# Bridging the Vocabulary Gap: from EGP to EAP 

CHUJO Kiyomi (Nihon University) NISHIGAKI Chikako (Chiba University)

### 1.0 Introduction

### 1.1 Background

The impact of IT technology on communication has been both explosive and pervasive.
People worldwide now have the opportunity to participate in business and personal affairs on a global level. As such, proficiency in English has become a requirement for many, including Japanese, who want to thrive in this highly technological world.

In Japan, the Test of English as a Foreign Language (TOEFL) has surged in popularity as a method for measuring practical English proficiency with both individuals and organizations. The Japanese Ministry of Education, Culture, Sports, Science and Technology (Mombukagakusho), for example, recently announced a target score for Japanese English teachers as 550 points on the TOEFL (Mombukagakusho, 2002). Many colleges and universities give English credits to students who meet TOEFL score requirements. Consequently, both educators and learners are acutely aware of the importance of improving TOEFL scores as a measure of communication skill.

In spite of the fact that most Japanese students study English for six years in junior and senior high school and another four years, at the most, in college or university, their practical English proficiency does not improve as much as expected. In fact, compared to other countries, the average Japanese TOEFL score ranks as one of the lowest in the world. One of the reasons TOEFL represents a challenging task for Japanese learners is their insufficient vocabulary. Because the TOEFL test is used as an admission tool by colleges and universities worldwide, and because its vocabulary is necessary for formal study, it can be categorized as English for

Academic Purposes (EAP) (Jordan, 1997). In contrast, Takefuta’s research (1981) indicates that the English used in high school textbooks can be categorized as English for General Purposes (EGP). The apparent gap between the EAP vocabulary used in TOEFL tests and the EGP learned through Japanese high school textbooks has made TOEFL tests difficult for Japanese learners. Consequently, Japanese learners need to overcome this gap.

This study was designed to address this issue by identifying, lemmatizing and comparing the vocabulary of various types of English language learning materials and proficiency tests. First, TOEFL vocabulary was compared to the vocabulary found in school textbooks and English tests used in Japan. The degree to which EGP textbooks and materials cover the vocabulary presented within two selected sample TOEFL tests was measured. Finally, a supplemental word list was created that more closely connects EGP textbook vocabulary to that of the TOEFL tests.

### 1.1. 1.2 Review of the Literature

Chujo \& Genung (2003) measured a wide range of vocabulary levels of English texts and materials used at Nihon University, using the British National Corpus (BNC). ${ }^{1}$ Since the BNC reflects present day English usage for speech and publications in the UK (Leech, et al., 2001), this vocabulary was classified as EGP. A "BNC High Frequency Word List" (BNC HFWL) was created by: (a) downloading from Adam Kilgarriff's Web page (http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/) the 38,683 unlemmatized words in the BNC which occur 100 times or more; (b) excluding proper nouns and numerals to ensure its suitability as a criterion list; (c) lemmatizing the words into base word categories; (d) listing each part of speech (POS) form under the same base word (for example, 'answer' (noun) and

[^0]'answer' (verb) would appear only once under the base word 'answer'; and, (e) changing British spellings to American spellings. Using this newly created BNC HFWL, Chujo and Genung then defined the levels "comprehension coverage" for the targeted textbook vocabulary. Nation (2001) emphasizes that "learners would need at least a 95 percent coverage of the running words in the input in order to gain reasonable comprehension and to have reasonable success at guessing from context"(p.114). Because the threshold vocabulary required to comprehend a given text is approximately equal to a 95 percent coverage (Laufer, 1989, 1992), the researchers chose this level as the target. In other words, in order for a reader to understand a text, the reader should be able to recognize 95 percent of the vocabulary, or not recognize one in every twenty words. In comparing the BNC HFWL to the English language materials used at Nihon University, the researchers counted how many words from the top of the list which were needed to achieve a 95 percent 'coverage' of the targeted texts.

Another way to clarify the vocabulary level of a targeted text is to calculate the extent to which the vocabulary in junior and senior high school textbooks (JSH), considered EGP, matches with the vocabulary of the targeted text. We can assume that the words that match the JSH vocabulary were learned by learners and are known to them. Chujo, et al. (1993) investigated the word matches between JSH vocabulary and other educational materials in order to obtain an estimate of the percentage of the amount of vocabulary learned by students at each stage as they advanced from junior high school toward senior high school. They discovered that the JSH textbook vocabulary was insufficient for 'practical' language activities.

### 2.0 Purpose of the Study

The purpose of the study is (a) to measure, utilizing the BNC HFWL, the graduations among vocabulary levels within English Language Teaching (ELT) materials such as junior
and senior high (JSH) texts, college freshman and sophomore teaching materials, and the TOEFL and other tests; (b) to measure the degree in percentages to which JSH textbooks, college textbooks, and other materials cover the vocabulary used in the selected sample TOEFL tests; and, (c) to create a word list to bridge the gap between the vocabulary found in the JSH and college ELT textbooks and materials to the vocabulary found in the TOEFL tests.

### 3.0 Procedure

First, samples of TOEFL tests and a variety of ELT materials and other English tests were collected. Second, word lists were created and vocabulary levels were assessed. Next, the rate of vocabulary coverage provided by the ELT materials (vis-à-vis the sample TOEFL tests) was identified. A specialized TOEFL word list was created, and finally, the efficacy of this list was examined. Each step is outlined below.

### 3.1 Targeted TOEFL Tests

Table 1 shows the two TOEFL sample tests used to measure the TOEFL vocabulary level.
Table 1 also includes tokens (total number of words), and types (number of different words) in the texts.

Table 1 Targeted TOEFL Tests

| Name of Tests | Tokens | Types |
| :---: | :---: | :---: |
| TOEFL Preparation Kit Vol 2. Test A | 7,174 | 1,464 |
| TOEFL Preparation Kit Vol 2. Test B | 7,014 | 1,476 |

### 3.2 ELT Textbooks and Materials Used for Comparison

Two types of material were collected for comparison to the TOEFL vocabulary: ELT materials used at junior and senior high school and college, and various types of English tests used extensively in Japan. Table 2 lists the ELT textbooks and materials used in this study. The junior and senior high school textbook series New Horizon 1, 2, 3 and Powwow I, II, Reading were selected since they are the most widely used textbooks in Japanese schools from the $7^{\text {th }}$ to the $12^{\text {th }}$ grade. ${ }^{2}$ The authors also collected ELT materials used at Nihon University and Chiba University. First Listening and Introduction to College Life are Computer-Assisted Language Learning (CALL) CD-ROM programs used in freshman listening courses. American Ideas in Japan is a textbook used in sophomore reading courses, and Wonderful USA is a textbook accompanied by a video and is used in sophomore listening courses. The number of tokens and types in these textbooks are also included in Table 2.

Table 3 shows the test materials used in this study. In addition to the TOEFL, the authors collected four different types of tests which many Japanese learners are likely to encounter as proficiency or college qualification tests. Using a wide range of tests will make it possible to provide a more precise comparison. As is shown in Table 3, two different sets of each test were

[^1]collected for all five types of tests. In total, ten different tests were examined. Table 3 also shows the average tokens and types of those test sets.

Table 2 ELT Textbooks and Materials Used

| Grade | Teaching Materials | Tokens | Types |
| :--- | :--- | :---: | :---: |
| Junior \& Senior High School | New Horizon 1, 2, 3 | 9,440 | 1,124 |
|  | Powwow I, II and Reading | 27,221 | 2,857 |
| College Freshman | First Listening (CALL CD-ROM) | 3,699 | 807 |
|  | Introduction to College Life (CALL) | 2,766 | 590 |
| College Sophomore | American Ideas in Japan (Units 1-14) | 6,322 | 1,073 |
|  | Wonderful USA | 6,322 | 1,091 |

Table 3 Tests Used

| Name of Tests | Tokens | Types |
| :--- | :---: | :---: |
| Daigaku Center Nyushi 2001 | 3,072 | 686 |
| Daigaku Center Nyushi 2002 | 2,882 | 693 |
| TOEIC Official Guide Vol. 1 Practice Test | 7,642 | 1,411 |
| TOEIC Official Guide Vol. 2 Practice Test | 7,035 | 1,552 |
| Eiken 2 ${ }^{\text {nd }}$ Grade (2000) | 4,057 | 833 |
| Eiken 2 ${ }^{\text {nd }}$ Grade (2001) | 4,351 | 849 |
| Eiken Pre 1 ${ }^{\text {st }}$ Grade (2000) | 6,087 | 1,493 |
| Eiken Pre 1 ${ }^{\text {st }}$ Grade (2001) | 5,990 | 1,397 |
| Eiken 1 ${ }^{\text {st }}$ Grade (2000) | 7,307 | 1,780 |
| Eiken 1 ${ }^{\text {st }}$ Grade (2001) | 7,249 | 1,740 |

### 3.3 Creating Word Lists from Collected Materials

A 'word list' is an alphabetized list of all the different words that occurred in a targeted text, accompanied by the frequency of occurrence. In this study, the authors lemmatized all words in the word lists. For example, inflectional forms such as 'cat-cats' and 'go-goes-went-gonegoing' were listed under the base word forms of 'cat' and 'go'. In addition, proper nouns and numerals were excluded from each text or material for "they are of high frequency in particular texts but not in others, ... and they could not be sensibly pre-taught because their use in the text reveals their meaning" (Nation, 2001, pp.19-20). To create the word lists, all the text data from the collected tests and ELT materials were inputted using a scanner. This data was proofread and then lemmatized by inflectional form. Next, part of speech (POS) tags were added to each word by using the Tree Tagger Program (http://www.ims.unistuttgart.de/projekte/corplex/TreeTagger/index.html ). Proper nouns and numerals were excluded from each list manually according to the POS. Finally, the data from each test or ELT material was collated into a word list using software programs developed by Takahashi (1999) and Takefuta (1986).

### 3.4 Assessing TOEFL, ELT Material, and English Test Vocabulary

The authors next assessed the vocabulary level of each text shown in Tables 1, $\mathbf{2}$ and $\mathbf{3}$ by comparing it with the BNC HFWL discussed earlier. Each targeted text or transcript vocabulary level was defined in terms of the number of words from the BNC HFWL that equaled 95 percent coverage of that text. Thus, the BNC HFWL was used to identify the graduations among the diverse vocabulary levels contained within the TOEFL samples, the ELT materials, and the English tests.

### 3.5 Vocabulary Matches Between TOEFL Tests and ELT Materials

In the next step in the study, a calculation was made on the increase in the rate of coverage provided by each ELT material (vis-à-vis the sample TOEFL tests) as one progressed from the lower-course texts or materials to the higher ones. Word matches between the sample TOEFL tests and the ELT materials were determined in order to obtain an estimate of the percentage of increase in the amount of vocabulary learned at each stage by the students.

### 3.6 Creating a TOEFL Vocabulary List

Nation (2001) states an effective word list should provide reasonable frequency of occurrence of words and encompass a wide range of texts. In order to create a specialized TOEFL vocabulary list with Nation’s criteria in mind, the authors collected 18 sets of TOEFL practice tests in addition to the two sets of TOEFL practice tests that are already listed in Table 1. Each test was processed into a word list by the same procedure described in 3.3. In total, twenty sets of practice tests were collected and processed. Table 4 shows the titles of the TOEFL practice tests used to create a specialized TOEFL vocabulary list, along with the tokens and the types in the tests.

The next step was to subtract the JSH textbook vocabulary from each of the twenty TOEFL practice tests in order to exclude the vocabulary that college students supposedly learned before entering college. The remaining 4,848 words were then arranged in order of frequency, resulting in a list of high-frequency TOEFL vocabulary items. Words on this list with a frequency of occurrence of less than four were excluded. To keep the sample well within an acceptable range, those words which appeared in only one or two of the TOEFL practice tests were deleted from the 1,412 remaining high frequency words. The resulting "specialized

Table 4 TOEFL Tests Used for Creating a TOEFL Vocabulary List

| Name of Practice Tests | Tokens | Types |
| :---: | :---: | :---: |
| TOEFL Test Preparation Kit Vol. 1 A | 7,078 | 1,401 |
| TOEFL Test Preparation Kit Vol. 1 B | 5,768 | 1,324 |
| TOEFL Test Preparation Kit Vol. 2 A | 7,540 | 1,502 |
| TOEFL Test Preparation Kit Vol. 2 B | 7,374 | 1,476 |
| TOEFL Test Preparation Kit Vol. 2 C | 7,176 | 1,410 |
| TOEFL Test Preparation Kit Vol. 2 D | 7,784 | 1,496 |
| TOEFL Practice Tests 2nd Edition Test A | 7,207 | 1,485 |
| TOEFL Practice Tests 2nd Edition Test B | 7,012 | 1,488 |
| TOEFL Practice Tests 2nd Edition Test C | 7,300 | 1,425 |
| TOEFL Practice Tests 2nd Edition Test D | 7,409 | 1,521 |
| TOEFL Practice Tests 2nd Edition Test E | 7,304 | 1,450 |
| TOEFL Practice Tests 2nd Edition Test F | 7,316 | 1,475 |
| Barron's How to Prepare for the TOEFL Test 1 | 7,257 | 1,303 |
| Barron's How to Prepare for the TOEFL Test 2 | 7,396 | 1,330 |
| Barron's How to Prepare for the TOEFL Test 3 | 7,353 | 1,270 |
| Barron's How to Prepare for the TOEFL Test 4 | 7,346 | 1,341 |
| Barron's How to Prepare for the TOEFL Test 5 | 7,451 | 1,271 |
| Barron's How to Prepare for the TOEFL Test 6 | 7,392 | 1,294 |
| Barron's How to Prepare for the TOEFL Test 7 | 7,566 | 1,275 |
| Barron's How to Prepare for the TOEFL Test 8 | 7,362 | 1,284 |
| Total | 145,391 | 6,839 |

TOEFL vocabulary list" contained 1,023 words. Among these 1,023 words, 322 words appear in the selected ELT texts and materials. The remaining 701 words are the vocabulary words that the researchers assume students would not necessarily learn from the ELT material, although some words in the list are familiar loan words which can be excluded by teachers who may wish to use this list. Due to the length of this specialized vocabulary list, it is not possible to include it in this publication, but it is available at http://www5d.biglobe.ne.jp/~chujo/.

### 3.7 Measuring the Increase in Coverage of TOEFL Vocabulary

Finally, the efficacy of the specialized TOEFL vocabulary list was examined by calculating the increase of coverage over the TOEFL sample tests. The researchers measured the percentage of increase in the known TOEFL-test vocabulary that students might reasonably be expected to obtain along with the learning of (a) JSH textbooks; (b) JSH and college freshman ELT material; (c) JSH, and college freshman and sophomore ELT materials; and (d) JSH, college freshman and sophomore ELT materials, plus the specialized TOEFL vocabulary list. The JSH and college ELT materials used for this investigation are those shown in Table 2.

### 4.0 Results and Discussion

### 4.1 Graduation of Vocabulary Levels among ELT Materials and Tests

The overall vocabulary levels of the diverse ELT materials and tests were measured by using the BNC HFWL as a scale. Figure 1 shows the vocabulary levels of all the ELT materials and tests investigated in this study (Tables 1-3). The vertical bars on the graph indicate the number of words from the BNC HFWL which are needed to cover 95 percent of each textbook and test. For example, in the Eiken $2^{\text {nd }}$ grade tests, 2,950 words from the BNC HFWL are required in order to comprehend 95 percent of the words used in the tests. The reader may recall that two
samples of each type of material or test were used; these are averaged together in the graph below.


Figure 1 Vocabulary Levels Measured by the BNC High Frequency Word List

Looking at the graph in Figure 1, we can see that the graduation of vocabulary levels among each ELT material and test appears as one might expect. The Eiken $2{ }^{\text {nd }}$ Grade test, which is said to be a desirable target level of English proficiency for high school graduates, ranks the lowest, followed by JSH textbooks, the Daigaku Center Nyushi test, then ELT CALL materials for freshmen, and then the Test of English for International Communication (TOEIC). Vocabulary levels continue to increase gradually from the TOEIC to the ELT texts for sophomores, and rise sharply to the TOEFL test and Eiken Pre $1^{\text {st }}$ Grade test levels, finally reaching the highest level with the Eiken $1^{\text {st }}$ Grade test. The vocabulary level of the Eiken $1^{\text {st }}$ Grade test corresponds to the expectations of Japanese English learners, and reaches the highest level among the tests examined in this study. Chujo and Takefuta (1994) estimated a necessary vocabulary size of about 7,000 to 8,000 words for Japanese English learners to attain their
various communicative goals. Their estimate coincides with the Eiken $1^{\text {st }}$ Grade test vocabulary level shown in Figure 1.

The graduation seen in Figure 1 indicates several interesting results. First, we can see that the TOEFL tests require more vocabulary than the TOEIC tests, which are currently enjoying a surge in popularity among business and industry employees. Second, we can see that the vocabulary levels of the Daigaku Center Nyushi test and JSH texts are very similar. Thus, quantitatively and qualitatively, the vocabulary of the Daigaku Center Nyushi test seems to be a reasonable target for college entrance examinations.

Figure 1 also shows that there is a big gap in terms of vocabulary level between the ELT texts and materials and the TOEFL tests. This shows that an understanding of the 6,150 most frequently occurring words in the BNC HFWL is needed in order to gain a 95 percent coverage for the TOEFL tests. This result indicates that if a student intended to gain a high score on a TOEFL test, he/she would need to make a determined and conscious effort to expand his/her EAP vocabulary during his/her college years.

On the other hand, we can see from Figure 1 that the vocabulary level of the TOEIC test falls between that of the ELT material for freshmen and that for sophomores. Therefore, with regard to vocabulary level, we might assume that college sophomores are quite ready to take TOEIC tests. Furthermore, the difference between the TOEIC and the JSH textbooks is about one thousand words as measured by the BNC HFWL. Such a result would indicate that the TOEIC, in terms of vocabulary level, is within the appropriate range for college students. In other words, college students could reach the TOEIC vocabulary level with a suitable and steady regimen of vocabulary learning.

When TOEIC vocabulary is compared to that of the JSH and ELT materials, it is clear that with the addition of approximately 1,000 EGP words, most Japanese English learners would be
ready to take the TOEIC test. On the other hand, the TOEFL vocabulary is categorized as EAP and is therefore somewhat different from the vocabulary usage contained in the BNC HFWL. As such, a greater knowledge of vocabulary would be required for the TOEFL.

Nation (2001) suggested that, after the 2,000 EGP word level, further study of the Coxhead's Academic Word List (cited in Nation, 2001) improves the coverage of academic articles considerably. This suggests that a specialized vocabulary may be the key to bridging the large gap in vocabulary between the EGP that appears in many ELT materials and the EAP vocabulary of TOEFL tests.

### 4.2 Insufficient Preparation for TOEFL Vocabulary Comprehension

In the next part of this study, the authors grouped the ELT materials used in the investigation above into three categories which students learn as their grades advance, adding each successive grade level to the former category. The first category is JSH texts; the second, JSH texts plus ELT materials for college freshmen; and the third, JSH texts, ELT materials for freshmen, and ELT textbooks for sophomores. The authors then calculated the increase in the text coverage as it rose with the addition of each subsequent category (Figure 2). The three categories of ELT texts and materials that were grouped together at each stage are shown in ascending order on the left-hand side of the bar graph. The bar graph itself shows to what percent each category of grouped educational materials cover the TOEFL test vocabulary. The figures on the graph (i.e., 85.5, 87.1, etc.) represent the coverage figures for the two sample TOEFL tests averaged together.

The graph demonstrates that a knowledge of JSH textbook vocabulary is insufficient for covering the TOEFL tests, and the level remains insufficient even if students master all the vocabulary which appears in ELT textbooks and materials for freshmen and sophomores. Counter to the authors' initial expectations, the ELT textbooks and materials did not adequately
bridge the gap between the ELT materials and the TOEFL tests. In other words, the bridge is short. Researchers (Laufer, 1989, 1992; Nation, 2001) point out that learners would need a 95 percent coverage that equals one unknown word in every 20 words. An 89.2 percent coverage in Figure 2 implies that there is one unknown word in every 9.3 running words. Such a ratio of known to unknown words would lead to a tremendous amount of dictionary work and would mean, in effect, that a student had not reached a level of knowledge that would allow for comprehension of the TOEFL test.


Figure 2 Percentage of TOEFL Test Vocabulary Covered by Categories of Grouped Educational Materials

### 4.3 Efficacy of a Specialized TOEFL Vocabulary List

Finally, the authors discovered that the addition of the specialized TOEFL vocabulary list, created to fulfill the criteria of 'range' and 'frequency' defined by Nation (2001), led to a marked improvement in vocabulary coverage on TOEFL tests (Figure 3). As is shown in Figure 3, adding the specialized vocabulary list to the normal ELT texts and materials improves the coverage of TOEFL test vocabulary from 89.2 percent to 95.0 percent. This figure meets the primary target coverage ratio of 95 percent that researchers point out as
necessary for achievement. Therefore, we can conclude that the gap in vocabulary level existing between ELT materials and TOEFL tests could be further reduced and that the threshold vocabulary level required to comprehend a given text can be reached by supplementing the ELT materials with the specialized TOEFL vocabulary list developed in this study.


## Figure 3 Coverage Maximized by Use of a TOEFL Vocabulary List

### 5.0 Conclusion

This research clarified some of the uniquely specific vocabulary features of the TOEFL. First, it was observed that a knowledge of 6,150 most frequently occurring words in the BNC HFWL, which represents EGP vocabulary, is needed in order to gain a 95 percent coverage of the sample TOEFL tests investigated. Second, this study confirms the existence of a large gap in vocabulary level between the selected ELT materials and the sample TOEFL tests. This means students need to build more vocabulary after they enter college to close the existing gap between the two. Third, the calculation of the extent in percentages to which the ELT materials
cover the TOEFL tests reveals that the textbooks and materials used in college English courses for freshmen and sophomores, which were initially meant to bridge this gap, have only a limited efficacy in doing so. Finally, creating a specialized TOEFL vocabulary list based on the criteria of "range and frequency" and adding this to the normal ELT materials used at colleges would lead to a marked improvement in the vocabulary coverage on TOEFL tests. In general, TOEFL vocabulary is categorized as falling within the domain of EAP and is different from EGP. This research suggests one method for bridging the gap between EGP and EAP. If a Japanese English learner learns the specialized TOEFL vocabulary list produced in this study along with normal college English texts and materials, it may be possible to achieve a 95 percent coverage. Since vocabulary is the foundation for language skills, the achievement of 95 percent is desirable for those students who take the TOEFL test.

It should be noted, however, that further research which includes more sample TOEFL tests would increase the effectiveness of both this study and of the specialized vocabulary list developed in this study. Such an expansion of this research may support the finding that the TOEFL specialized vocabulary list appears to be an effective supplementary tool in bridging the gap between ELT texts and materials and the TOEFL course. Once this is accomplished, the specialized TOEFL vocabulary list can be integrated with a set of vocabulary software already created by the authors (Chujo, et al., 2002). This will provide students with expanded opportunities for vocabulary acquisition and practice as well as prepare them to be more successful in their TOEFL coursework.

Acknowledgements: The authors wish to thank Adam Kilgarriff, whose website and BNC frequency list were indispensable to this study, and Dr. Takefuta and Dr. Takahashi for their valuable and time-saving Word Analysis software program.

## References

Asano, H., et al. (1999). New Horizon English Course 1, 2, 3. Tokyo: Tokyo Shoseki.
Barron's Educational Series, Inc. (2001). Barron's How to Prepare for the TOEFL.
Chujo, K., \& Genung M. (2003). Vocabulary-level assessment for ESP texts used in the field of industrial technology. Proceedings of the KATE 2003 Winter International Conference, Feb.15, 32-36.

Chujo, K., \& Takefuta, Y. (1993). Gakushugoi-no yukodo. Working Papers in Language and Speech Science, Chiba University, No.3, 116-122.

Chujo, K., \& Takefuta, Y. (1994). Gendai-Eigo-no Keyword Plus $\alpha$ 2,000. Chiba-Daigaku Kyoiku Jissen Kenkyu, 1, 253-267.

Chujo, K., Shiina K., \& Nishigaki C. (2002). Development of a CALL system to teach vocabulary for the TOEIC test. AILA 2002 Singapore, Association Internationale de Linguistique Appliquee, Dec. 17.

Educational Testing Service (1998). TOEFL Practice Tests Volume 1.
Educational Testing Service (1999). TOEFL Practice Tests Volume 2.
Finocchiaro, M. (1964). English as a Second Language. Tokyo: Kinseido.
Jordan, R. R. (1997). English for Academic Purposes. Cambridge: Cambridge University Press.
Kiggell, T. (2000). Wonderful USA. Tokyo: MacMillan Languagehouse Ltd.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren and M. Nordman (Eds.), Special language: from humans thinking to thinking machines. Clevedon: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaud and Bejoint (Eds.), Vocabulary and Applied Linguistics. London: Macmillan.

Leech, G., Rayson, P., \& Wilson, A. (2001). Word Frequencies in Written and Spoken English. Harlow: Pearson Education.

Mombukagakusho (2002). http://www.mext.go.jp/b_menu/shingi/chousa/shotou/020/sesaku/020702.htm

Nation, I.S.P. (1994). New Ways in Teaching Vocabulary. Virginia: TESOL, Inc.
Nation, I.S.P. (2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

Suenaga, K. et al. (2001). Powwow English Course I, II, Reading. Tokyo: Buneido.
Takahashi, H. (1999). A development of computer programs for vocabulary frequency count and analysis. Papers on Language and Cultures, Center for Foreign Languages, Chiba University, 5, 57-70.

Takefuta, Y. (1980). Computer-no Mita Gendai Eigo. Tokyo: Educa.
Takefuta, Y. (1986). Eigo-Kyoshi-no Pasokon. Tokyo: Educa.
Takefuta, Y. et al. (2001) Listen to Me! First Listening. Chiba University.
Takefuta, Y. et al. (2002). Listen to Me! Introduction to College Life. Chiba University.
Takesue, Y. \& Miller, K. (1996). American Ideas in Japan. Tokyo: Seibido.


[^0]:    ${ }^{1}$ With more than 100 million words, the BNC is one of the largest corpus resources in the world.

[^1]:    ${ }^{2}$ Initially, the researchers looked at three sets of textbooks: the junior high New Horizon Series, the intermediate level senior high Powwow Series, and an advanced senior high Unicorn Series. Because of insignificant differences in vocabulary levels in the intermediate and advanced level senior high textbooks (3,000 and 3,100 word levels, respectively), the authors felt one senior high school text series was adequate for the purposes of this study. Junior high school texts are available at only one level.

