning, students were required to establish the communication breakdown in both the reading and listening tasks and make a list of the words they were most interested in learning. This process enabled students to develop sensitivity to the use of words in different language styles and in diverse contexts. In the “observe” stage, students were required to use the iWeb corpus to investigate these words by examining high-frequency collocates, clusters, and topics, so that when these words were clearly presented in real world contexts, the students could better perceive and comprehend the language habits of native speakers (see Appendix 2). In the “discover” stage, students were required to explore the rules and patterns of language use by reading the iWeb concordancers of these words and phrases.
through both direct corpus search and indirect corpus search, using concordance lines extracted by
the teacher. Students were required to take note of patterned language (lexical chunks). In the
“internalize” stage, students were required to participate in both recognition and production
activities. Recognition activities, including cloze tests (see Appendix 3), paraphrasing, dictation,
correcting and collocation matching, were used to test how well the students recognized the
lexical chunks. Production activities involved summary writing (see Appendix 4) and public speak-
ing. Students were required to write summaries of news stories and magazine articles and were
assessed on their use of lexical chunks. Well-written summaries were shared on the QQ social
media group of which all participants were members. Public speaking took the form of a group
discussion in which students were required to use newly learned chunks from the TED talk to
express their understanding and opinions of the talk.

The EIODI teaching pattern was used both in and outside of class. Textbook articles and text-
book VLS materials were predominantly used in class. The teacher helped students to identify
chunks and to develop sensitivity to the functions of chunks in diverse contexts and in various
language styles (see Appendix 5). Direct concordancers from iWeb corpus were used as real-world
examples and the students observed language patterns in order to improve their ability to perceive
native English language use. Meanwhile, students were required to follow the EIODI pattern to
learn from other authentic materials outside of class and to accomplish tasks, such as paraphras-
ing, correcting, collocation testing, public speaking, summary writing, and group sharing.

4. Results and analysis

4.1. Data collection

Student feedback on course content, teaching quality, and overall learning experience has been
found to be an important guide for course design and development, and plays a significant role in
the delivery of high-quality, student-centered education (Leckey & Neil, 2001; Steyn et al., 2018).
In order to acquire a clearer understanding of the role of DDL activities in increasing learners'
awareness of lexical chunks, how language awareness influences English proficiency, and learners' attitudes toward this DDL approach, two anonymous questionnaires were used to collect data. The first was a feedback questionnaire on the teaching approach used (see Appendix 6), distributed by the teacher on the last day of class. The second was a course evaluation questionnaire conducted by the university administration center to examine student progress throughout the course.

The teaching approach feedback questionnaire consisted of 30 statements in four categories:
awareness of lexical chunks, enhancement of language proficiency, difficulty using iWeb corpus,
and attitude toward the pedagogical approach. Some items were adapted from existing research
(Asuman et al., 2016; Geluso & Yamaguchi, 2014). This was a hard copy paper questionnaire and
all statements were presented in English. The students were required to indicate their degree of
agreement with each statement on a 5-point Likert scale. All questionnaires were completed and
158 valid sets of data were collected. Data analysis was conducted using SPSS 16.0. Table 1 – 4
present results for number of participants, percentage of agreement based on the population
choosing 4 and 5 (agree and strongly agree), agreement range (scale minimum and maximum),
means, and standard deviations for each category. The university administration center’s course
evaluation questionnaire was conducted at the end of the semester. Table 5 reports the results
obtained from Zhejiang University’s educational administration system (http://jwbinfosys.zju.
edu.cn).

4.2. Data analysis

As Table 1 demonstrates, students felt strongly about the general increase in their lexical aware-
ness, as shown in the response to item one: My awareness of lexical chunks has developed (m:
4.54). Following intensive exposure to lexical chunks, including the introduction of basic informa-
tion about the lexical approach, exploratory learning of lexical chunks through the iWeb corpus,
and task-based activities, increased sensitivity to lexical chunks was reported. The statement with
Table 1. Lexical awareness through DDL

| Item Category                                                                 | N   | Percentage of Agreement | Min. | Max. | Mean | SD  |
|--------------------------------------------------------------------------------|-----|-------------------------|------|------|------|-----|
| Through DDL tasks, my awareness of                                              |     |                         |      |      |      |     |
| 1. Lexical chunks has developed                                                 | 158 | 94.9%                   | 3    | 5    | 4.54 | 0.593 |
| 2. The importance of word frequency has developed                              | 158 | 86.7%                   | 2    | 5    | 4.37 | 0.727 |
| 3. The use of words in different registers has developed                        | 158 | 64.6%                   | 2    | 5    | 3.8  | 0.788 |
| 4. Idiomatic collocations has developed                                         | 158 | 91.2%                   | 2    | 5    | 4.42 | 0.670 |
| 5. The differences between near synonymous phrases has developed               | 158 | 72.8%                   | 2    | 5    | 3.95 | 0.772 |
| 6. Active participation in the learning process has developed                  | 158 | 80.4%                   | 1    | 5    | 4.22 | 0.824 |

1: strongly disagree 2: disagree 3: uncertain 4: agree 5: strongly agree

The second highest levels of agreement were My awareness of idiomatic collocations has developed (m: 4.42), matching item two in Table 2: Useful for finding idiomatic collocations (m: 4.54) and item one in Table 3: I trust the collocations I find in the corpus (m: 4.30). Item 2, My awareness of the importance of word frequency has developed (m: 4.37), also scored highly, as did item 6, My awareness of active participation in the learning process has developed (m: 4.22). The development of word frequency awareness may be closely related to exposure to the iWeb corpus, which highlights word frequency. Item 6 matched item 3 in Table 3: I enjoy being able to direct my own learning (m: 4.03). There was less agreement on statements such as those concerning word use in different registers (item 3, m: 3.8) and the difference between near synonymous phrases (item 5, m: 3.95). For many EFL learners, word use precision is a problem. DDL activities and the lexical approach offer new perspectives on learning vocabulary and more emphasis should be given to the in-depth study of words.

Table 2 shows that students were positive about the role of DDL in learning vocabulary (m: 4.57), finding idiomatic collocation (m: 4.54), finding new ways of using familiar words (m: 4.45), finding different uses for the same word (m: 4.36), learning new phrases through familiar vocabulary (m: 4.35), improving reading (m: 4.27), using collocations in writing (m: 4.21), and using more idiomatic expressions in speaking (m: 4.04). Items 1–6 concerned vocabulary learning, and all achieved mean agreement greater than 4, with the exception of item 5, Helpful for understanding the difference between near synonymous phrases (m: 3.99). Items 7–10 concerned improvements to basic language skills, and all achieved mean agreement greater than 4, with the exception of item 10, Helpful for improving listening (m: 3.48).

With regard to the iWeb corpus, Table 3 indicates three noteworthy problems with its use: some learners felt that there were too many sentences (m: 3.46); some agreed with the statement that it
was time-consuming to search for words (m: 3.27); and some stated that the cut-off sentences were difficult to understand (m: 3.11). Similar results were found in Geluso and Yamaguchi (2014) which indicated that without appropriate support, both teachers and students might find DDL approach time-consuming and difficult in dealing with concordance lines. Items for which there was less agreement were, It is hard to learn to use iWeb corpus (m: 2.49), Access to the iWeb page is limited (m: 2.19), and It is hard to understand the concordance lines (m: 2.58). Based on the general standard deviation of this category (>1), there was a greater dispersion of agreement in this category compared to others (<1), demonstrating a range of abilities in using the corpus, which implied that students with low self-motivation and greater dependence on guidance from the teacher might require more focused instruction on corpus use.

Table 4 shows that all items scored a mean agreement of above 4, with the exception of item 2, Using the corpus has motivated me to learn words (m: 3.69). This indicates that most students had

| Item Category | N   | Percentage of Agreement | Min. | Max. | Mean | SD   |
|---------------|-----|-------------------------|------|------|------|------|
| 1 Useful for learning vocabulary | 158 | 97.5%                   | 3    | 5    | 4.57 | 0.546 |
| 2 Useful for finding idiomatic collocation | 158 | 95%                     | 3    | 5    | 4.54 | 0.593 |
| 3 Helpful to find new ways of using familiar vocabulary | 158 | 89.2%                   | 2    | 5    | 4.45 | 0.701 |
| 4 Useful for learning new phrases through familiar vocabulary | 158 | 89.2%                   | 2    | 5    | 4.35 | 0.686 |
| 5 Helpful for understanding the differences between near synonymous phrases | 158 | 76.6%                   | 1    | 5    | 3.99 | 0.757 |
| 6 Useful for finding different usages of the same word | 158 | 88%                     | 2    | 5    | 4.36 | 0.725 |
| 7 Helpful for using more collocations in writing | 158 | 82.3%                   | 2    | 5    | 4.21 | 0.774 |
| 8 Helpful for using more idiomatic expressions in speaking | 158 | 75.3%                   | 1    | 5    | 4.04 | 0.793 |
| 9 Helpful for improving reading | 158 | 73.5%                   | 2    | 5    | 4.27 | 0.777 |
| 10 Helpful for improving listening | 158 | 51%                     | 1    | 5    | 3.48 | 0.813 |

1: strongly disagree 2: disagree 3: uncertain 4: agree 5: strongly agree
Table 3. Difficulties using iWeb corpus

| Item Category                                                                 | N  | Percentage of Agreement | Min. | Max. | Mean  | SD    |
|-------------------------------------------------------------------------------|----|-------------------------|------|------|-------|-------|
| 1. It is hard to learn to use iWeb corpus.                                    | 158| 21.5%                   | 1    | 5    | 2.49  | 1.155 |
| 2. It is time-consuming to search for words.                                  | 158| 46.9%                   | 1    | 5    | 3.27  | 1.131 |
| 3. There are too many sentences.                                              | 158| 55.1%                   | 1    | 5    | 3.46  | 0.935 |
| 4. The cut-off sentences are difficult to understand.                          | 158| 38.6%                   | 1    | 5    | 3.11  | 1.151 |
| 5. Access to iWeb page is limited.                                            | 158| 14.6%                   | 1    | 5    | 2.19  | 1.083 |
| 6. It is hard to understand the concordance lines.                            | 158| 22.8%                   | 1    | 5    | 2.58  | 1.096 |

1: strongly disagree 2: disagree 3: uncertain 4: agree 5: strongly agree

a positive attitude toward this lexical approach using DDL activities. There were high levels of agreement for the statements I think corpus use is better suited to intermediate learners than beginners (m: 4.31), I trust the collocations I find in the corpus (m: 4.30), I would recommend DDL to other English learners (m: 4.16), and I believe that grammar and vocabulary are more closely related (m: 4.16). Here, item 8 is a reminder that English proficiency levels among students may be an important factor contributing to the effectiveness of this approach. Item 7 shows that the lexical approach can help learners to understand that grammar and vocabulary are not necessarily acquired separately. Surprisingly, item 2, Using the corpus has motivated me to learn words, was the statement upon which respondents showed least agreement. Table 1 and Table 2 demonstrate that DDL helps students to increase their awareness of lexical chunks and improve their understanding of vocabulary. However, a question remains as to why the students did not feel strongly motivated to learn words. The reason for this may be attributed to the difficulties confronted in using the corpus, such as the students’ belief that it was time-consuming and contained too many sentences.

Table 5 provides further evidence of approval for the course design based on the course evaluation conducted by the university educational administration system. This evaluation removed both the top and bottom 5% of the population, resulting in an average satisfaction score for the course of 4.88. This reveals that, in general, students were satisfied with the course design.

In order to assess the students’ perceptions of their improvements in English proficiency with their actual improvements, Table 6 compares Finals for College English Course test scores for two successive academic years, 2018–2019 (18–19) and 2019–2020 (19–20). Both groups consisted of first-year non-English majors at Zhejiang University, classified into Band 3 based on their College Entrance Examination English test score. Both groups took the same course taught by the same teacher, and both groups took final tests of the same difficulty level. The two groups differed only with regard to teaching method. A traditional 3P teaching mode was adopted in the 18–19 course, while the new Lexical Approach with DDL activities was trialed in 19–20 course. As Table 6
demonstrates, the 19–20 students, who were taught with the new method, outperformed the 18–19 students, who were taught with the traditional method. There was an almost 3% increase in average scores and, of the tested language skills, the biggest improvements were in vocabulary and listening, while reading and writing showed no noticeable improvement. These results support the students’ perceptions of overall improvement in their English proficiency, especially with regard to vocabulary learning (Table 2).

5. Discussion

5.1. The importance of being “trained to notice”

The common language units traditionally taught to students are words, phrases, and sentences. For the teaching of grammatical structure, these units are useful because words and phrases are used to build sentences. However, these units may not be the best choice for practical language use, because EFL learners usually have little understanding of how words are put together in real use, even though they can construct grammatically correct sentences with a large vocabulary and according to grammatical rules. In this study, lexical chunks were introduced as salient units for language use in the belief that language is made up of chunks.

Table 4. Attitudes toward the DDL approach

| Item Category | N  | Percentage of Agreement | Min. | Max. | Mean | SD  |
|---------------|----|-------------------------|------|------|------|-----|
| 1             | 158| 91.1%                   | 3    | 5    | 4.30 | 0.623|
| 2             | 158| 60.1%                   | 1    | 5    | 3.69 | 0.829|
| 3             | 158| 74.7%                   | 1    | 5    | 4.03 | 0.852|
| 4             | 158| 77.2%                   | 1    | 5    | 4.06 | 0.827|
| 5             | 158| 76.3%                   | 2    | 5    | 4.10 | 0.808|
| 6             | 158| 82.3%                   | 1    | 5    | 4.16 | 0.828|
| 7             | 158| 79.7%                   | 2    | 5    | 4.16 | 0.753|
| 8             | 158| 84.8%                   | 1    | 5    | 4.31 | 0.844|

1: strongly disagree 2: disagree 3: uncertain 4: agree 5: strongly agree
of meaning. When these chunks are put together, they produce coherent meaning, and only a small proportion of spoken sentences are entirely novel creations (Moudraia, 2001). Therefore, this lexical approach pays special attention to lexical chunks, including polywords, collocations, institutionalized utterances, and sentence frames and heads. “Instead of words, we consciously try to think of collocations and to present these in expressions. Rather than trying to break things into ever smaller pieces, there is a conscious effort to see things in larger, more holistic ways” (Lewis, 1997a, p. 204). As the words that co-occur in natural text with greater than random frequency, lexical chunks are not determined by logic. Instead, arbitrary clusters become the prefabricated units of language use. Nattinger (1980) suggests that language teaching should be based on the idea that language production is the piecing together of ready-made units appropriate for particular situations.

However, English language education has, for a long time, failed to properly emphasize the important role of these ready-made units in language teaching and learning. The EIODI teaching pattern with effective DDL activities used in this study has raised EFL learners’ lexical awareness through the identification of lexical chunks in reading and listening, the development of sensitivity of English collocation with adequate exposure to authentic language data in iWeb corpus as well as other tested materials, and the improvement of the confidence in language use in recognition and production activities. Upon completion of the 16-week course, learners reported increased awareness of lexical chunks and a positive effect on their vocabulary learning, reading, writing, and speaking. In addition, students expressed a positive attitude towards self-directed chunk learning and a deepened understanding of grammar and vocabulary while the traditional 3P mode always makes Chinese EFL learners “feel frustrated with remembering English words” for having inadequate vocabulary learning strategies (Ma, 2012, p. 1199). Therefore, the importance of being “trained to notice” should be stressed in English vocabulary teaching.

5.2. Lexical awareness and language proficiency
For decades, researchers have explored the correlation between language awareness and language proficiency. James and Garrett (1992) argued that “language awareness needs no justification in terms of improvement in skill, just as biology does not have to prove that it has led to improved crop or stock production.” As for foreign language learning, research has found that information processing speed is an important factor in determining the success of information reception in a foreign language. This ability, in turn, is entirely dependent on the readiness with which patterns are recognized, and meaning decoded, while the message is retained in the listener’s short-term memory (Ilaria, 2015). Therefore, the quick recognition of language patterns

| Table 5. Course Evaluation |
|---------------------------|
| Item Category             | N   | Mean | Very Satisfied | Satisfied | Neutral | Dissatisfied | Very Dissatisfied |
|---------------------------|-----|------|----------------|-----------|---------|--------------|------------------|
| 1 Rich and helpful resources | 81  | 4.87 | 71             | 10        | 0       | 0            | 0                |
| 2 Reasonable course requirements | 81  | 4.91 | 74             | 7         | 0       | 0            | 0                |
| 3 Valuable learning experience | 81  | 4.86 | 71             | 8         | 2       | 0            | 0                |
| 4 Overall impression on the course | 81  | 4.89 | 73             | 7         | 1       | 0            | 0                |

1: Very Dissatisfied 2: Dissatisfied 3: Neutral 4: Satisfied 5: Very Satisfied
Source: http://jwbinfosys.zju.edu.cn
is critically important in language learning and should be a skill taught to EFL learners who otherwise cannot naturally develop this ability.

This study examined the correlation between lexical awareness and language competency. First, the lexical approach increases learners’ awareness of co-occurrence words which account for the majority of English text. An increase in the number of collocations and clusters learned through DDL activities helps students to reduce reading processing time. Second, corpus data increase learners’ awareness of word frequency, enabling them to pay special attention to high-frequency words. An in-depth study of these words, including knowing high-frequency collocates, clusters, and closely related topics deepens learners’ understanding of word use, which may lead to enhanced reading comprehension. Third, the increased awareness of both word frequency and word environment through DDL activities improves word discrimination and idiomatic collocation abilities, leading to a reduction in Chinglish expressions when writing. As J. Li (2018, p. 498) states, “when learners value lexical chunks and take lexical, grammatical and semantic occurrence seriously, their vocabulary knowledge of syntagmatic relations will be substantially improved.”

5.3. Lexical awareness and language learning attitude

English teachers in China have, for a long time, used a traditional deductive approach, that focuses on teaching grammatical rules, sentence structures, and the use of words followed by examples of use. This approach is considered logical, efficient, and time-saving (Fischer, 1979; Hammerly, 1975). However, many students familiar with language rules and in possession of a broad vocabulary, fail to develop the ability to apply this knowledge to their own language use because their grammar and vocabulary have not been acquired through real language exposure. In other words, even those students with knowledge of the language and an awareness of language structure may find it difficult to use the language in real-life contexts.

Lexical awareness focuses on both language knowledge and language use, and encourages intentional and explicit reflection on language and language learning through the learner’s active involvement. In this study, DDL activities were used for two purposes. First, through the use of self-accessible authentic text in a concordancing program, DDL was found to improve the analytical power and independence of English learners (Asuman et al., 2016). This student-centered exploratory learning approach enables students to explore language use by observing authentic language data. Thus, learners become active participants in the learning process in which they discover and formulate the rules of language use for themselves. Second, DDL is a more interesting and impressive way of learning than memorizing words and language rules through direct teaching. This is inferred from the results of this study which indicates that most participants enjoyed self-directed learning (74.7%) and most said they would continue to use this approach (76.3%). Thus, this approach enhances the autonomy of language learners and has a long-term effect on language development. Both of these are of great importance to intermediate and advanced learners who are more likely to lose the motivation to learn and who may have fewer opportunities to take formal language courses. This bottom-up approach is perhaps more suitable for intermediate learners, such as the participants in this study, and advanced learners, as they already
possess the basic language skills, and when exposed to real language data, their understanding of the English language will deepen and their awareness of language use will improve.

6. Conclusion and implications

6.1. Conclusion
This study examined the effects of data-driven learning activities on increasing lexical awareness among intermediate EFL learners. Following a semester-long College English Course, in which iWeb corpus and other real communicative materials were used, and EIODI teaching patterns were adopted for both in-class and after-class activities, students reported increased awareness of lexical chunks and a positive impact on English reading, writing, and speaking skills. Questionnaire data revealed that most learners developed positive attitudes toward this approach and were willing to continue using this approach in the future. Test results also suggest that the lexical approach with DDL activities adopted during the course has a positive effect on overall English proficiency.

One limitation of this study, however, is that the test results reported here compare the final test scores of two groups of students, while the progress made by the students involved in the study was not examined by way of pre- and post-course tests. Therefore, in order to substantiate students’ perceptions of their improved abilities, appropriately designed pre- and post-course tests should be conducted in the follow-up study.

6.2. Implications
The findings of this action research have pedagogical implications for deeper and longer acquisition of lexical chunks. First, certain DDL activities prove effective in adopting lexical approach. In this study, the teaching pattern of EIODI was designed and at each stage, specific activities and tasks were carried out to internalize lexical chunk learning. Pereyra (2015) found that improved knowledge of lexical chunks did not necessarily follow the acquisition of lexical chunks, and the blending of extensive reading with a lexical approach was encouraged in order to promote lexical chunk acquisition. In the current study, apart from recognition tasks, production tasks such as public speaking and summary writing were also used to help learners to internalize what they had learned. Unlike recognition tasks, such as dictation, matching, or correcting, production tasks require learners to flexibly use lexical chunks in their own speech and writing. Corpus research has already shown that foreign language learners face difficulties in freely using collocations and even controlled productive use can pose a serious challenge (Peters, 2015). Second, in this teaching experiment, the students’ lexical awareness could be raised after intensive and explicit exposure to chunks in real language use, but strengthening lexical awareness and promoting language use in the long term remain a challenge. For EFL learners, lexical chunks cannot be acquired as naturally and easily as those of their native language and more memory work and processing time are required. Therefore, in order to keep the learners’ enthusiasm in DDL, two types of data should be organically integrated, that is, the use of corpora should be accompanied by real communicative contexts. In addition, new teaching strategies such as flipped classroom, the use of mobile application and even the use of humor (Nagy, 2020; Özcan & Kert, 2020; Suranakkharin, 2017) could be absorbed into the course design.

Funding
This work was supported by the Teaching Reform Research Projects of Zhejiang University 2018 [NO. Zdjg08044].

Author details
Liya Xue
E-mail: juleah@163.com
ORCID ID: http://orcid.org/0000-0001-5751-2099
1 Institute of Linguistics and Applied Linguistics in Foreign Languages of Zhejiang University, Hangzhou, China.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Citation information
Cite this article as: Using data-driven learning activities to improve lexical awareness in intermediate EFL learners, Liya Xue, Cogent Education (2021), 8: 1996867.

References
Asuman, A., Arzu, Ş. V., & Kodziye, D. A. (2016). Lexical awareness and development through data driven...
learning: attitudes and beliefs of EFL learners. Journal of Education & Training Studies, 4(3), 87–96. https://doi.org/10.1111/jets.v4i3.1223
Boers, F. (2008). Formulaic sequences and perceived oral proficiency: putting a lexical approach to the test. Language Teaching Research, 10(3), 245–261. https://doi.org/10.1191/136216808sr195oa
Bolitho, R., Carter, R., Hughes, R., Ivančič, R., Masuhara, H., & Brian Tomlinson, B. (2003). Ten questions about language awareness. ELT Journal, 57(3), 251–259. https://doi.org/10.1093/elt/57.3.251
Bolitho, R., & Tomlinson, B. (1995). Discover English. Heinemann.
Boontam, P., & Phocharaensil, S. (2018). Effectiveness of English preposition learning through data-driven learning (DDL). The Southeast Asian Journal of English Language Studies, 24(3), 125–141. https://doi.org/10.17576/iel-2018-2403-10
Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. ReCALL, 21(1), 37–54. https://doi.org/10.1017/S0958344009000668
Boulton, A. (2010). Data-driven learning: taking the computer out of the equation. Language Learning, 60(3), 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x
Carter, R. (2003). Language awareness. ELT Journal, 57(1), 44–46.
Chen & Cheng. (2018). Corpus-based approach of innovative construction of second language. Foreign Language Research, 35(3), 63–67. 169.
Ding, Y., & Yi, Y. (2005). Use of formulaic language as a predictor of L2 oral and written performance. Journal of PLA University of Foreign Languages, 28(3), 49–53.
Ellis, R. (1998). Teaching and research: options in grammar teaching. TESOL Quarterly, 32(1), 39–60. https://doi.org/10.2307/3587901
Figueiras, N. (2012). The impact of the CEFR. ELT Journal, 66(4), 477–485. https://doi.org/10.1093/elt/ccs037
Fischer, R. A. (1979). The inductive-deducitive controversy revisited. The Modern Language Journal, 63(3), 98–105. https://doi.org/10.1111/j.1540-4781.1979.tb02433.x
Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: student attitudes and efficacy. Recall, 26(2), 225–242. https://doi.org/10.1080/1040989X.2014.905004
Hammerly, H. (1975). The deduction/induction controversy. Modern Language Journal, 59(1–2), 15–18. https://doi.org/10.1111/j.1540-4781.1975.tb03619.x
Hassan, M. (2018). The Impact of Using Concordances on Enhancing EFL Students’ Use of Collocations in Writing English. Cairo University.
He, X. (2016). Non-english majors’ listening teaching based on lexical chunks theory and schema theory. English Language Teaching, 9(2), 142. https://doi.org/10.5539/elt.v9n2p142
Hsu, J. Y., & Chiu, C. Y. (2008). Lexical collocations and their relation to speaking proficiency of college EFL learners in Taiwan. Asian EFL Journal, 10(1), 181–204.
Hua, T. K., & Azmi, N. A. (2021). Collocational competence as a measure of esl/efl competency: A scoping review. 3l:The Southeast Asian Journal of English Language Studies, 27(1), 115–128. https://doi.org/10.17576/iel-2021-2701-09
Ilaria, V. (2015). Metalinguistic knowledge, language awareness and language proficiency. US-China Foreign Language, 13(1), 34–43. https://doi.org/10.17265/1539-8080/2015.01.004
James, C., & Garrett, P. (eds.). (1992). Language awareness in the classroom. Longman.
Johns, T. (1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. English Language Research Journal, 4, 27–45.
Kazemi, M., Khatraei, S., & Rosekh, A. E. (2014). The impact of teaching lexical bundles on improving Iranian EFL students’ writing skill. Procedia - Social and Behavioral Sciences, 98(special issue), 864–869. https://doi.org/10.1016/j.sbspro.2014.03.493
Leckey, J., & Neill, N. (2001). Quantifying quality: the importance of student feedback. Quality in Higher Education, 7(1), 19–32. https://doi.org/10.1080/13538320120045058
Leech, G. (1983). Principles of pragmatics. Longman.
Levee, W. (1989). Speaking: From intention to articulation. MIT Press.
Lewis, M. (1993). The lexical approach: The state of ELT and a way forward. Language Teaching Publication.
Lewis, M. (1997). Implementing the lexical approach: Putting theory into practice. Language Teaching Publications.
Lewis, M. (2000). Chapter 9. Materials and resources for teaching collocation. In M. Lewis (Ed.), Teaching collocation: Further development in the lexical approach (pp. 186–203). Oxford University Press.
Li, J. (2018a). An empirical study on teaching prefabricated chunks in English listening comprehension. Education Teaching Forum, 38, 217–218
Li, X. (2018b). Exploring Chinese EFL learners’ acquisition development of depth of vocabulary knowledge based on input corpus and output corpus. Proceedings of the international conference on contemporary education, social sciences and ecological studies (CESSES 2018), 283. Moscow, Russia. https://doi.org/10.2991/ceses-18.2018.110
Lin, M. H., & Lee, J.-Y. (2015). Data-driven learning: changing the teaching of grammar in EFL classes. ELT Journal, 69(3), 264–274. https://doi.org/10.1093/elt/ccv010
Luo, Q. (2016). The effects of data-driven learning activities on EFL learners’ writing development. Springerplus, 5(1), 1255. https://doi.org/10.1186/s40064-016-2935-5
Mo, R. (2012). Vocabulary proficiency instruction for Chinese EFL learners. Theory and Practice in Language Studies, 2(6), 1199–1205. https://doi.org/10.4304/tpls.2.6.1199-1205
Mominlou, M., & Enayati, B. (2018). The effects of lexical chunks teaching on EFL intermediate learners’ speaking fluency. International Journal of Instruction, 11(3), 179–192. https://doi.org/10.12973/jii.2018.11313a
Moudraio, O. (2001). Lexical approach to second language teaching. ERIC Digest, 06(00), 1–8.
Nagy, T. (2020). Raising collocational awareness with humour. Acta Universitatis Sapientiae, Philologica, 12(2), 99–113. https://doi.org/10.2478/ausq-2020-0015
Nattiging, J. R. (1980). A lexical phrase grammar for ESL. TESOL Quarterly, 14(3), 337–344. https://doi.org/10.1006/tequ.2001.2203;3586598
Nattiging, J. R., & DeCarro, J. S. (1992). Lexical phrases and language teaching. Oxford University Press.
Newell, A. (1990). Unified theory of cognition. Harvard University Press.
Özcan, M. S., & Kert, S. B. (2020). Use of mobile applications in collocation teaching. Cyprus Journal of Educational Sciences, 15(1), 1176–1190. https://doi.org/10.1884/j.cjes.v15i1.S.629
Pang, J. (1996). Language awareness in English teaching. Foreign Language Circle, 63(3), 11–14.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency.
Appendices

1. Example of iWeb corpus home page for the word “personality”

(1) Example of a student’s in-depth study of words using iWeb corpus
   (1) Example of cloze test, with all sentences extracted by the teacher from iWeb concordances
   (1) Example of student summary writing and chunk collection

(2) Extract of a passage marked by the teacher (red lines are lexical chunks; blue lines are parallel patterns; green lines are passive voice)

   (1) The questionnaire

Questionnaire on “Data-Driven Learning“

Hello, class! Our College English course (Band 3) is coming to an end. Thank you very much for your active participation. This semester we have been experimenting with “Data-Driven Learning“ (DDL) to improve lexical chunk awareness and language use ability. Now you are invited to complete the following questionnaire including four sections on DDL problems based on your own learning experience to improve my future teaching and research. Thank you!

(1) Through DDL tasks my awareness of

   () 1. Lexical chunks has developed (语块意识)

   () 2. The importance of word frequency has developed (词频意识)

   () 3. The use of words in different registers has developed (词汇的文体风格意识)

   () 4. Idiomatic collocations has developed (地道搭配意识)
(5) The differences between near synonymous phrases has developed (近义短语差别意识)
(6) Active participation in learning process has developed (主动学习意识)

(1) **Positive aspects of DDL**

(1) Useful for learning vocabulary
(2) Useful for finding idiomatic collocations
(3) Helpful to find new ways of using familiar vocabulary
(4) Useful for learning new phrases through familiar vocabulary
(5) Helpful for understanding the differences between near synonymous phrases (近义短语)
(6) Useful for finding different usages of the same word
(7) Helpful for using more collocations in writing
(8) Helpful for using more idiomatic expressions in speaking
(9) Helpful for improving reading
(10) Helpful for improving listening

(1) **Difficulties confronted while using iWeb corpus**

(1) It is hard to learn to use iWeb corpus.
(2) It is time-consuming to search for words.
(3) There are too many sentences.
(4) The cut-off sentences are difficult to understand.
(5) Access to iWeb page is limited.
(6) It is hard to understand the concordance lines (索引行).

(1) **My attitude towards DDL instruction**

(1) I trust the collocations I find in the corpus
(2) Using corpus has motivated me to learn words
(3) I enjoy being able to direct my own learning
(4) I think corpus use should be taught in English class more regularly
(5) I will use the corpus in future English learning
(6) I would like to recommend DDL to other English learners
7. I believe that grammar and vocabulary are more closely related

8. I think corpus use is better suited to advanced learners than beginners