

Is Incidental Vocabulary Acquisition Feasible to EFL Learning?

Luo Jian-ping¹

¹Guangdong University of Petrochemical Technology, China

Correspondence: College English Department, Guangdong University of Petrochemical Technology, Maoming, Guangdong, 525000, China. E-mail: lm2923302@126.com

Received: July 2, 2013 Accepted: July 24, 2013 Online Published: September 4, 2013

doi:10.5539/elt.v6n10p245 URL: <http://dx.doi.org/10.5539/elt.v6n10p245>

Abstract

For learning English as a foreign language, the efficiency of the approach of incidental vocabulary acquisition depends on the word frequency and text coverage. However, the statistics of English corpus reveals that English is a language that has a large vocabulary size but a low word frequency as well as text coverage, which is obviously not in favor of the approach. Also, the statistics reveals that, by learning English words from reading incidentally as the approach claims, the learners will have to add up their reading out of class to a large quantity as ten times as that in class. Accordingly, it will be too much for them to do and clearly it is unfeasible to these non-English major learners, who could not probably have time to spend in reading so much just for picking up new words incidentally. This paper aims to prove that the approach is not feasible at all to EFL learning in China.

Keywords: incidental vocabulary acquisition, intentional language learning, word frequency, text coverage, college English

1. Introduction

1.1 Some Poor Data from College English Test Band 4

Twice a year, College English Tests of Band 4 and 6 are held for all students of universities and colleges in China, and the feedback data of the results always show that most students have got a poor grade from the tests, which can be seen in the following tables (The data are only a small part from College English Test Band 4 and 6 Service.). See Table 1 first:

Table 1. The mean scores of CET4 in the recent years

Test Date	All-M (SD)	Un-M (SD)	211-M (SD)	Non-M (SD)
2012.6	391 (63)	400 (65)	439 (82)	396 (61)
2011.6	390 (62)	399 (65)	433 (79)	396 (62)
2010.12	386 (66)	396 (69)	436 (85)	391 (65)
2010.6	387 (69)	398 (73)	436 (88)	394 (69)

In table 1, All-M is the *mean of all test-takers*, SD is *standard deviation*, Un-M is the *mean of those undergraduates among all test-takers*, 211-M is the *mean of the undergraduates only from 211-universities* (key universities), and Non-M is the *mean of the undergraduates from the non-211-universities* (non-key universities). For CET4 score system, the test pass line is 425, the highest score is 710, and the lowest score is 220. All CET 4 and 6 scores are given according to the Z-score system of Normal Distribution, in which the 425 is in -2 zone, the 710 in +3 zone, and the 220 in -4 zone, with the 500 on the 0 point as the norm mean.

It is not difficult to see from Table 1 that the English teaching and learning in China (ETLC) have long been staying in a low level, since the mean scores of all test-takers are below 400, except those of 211-M, which are just slightly better than the others. However, it would be more surprising to see the number and rate of those college students who have learned that foreign language for at least eight years, but failed the test again and again. See table 2:

Table 2. Some comparative data about the test

Test Date	All U-takers	≥430 / %	630-710 / %	330-220 / %
2012.6	3614882	323764 / 29.5%	3898 / 0.1%	456826 / 12.6%
2011.6	3420565	286042 / 28.3%	4277 / 0.1%	430153 / 12.6%
2010.12	3572224	277713 / 28.6%	3107 / 0.1%	594345 / 16.6%
2010.6	3313653	256240 / 30.3%	4792 / 0.1%	582239 / 17.6%

The data in table 2 are all about the undergraduates with a non-zero score from the test. In the table, “*All U-takers*” refers to the number of test takers who are four-year undergraduates from universities and colleges with scores above zero, and “*≥430 / %*” refers to the number and rate of the four-year undergraduates whose scores are above 430, and means that those test takers have passed the test, and “*630-710 / %*” refers to the number and rate of the four-year undergraduates who get the top score, and “*330-220 / %*” refers to the number and rate of the four-year undergraduates whose scores stay on the bottom in the grade system.

1.2 Something Crucial about English Teaching and Learning in China

The tables above tell something more crucial about ETLC. As the CET4 pass line is 425, we find that, for each time of the test, there is nearly 70% of the four-year university students who could not pass the test, and the total number of the students can be inferred to far more than two million from the data of the “*≥430*”. Besides, since the score system tells that the 330-220 means that the test takers have almost achieved nothing from the test, we are surprised to find that there are so many poor students of far more than four hundred thousand and they might probably know little about English even though they have been learning that language for eight years long.

CET 4 was designed to measure and guide ETLC from the very beginning when it came into being in 1980s, but it is hard to find any significance from its practice of so many years, and therefore it could not be by any means defined as “It has already produced a good social benefit” by Yang Huizhong (2004). In fact, ETLC is “the most time-consuming course” (Luo, 2013), for there is nearly 70% of the Chinese students who have spent so long a time on learning English from primary school to college but still can hardly read any thing from that language. So, questions would come out: What is wrong with ETLC? What is the cause of such poor test performance?

1.3 Small Vocabulary Size and Poor Learning

Generally believed in the world of English teaching and research in China, one’s poor performance is due to his or her too small English vocabulary. In the past decades, many scholars have conducted English vocabulary surveys and found that the Chinese students who were then studying English in universities and colleges usually had a small English vocabulary. A latest survey, for example, shows that their vocabulary is of only from 2300 to 6800 English words (Dai Junhong, 2013). But, it can be inferred, according to the feedback data mentioned above, that a lot of undergraduates might have a far smaller vocabulary than that they claimed, perhaps not more than 1500 English words. The number is what a junior middle school student is required to have at least, because any one good junior middle school student would learn English better than these poor college students whose score always stay on that bottom.

Almost all scholars agree that, for learning English as a foreign language (EFL), a small vocabulary would certainly restrict the improvement of learning, but they differ on how the learners should acquire English vocabulary. Some argue that the EFL learners should try their best to learn and memorize English words intentionally (Allen, 1983), while most other scholars hold that a reading in large quantity is the best way to pick up English words incidentally for EFL learning, which is known as *Incidental Vocabulary Acquisition (IVA)*. They believe a large vocabulary only comes from a large reading. Is it really true?

1.4 Questions of This Research

However, facing such poor results of CET4 mentioned above, shouldn’t we ask ourselves such questions as that, if IVA were really a help for ETLC, why so many CET4 takers, more than the half, could not give themselves a good performance but failed, having learned that foreign language for so long a time? Isn’t their failure related to a too small vocabulary? Shouldn’t we rethink it over profoundly? It is believed that something must be wrong with ETLC, especially the approach of vocabulary learning. So, this paper is going to discuss the following three questions: first something about IVA researches and problems; then, the features of English vocabulary; last, the unfeasibility of IVA in EFL learning.

2. About IVA Researches and Problems

2.1 About IVA Research

It is generally held that so-called IVA was first mentioned in 1985 by Nagy, Heman and Anderson (1985). But not until about 2000, did it attract much attention of the field of English teaching and research in China. IVA is widely defined that the learners can incidentally pick up the new words when they come across them for many times in their reading just for information or message rather than for learning English. Scholars try to prove that, for EFL learning, one could acquire a new word after many times encountering it during their practice of listening, speaking, reading and writing (Laufer, 1998). They call the new word learned with IVA a by-product, and claim it is different from the approach of intentional language learning (ILL) (Laufer & Hulstijn, 2001).

Since about 2000, quite a few IVA research papers in China have gradually got more and more notice, like those by Gai Shuhua (2003), Duan Shiping & Yan Chensong (2004), Li Hong & Tian Qiuxiang (2005), and so on. Gai Shuhua is one of the earlier scholars who introduced the IVA researches inside and outside China (2003), and also conducted an empirical study on English major students. Duan & Yan (2004) drew a conclusion from their research that it would be better to choose and compile the reading materials with multiple choices glossing, because they got a better result with IVA. Li & Tian (2005) pointed out that IVA only means to pick up new words unintentionally from reading, and there is no reason to put it against ILL, and for EFL learners, it is better to focus on learning and memorizing new words intentionally while reading. Li and Tian's view wins much acceptance from other scholars.

Many English vocabulary teaching researches focus more on the effectiveness of IVA and strongly claim that it is good to ETLC, even though most only carried out an empirical study on reading just one or two passages or novels. Only a few researches have noticed the restriction of word frequency on the efficiency of IVA which was warned by Laufer (2003), but almost failed to probe it further.

2.2 About Two Restrictions on the Efficiency of IVA

To acquire a new word from IVA, some scholars (Zahar, Cobb & Spada, 2001) find that the word frequency needs from 6 to 20 times' encountering. The average is 10 times (Saragi, Nation & Meister, 1978), and at least 8 times is needed (Horst, Cobb & Meara, 1998; Waring & Takaki, 2003). Accordingly, some Chinese scholars emphasize that it will be better to consult a dictionary or combine IVA with ILL (Gao Xinhua, 2010).

Some scholars also notice another problem, the learner's vocabulary size, which would become another restriction on the effectiveness of IVA beside the word frequency. Li & Tian (2005) claim that the application of IVA requires the student to learn at least 2000 words first; Gai (2003) believes that "2000 to 3000 words are needed first, and for College English learning, 5000 to 6000 words must be the base for IVA". Some hold that, in all, the first thing is to learn the first several thousand words before using IVA" (Nagy, Hermann & Anderson, 1985; Nation, 2001), and the EFL learners could not have a good understanding of what they read nor acquire any new words from reading as the English native do until they have mastered at least 5000 word families first (Coady, & Huckin, 1997; Nation, 2001).

Of all researches, there is hardly any voice saying "no" to IVA but one from Luo Weihua & Deng Yaochen (2009), who use corpus and find that, if 10 times encountering is the average to pick up a word incidentally, then an EFL learner needs to finish a reading of four hundred thousand words, the same as 400 texts with 1000 words each, but only to pick up 2600 English words. So, the required reading quantity is too large and heavy for a Chinese EFL learner to have time to finish, which then might actually constitute the main factor to cause such a small vocabulary and the poor learning mentioned above.

Another research by Wu Wei and Xu Hong (2006) also notices the problem and calculates that a college student needs to read 7400 words every day so as to pick up the required vocabulary during his or her College English learning years. Though their research lacks something scientific, it tells the truth about IVA. That is, a much large reading quantity needed, but a much low efficiency of acquiring words gained.

2.3 About Major Opinions of IVA Researches

However, quite a few scholars out and in China, insist that IVA is feasible to EFL learning. They take for granted that a considerable part of the EFL learner's vocabulary is made up of the by-products of reading, and IVA is the only way to enlarge their vocabulary (Nagy, Hermann & Anderson, 1985; Nation, 2001; Wu Wei & Xu Hong, 2006). They seem to have neglected something about the features of the English vocabulary and failed to realize that the English word frequency was actually an insurmountable obstacle to IVA in EFL learning.

So, there is a need to continue the discussion of features of English vocabulary first in the following.

3. Features of English Vocabulary

3.1 A Comparative Study of English and Chinese Vocabulary

How do we know about the major features of English vocabulary? The best way is to make a comparison with other languages, here take Chinese for example. We have collected two corpora of one million words each, one is English, and the other is Chinese. They both consist of texts, news, stories, novels, and academic literatures. For English statistics, we count word types, instead of lemmas or word families. For Chinese statistics, we count Chinese characters, not words or expressions. “The Chinese character is a basic unit of Chinese structure”, Xu Tongqiang (2005) argues, a professor of Beijing University. “The character frequency is one important attribute for use of Chinese” (Li Guoying & Zhou Xiaowen, 2011). So, “the statistics of Chinese characters is of a much value to language teaching” (Fu Yonghe, 1985). The following is a comparison of word frequency and text coverage between English and Chinese:

Table 3. Frequency of English words and Chinese characters

English	Chinese	
Tokens	1 015 941	1 025 527
Types	35 416	4 513
F ≥ 50	2108 (5.9%)	1704 (37.7%)
F ≥ 15	5494 (15.5%)	2574 (57.0%)
F ≥ 10	7370 (20.8%)	2887 (63.9%)
F ≥ 5	12034 (33.9%)	3392 (75.1%)
1-f ws	13096 (36.9%)	542 (10.1%)

In table 3, the percentage in the bracket is the rate of the types used in the English or the Chinese corpus, and both numbers and rates are the types accumulated except those in the line “1-f ws”, which means “Hepax Legomena” (Gui Shichun, 2010), a word of just one frequency in the corpus. For text coverage, now see table 4:

Table 4. Text coverage of English and Chinese

Ws or chs	En txt coverage	Ch txt coverage
First 50	43.3%	29.1%
First 100	50.9%	40.5%
First 1000	74.1%	90.3%
First 1500	78.4%	95.4%
First 2000	81.4%	97.8%
First 3000	85.5%	99.5%
First 4000	88.1%	99.9%
First 5000	90.0%	—
First 10000	94.9%	—

In table 4, “Ws or chs” refers to English words or Chinese characters, “En txt coverage” refers to English text coverage, and “Ch txt coverage” to Chinese text coverage. The data are also from the same corpora as table 3, with about one million words or characters in each.

3.2 Discussion of Major Features of English Vocabulary

From the above tables, it is not difficult to find that, compared with Chinese, English has a much larger vocabulary, and therefore, because of so many words, English word frequency is then unavoidably much lower than the Chinese one. Averagely, for example, every English word shares a frequency of 28.6 times, while every Chinese character has a frequency of 227.2 times. Seemingly, there are more than 7300 words with a frequency

of above 10 times from the English corpus, which makes it possible to pick up from reading, but they only take up so small a part of the vocabulary (20.8%) that it could not meet the least required ability of reading. On the other hand, there are so many words that only appear one time (36.9%) that it would be impossible to pick them up just from reading. Then, how could one get a good understanding of what he or she is reading without a good understanding of these words?

A low text coverage is also another feature of English. According to the comparison shown in Table 4, the first 100 English words have a much larger coverage than the Chinese ones, but unfortunately, the other words soon slow the coverage down, and as more and more words come in, the coverage comes up more slowly. For the claimed best coverage of 95% (Schmitt & McCarthy, 1997), there are at least ten thousand words required to learn, which would seem undoubtedly to be hard for an EFL learner, while for Chinese, there are only 1500 characters enough.

Besides, from above statistics, it can be seen that English word frequency is only distributed highly among the first one thousand words, but for the second and the third and the other thousand words, it scatters quickly away and drops sharply, which therefore would be worse to the IVA approach.

4. Unfeasibility of IVA in EFL Learning

4.1 Too Small a Vocabulary Able to Be Picked Up with IVA

Let's take the course called *New Horizon College English* (NHCE) as a concrete example. The course has four books with ten units each, and each unit is made up of two texts, so 80 texts totally, and in each text there are about 800 words averagely. The following are the word frequency of NHCE. See Table 5:

Table 5. Word frequency of NHCE

Tokens: 65635 Types: 7812		
Frequency	Words	Type rate
F ≥15	561	7.1%
F ≥10	865	11.0%
F ≥5	1874	23.9%
1-f ws	3427	43.8%

In this table, the number and rate of the columns "Words" and "Type rate" are accumulated except those in the line of "1-f ws". Suppose that word frequency of 10 times is accepted for picking up a word incidentally by means of IVA, then there are only a much small vocabulary of 865 words with the frequency above 10 times, and other more than 6800 words could not probably be picked up until the learners enlarge their reading greatly to meet them for more enough times, or use the ILL approach instead.

4.2 Too Large a Vocabulary Needed for IVA

It is generally believed that the proper vocabulary size for IVA is the one that can reach the claimed best coverage of 95%. And the needed vocabulary for NHCE course books could be found from table 6 in the following:

Table 6. Text coverage of NHCE

Words	Txt cvrg	Freq.
First	1000	76.4% ≥ 8
Second	1000	85.4% ≥ 4
Third	1000	90.4% ≥ 3
Fourth	1000	93.6% ≥ 2
Fifth	1000	95.7% ≥ 1
Sixth	1000	97.2% ≥ 1

In table 6, "Txt cvrg" is Text coverage, and "Freq." is the short of Frequency. According to the table, the large

vocabulary needed is 5000 words at least, which would be a hard job to the learners. Also, since the word frequency drops so sharply from 8 to 1 with every thousand words, how could the learners pick up so many words from the second to the fifth thousand words incidentally? Apparently, they couldn't until they enlarge their reading quantity greatly.

4.3 Then Too Heavy a Reading Load for EFL Learning

Then, under the impacts of word frequency and text coverage, how large is a reading needed for picking up so many words incidentally?

Suppose that the frequency of 10 is the lowest frequency to pick up a new word, and also suppose that the learners must add up their vocabulary to 5000 words, then, according to the statistics of the English corpus in Table 3, they have to finish so large a reading as more than 660,000 words out of class, eleven times as much as that of NHCE course books. That is to say, the NHCE learners have to read 800 more texts with 800 words each after class, or when they finish reading one text in class, they still have other ten texts waiting to be read. Is it possible for them to do that? No. Even though they are interested in reading so much, they could not have so much time to complete that job. For, as non-English majors, these NHCE learners have many other subjects to study, and could not possibly spend any more time on this non-major course.

Quite different from English, Chinese characters have a much more high frequency and only with 1500 characters does the text coverage reach 95%, and therefore, a Chinese learner may be able to learn the required words with IVA. But for EFL learning, IVA may not. It is unfeasible without much more reading as a precondition.

5. Conclusion

For learning vocabulary, the IVA approach is much restricted by word frequency and text coverage of the language. English is such a language with a low word frequency as well as text coverage, and so for College English learning in China, IVA is just merely one strategy to pick up words from reading. Since IVA asks for too much a reading but only gives a much low efficiency of vocabulary acquisition, it could not have been taken as a guide for ETLC. On the contrary, ILL would be a better one supposed in EFL learning, because "much of the English vocabulary must be learned", and "students should be encouraged to take more responsibility for their own vocabulary learning" (Allen, 1983), and also because learning English words intentionally "can give a sense of progress, and a sense of achievement" (Schmitt & McCarthy, 1997).

References

- Allen, Virginia, French. (1983). *Techniques in Teaching Vocabulary* (pp. 5-7). Oxford: Oxford University Press.
- Coady, J., & Huckin, T. (1997). *Second Language Vocabulary Acquisition A Rationale for Pedagogy* (pp. 225-237). Cambridge: Cambridge University Press.
- Dai, Junhong. (2013). A Study on the Vocabulary Size of Non-English Majors at CET-4 Level. *Journal of Chongqing University of Technology (Social Science)*, 1, 118-122.
- Duan, Shiping, & Yan, Chensong. (2004). Multiple Choice Glossing on Incidental English Vocabulary Acquisition. *Foreign Language Teaching and Research*, 3, 213-218.
- Fu, Yonghe. (1985). The New Results of The modern Chinese Character frequency Statistics. *Language Planning*, 3, 44-45.
- Gai, Shuhua. (2003). A Review of Incidental Vocabulary Acquisition. *Journal of PLA University of Foreign Languages*, 2, 73-76.
- Gao, Xinhua. (2010). An Empirical Study of the Dual Unity between Conscious and Incidental Vocabulary Acquisition. *Shandong Foreign Language Teaching Journal*, 5, 55-59.
- Gui, Shichun. (2010). *A Corpus-based Analysis of the Register of English Linguistics* (pp. 32). Beijing: Foreign Language Teaching and Research Press.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207-223.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task induced involvement. *Applied Linguistics*, 22, 12-26.
- Laufer, B. (1998). The development of passive and active vocabulary in second language: Same or different? *Applied Linguistics*, 19, 255-271.

- Li, Guoying, & Zhou, Xiaowen. (2011). Improvement in Statistic Method to Chinese Character Frequency Study. *Journal of Beijing Normal University (Social Sciences)*, 6, 45-50.
- Li, Hong, & Tian, Qiuxiang. (2005). A Study of Second Language Incidental Vocabulary Acquisition. *Foreign Language Education*, 3, 52-56.
- Luo, Jian-ping. (2013). An Action Research on Improvement of Reading Comprehension of CET4. *Englsh Language Teaching*, 4, 89-96.
- Luo, Weihua, & Deng, Yaochen. (2009). A Study of English Lexical Repetition Pattern Based on BNC Texts. *Foreign Language Teaching and Research*, 3, 224-229.
- Nagy, W., Heman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quaterly*, 20, 233-253.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press, 2001.
- Saragi, T., Nation, P., & Meister, G. (1978). Vocabulary Learning and Reading. *System*, 6, 72-78.
- Schmitt, Norbert, & McCarthy, Michael. (1997). *Vocabulary: Description, Acquisition and Pedagogy* (pp. 12). Cambridge: Cambridge University Press.
- Waring, R., & Takaki, M. (2003). At What Rate do Learners Learn and Retain New Vocabulary from Reading a graded reader? *Reading in a Foreign Language*, 15, 130-163.
- Wu, Wei, & Xu, Hong. (2006). Effects of Frequency on Incidental Vocabulary Learning through Reading. *Journal of Chongqing University (Social Science Edition)*, 4, 116-121.
- Xu, Tongqiang. (2005). Chinese Character-centered Theory and Language Study. *Language Teaching and Linguistic Studies*, 6, 1-11.
- Yang, Huizhong. (2004). An Analysis of the English Proficiency of the Chinese Students as Reflected in the National CET Test. *Foreign Languages in China*, 1, 56-60.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring Vocabulary through Reading: Effects of Frequency and Contextual Richness. *The Canadian Modern Language Review*, 57, 541-572.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).