

Comparing the Three Specialized Vocabularies Used in ‘Business English,’ TOEIC, and British National Corpus Spoken Business Communications

Kiyomi Chujo (Nihon University) Michael Genung (Nihon University)

This paper describes the differences among three business-oriented vocabularies collected from the following three sources: (1) textbooks used in a radio Business English program—“NHK Business Eigo;” (2) TOEIC practice tests; and (3) British National Corpus (BNC) spoken data taken from business meetings, consultations, and interviews. After generating three word lists from each corpus, the following four comparisons were made: (1) computation of the percentage of common words among the three vocabularies; (2) examination of the three specialized vocabularies extracted from each corpus by use of statistical measures such as log-likelihood ratio (LLR) and mutual information (MI); (3) assessment of each vocabulary level; (4) and calculation of the percentage of known words in each vocabulary learned from English textbooks in junior and senior high schools. This study revealed interesting and beneficial features of the three business-oriented vocabularies both qualitatively and quantitatively.

1. Background

In recent years, as English increasingly functions as a lingua franca of international business and communications, there has been a growing interest in and necessity for Business English: i.e., the English language used in the context of business. Especially in Japan, this kind of English is important not only in the business community but also in universities, which, increasingly, find themselves responsible for providing business-related English skills to an ever-expanding population. In Japanese daily life, there are two things instantly associated with Business English which are easily accessible to anyone hoping to improve his or her Business English skills — the radio program “NHK Business Eigo” and the Test of English for International Communication (TOEIC)¹⁾. The former started in 1987 and is broadcast by NHK (Japan Broadcasting Corporation). The secret to its popularity is its easy access through radio to listeners and learners, as well as its reasonably priced textbooks. The latter, the TOEIC test, started in 1979 in Japan and was designed to evaluate communicative skills for both everyday and business situations. Currently, it is enjoying a surge in popularity in both the workplace and universities due to its adoption by many companies and organizations as one of the criteria for selecting new recruits and for promoting employees.

According to “the tree of ELT” (English language teaching — Hutchinson & Waters, 1987), “Business English” is categorized as one of the branches of English used for specific purposes (ESP), all of which are differentiated from General English (GE). ESP is usually divided into three large categories according to the general nature of the learners’ specialization: EST

(English for Science and Technology), EBE (English for Business and Economics), and ESS (English for Social Sciences) — (Hutchinson & Water, 1987). In this article, business-oriented vocabulary (assumed to be classified as EBE) will hereafter be designated as Business English.

In ESP one characteristic of the linguistic knowledge needed to comprehend specialized texts is the heavy load of corresponding specialized vocabulary or “technical words that are recognizably specific to a particular topic, field, or discipline” (Nation 2001:198), for these words convey the import of the subject knowledge. According to Nation (2001:18), “there has been no survey done of the size of technical vocabularies and little research on finding a consistently applied operational definition of what words are technical words.” Looking at dictionaries such as *A Dictionary of English Usage for Business and Finance* (Hashimoto, 1991) or *Longman Business English Dictionary* (2000), one finds that the first contains 5,300 words and the second over 20,000 words and phrases. This is a huge number of vocabulary items. Thus, we can easily see that language teachers need to prepare learners to deal with tremendous amounts of specialized vocabulary.

Naturally, one wonders just how large the two above-mentioned familiar business-oriented vocabularies used in “NHK Business Eigo” and TOEIC tests are; and, furthermore, how similar or dissimilar they are vis-à-vis the one actually used in real business communications. Fortunately, for the purposes of this study, authentic business-related spoken transcripts from the British National Corpus (BNC) are available. These corpus items make it possible to compare both qualitatively and quantitatively the features among the three business-oriented vocabularies that are used in authentic business communications, in educational business texts, and in tests for measuring communication skills.

2. Review of the Literature

In reviewing the studies on Business English vocabulary, two main types of language materials — business letters and business English textbooks — were used to identify this kind of grouping of words. Examples of the first type of materials are Mellinger (1970), Matsumoto (1982), and Tribble (2004), all of which analyzed written business communications. Examples of the second type of materials are Kadota & Kamewari (1976) and Ishikawa et al. (1987), all of which analyzed Business English textbooks. Most of these studies collected texts in order to create word lists, some of them comparing the completed lists to school textbook vocabulary for the purpose of distinguishing Business English

vocabulary from ordinary vocabulary use. As for TOEIC vocabulary, Chujo (2003) researched the type of words used in TOEIC practice tests and selected two sets of specialized vocabulary through a conventional method using 'frequency' and 'range.' This study also used 'percentage of text coverage' as an indicator to determine how the two sets of specialized vocabulary affect the language learning goal, particularly the number of words that need to be known in order to be able to comprehend the test questions effectively. As for the research concerning the vocabulary of educational programs broadcast on the airwaves, Hasegawa (1971) examined the spoken vocabulary used in twelve monthly TV English program texts.

As a unique way to assess the vocabulary levels of various texts, Chujo & Genung (2003) used a lemmatized "British National Corpus High Frequency Word List" (BNC HFWL) of 14,004 words as a criterion. In establishing this criterion, they noted that researchers such as Laufer (1989, 1992) and Nation (2001) point to the fact that learners would need a coverage level of 95 percent to understand the meaning of texts. Accordingly, they chose this 95 percent coverage level as their target. Additionally, each vocabulary level of a targeted text was defined in terms of the number of words from the BNC HFWL that equaled 95 percent coverage of that text; that is, the authors counted how many words from the top of the BNC HFWL were needed to achieve 95 percent coverage of the targeted text.

As suggested by Nation (2001:18), "one way of making a technical vocabulary is to compare the frequency of words in a specialized text with their frequency in a general corpus." Putting such a suggestion into practice, Chujo & Utiyama (2004) and Utiyama et al. (2004) proposed using multiple statistical measures for comparing the above-mentioned two kinds of frequencies and for extracting various levels of specialized lists. Such measures include 'log-likelihood ratio,' or LLR, (Dunning, 1993) and 'mutual information,' or MI, (Church & Hanks, 1989) scores. They suggested that LLR, for example, identifies appropriate level words for intermediate-level or sub-technical words, and MI for upper-intermediate-level or technical words. Currently, LLR and MI are the two most commonly used popular statistical measures in the field of corpus linguistics.

Summing up the results of these studies, we can draw the following conclusions. Firstly, few studies were conducted on the vocabulary of spoken business communications and "Business Eigo" texts. Secondly, there are two useful means to assess the vocabulary level of a targeted text. Method number one is to calculate the percentage of coverage of vocabulary learned through targeted texts. Method number two is to assess the vocabulary level of each targeted text by calculating the number of words from the BNC HFWL that equals 95 percent coverage of that text. Finally, in order to acquire each specialized vocabulary, we can utilize

statistical measures such as LLR and MI in addition to the conventional criteria, such as frequency and range.

3. Purpose of the Study

The purpose of this study is to describe both quantitatively and qualitatively the outstanding features of three sets of business-oriented vocabulary collected from the following three sources: (1) monthly textbooks used in a radio Business English program—“NHK Business Eigo;” (2) TOEIC practice tests; and (3) British National Corpus (BNC) spoken data taken from business meetings, consultations, and interviews.

4. Procedure

First, fifty-eight sets of BNC dialogues actually used in a business context, six monthly textbooks used in “NHK Business Eigo” (Sugita, 2001), and sixteen TOEIC practice tests were collected. Second, a word list was created for each and then used to calculate the extent and overlap of three business-oriented vocabularies. Third, using statistical measures, three specialized ESP word lists were created to observe the features of each business-oriented vocabulary. Fourth, each vocabulary level was assessed. Finally, the rate of vocabulary coverage provided by the school textbooks over the three business-oriented vocabularies was identified. Each step is outlined below.

4.1 Three Business-Oriented Language Materials

The authors first collected fifty-eight sets of spoken communications (mainly dialogues) set in a business context — these included transcriptions of 40 business meetings, 9 consultations, 7 interviews, and 2 presentations selected from 136 spoken components of the British National Corpus (hereafter, referred to as BNC dialogues). Second, they gathered six monthly textbooks used in the radio program “NHK Business Eigo” from April to September, 2001 (hereafter, referred to as Business Eigo). Finally, they acquired sixteen sets of TOEIC practice tests. Six were retired tests available to the public (hereafter, referred to as TOEIC tests), including one TOEIC Bridge practice test (hereafter, referred to as TOEIC Bridge), a test recommended for examinees who scored less than 450 point in TOEIC. The remaining ten were practice tests published by various authors and publishers²⁾. **Table 1** shows the numerical data relating to the three sets of business-oriented vocabularies, as well as the

tokens (total number of words) and the types (number of different words) which appear in each of the selected groups of texts.

Table 1 The Three Business-Oriented Vocabularies Analyzed

| Source | Number of Texts | Tokens | Types |
|---|-----------------|---------|-------|
| Business Dialogues from BNC (BNC dialogues) | 58 | 474,613 | 6,878 |
| NHK Business Eigo texts (Business Eigo) | 6 | 30,458 | 3,042 |
| TOEIC Practice Tests (TOEIC tests) | 16 | 107,081 | 5,016 |

The authors also collected the top selling series of junior and senior high school (hereafter, referred to as JSH) textbooks from which college students or college graduates studied English before entering the university. These are shown in **Table 2**, along with number of tokens and types. The *Unicorn I, II* and *Reading* series is one of the most advanced-level series of textbooks available to senior high schools in Japan.

Table 2 Junior and Senior High School Textbook Vocabulary

| Textbooks | Tokens | Types |
|---|--------|-------|
| <i>New Horizon 1, 2, 3</i> | 9,440 | 1,124 |
| <i>Unicorn I, II</i> and <i>Reading</i> | 36,678 | 3,478 |
| Total (JSH textbook vocabulary) | 46,118 | 3,747 |

All of the collected data were lemmatized; i.e., for each item selected all inflected word forms having the same stem were listed under a base form and alphabetized with frequency of occurrence information. Proper nouns and numerals were manually excluded from each 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: 19-20).

4.2 Calculating the Extent and Overlap of Business-Oriented Vocabularies

The authors then totaled the three vocabularies in order first to determine the number of words that appeared at least one time in any of the three; and, second, to calculate the number of words occurring in common so as to identify those items appearing in all three business-oriented vocabularies.

4.3 Extracting Specialized Vocabularies from the Three Business-Oriented Vocabularies

In order to extract specialized vocabularies and to observe the characteristics of each one, two statistical measures, LLR and MI, were applied to identify the specialized words in each of the three corpora, using the methodology described in Chujo & Utiyama (2004) and Utiyama et al. (2004). These measures identify words whose frequency is significantly higher in a small text of interest, i.e., the BNC dialogues, Business Eigo, and the TOEIC tests, rather than in a large reference corpus, i.e., the BNC HFWL mentioned earlier.

4.4 Assessing Vocabulary Level within the Selected Texts

The authors next assessed the vocabulary level of each text shown in **Table 3** by comparing it with the BNC HFWL, which was used as a criterion for this research. Each targeted text vocabulary level was defined in the following terms: namely, by identifying and quantifying the number of words from the BNC HFWL that equaled 95 percent coverage of that text. Thus, the BNC HFWL was used to calibrate the graduations among the diverse vocabulary levels contained within the BNC dialogues, Business Eigo, the TOEIC Bridge test, and the TOEIC tests. Also, the vocabulary levels of each group of texts were averaged.

Table 3 Texts Measured by Vocabulary Level and Coverage

| Source | Number of Texts | Average Types | Average Tokens |
|---------------|-----------------|---------------|----------------|
| BNC Dialogues | 58 | 833 | 8,183 |
| Business Eigo | 6 | 1,015 | 5,078 |
| TOEIC Bridge | 1 | 637 | 2,358 |
| TOEIC Tests | 5 | 1,366 | 6,836 |

4.5 Measuring JSH Textbook Coverage over the Selected Texts

Next, the authors calculated the extent to which the vocabulary in JSH texts covers the vocabulary used in each of the three business-oriented texts shown in **Table 3** above. This constitutes an alternative way of obtaining an accurate estimate of the vocabulary level of each text; and, furthermore, indicates the amount of vocabulary increase the learners need to acquire in order to attain each of the three business-oriented communication goals. The percentages of coverage indices of each group of texts were averaged.

In addition to this, the authors combined the Business Eigo texts' vocabulary with the JSH texts' vocabulary, and then calculated the increase in the rate of coverage over BNC business dialogues and TOEIC practice tests. This was done in order to obtain an estimate of the percentage of increase in the foundational vocabulary acquired by learners in JSH after thoroughly assessing the impact of the additional items absorbed from the radio program.

5. Results and Discussion

5.1 The Extent and Overlap of Business-Oriented Vocabularies

The total number of words among the three sets of vocabulary was 9,108, while their overlap consisted of 1,798 words. Out of these 1,798 words, 535 words do not appear in JSH textbook vocabulary. Assuredly, then, these 535 words may be considered as being the most important words needing to be acquired for an adequate comprehension of Business English, considered from the standpoint of authentic business communications, educational business texts, and tests for measuring communication skills. Excluding JSH vocabulary, the most frequently appearing 100 business-oriented words common to the three business-oriented vocabularies are attached in the **Appendix**.

5.2 Specialized Vocabularies of the Three Business-Oriented Vocabularies

The above observation reveals that the words used in common among the three business-oriented vocabularies are very few in number; it also suggests that each vocabulary is different from the other two. Considering these facts, the authors proceeded to the next step.

In order to identify the specialized vocabulary in each of the three business-oriented vocabularies, two statistical measures, LLR and MI, were applied to each corpus. Due to space limitations, **Table 4** shows only the excerpts of the top-20-word extractions by MI. Interesting contrasts exist among these three lists. MI was able to identify the different types

of ‘technical business words’ characteristically used in each kind of business communication.

BNC dialogues contain such technical words as *pallet*, *costing*, *accrue*, and *turnaround*, which are used in the business activities of trade, distribution, and finance. They also contain such words as *photocopy*, *erase*, *byte*, *spreadsheet*, *worksheet*, and *folder*, which are used in office work, particularly computing. Finally they contain words related with personnel, such as *divisional*, *overtime*, *seconder*, and *empowerment*.

Business Eigo vocabulary shows the influence of the various program topics chosen to attract the listeners’ attention; examples are *healthcare* and *gourme*. In addition, words such as *acupuncture*, *yoga* and *chemotherapy* were used in the topic of alternative medicine. One of the program’s favorite topics seems to be IT, for such words as *online*, *high-tech*, *high-speed*, *surf*, and *gadget* are among the top 20 characteristic words. Of course, there are also personnel related words such as *perk*, *severance*, and *teamwork*.

TOEIC tests also show broad topic coverage connected to business or daily communications with such examples as *supervisor*, *receptionist*, *payroll*, *relocate*, and *typist* being related to personnel; *photocopy*, *bookcase*, *memo*, and *fax* to office work; *refund* and *reimbursement* to money; and *pharmacy*, *renovation* and *dishwasher* to daily life.

Table 4 Excerpts of the Top 20 Characteristic Word Comparisons in the Three Business-Oriented Vocabularies Extracted by Mutual Information (MI).

| Rank | BNC Dialogues | Business Eigo | TOEIC Tests |
|------|---------------|---------------|---------------|
| 1 | photocopy | healthcare | refund |
| 2 | pallet | online | merchandise |
| 3 | erase | acupuncture | consulate |
| 4 | byte | high-tech | photocopy |
| 5 | spreadsheet | gourmet | supervisor |
| 6 | bearing | clout | bookcase |
| 7 | advertiser | fickle | pharmacy |
| 8 | worksheet | yoga | reimbursement |
| 9 | conductivity | well-informed | memo |
| 10 | raffle | severance | instructional |
| 11 | costing | teamwork | receptionist |
| 12 | divisional | high-speed | dishwasher |
| 13 | folder | gadget | payroll |
| 14 | accrue | burner | incorporated |
| 15 | assignment | surf | fax |
| 16 | turnaround | carnation | relocate |
| 17 | chute | enroll | typist |
| 18 | overtime | perk | banquet |
| 19 | seconder | overboard | renovation |
| 20 | empowerment | chemotherapy | enroll |

5.3 Graduation of Vocabulary Level among the Business-Oriented Texts and Tests

In the following step, the differences among the three vocabularies were observed from the standpoint of vocabulary level. First, the authors measured the overall vocabulary level of the targeted business-oriented vocabularies by using the BNC HFWL as a scale. The results are shown in **Figure 1**. The vertical bars on the graph indicate the number of words from the BNC HFWL which are needed to cover 95 percent of each text collected in this study. For example, considering the five TOEIC practice tests, 3,714 words on average are required from the BNC HFWL in order to comprehend 95 percent of the words used in each test. The levels of fifty-eight BNC business dialogues, five TOEIC tests, and six Business Eigo texts were averaged together for exhibition on the graph.

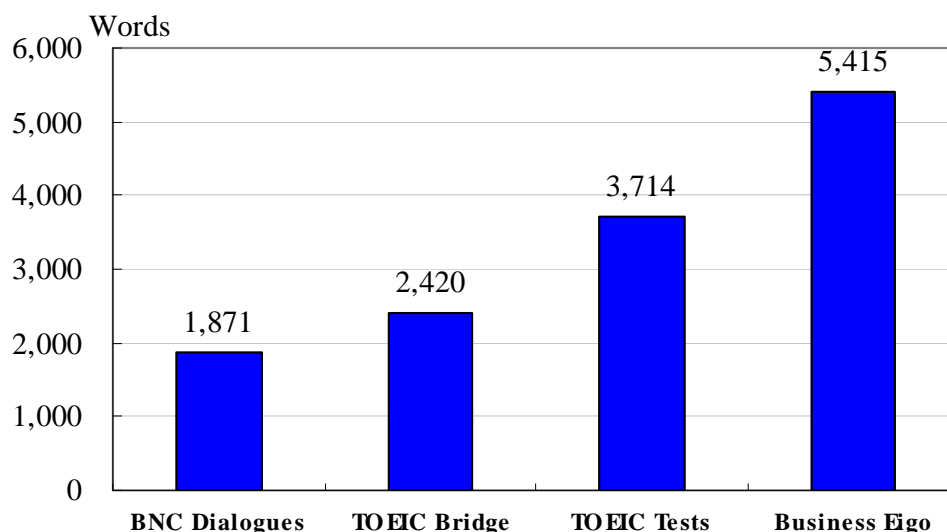


Figure 1 Vocabulary Levels of Business-Oriented Vocabularies Measured by the BNC High Frequency Word List

After looking at the graph in **Figure 1**, we can see that the graduation of vocabulary levels among each type of vocabulary appears as one might expect. Authentic conversation is at the lowest level, increasing gradually to the level of the TOEIC Bridge test (which is a kind of “baby” TOEIC), then increasing again to the level of the normal TOEIC, and finally rising sharply to the level of radio’s Business Eigo.

The graph reveals that the vocabulary level of BNC dialogues is the lowest among all the vocabularies. One possible way to account for this result is to suppose that the business dialogues do not differ greatly from the general vocabulary usage represented by BNC HFWL,

which represent the most frequently occurring items among 100 million words³⁾. We can easily imagine that, inevitably, the spoken language used in the conduct of actual business communications is efficiently selected by referring to a minimum number of items within a shared lexical core. Through these observations, along with the ones offered by **Table 4**, we can confirm that the major concern of real business communications is to attain very strictly the business communication goals relating to product presentation, information exchange, offer proposal and acceptance, and contract negotiation, as well as those concerned with the discussion of such general business topics as management skills, promotional strategies, training courses, business services and the like.

The graph in **Figure 1** also shows that knowledge of the 5,415 most frequently occurring words in the BNC is needed in order to gain 95 percent coverage of the Business Eigo texts investigated — this is the highest level of knowledge required among any of the lists of frequently occurring words appearing on the chart. Furthermore, the vocabulary of Business Eigo seems to be a little different from the normal vocabulary usage represented by the BNC. From observation of **Table 4**, we see that besides such typical business topics as job interviews, etc., the program contains diverse social and economic ones too, such as digital divide and teen marketing. It also features topics like ‘personal weight control’ and ‘customized travel itineraries’ which, while not typically considered as being business-related activities, have become, in the present era, issues of concern for modern business people. In terms of everyday language usage then, knowledge of a larger amount of words is needed to gain 95 percent coverage of the Business Eigo texts. This is what accounts for the vocabulary level of the Business Eigo being measured so high.

5.4 Insufficient Preparation for Business-Oriented Texts and Tests

In the next part of this study, the authors calculated the extent to which the vocabulary in JSH texts cover the vocabulary used in each of the three business-oriented texts and displayed the results in **Table 5**. This is another way to obtain a good estimate of the vocabulary level of each text and also shows the amount of vocabulary increase the learners need in order to attain each of the three business-oriented communication goals. The figures in the second row in **Table 5** represent the percentages of coverage indices of JSH texts over each business-oriented text averaged together. Additionally, the authors combined the Business Eigo texts’ vocabulary with the JSH texts’ vocabulary, and then calculated the increase in the rate of coverage over BNC business dialogues and TOEIC practice tests. This was done in

order to obtain an estimate of the percentage of increase in the foundational vocabulary learned at junior and senior high schools through addition of the items acquired from the radio program by the learners. The results are shown in the third row of **Table 5**.

Table 5 Coverage Increased by Combinations of JSH Texts and Business Eigo

| | BNC Dialogues | TOEIC Bridge | TOEIC Test | Business Eigo |
|---------------------|---------------|--------------|------------|---------------|
| JSH Texts | 93.1 % | 94.7% | 88.7% | 89.0% |
| JSH + Business Eigo | 96.1% | 97.6% | 94.0% | - |

Table 5 demonstrates that knowledge of JSH textbook vocabulary is insufficient for covering all of the business-oriented texts and tests investigated in this study. Since the JSH texts are classified as General English and supposedly represent normal vocabulary usage as distinct from the ESP/EBE vocabularies represented by these three business-oriented vocabularies, we can conclude that to one degree or another the technical vocabulary occurring in the area of business would be a necessary supplement for any learners wanting to expand their foundational JSH vocabulary for the specific intent of acquiring an ability to communicate at a Business-English level. Thus, Business Eigo was added to the JSH texts in order to simulate the usual case in which learners complete JSH texts and then use radio programs to improve their English skills.

Coincident with the authors' initial expectations, Business Eigo adequately bridges the gap between the level of JSH texts and the levels of the BNC dialogues and the TOEIC Bridge. However, the TOEIC test level remains unattainable even if students master all the vocabulary appearing in both JSH and Business Eigo textbooks. Business Eigo did not adequately bridge the gap between the JSH texts and the TOEIC tests. In other words, the bridge is too short. As mentioned earlier, researchers point out that learners would need a 95 percent coverage, which equals one unknown word in every 20 words. The 94.0 percent coverage in **Table 5**, however, implies that there is one unknown word in every 9.5 running words. Such a ratio of known to unknown words would mean, in effect, that a learner had not reached a level of knowledge that would allow for comprehension of the TOEIC test. Chujo & Genung (2003) found that creating a specialized vocabulary list and adding this to the normal ELT materials used at colleges would lead to a marked improvement in the vocabulary coverage of the targeted ESP text. Such facts suggest that a specialized vocabulary list may be the key to

bridging the large gap in vocabulary between ELT texts and ESP-related materials, which, in turn, might lead to a more optimum performance on the TOEIC test. For example, the authors have confirmed that by adding the top 300 items characteristic to the TOEIC test specialized vocabulary extracted by MI (shown in **Table 4**) to the vocabularies of the JSH and Business Eigo texts, the coverage of TOEIC test vocabulary is improved from 94.0 percent to 95.2 percent. This figure meets the primary target coverage ratio of 95 percent that researchers point to as being necessary for achievement.

6. Conclusions

This study clarified some of the uniquely specific vocabulary features of the three business-related vocabularies. First, we were able to obtain the 535 most important business-oriented words common to the three vocabularies. Second, through the observation of words extracted by use of statistical measures, we were able to identify the three different types of technical business words. Third, we observed that a knowledge of the 1,871 most frequently occurring words in the BNC HFWL (representing the General English vocabulary) is needed in order to gain a 95 percent coverage of each of the authentic business dialogues, while a knowledge of 3,714 words is needed for each TOEIC practice test and a knowledge of 5,415 words for each monthly Business Eigo text. In addition, the results of this vocabulary level assessment were thoroughly confirmed by measuring JSH text coverage over each text within the three vocabularies. These calculations reveal that insufficient knowledge of JSH textbook vocabulary can be supplemented by combining it with the Business Eigo vocabulary, and that their combination is efficacious in covering the BNC dialogues and the TOEIC Bridge practice test. This research also suggests that if a Japanese English learner acquires a specialized TOEIC vocabulary along with the above-mentioned JSH-Business Eigo combination, it may be possible to achieve 95 percent coverage.

As an extension of this study of spoken business-oriented vocabularies, the authors are exploring the items within a written business-oriented vocabulary in a 'commerce and finance' corpus, which is a component of the BNC. It is hoped that the observation of both spoken and written business-oriented vocabularies using the same approach suggested in this paper will further contribute to preparing a steady and effective method for meeting the increased need for pedagogical materials, as well as address the demands of a rapidly growing global business community.

References

- Asano, H., et al. (1999) *New Horizon English Course 1, 2, 3*. Tokyo: Tokyo Shoseki.
- British National Corpus World Edition 2000.
- Chauncey Group International (2000) *TOEIC Koushiki Guide & Mondaishuu*. Tokyo: IIBC.
- (2002) *TOEIC Koushiki Guide & Mondaishuu Vol. 2*. Tokyo: IIBC.
- (2003) *TOEIC Bridge Koushiki Guide & Mondaishuu*. Tokyo: IIBC.
- Chujo, K. (2003) 'Eigo Shokyuusha-muke TOEIC Goi 1 & 2 no Sentei to Sono Kouka [Selecting "TOEIC Vocabulary 1& 2" for Beginning-Level Students and Measuring Its Effect on a Sample TOEIC Test]. *J. of the College of Industrial Technology Nihon University*, 36, 27-42.
- Chujo, K., and M. Genung (2003) 'Vocabulary-Level Assessment for ESP Texts Used in the Field of Industrial Technology,' *English Teaching*, 58, 3, 259-274.
- Chujo, K. and Utiyama, M. (2004) Toukeiteki Shihyou wo Shiyoushita Tokuchougo Chuushutsu ni Kansuru Kenkyuu [Using Statistical Measures to Extract Specialized Vocabulary from a Corpus]. *KATE Bulletin*, 18, 99-108.
- Church, K.W. and Hanks, P. (1989) 'Word Association Norms, Mutual Information, and Lexicography,' *Proceedings of ACL-89*, 76-83.
- Dunning, T. E. (1993) 'Accurate Methods for the Statistics of Surprise and Coincidence,' *Computational Linguistics*, 19, 1, 61-74.
- Hasegawa, K. (1971) 'Speaking Vocabulary to Kyoukasho no Goi,' *Eigokyouiku*, 20, 3, 38-41.
- Hashimoto, M. (2000) *Bijinesu Eigo – Kenkyuu to Kyouiku* [Business English – Studies and Education]. Tokyo: Elco.
- Hashimoto, M. (1991) *Keizai-Eigo Eiwa-Katsuyou Jiten* [A Dictionary of English Usage for Business and Finance]. Tokyo: Nihon Keizai Shimbunsha.
- Hutchinson, T. and Waters, A. (1987) *English for Specific Purposes*. Cambridge: CUP.
- Ishikawa, Y., Tanaka, K., Takahashi, H. & Takefuta, Y. (1987). Bijinesu-eigo no Goi [The Vocabulary of English in Business Communication]. *Gogakukyouiku-Kenkyuusho Kiyou*, 1, 53-66.
- Kadota, A. and Kamewari, T. (1976) 'Business English Kyouiku Kihon 500-go' *Bulletin of Kagoshima Prefecture Junior College*, 27, 113-134.
- 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. 316-323.
- (1992) 'How Much Lexis Is Necessary for Reading Comprehension?' in L. Arnaud

- and H. Bejoint (eds.), *Vocabulary and Applied Linguistics*. London: Macmillan. 126–132.
- Matsumoto, S. (1982) ‘Shougyou Tsuushinnbun ni Shiyousareta Goi no Computer niyoru Bunseki,’ *Kitakyuushuu-daigaku Gaikokugakubu Kiyou*, 48, 43-101.
- Mellinger, M. (1970) *Basic Vocabulary for Written Business Office Communications*. Chicago: Chicago State College.
- Nation, P. (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Pearson Education Ltd (2000) *Longman Business English Dictionary*, Harlow: Pearson Education Limited.
- Suenaga, K. et al. (2002) *Unicorn English Course I, II, Reading*. Tokyo: Buneido.
- Sugita, S. (2001) *Yasashii Business Eigo*, NHK, 15, 1-6.
- T. F. Communications. (1997). ‘Practice TOEIC,’ *TOEIC Friends*, 3(4), 24-53.
- TOEIC Un’ei linkai. (1981). *Dai-Ikkai TOEIC Mondaishu*. Tokyo: Eibun Asahi.
- TOEIC Un’ei linkai. (1982). *Dai-Sankai TOEIC Mondaishu*. Tokyo: Eibun Asahi.
- Tribble, C. (2004) ‘Business and Professional Communication: Managing Relationships in Professional Writing,’ *Proceedings of the Sixth Teaching and Language Corpora Conference*, Granada. 48.
- Utiyama, M., Chujo, K., Yamamoto, E. and Isahara, H. (2004). ‘Eigokyouiku no Tameno Bunya Tokuchou Tango no Sentei Shakudo no Hikaku,’ [A Comparison of Measures for Extracting Domain-Specific Lexicons for English Education]. *J. of Natural Language Processing*, No.484, 165-197.

¹⁾ TOEIC is a registered trademark of Educational Testing Service (ETS).

²⁾ Six TOEIC retired tests were opened to the public by TOEIC: Un’ei linkai, 1981, 1982; T. F. Communications, 1997; The Chauncy Group International, 2000, 2002. The rest were collected from T. F. Communications, 2003, and from other publishers such as Nagase, 2000 and ALC, 2002.

³⁾ The BNC HFWL represents the entire 100,106,029 words included in the BNC. In order to calculate the vocabulary level for each BNC dialogue, the BNC HFWL vocabulary was compared to the dialogue. Of course, since each of the BNC dialogues is a component of the entire BNC (BNC HFWL), each has obviously contributed data to the entire BNC HFWL. Nevertheless, since on average the size of each BNC dialogue text is a mere 8,183-words (when compared to the 100,106,029 words of the BNC HFWL), the data

contained within a single text could not possibly have a significant affect on the overall content of the BNC HFWL. Very simply, the size difference between the two is so huge as to render any effect insignificant — for example, a single dialogue covers only 0.008% of the entire BNC. Thus, the authors consider the BNC dialogue texts not to affect the result of the comparisons between the BNC HFWL and each of the BNC dialogues.

Appendix Most Frequently Appeared 100 Words among Three Business-Oriented Vocabularies

| | | | | |
|-------------|-------------|-------------|--------------|-------------|
| access | commission | employment | management | research |
| account | complaint | estimate | objective | reservation |
| advantage | conference | executive | operation | review |
| agency | consultant | exercise | option | row |
| agent | contract | expand | organization | scheme |
| airline | copy | expense | overall | secretary |
| appointment | corporate | facility | package | sector |
| appropriate | corporation | fairly | passenger | seminar |
| arrange | credit | fax | payment | session |
| assignment | current | fee | personnel | sponsor |
| available | customer | file | print | staff |
| bill | data | financial | procedure | statement |
| bonus | delay | fund | profit | stuff |
| booklet | delivery | golf | propose | survey |
| budget | direct | identify | purchase | target |
| cancel | division | impact | reference | tax |
| cash | doubt | improvement | region | term |
| claim | earn | insurance | register | trend |
| client | employ | investment | rent | update |
| coffee | employee | location | request | worth |

Acknowledgements: This research is partly supported by the TOEIC Research Proposal grant. We thank Dr. Masao Utiyama, National Institute of Information and Communications Technology, for his accurate and timesaving computer programs. We are also grateful to Atsushi Yamazaki, an MA candidate with the Graduate School of Nihon University, for his data analysis assistance.