Paper-based Data-driven Learning for Beginners in a Japanese University EFL Class
日本の大学における英語初級者のための紙媒体データ駆動型学習

Edward McShane

マクシェイン・エドワード

Abstract: Despite growing empirical evidence of the effectiveness of data-driven learning (DDL), its popularity is hindered by perceptions that it is unsuitable for beginners, and requires student and teacher training as well as technology in the classroom. This pilot study measures the effectiveness of a paper-based DDL activity aimed at teaching vocabulary to low-level EFL learners. It also demonstrates how a DDL activity can be implemented in a technology free classroom, without the need for student or teacher training. Beginner level students in an intact university EFL class received DDL vocabulary instruction combined with more traditional methods. Pre and post-tests revealed significant gains for vocabulary items that received DDL instruction, while items that only received traditional instruction and control items showed no significant gains.

Keywords: Data-driven learning, paper-based, beginner, vocabulary

要旨：データ駆動型学習（DDL）については、その効果に関する実証研究が進んでいるものの、初修者には不向きであるとか、IT 機器そのものの必要性や、学習者と教員の機器を使用するための訓練が必要であるという認識により、広く行われるまでには至っていない。本稿は、外国語として英語を学ぶ初級レベルの学習者への語彙指導に紙媒体の DDL 活動を取り入れて、その効果を測定するパイロットスタディーである。その中で、IT 機器や学習者・教員の訓練も要さずに DDL は実行可能であることも示す。大学の初級英語のひとクラスを対象に、伝統的な手法と組み合わせてデータ駆動型の語彙指導を行った。指導の前後に行ったテストの結果では、DDL 活動を行った語彙項目にはかなりの得点上昇が見られたが、伝統的な方法で指導した語彙項目、およびコントロール群の語彙項目では得点の上昇は見られなかった。

キーワード：データ駆動型学習、紙媒体、初級、語彙

1. Introduction

With the increasing power of computers and easier access to electronic corpora and concordancing software, corpus linguistics has been slowly influencing language teaching. However, indirect applications of corpora, such as developing teaching materials and reference publishing, have outweighed direct applications in the classroom. Direct applications, known as data-driven learning (DDL), are those that put learners in immediate contact with the corpus data, allowing students and teachers quick

access to reliable information about how language is and can be used (Chambers, 2010; McEnery & Xiao, 2011; Römer, 2006).

Tim Johns, who coined the phrase and is credited with increasing its popularity (Chambers, 2010), defined DDL as:

> '..the use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output' Johns and King (1991:iii, cited in Boulton, 2011: 564).

Using corpora in the classroom can 'cut out the middle man' and allow students to become a 'Sherlock Holmes' to discover how to use language by themselves (Johns, 1991:30, 1997:101, cited in Boulton, 2011: 565). DDL activities will usually present corpus data to learners in the form of a *key word in context* (KWIC) concordance (Boulton, 2011; Lessard-Clouston & Chang, 2014). Such activities can involve learners accessing corpora directly using a computer (direct-access DDL), or corpus data printed on paper (paper-based DDL) to highlight a particular language problem (Römer, 2006).

Introducing corpora to the classroom, however, presents a number of problems for teachers and students. Technology in the form of personal computers is required during lessons (Oghigian & Chujo, 2010); expertise in corpus linguistics tools and methods is required of teachers and students (Boulton, 2010; Campoy et al., 2010; Chambers, 2010; Gilquin & Granger, 2010; Lessard-Clouston & Chang, 2014); and it can be too difficult for low-level students (Chujo et al., 2012a).

While empirical evidence of the effectiveness of DDL has been growing, there has been a lack of research with beginner level students (Mizumoto & Chujo, 2015). Boulton (2008a, 2008b), and Mizumoto and Chujo (ibid) wish for more empirical evidence of how DDL can be effective, especially with beginners, and Römer (2006: 128) wishes for researchers to help 'spread the word' to practitioners of language teaching so that student teachers, professional teachers, students, materials writers etc. can be convinced of the value of corpus linguistics to language learning and its use can gradually be spread.

This paper aims to help 'spread the word' by testing the effectiveness of a DDL activity aimed at teaching vocabulary to low level students. The activity is one that seeks to respond to the challenges of DDL by requiring no corpus linguistics expertise and no computer equipment in the classroom.

2. Data-driven learning for beginners

DDL is seen by some as being unsuitable for low-level students because the large amount of data corpora provide can be overwhelming for even advanced learners (Chujo et al., 2012a); however, some researchers have shown that DDL can be effective for low-level students when it is tailored to meet their needs using parallel bilingual corpora, or paper-based materials. This section will describe a number of DDL studies involving low-level learners.

Motivated by a gap found between what is taught in Japanese secondary textbooks and what is tested on the Test of English for International Communication (TOEIC), Chujo and her colleagues (Chujo et al., 2006, 2009, 2012a; 2012b, 2013; Oghigian & Chujo, 2012) began an annual study in 2005 in low-level EFL classes for university engineering students. Students were introduced to DDL through activities involving a parallel bilingual (Japanese and English) corpus sourced from a newspaper. Paper-based activities involving vetted data were later included, and in response to difficulties with software, they developed their own concordancer.

These studies consistently resulted in gains in identifying and producing noun phrases and verb phrases, with DDL groups improving more than the non-DDL groups in these areas (Chujo et al., 2009, 2012a, 2012b); however, improvement in answering more complex TOEIC type questions targeting these forms was less consistent (Chujo et al., 2009, 2012a). A non-DDL group, which used a listening course instead of DDL gained more in the proficiency testing TOIEC Bridge Test, especially on the listening section (Oghigian & Chujo, 2012). Whether DDL was paper-based, direct-access, or a combination of both, there was no significant difference in gains (Chujo et al., 2012a).

Feedback from students regarding working with corpora (paper-based and direct-access) was mostly positive, with students indicating that they would like to use concordancing software in place of a dictionary (Chujo et al., 2006), and that tasks were useful and accessible (Chujo et al., 2009). Students indicated that mother tongue (L1) translations provided by the parallel corpus were necessary, but became less reliant on L1 translations over time (Chujo et al., 2009), with those using more paper-based materials with vetted concordance lines becoming less reliant than those doing direct-access activities (Chujo et al., 2012a). Students felt that paper-based activities saved time, allowing more tasks to be completed, and were less worried about making mistakes because they could not go astray. They indicated a number of advantages to using computers, welcoming the Japanese translations provided by the parallel

concordancer, along with longer concordance lines - paper based concordances had to be truncated to fit on paper. Computer work was also seen as more active than paper-based. Students felt that carrying out the search by themselves helped them to memorise spelling and forced them to think about grammar and vocabulary more carefully (Oghigian & Chujo, 2010). Despite the number of differences between the two methods, students had no preference between paper-based, computer-based or a combination of the two (Chujo et al., 2012a).

Boulton (2008a) compared the effectiveness of DDL instruction with a traditional teaching method with lower-intermediate level (TOEIC 405-600) students at a university in France. The focus of the DDL activities was on grammar and usage involving problematic language items. In another study intended to demonstrate how easy it can be to gather empirical evidence of the effectiveness of DDL (2008b), he had beginner level students choose between phrasal verbs and their respective base verbs in edited concordance lines.

Boulton (2008a) found that the DDL group gained more than the traditional method group in answering TOEIC type questions, but that it was not significant. He found that the highest level learners gained more than the other learners from the dictionary activities, while all levels benefited equally from the corpus work. In the study involving verbal phrases (2008b), significant gains were made in the post-test for choosing the correct verb form, with the phrasal verbs gaining more than the base verbs. Students preferred learning grammar and usage from a corpus than a dictionary, finding it easier, more useful and better for helping prevent future errors, with all levels of learner feeling similarly. Students also indicated they would prefer paper-based activities in the future (Boulton 2008a).

While these studies do no present a large amount of evidence of the effectiveness of DDL with low-level learners, the results do indicate that low-level learners are in favour of DDL activities, and that either paper-based or direct-access can be implemented depending on the teaching environment. Although parallel bilingual corpora seem to be useful, they are difficult to obtain, and may only be suitable in classes where students share a mother tongue. While it has been demonstrated that DDL can lead to improvements in target areas with beginners if presented suitably, the greater improvement in the proficiency test of the non-DDL group in Chujo et al., (2009, 2012a) serves as a warning that DDL should not replace traditional methods, but be used alongside.

3. Method

3.1 Participants

The research took place in a Japanese university in an intact EFL reading class with a focus on vocabulary. The participants were 1st year students and were placed in a class based on their TOEIC scores at the beginning of the academic year, which ranged from 260 to 295.

3.2 Language items

For the research, 20 vocabulary items were selected from target words in the course textbook, Reading Explorer 1 (Douglas, N. and Bohlke, D., 2015). The units containing these words were encountered as part of the course during the research period. Each unit in the textbook is divided into two sub-units, each containing ten target vocabulary words. Three sub-units were covered during the research period, thus the 20 words were selected from 30 available. Using an online Word Level Checker tool (Someya, 2009), the JACET level of target words in the textbook were identified. JACET 8000 is a 'word list designed for all English learners in Japan' (Uemura & Ishikawa, 2004: 333). The majority of the words were found to be JACET level 2, i.e. their frequency rank was between 1000 and 1999. With the intention of making the vocabulary items similar in difficulty for students, the 20 selected words were of a frequency rank between 1500 and 3500. This was as narrow a range that would allow for 20 words to be selected from the 30 available. The procedure section will describe how these words were divided into two groups. An additional ten words were selected to form a control group. These words were taken from the same JACET range as before; however, they were not among the target words in the textbook.

3.3 Test instruments

A pre-test (Appendix A) was made that consisted of multiple choice gap-fill questions: one for each of the thirty vocabulary items. Each question contained an example sentence with the target word removed, accompanied by four choices: the correct answer and three distractors, which were of the same word class as the target word and of a similar JACET frequency rank.

The example sentences were sourced from a learner's dictionary website (Merriam-Webster.com, 2017). It was felt that these sentences would be comprehensible for students as they mostly contained words at or below the level of the

target word. It should be noted that these sentences are probably manufactured, therefore likely to include different collocation patterns from authentic text (Smith et. al., 2010), whereas the DDL instruction, which will be described later, consisted of authentic sentences. Before choosing to use sentences from the learner's dictionary, corpora were used to generate sentences for the test; however, it was difficult to find examples that met two important criteria: that the words in the sentence were no more difficult than the target word, and that the context was clear enough that the target word could reliably be chosen.

This style of question was chosen because it was felt that it would be sensitive to the kind of DDL treatment that was implemented. Also, such questions are similar to those used in part 5 of the listening and reading section of the TOEIC test. Students at this university are required to achieve a TOEIC score of 450 to graduate, therefore improvement in such a question type has added value. Figure 1 shows an example question from the pre-test. Vocabulary questions such as these are a useful measure of proficiency that 'allow learners to demonstrate that they understand vocabulary in context' (Smith et. al., 2010:1).

3	I didn't ------ you at first with your new haircut.
a)	recover
b)	recognize
c)	shift
d)	comment

Figure 1: Pre-test question

Because the research was conducted in an intact class, and not a controlled experiment, it was felt that a thirty-item test was as large as the test could be without causing disruption. This decision also determined the total number of vocabulary items in the study. The post-test used the same questions as the pre-test with the order of the questions changed and the order of the answer choices randomised.

3.4 Instruction

During the research period, the 20 target words received a combination of instruction methods that are described in this section.

Textbook instruction: Each sub-unit of the textbook contains a reading in which ten target words are highlighted. The readings are followed by comprehension

questions and then by vocabulary exercises for the ten target words including gap fills, definition matching, and selecting the best word from a binary choice.

Self-study: As part of the students' summary assessment they are to add a certain number of words to their vocabulary notebooks by the end of the semester. The textbook target words are assigned as homework to be entered into their vocabulary notebooks, along with each word's class, a definition in English or their mother tongue, and an example sentence.

DDL instruction: Prior to the beginning of research, a number of DDL activities and tools were introduced to the class. This helped familiarise students with the methods and determined what complexity and type of activity suited the students. To help students write example sentences in their vocabulary notebooks, they were introduced to Skell (Baisa, V. & Suchomel, V., 2014), a free online concordancer that outputs full sentence concordances of search terms in a *Good Dictionary Example* (GDEX) order. Students seemed very positive about Skell, with many of them using it for self-study; however, many seemed to misuse it, writing example sentences they could not understand, and also seemed to misunderstand Skell's *similar word* function, which produces a list of words with similar colocation patterns. The list of *similar words* includes non-synonymic words and antonyms, which some students wrote as the definition of the target word. For example, one student defined *male* as *female*.

Paper-based data-driven learning activities were also trialled in class before research began. KWIC concordances produced using tools from the Compleat Lexical Tutor website (Cobb, 2017), which from now on will be referred to as Lextutor, and full sentence concordances copied and pasted from Skell were trialled, with students indicating they preferred full sentences to KWIC concordances. It was felt that Skell provided more suitable examples for the students because of the GDEX ordering. Lextutor includes graded reader corpora, but these did not feature some of the target words, and it was intuitively felt that they did not include typical usage patterns for some words, despite Allan's (2008) findings that graded readers can provide authentic patterns. Other corpora available in Lextutor, such as Brown, and the British National Corpus (BNC), were deemed too difficult.

For the research treatment (Appendix B), short five-line concordances were printed with the target word removed. Sentences were copied and pasted from Skell then edited to be presentable on a worksheet. For each word, approximately 10 lines were copied and reduced to five. Reasons for eliminating lines included: the line being deemed too difficult; the content relating to sensitive issues; or the sense of the word

differing from that in the textbook. Students selected a word from a list that could complete all sentences in a concordance. For concordances where the target word was a noun or verb, after confirming the correct answers, students wrote the correct verb or noun form for each example sentence.

3.5 Procedure

The results of the pre-test were used to divide the 20 vocabulary items from the textbook into two groups: the DDL group and the non-DDL group. The vocabulary items were sorted first by the textbook lesson they appeared in, then by the number of students that answered them correctly. Vocabulary items were then alternately assigned to the DDL and non-DDL groups. This meant that each group would contain a similar mix of difficulty, and timing of instruction would be spread evenly across the research period. The non-DDL group items received textbook and self-study instruction; the non-DDL group received DDL instruction in addition to textbook and self-study. The third group was a control group that received no instruction and contained words that were not targeted in this class or in textbooks in the students' other EFL classes. The three groups of vocabulary items can be seen in Appendix C.

Beginning with the pre-test and ending with the post-test, the research was conducted over six 90-minute lessons over a three-week period. During this period the textbook was used in class, with target words being assigned for homework after the textbook sub-unit had been completed in class. In the following lesson a DDL activity was then be conducted. The order and timing of the procedures can be seen in Table 1. The sub-units of the textbook are 8B, 9A, and 9B. For each sub-unit three previously described instruction types are shown: textbook (Text), self-study (Self), and DDL instruction (DDL).

Lesson #	1	2	3	4	5	6
Date	25 Oct	30 Oct	1 Nov	6 Nov	8 Nov	13 Nov
Relevant content	Pre-Test 8B Text 8B Self	8B DDL 9A Text 9A Self	9A DDL 9B Text 9B Self	NO LESSON	9B DDL	Post-test

Table 1: Order and timing of procedures

It was not possible to conduct a delayed post-test in this research, as it took place in an intact class using language items that were part of the course, therefore the language items received further instruction after the post-test invalidating further testing.

4. Results and Discussion

Table 2 shows the descriptive statistics for the pre and post-tests. Each group included 10 vocabulary items, and 15 subjects' test results were included in the results. The mean pre-test scores were similarly low for all three groups, with the non-DDL group scoring highest with a mean score of 3.8 out of 10, and the DDL group scoring the lowest with 3.2 out of 10. Although pre-test scores were used to create comparable DDL and non-DDL groups, some students' results were later eliminated from the research because of absences, resulting in a noticeable difference in their pre-test means. The post-test score for the control group was lower than the pre-test, while both the DDL and non-DDL groups improved on the post-test, with the DDL group's improvement larger than the non-DDL group.

Group	# students	# items	Pre Mean	S.D	Post Mean	S.D
DDL	15	10	3.20	1.78	5.87	1.73
non-DDL	15	10	3.80	1.15	4.13	2.00
Control	15	10	3.73	1.75	3.40	1.59

Table 2: Descriptive statistics

For each group of vocabulary items pre and post-test scores were compared using the Wilcoxon Signed Rank test (Table 3). This test was chosen because pre and post-test scores of groups with identical subjects that had not been randomly chosen were being compared (Turner, 2014). The test was carried out using the statistics software R, following the instructions in Turner (ibid:). Results with a p-value less than 0.05 were considered significant and are shown in bold. On the basis of this small study, there is a 95% certainty that the DDL activities combined with traditional instruction helps to improve students' vocabulary knowledge (W= 0; p = 0.001534).

As the students showed significant gains in the vocabulary items in the DDL group while showing no improvement in the non-DDL group and control group, the DDL instruction combined with traditional treatment seems to have helped the students improve their vocabulary knowledge. However, these results lack validity as the study

involved only a small number of participants and vocabulary items. Also, the post-test was conducted very soon after treatment so does not indicate long-term retention of vocabulary.

Group	W	p
DDL	**0**	**0.001534**
non-DDL	25.5	0.529
Control	24.5	0.4687

Table 3: Wilcoxon signed rank test

That these students have very low TOEIC scores despite having had at least six years of English education at secondary level, combined with the lack of improvement in the non-DDL group may suggest they do not have effective self-study skills or that they lack motivation. If the problem is lack of ability to study, then teaching DDL skills can help them to improve their ability to study alone and become more effective autonomous learners. If the problem is lack of motivation, students may be motivated by an approach to vocabulary learning such as this, which has been shown to be seen as novel and useful by similar groups of learners (Oghigian and Chujo, 2010).

The DDL activity described in this paper was successful in its goals of overcoming some of the challenges of DDL. As well as producing positive results, it required very little time to prepare, and no expertise was required to produce it. For research purposes the activity type was kept consistent; however, very little extra work would be required to produce DDL activities aimed at different aspects of vocabulary knowledge. For example, word class knowledge could be focused on by deliberately choosing words that contain a variety of forms. Oghigian and Chujo (2010) show such an activity for the word *development* using a KWIC concordance.

Skell was chosen as the concordancer because of its full sentence concordances, GDEX ordering, and ease of transferring data to paper. However there are other tools freely available that allow other activity types to be produced. Important features lacking from Skell that are available with other tools include the ability to produce KWIC concordances, use wildcards to search for desired patterns, and the ability to sort to the left or right of the search term. Most concordancing software produce KWIC concordances, therefore it is necessary to familiarise students with them as well as full sentence concordances. While KWIC concordances may be daunting at first for learners, they have advantages. As KWIC concordances consist of truncated lines aligned by the

target word or phrase learners are more likely to draw their attention to the words on either side of the search term, making them more suitable to learning colligation and collocation patterns. Lextutor provides a suitable suite of tools that are easy to use and print friendly, making them suitable for paper-based DDL for beginner learners and teachers. Lextutor allows for searching by word, lemma or phrase, and also importantly allows KWIC concordances to be sorted by words to the left or the right of the search term.

5. Conclusion

This research demonstrated a data-driven learning activity that responded to some of the criticisms made of DDL by requiring no expertise to produce, requiring no computers in the classroom, and being suitable for low-level learners. The effectiveness of the activity was measured with a group of low-level students with the results indicating that when combined with instruction provided by a course textbook and students' self-study, it was significantly more effective in teaching vocabulary than self-study and textbook instruction alone.

While introducing paper-based DDL to the classroom is seen as a step towards providing learners with the skills to use direct-access DDL for autonomous learning, it could also be a stepping stone for teachers to become accustomed to the tools and methods of corpus linguistics, allowing them not only to conduct effective DDL activities with their students but also to enjoy the other benefits corpus linguistics has to offer language teachers, such as the ability to investigate language themselves and use corpora for materials development.

References

Allan, R. (2008). Can a graded reader corpus provide "authentic" input? ELT Journal, 63 (1), 23–32. doi:10.1093/elt/ccn011

Baisa, V., & Suchomel, V. (2014, December). SkELL: Web interface for English language learning. In Eighth Workshop on Recent Advances in Slavonic Natural Language Processing (pp. 63-70).

Boulton, A. (2008a). DDL: Reaching the parts other teaching can't reach? Teaching and language corpora 8, 38-44. <hal-00326706>

Boulton, A. (2008b). Looking (for) empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk, (ed.) Corpus Linguistics, Computer Tools, and Applications: State of the Art (pp. 581-598). Frankfurt: Peter Lang. <hal-00326983v2>

Boulton, A. (2010). Data-Driven Learning: Taking the Computer Out of the Equation. Language Learning, 60(3), 534–572. doi:10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2011). Data-driven learning: the perpetual enigma. In S. Gozdz-Roszkowski (ed). Explorations Across Languages and Corpora (pp. 563-580). Frankfurt: Peter Lang, <hal-00528258v2>

Campoy, M. C., Belles-Fortuno, B., & Gea-Valor, M. L. (Eds.). (2010). Corpus-based approaches to English language teaching. London: Continuum.

Chambers, A. (2010). What is data-driven learning? In A. O'Keefe & M.J. McCarty, (eds), The Routledge Handbook of Corpus Linguistics (pp. 345-58). Abingdon, Oxon: Routledge.

Chujo, K., Utiyama, M., & Miura, S. (2006). Using a Japanese-English parallel corpus for teaching English vocabulary to beginning-level students. English Corpus Studies, 13, 153-172.

Chujo, K., Anthony, L., Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. In M. Mahlberg, V. González-Díaz, and C. Smith (Eds.) Proceedings of the Corpus Linguistics Conference (CL 2009). July 20-23, 2009. Liverpool, UK.

Chujo, K., Anthony, L., Oghigian, K., & Uchibori, A. (2012a). Paper-Based, Computer-Based, and Combined Data-Driven Learning Using a Web-Based Concordancer. Language Education in Asia, 3(2), 132–145. doi:10.5746/LEiA/12/V3/I2/A02/Chujo_Anthony_Oghigian_Uchibori

Chujo, K., Oghigian, K., & Nishigaki, C. (2012b). Beginner level EFL DDL using a parallel web-based concordancer. Proceedings of the FEELTA 2012, Far Eastern Federal University, Vladivostock, Russia (pp. 1-5).

Chujo, K., Anthony, L., & Oghigian, K. (2013). Teaching remedial grammar through data-driven learning using AntPConc. Taiwan International ESP Journal, 5(2), 65-90.

Cobb, T., Compleat Lexical Tutor v 8.3 [computer program]. Accessed 2017 at https://www.lextutor.ca/

Douglas, N. and Bohlke, D., (2015). Reading Explorer 1 (2nd ed.). Boston, National Geographic Learning.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. In A. O'Keefe & M.J. McCarty, (eds), The Routledge Handbook of Corpus Linguistics, (pp. 359–370). Abingdon, Oxon: Routledge.

Johns, T. (1991b), 'From printout to handout: grammar and vocabulary teaching in the context of data-driven learning.' In: T. Johns & P. King (Eds.), Classroom Concordancing. English Language Research Journal, 4: 27-45.

Johns, T. (1997a), 'Contexts: the background, development and trialling of a concordance-based CALL program.' In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), Teaching and Language Corpora. Harlow: Addison Wesley Longman. 100-115.

Johns, T. & P. King (Eds.), (1991), Classroom Concordancing. English Language Research Journal, 4.

Lessard-Clouston, M., & Chang, T. (2014). Corpora and English Language Teaching: Pedagogy and Practical Applications for Data-Driven Learning. TESL Reporter 47(1-2), 1-20.

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning? In E. Hinkel (ed), Handbook of Research in Second Language Teaching and Learning (pp. 364–380). London: Routledge.

Merriam-Webster.com (2017). Learner's Dictionary. Accessed November 2017 from http://learnersdictionary.com/

Mizumoto, A., & Chujo, K. (2015, in press). A meta-analysis of data driven learning approach in the Japanese EFL classroom. English Corpus Studies, 22.

Oghigian, K., & Chujo, K. (2010). An effective way to use corpus exercises to learn grammar basics in English. Language Education in Asia 1(1), 200-214.

Oghigian, K., & Chujo, K. (2012). DDL for EFL beginners: A report on student gains and views on paper based concordancing and the role of L1. In J. Thomas, A. Boulton (eds), Input,

Process and Product: Developments in Teaching and Language Corpora, (pp. 130-143). Brno: Masaryk University Press.

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.

Smith, S., Avinesh, P., and Kilgarriff, A. (2010). Gap-fill tests for language learners: Corpus-driven item generation. In Proceedings of ICON-2010: 8th International Conference on Natural Language Processing, pages 1–6.

Someya, Y. (2009). Word level checker. Accessed October, 2017, from http://www.someya-net.com/wlc/

Uemura, T., & Ishikawa, S. I. (2004). JACET 8000 and Asia TEFL vocabulary initiative. The Journal of AsiaTEFL, 1(1), 333-347.

Appendix A: Pre-test

Practice Vocabulary Test #1

1   She was ——— to pass the exam.
a)   tired
b)   determined
c)   deserted
d)   confused

2   She ———- under the bed.
a)   hid
b)   tricked
c)   rescued
d)   instructed

3   I didn't ———- you at first with your new haircut.
a)   recover
b)   recognize
c)   shift
d)   comment

4   The attack ———- the world.
a)   adopted
b)   recovered
c)   shocked
d)   taped

5   An elephant ——— more than a mouse.
a)   weighs
b)   declares
c)   repairs
d)   prays

6   Can you ———- your town on the map?
a)   bother
b)   relieve
c)   rely
d)   locate

7   This diamond was very rare and quite ——-.
a)   brief
b)   moral
c)   valuable
d)   blind

8   The things in the museum are very strange and ———.
a)   rare
b)   wise
c)   prime
d)   frequent

9   An oven-toaster is a useful piece of ———- for the kitchen
a)   identity
b)   equipment
c)   household
d)   perception

10   He is not ———- of making those medical decisions himself
a)   terrible
b)   capable
c)   favorite
d)   direct

11   He is thinking about changing ——— and becoming a police officer.
a)   occupations
b)   conclusions
c)   assistant
d)   agency

12   The entire town was ———- by their deaths
a)   regarded
b)   employed
c)   invited
d)   affected

13   He's ———- at a local supermarket.
a)   defined
b)   explored
c)   employed
d)   reserved

14   Central Park is one of New York City's many ———-
a)   articles
b)   departments
c)   lessons
d)   treasures

15 Her personal ----- is estimated to be around
    $10 billion.
a) wealth
b) exhibition
c) factory
d) version

16 The company bottles only the ---- water.
a) saddest
b) thinnest
c) purest
d) slowest

17 The customer complained in a --- voice.
a) pure
b) loud
c) clever
d) brief

18 Plant roots ---- water.
a) absorb
b) spot
c) correct
d) labor

19 He was ----- so he is now the manager.

a) urged
b) obtained
c) invented
d) promoted

20 She ----- up from her book when he entered
    the room.
a) referred
b) burned
c) glanced
d) reminded

21 We need to leave ----- after the game or we
    will miss the train.
a) sadly
b) softly
c) daily
d) immediately

22 In this state, it is ---- for anyone under the
    age of 20 to drink alcohol.
a) civil
b) illegal
c) cheap
d) soft

23   ----- the lights went out

a) Directly
b) Relatively
c) Originally
d) Suddenly

24   ----- traveled hundreds of miles to trade in the
    city.
a) Nurses
b) Pupils
c) Detectives
d) Merchants

25 How long will Kyoto be able to ----- its
    traditions?
a) contract
b) joke
c) preserve
d) earn

26 The storm caused ----- damage to the roof.

a) ethnic
b) brief
c) severe
d) precious

27 Unfortunatley, all the files were ---

a) warned
b) complained
c) poured
d) destroyed

28 I have been ----- by every university I applied
    to.
a) rejected
b) translated
c) pretended
d) marched

29 What's the ----- of the building?

a) wealth
b) debt
c) height
d) cloud

30 I got only a --- better grade when I retook
    the test.
a) gradually
b) constantly
c) sadly
d) slightly

Appendix B: Data-driven learning activity

Vocab Unit 9B

a) equipment	b) destroy	c) capable	d) limit	e) majority

1 a) The building was _____ while being moved.
   b) The basketball hoop has already been _____.
   c) A large grinding stone was _____ intentionally.
   d) Each robot _____ is worth 50 points.   ____
   e) The fire subsequently _____ three surrounding properties.
   f) Dutch pilots claimed 55 enemy aircraft _____.

2 a) This company is still selling _____ online.
   b) But many professions require expensive basic _____.
   c) His eye control _____ is always latest technology.
   d) Neither strenuous exercises nor _____ are involved.   ____
   e) Cross country running involves very little specialized _____.
   f) Workers using personal protective _____ while painting poles.

3 a) Private _____ companies are often family businesses.
   b) The national consensus favoring _____ abortion rights remains intact.
   c) Parking is often very _____ –please observe local regulations.
   d) Their _____ nature makes rounds more tense.   ____
   e) The common _____ access freeway speed limit is 65 mph.
   f) Storage space was _____ while more nuclear wastes were produced.

4 a) A 55 percent _____ supports marriage equality.
   b) The silent _____ has been silent too long.
   c) The _____ received prison sentences although several hundred were executed.
   d) The vast _____ were released within days.   ____
   e) A scary idea in politics is " _____ rules".
   f) The vast _____ of barn fires are preventable!

5 a) But those problems are _____ of resolution.
   b) But neither is _____ of saving free society.
   c) Effective – they are usually _____ individuals.
   d) But many bacteria are _____ of causing severe infections.   ____
   e) Her employer said she was very _____.
   f) A human being is _____ of vastly more complicated behaviour still.

Appendix C: Vocabulary items

DDL	non-DDL	Control
affected	determined	absorb
capable	employed	glanced
destroy	height	loud
equipment	illegal	merchants
hid	locate	promoted
immediately	occupation	purest
preserve	rare	rejected
recognize	suddenly	severe
shock	treasures	slightly
weigh	valuable	wealth