Data-driven Vocabulary Learning vs Traditional Instruction at a High School in Uganda

Bilal TEKIN *

Abstract

The importance of vocabulary knowledge has gained much more attention in recent years. In order to increase the lexical knowledge, there have been many attempts by various researchers to find new vocabulary learning strategies. Data-Driven Learning (DDL) was introduced as a new vocabulary learning strategy by Tim Johns (1991) by taking advantage of online corpora. The present study aims to investigate whether this new approach is effective in vocabulary learning compared to traditional instruction. 4 months of experimental study involved two different groups as experimental (DDL) and control (traditional instruction) group. The analyses of the end of pre and post-tests showed that the experimental group performed better than the control group. The analyses showed that the difference in performance between the two groups was statistically significant.

Keywords: Corpus, Corpora, Concordance, Data-driven Learning, Vocabulary Learning

Introduction

Vocabulary learning has become one of the biggest interests of researchers in recent years. In order to convey the message to others, being proficient in grammar is not enough. In most circumstances, if one does not know the appropriate word or collocation, the message will not be delivered. Laufer (1986) and Nation (1990) stress that the lack of vocabulary is one of the main reasons for the difficulties of both receptive (listening and reading) and productive skills (speaking and writing). Wilkins (1972) stated that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p.23). Likewise, Folse (2004) emphasizes that "people can generally communicate their meaning with less than perfect grammar whereas incorrect vocabulary can substantially impede communication" (p. 26).

Student-centered language teaching has taken an important role in education and the learners have been required to discover word meanings and learn collocations independently. So, to find out how they learn and *what affects* their learning has become as crucial as what is learned. In this respect, there have been many researches on vocabulary learning strategies (Oxford, 1990; Sanaoui, 1995; Schmitt, 1997; Nation, 2001).

O'Malley and Chamot (1990) defined language learning strategies as "the special thoughts or behaviors that individuals use to help them comprehend, learn, or retain new information" (p. 1). Similarly, Oxford (1990) defined language learning strategies as "specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations" (p. 8).

Data-Driven Learning (DDL) emerged as a new language learning strategy in Tim Johns' (1991) works. DDL aims to enable learners learn the meaning of the desired word, structure and collocation using discovery skills on corpora. Since corpus is a collection of written or spoken transcripts of authentic texts in electronic form (Partridge, 2006, p. 1), by means of DDL, learners benefit from it not only in classes, but also out of classrooms because DDL stimulates noticing, discovery skills and better independent language learning.

In DDL, the information in corpus is presented by means of concordance lines (Tribble and Jones, 1997) in which a particular word or phrase is listed in all sentences vertically usually a few words to the left and right.

Today, the improvement in the field of Computer Assisted Language Learning (CALL) has encouraged language teachers and corpora users to provide hands-on experience and authentic data through concordance lines by means of software programs such as KWIC, CTAGS etc. or free online concordance websites like BNC, Lextutor, etc. Thanks to these developments, one can reach and find out how collocations of a particular word work together.

Students undergo stages in Data-Driven Learning which allows them to absorb the information required. These stages

* PhD student at the Faculty of Education, International Black Sea University, Georgia. Teacher at Galaxy International School, Kampala, Uganda. E-mail: bilaltekin66@gmail.com

are motivational, diagnostic, and empirical.

I. Motivational Stages of DDL Application

Nowadays learners are mostly comfortable and relaxed with the usage of computers. Moreover, they are happy and more motivated, when they use and integrate technology into their education. Not only do learners learn in this way at school, but also on their own.

Motivational stages of DDL application can be categorized as:

a)   Stages which involve technology: Since DDL is implemented by learners autonomously and because they feel more comfortable using computer on their own, students' motivation level is mostly high applying DDL.

b)   Stages which involve real life experiences: As a result of access to the online corpuses such as BNC, Google, US TV Talk, etc., learners have an opportunity to reach authentic rather than artificial contexts.

c)   Stages which involve self-learning: Today technology is everywhere and the usage and accessibility of internet is increasing day by day. Therefore, learners may use and apply DDL when they are at home or in any other environment where there is internet access.

d)   Stages which involve cooperation: Through DDL, learners find out the method of discovery themselves and share their experience or deductions with others.

II. Diagnostic Stages of DDL Application

In diagnostic stages, learners discover and deduce what they initially intended to search.

Diagnostic stages can be categorized as:

a)   Stages which involve discovery: Learners using their own potential learn, use and improve their discovery skills.

b)   Stages which involve deduction: Learners read sentences as if they are investigating a crime and when they deduce the meaning, they are happy like they have arrested the criminal.

III. Empirical Stages

Throughout empirical stages, learners are exposed to language and they observe, analyze and make conclusions on the target unknown word and the language used in online corpuses.

Empirical stages can be categorized as:

a)   Stages which involve integration: Learners have a chance to compare what they knew and what they are learning by integrating their knowledge of language with the language they discover.

b)   Stages which involve memory skills: Learners relate the meaning of word with the incidents in online corpuses so they remember connecting the meaning of target unfamiliar word.

c)   Stages which involve language use: Learners observe and analyze the collocation, syntax and semantics by means of reading concordance lines (2-3 words left from the particular word and 2-3 words right). By doing this, they learn not only the meaning of the word, but also its collocation.

d)   Stages which involve exposure: Even if there is no intentional focus, the structure of the language together with the collocations are observed and analyzed unconsciously by learners. The more sentences students read, the more they are exposed to the language use and they subconsciously learn the usage of the unfamiliar word with other words.

Although being a new technology, DDL is not too difficult to use, and can be mastered both by teachers and students, if it is presented to them, so it should not be seen as a radical strategy, but acknowledged as ordinary practice (Boulton, 2010). Bernardini (2001) states that "the difficulties should not be overestimated; learners should quickly acquire the skills needed" (p.243). Likewise, Sinclair (2004) declares that "any teacher or student can readily enter the world of the corpus and make the language useful in learning" (p. 297).

In the English class the steps of mastering DDL can be categorized as: training, execution, deduction, cooperation and assessment.

In the first step, the teacher explains to the students how to apply and derive benefit from online corpus. Sinclair (2004) emphasizes training by stating "both the teacher and the student can make use of a corpus right away, with only a modest few hours of orientation" (p. 288).

In the second step, the teacher introduces the first online corpora "British National Corpus" (BNC) to the students. The students enter the web site with the teacher's help. When the students are ready and the website is uploaded, the first word is entered to the website and with the help of "find" or "ctrl + f" command. The target word in each sentence is highlighted as in Figure 1.

In the execution step, the students perform what they are trained for. If the students are directed to learn the meaning of an unfamiliar word, they read all sentences focusing on it. Students read as many sentences as they can read within the time allocated to them by the teacher.

In the deduction step, the students try to infer the meaning of the word from sentences. They relate the sentences and context they come across and infer the meaning. If the allocated time is not enough, the teacher provides some

more minutes for them to read more sentences.

In cooperation step, the students first read the sentences in online corpora and after the allocated time is finished, they compare what they deduced with their partner(s). This activity is held in pairs or in groups upon the teacher's decision. After the students share their deductions, they also share their ideas about how they inferred the meaning. Students find out the way of their friend(s), so that in the second run, they can use their friends' strategies as well.

In the assessment step, the teacher asks the students what they inferred from the sentences they read individually or in group. After the teacher gets feedback from the students, he/she gives the exact meaning and definition of the unfamiliar target word.

After the target word's meaning is deduced, the teacher introduces the second website which is "Corpus Concordance English - Lextutor". The students enter the website with the teacher's help and choose a corpus or a number of corpuses from the list. The corpuses are Brown, BNC Written, BNC Spoken, Academic General, BNC Med, BNC Commerce, BNC Humanities, BNC Law, BNC Social Science, BlaRC Brit Law Reports, Electrical Engineering, US TV Talk, Univ. Word List, TV Marlise, RAC Academic, Academic Abstracts, Call of the Wild, JPU Learner, BNC Speech, and BNC COCA. The teacher guides the students to choose a corpus (or more than one corpus at a time) related with the word.

As soon as the students enter the target word, a number of sentences are enumerated in a concordance style in which the target words are listed vertically. The concordance

A06 **2070** One of the great things about drama school is what you learn to reject as much as what you actually learn.

A30 **335** The conference voted by 4,592,000 votes to 1,443,000 to reject a motion advocating the setting up of a working party to consider options for electoral reform.

AHN **1547** We British do indeed believe we should all be equal under the law, but in practice we reject social engineering which tries to make us equal in other respects.

ANA **406** If the parents of a mentally handicapped child immediately reject the child on realizing that it is handicapped, or find that life with the child is so difficult that they feel unable to cope, it falls upon the state to find an alternative place of residence, in particular on local authorities.

ASK **719** But this is to introduce a utilitarian calculus, which some may reject as a basis for choosing how to allocate resources.

Figure 1. *A sample of highlighted word ("reject") in BNC (British National Corpus, 2010).*

```
right, so you would actually  REJECT  this model as it stands, alright
it once and once and for all  REJECT  this philosophy and send it back
ion submitted by Mr . I also  REJECT  each of the additions submitted
stand up and to renounce and  REJECT  your statements and weaken your
hey go out we all come back,  REJECT  work, such and such a component
d er er we don't hide behind  REJECT  letters. If I don't think you're
what we don't do hide behind  REJECT  letters. Mm mm. Er regret letter
reference is, don't know? But  REJECT  the option of expansion of exist
bably reject, oh yes, we can  REJECT  the null at ten percent of norma
Victorian values. Colleagues,  REJECT  the ideas of workfare, make sure
```

Figure 2. *A sample of concordance lines in BNC for "reject" (British National Corpus, 2010).*

lines consist of minimum 2-3 words left and 2-3 words right from the target word.

The teacher asks the students to read the concordance lines beginning from 2-3 words left to 2-3 words right from the target word. The students analyze the collocations and observe the word and its collocations. The teacher also highlights the collocations by repeating them loudly. Moreover, the teacher helps the students recognize the structure of sentence with the target word.

IV. Methodology

As the impact of DDL in foreign language teaching has been controversially viewed in recent years (there are some doubts about its difficulty as well as its efficiency), in the present study the effectiveness of DDL in vocabulary learning will be analyzed.

The research question of the study was as follows:
Is there a statistically significant difference between the experimental (data-driven vocabulary learning) and the control (traditional vocabulary instruction) group at high school level?

The study was performed in one of the private boy dormitory high schools in Uganda. The participants were studying in secondary section in Year 10, aged between 14 and 15. The English proficiency level of the participants was upper-intermediate as they took reliable Michigan Proficiency Vocabulary Test before getting education, so the groups were homogenous.

There were 84 participants, who were chosen randomly and divided into two groups. The first group was the Experimental group (EG) which applied DDL, on the other hand, the second group was the Control group (CG) which was taught vocabulary in the traditional way (without any DDL).

Some of the learners failed to complete some lessons and post-tests, so 12 students' results were removed from the analyses. Therefore, the final N size was 36 for Experimental group and 36 for Control Group at the end for the analysis of the results.

32 words out of 50 were chosen according to preliminary testing that more than 60% of the students did not know (failed to fulfil corresponding items) in order to be taught. Eight words were taught in each session which was composed of two lessons totally lasting 80 minutes.

The duration of the experiment was three months and during this time, both EG and CG took a pre-test before the experiment (for 50 words) and a post-test (for 32 words) at the end of experiment.

The tests held involved multiple choices, gap-filling, matching the words with their definitions, interpretation/synonym, and words in context tasks. The whole test was assessed out of 100 points.

Throughout the experiment, the control group was taught the unfamiliar words by means of 12 traditional vocabulary activities in order. The activities were as follows:

1. The teacher reads the words aloud and then the students repeat after him/her loudly.

2. The teacher presents the definition of the unfamiliar words one by one.

3. The teacher asks the students to repeat the words loudly in group and alone.

4. The teacher asks the students to write each of the target words 5 times to their notebook.

5. The teacher presents the synonyms of target words one by one.

6. The teacher reads the words aloud and asks the students to find and give the synonym of the target word verbally.

7. The teacher reads the synonym of the word aloud and asks the students to find and announce the target word loudly in groups or alone.

8. The teacher presents the antonyms of target words one by one.

9. The teacher reads the words aloud and asks the students to find and give the antonym of the target word verbally.

10.The teacher reads the antonym of the word aloud and ask the students to find and announce the target word loudly in groups or alone.

11. The teacher asks the students to fulfill gap-filling and match drills, write sentences by using the target words and read one of them aloud in class.

12.Form (vocabulary)-focused speaking exercises were applied.

For each session, so as to teach the eight unknown words, the activities above were implemented and the teacher evaluated the results by means of post-test after each session.

The experimental group, on the other hand, was taught the unfamiliar words by means of 16 steps for each session. The steps were used in the order as follows:

1. The teacher asks the students to assemble in the computer lab with an internet connection and open a web page (British National Corpus).

2. The teacher gives the students the target words one by one.

3. The teacher asks the students to read first 10 examples of the target word and try to infer the meaning.

4. The teacher encourages them to read 10 more sentences to make their guess stronger.
5. The students do not announce what they found by reading sentences until the teacher asks them to utter.

6. Upon their guess, the teacher divides the students

into groups and guides them to compare and discuss their results of guess with their classmates.

7. As soon as all the students are certain about the meaning of the target word, they share with their teacher what they inferred.

8. If the students fail to find the exact meaning, the teacher encourages them to read more sentences.

9. The teacher asks the students to write the target word and choose a corpus or more than one corpus in the website (Corpus Concordance English).

10. After the results come out, the teacher shows the students how to read the concordance lines.

11. Upon demonstration, the teacher asks the students to read the target word by including 3-4 words on the left and 3-4 words on the right.

12. The students read all the concordance lines silently.

13. The teacher asks the students read the concordance lines loudly in group and alone.

14. The students see the target word and comprehend the usage of word with other word groups so data in concordance drives the students to learn.

15. The teacher asks the students to fulfil gap-filling and matching drills, write sentences by using the target words and read one of them aloud in class.

16. Form (vocabulary)-focused speaking exercises were applied.

With time, students became more independent in using DDL. If the first 4 weeks involved whole-class activities, later the students were working in pairs and small groups, while the teacher was available for help whenever needed. While doing homework, students were also recommended to use DDL when they came across unfamiliar vocabulary.

V. Results

The results revealed that the mean pretest scores in both groups seem close to each other (M=23.44; M=24.83) for CG and EG respectively as in Table 1 and Figure 3 below. The standard deviations are also close to each other, but a bit too high (showing that the groups' level of vocabulary skills is not very homogeneous).

Table 1. *Pre-test Scores for CG+EG*

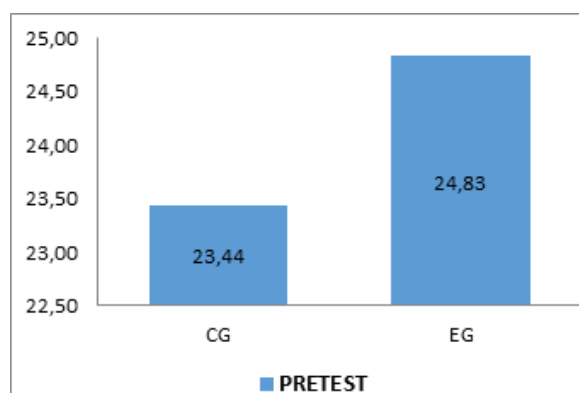| GROUP	Mean	N	Std. Deviation
CG	23.4375	36	11.14284
EG	24.8264	36	13.36190
Total	-	72	-

Figure 3. *Pre-test Scores out of 100 for Control and Experimental Groups*

After receiving the treatment, their scores in the post-test changed (M=55.90; M=65.19) for CG and EG group respectively as in Table 2 and Figure 4. The control group improved the result by 32. 4625, while the experimental group improved the result by 40.36.35. To evaluate whether there is significant difference or not, ANOVA test was implemented. There is no significant difference between the groups in pre-test, $F(1, 71) = 0.229$, $p = 0.633$, whereas in terms of posttest, the main effect comparing the two instructional types was found significant, $F(1, 71) = 6.254$, $p = 0.015$, thus showing difference in the effectiveness of the two instructional types as in Table 3.

Table 2. *Post-test scores out of 100 for Control and Experimental Groups*

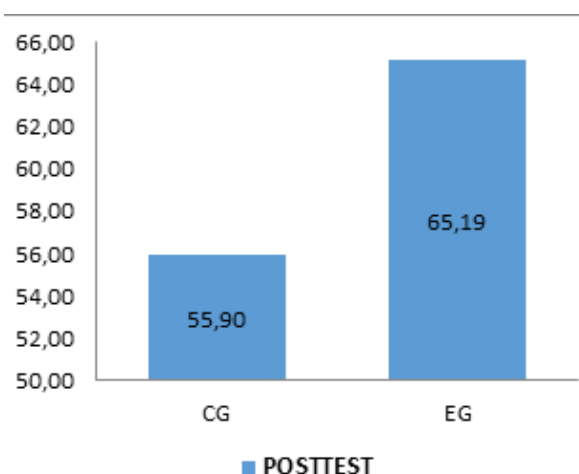| GROUP	Mean	N	Std. Deviation
CG	55.9028	36	17.22646
EG	65.1910	36	14.13726
Total	-	72	-

Figure 4. *Post-test Scores out of 100 for Control and Experimental Groups*

Table 3. *Pre-test and Post-test Scores effects for CG+EG (ANOVA)*

		Sum of Squares	df	Mean Square	F	Sig.
PRE-TEST	Between Groups	34.722	1	34.722	.229	.633
	Within Groups	10594.618	70	151.352		
	Total	10629.340	71			
POST-TEST	Between Groups	1552.870	1	1552.870	6.254	.015
	Within Groups	17381.456	70	248.307		
	Total	18934.326	71			

VI. Limitations of the Study and Directions for Future Research

It is difficult to come to generalized conclusions, because:

• the duration of the experiment was limited to three months;

• the experiment was held with one grade of one school in Uganda with 72 students.

Since Data-Driven Learning has not a long background and there are limited empirical studies, in order to come to really generalizable results, further empirical studies should be held with more participants, in more schools and countries.

VII. Conclusion

In the research, one of the recent vocabulary learning strategies, data driven vocabulary learning, was evaluated as being more effective compared to traditional way of vocabulary teaching. In the pre-test, it was found that there was no significant difference between the groups, however, in the posttest, the experimental group, which scored M=24.83 in the pre-test and M=65.19 in the post-test, had significant difference when compared to the control group, which scored M=23.44 in the pre-test and M=55.90 in the post-test.

The students using DDL became researchers using their own talents and discovery skills investigating the real meaning of the target word in a real context. It was also enjoyable, like a detective game, as they were working on the concordance lines to discover the language items. Furthermore, the students learned how to benefit from inductive explorations (Meyer, 2004) and became autonomous learners.

Finally, this study showed the higher effectiveness of DDL in vocabulary learning in class compared to the traditional way of learning. Moreover, by applying and learning data driven learning, the students got a 'portable teacher' out of classroom.

References

Bernardini S. (2001) 'Spoilt for choice': A learner explores general language corpora. In Aston G. (Eds.) Learning with Corpora. Houston: Athelstan, p. 220-249

Boulton, A. (2010). Data-driven learning: taking the computer out of the Equation. Language Learning, 60(3), p. 534–572 British National Corpus (2010). Retrieved February 13, 2015 from http://www.natcorp.ox.ac.uk/

Folse, K. S. (2004). Vocabulary Myths: Applying Second Language Research to Classroom Teaching. Ann Arbor, MI: University of Michigan

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. Revised version in Johns, T. & P. King (Eds.). Classroom Concordancing. English Language Research Journal 4, p. 27-45

Laufer, B. (1986). Possible changes in attitude towards vocabulary acquisition research. International Review of Applied Linguistics in Language Teaching, 24 (1), p. 69-75

Meyer, C. F. (2004). English Corpus Linguistics: An introduction. Cambridge University Press

Nation, I. S. P. (1990). Teaching and Learning Vocabulary. New York: Newbury House

Nation, I. S. P. (2001). Learning Vocabulary in Another Language. Cambridge, UK: Cambridge University Press

O'Malley, J. M., & Chamot, A. U. (1990). Learning Strategies in Second Language Acquisition. New York: Cambridge University Press

Oxford, R. L. (1990). Language Learning Strategies: What every the teacher should know. Boston: Heinle & Heinle Partridge, B. (2006). Discourse Analysis. New York, NY. Continuum

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. Modern Language Journal, 79(1), p. 15-28

Schmitt, N. (1997). Vocabulary learning strategies. In Schmitt, N. and McCarthy, M. (eds), Vocabulary: Description, Acquisition, and Pedagogy. Cambridge: Cambridge University Press. p. 199-227

Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. Sinclair (Ed.), How to Use Corpora in Language Teaching. Amsterdam: John Benjamins, p. 271-299

Tribble, C. & Jones, G. (1997). Concordances in the Classroom: A resource guide for the teachers. Houston: Athelstan Publications

Wilkins, D. A. (1972). Linguistics in Language Teaching. Cambridge, MA: MIT

My notes:

1. This is one of the best manuscript I read so far. Well written and clear to understand.
2. Minor correction is needed. My editing is on the text ("track change format").
3. At the end of "conclusion" before "references", It would be nice if the author can add a sentence what is the suggestion for the next after this research.
4. The text is in the order of scientific order and the literature in the "reference" matches with the body of the text.
HT (March 10, 2015)