Vocabulary Size Required for the TOEFL iBT Listening Section

Masaya Kaneko
Tokyo Denki University in Saitama, Japan

The aim of the present study is to estimate the vocabulary size target for the Listening Section of the TOEFL Internet-based Test (iBT). The lexical frequency levels of the listening passages from 5 real past tests were determined with the use of Nation's (2006) word-family lists. It was found that the first 3,000 word families plus proper nouns and marginal words yielded 95% coverage of the texts, and that it took the most frequent 6,000 word families to reach 98% coverage. Comparing the results of the present study with those reported in Kaneko's (2014) study, the Listening Section of the TOEFL iBT appears to require only half as large vocabulary size as the Reading Section. Pedagogical and research implications are discussed in detail.

本論の目的は、TOEFL　iBTリスニングセクションの目標語彙サイズを測定することである。5つのTOEFL　iBT過去試験問題を対象に、Nation (2006)のワードファミリーリストを用いて語彙頻度レベルを決定した。固有名詞など理解に影響を与えない語を含めた場合、最も頻度の高い3千語で文中の単語の95%、6千語で98%をカバーした。また、Kaneko (2014)の研究結果と比較すると、TOEFL iBTリスニング問題に求められる語彙サイズは、リーディング問題に求められる語彙サイズの半分で十分であることが示唆される。教育上および研究上の含意についても詳細に論じる。

A recent proposal made by the Japanese government has pushed some Japanese learners of English to consider taking the Test of English as a Foreign Language (TOEFL). The proposal was made by the education reform panel of the ruling Liberal Democratic Party of Japan (Yoshida, 2013). According to the proposal, all public and national universities in Japan should use the TOEFL test instead of the National Center Test, which has been used by all national and public universities in Japan since 1990, to help determine admission. Subsequently, it was decided that additional points are given to applicants for national public servants commencing from fiscal 2015 depending on their TOEFL test scores. Concerning the replacement of the National Center Test with the TOEFL, the decision remains pending. However, in response to the government's push for the TOEFL, several local governments have urged students to take the TOEFL test. For instance, the local government of Yokohama in Kanagawa Prefecture has decided to make all grade 11 students at eight designated public high schools in the city take the TOEFL test, with the test fee being paid by the city (Okada, 2014). In addition, Osaka city announced on November 11, 2013, that it is going to recruit special TOEFL lecturers called Super English Teachers to help high school students in the city prepare for the test.

However, some cast doubt on the use of the TOEFL test as a measurement of English proficiency levels of Japanese learners of English. In a newspaper article appearing in the Asahi Shimbun (Tonedachi, 2013), Erikawa argues that a vocabulary beyond the 10,000-word frequency level frequently appears in the TOEFL test. Citing Ishida's (2004) work, Erikawa claims that even the most frequent 10,000 lemmas on the Standard Vocabulary List (SVL), which was compiled by a Japanese publisher named ALC (n.d.), could account for only around 86% of the running words in several TOEFL practice tests. Sato (2013) also maintains that a vocabulary ranging from the 15,000 to 30,000-word frequency level is necessary for the TOEFL although he does not specify the source of the lexical size requirement.

A vocabulary beyond the 10,000-word frequency level is far beyond the minimum vocabulary size requirement for Japanese senior high school students. According to the current course of study guidelines made by the Ministry of Education, Culture, Sports, Science, and Technology (2009), high school students are supposed to acquire 3,000 word families through six-year English education at junior and senior high schools although which 3,000 words should be learned and the rationale for the vocabulary size target are not specified. Considering the claims made by Erikawa and Sato, the TOEFL test seems to be too demanding for Japanese high school students in terms of the lexical size requirement as they suggest.

However, a more recent study suggests that a vocabulary beyond the 10,000-word frequency level may not be necessary for the TOEFL test. Chujo and Oghigian (2009) estimated the vocabulary level

of two versions of the TOEFL test: TOEFL Paper-based Test (PBT) and Internet-based Test (iBT). The vocabulary appearing in an official TOEFL iBT practice test and six PBT practice tests was examined with the use of the lemma-based SVL (ALC, n.d.) and Nation's (2006) word-family lists. It was found that the most frequent 6,242 lemmas or 5,000 word families were needed to account for 95% of the running words in the six practice tests for the TOEFL PBT. They also found that the top 4,719 lemmas or 4,000 word families provided 95% coverage for the iBT practice test. These findings suggest that a vocabulary beyond the 10,000 frequency level is not necessary for the TOEFL test.

However, it should be noted that there is a methodological issue with Ishida's (2004) and Chujo and Oghigian's (2009) studies. They combined texts from different sections of the practice TOEFL tests into one corpus and calculated coverage figures for the mixed text. However, coverage figures derived from combined written and spoken texts can be misleading as research has revealed that spoken text employs greater use of high-frequency vocabulary than written text (e.g., McCarthy & Carter, 1997; Nation, 2006; Schonell et al., 1956). In order to estimate an accurate vocabulary size target for the TOEFL test, written and spoken texts should be separated. Also, in past text-coverage studies on the TOEFL test, *practice* tests were examined. However, analyses of authentic past tests should yield more accurate estimates. Since research has not explored the issue of whether the vocabulary level of official practice tests for the TOEFL iBT is similar to that of authentic tests, vocabulary size estimates derived from real past tests should be more reasonable. Addressing the methodological issue and the limitation, Kaneko (2014) examined the vocabulary appearing in reading passages from five past TOEFL iBTs (Educational Testing Service, 2013) using Nation's (2006) word-family lists. It was found that the most frequent 6,000 word families provided 95% coverage for the reading passages. He suggests that unless 98% coverage or more is desired, a vocabulary beyond the 10,000-word frequency level is not necessary for the reading section of the TOEFL iBT. Concerning the vocabulary size target for the listening section, however, there are no published studies available which address the two aforementioned issues.

Considering all the issues mentioned above, we have not yet reached the answer to the question of whether a vocabulary beyond the 10,000-word frequency level is required for the TOEFL test. Kaneko's (2014) findings suggest that a vocabulary below the 6,000-word frequency level should be sufficient for the listening section of the TOEFL test but this has not been fully explored.

There are three research questions in the present study: (1) to estimate the vocabulary size target for the listening section of the TOEFL iBT, (2) to answer the question of whether a vocabulary beyond the 10,000-word frequency level is necessary for TOEFL, and (3) to assess whether the vocabulary required for the listening section of the TOEFL iBT is beyond the vocabulary level of high school students.

Methodology

Materials

In the present study, spoken words appearing in listening passages from five past TOEFL iBTs (Educational Testing Service, 2013) were examined. Each listening test contains six passages; thus a total of 30 passages were analyzed with the RANGE program (Heatley, Nation, & Coxhead, 2002). Before analyzing the passages with RANGE, several modifications were made. First, contractions involving *'d* (represents either *had* or *would*) and *'s* (*has* or *is*) were split into their separate components (e.g., *I'd = I + would* or *I + had*) because RANGE cannot recognize such differences. Second, connected speech such as *kinda*, *y'know*, and *dunno* was also modified to the original separate word items. Without making these modifications, RANGE may categorize some words as words beyond a vocabulary of the 14,000-word frequency level. For instance, if *dunno* remains intact in the text, RANGE would regard it as a word outside the most frequent 14,000 words although all the separate components of *dunno* (i.e., do + not + know) are in fact the 1,000 word level. After making these modifications, the spoken words appearing in the listening passages were examined with RANGE. There were 20,953 tokens in total, and the average number of tokens for each test was 4,190.6. The average length per passage was 698.4 words.

Nation's Word-Family Lists

In order to ensure a meaningful comparison between findings of the present study and those reported in past text-coverage studies on the TOEFL test, the word lists used in the present study first need to be clarified. The lexical frequency level of the listening passages was measured with the use of Nation's (2006) word-family lists derived from the British National Corpus (BNC). The main rationale for the adoption of Nation's lists was that it enables findings of the present study to be comparable to those of past text-coverage studies on the TOEFL test. Except for Hirai's (2000) study,

in which West's General Service List (1953) and the University Word List (Xue & Nation, 1984) were used, word lists derived from the BNC were used in past-published text-coverage studies on the TOEFL test. Ishida (2004) and Chujo and Oghigian (2009) used the Standard Vocabulary List (SVL), which was compiled using the BNC. Mizumoto (2006) used the JACET 8,000 word list (Ishikawa et al., 2003), which was also derived from the BNC. In Kaneko's (2014) and Chujo and Oghigian's (2009) studies, Nation's (2006) word-family lists were used to measure the vocabulary level of the TOEFL iBT. It should be noted here that word lists compiled from American-English corpora such as the Corpus of Contemporary American English (COCA) (Davies, 2008) should yield better coverage counts since the TOEFL is based on American English. However, the use of American-English-based word lists would not ensure a meaningful comparison between findings of the present study and those derived from past text-coverage studies on the TOEFL. Therefore, Nation's BNC-based word-family lists were used.

Text Coverage and Adequate Comprehension

Research has demonstrated that text coverage, or the amount of known words in a text, affects listening comprehension in L1 (Van Zeeland & Schmitt, 2013) and in L2 (Bonk, 2000; Stahr, 2009; Van Zeeland & Schmitt, 2013). Van Zeeland and Schmitt found that the higher text coverage became, the better comprehension the participants achieved. In other words, a target coverage figure largely depends on a desired comprehension level.

The present study assumed 95% coverage as the target and that it allows adequate comprehension to occur in the listening section, which is in line with Chujo and Oghigian (2009). Research on L2 listening comprehension conducted by Van Zeeland and Schmitt (2013) showed that 95% coverage provided an average comprehension score of 76.5%. This comprehension level is considered to be advanced for the TOEFL iBT. According to the score scale for the listening section of the TOEFL iBT, 22 points or higher out of the maximum possible score of 30 points (i.e., 73% or more) is classified as *high*: the top out of the three score ranges. Achievement of the most advanced level should be considered adequate for the listening section of the TOEFL iBT.

Results

Table 1 illustrates cumulative coverage figures on each TOEFL iBT listening test using Nation's word-family lists. For coverage figures obtained by each frequency band, the most frequent 1,000 word families yielded an average of 85.09% coverage (*SD* = 1.34). The next most frequent 1,000 word families produced a mean of 6.62% additional coverage (*SD*

Table 1. Cumulative Coverage Figures on the Listening Passages from Five Past TOEFL iBTs by Nation's BNC Word-family Lists

Word Family	Test 1	Test 2	Test 3	Test 4	Test 5	Mean
Proper nouns	0.68	0.51	0.66	0.21	0.33	0.47
Marginal words	1.56	1.18	1.76	1.26	0.92	1.33
1,000	88.26	84.21	87.02	86.12	86.53	86.42
2,000	95.53	90.66	92.31	93	93.77	93.05
3,000	96.82	92.87	95.19	94.4	95.74	95
4,000	98.28	95.23	96.43	96.69	97.31	96.78
5,000	98.84	95.88	97.34	97.36	98.16	97.51
6,000	99.33	96.7	98.09	98.15	98.63	98.18
7,000	99.45	96.96	98.56	98.29	98.89	98.43
8,000	99.6	97.54	98.72	98.55	99.08	98.69
9,000	99.65	97.73	99.07	98.69	99.25	98.87
10,000	99.67	98.4	99.21	98.86	99.42	99.11
11,000	99.77	98.52	99.26	98.91	99.54	99.2
12,000	99.77	98.71	99.45	99.43	99.66	99.4
13,000	99.79	99	99.64	99.48	99.71	99.52
14,000	99.81	99.19	99.66	99.53	99.71	99.58
Not in the lists	100	99.98*	100.01*	100.01*	99.99*	99.99*

Note. Totals of percentages are not 100 because of rounding.

= 0.81), and the third most frequent 1,000 word families 1.95% (*SD* = 0.64). The fourth 1,000 word families yielded an average of 1.78% coverage (*SD* = 0.5). From the fifth most frequent 1,000 word families onwards, each frequency band produced less than 1% coverage. As Figure 1 depicts, text coverage figures obtained by each 1,000-word family rapidly declined as vocabulary became less frequent. Concerning words outside Nation's BNC lists, an average of 17.6 tokens were found in each passage, which constituted 0.41% coverage of the tokens on average. Most words outside Nation's lists were closely related to the topic of each passage. Therefore, the meaning of these low-frequency words can be guessed from the context. Proper nouns appearing in the listening passages constituted an average of 0.47% of the tokens. Marginal words such as exclamations and hesitation procedure consisted of 0.85% of the total running words on average. If the present study includes the coverage figures for proper nouns and marginal words in the cumulative coverage totals, as past text-coverage studies on spoken discourse did (Nation, 2006; Webb & Rodgers, 2009a, 2009b), then the most frequent 2,000 word families would provide 93% coverage for the listening passages. With a vocabulary of the most frequent 3,000 word families plus proper nouns and marginal words, 95% of the tokens were covered.

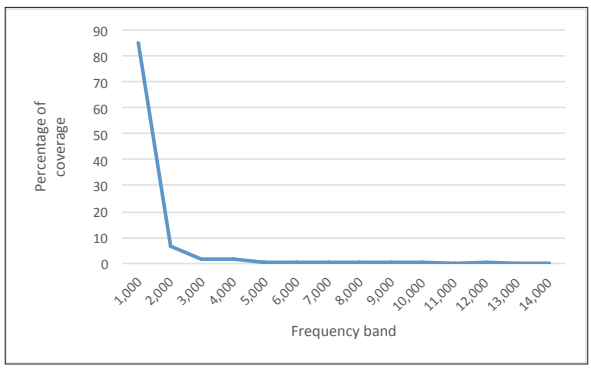

Figure 1. **Average Coverage Figures on Combined Five TOEFL iBT Listening Tests by Each 1,000 Word Families on Nation's BNC Lists**

Discussion

To summarize, the most frequent 3,000 word families plus proper nouns and marginal words provided 95% coverage for the listening passages. Table 2 compares mean cumulative coverage figures on the combined listening passages from the five past TOEFL iBTs with those on reading passages, as reported in Kaneko's study (2014). The results clearly demonstrate that the lexical level of the listening section is lower than that of the reading section as expected. Thus, it is highly likely that a vocabulary at the 6,000-word frequency level is sufficient for the reading and listening sections of the TOEFL iBT, indicating that a vocabulary beyond the 10,000-word frequency level is not necessarily required. Regarding the third research question, findings of the present study suggest that the lexical requirement of the listening section of the TOEFL iBT is not beyond the level of high school students. Pedagogical implications, limitations of the present study, and future research suggestions are discussed in detail in the following section.

Table 2. Average Cumulative Coverage Figures on the Listening and Reading Passages from Five Past TOEFL iBTs by Nation's BNC Word Family Lists

Word Family	Coverage on the listening passages	Coverage on the reading passages*
Proper nouns	0.47	2.37**
Marginal words	1.33	N/A
1,000	86.42	73.43
2,000	93.05	84.2
3,000	95	88.51
4,000	96.78	91.56
5,000	97.51	93.81
6,000	98.18	95.09
7,000	98.43	96.04
8,000	98.69	96.9
9,000	98.87	97.34
10,000	99.11	97.8
11,000	99.2	97.99
12,000	99.4	98.26
13,000	99.52	98.6
14,000	99.58	98.92
Not in the lists	99.99***	99.99***

Note. Adapted from "Is the Vocabulary Level of the Reading Section of the TOEFL Internet-Based Test Beyond the Lexical Level of Japanese Senior High School Students? " by M. Kaneko, 2014, Vocabulary Learning and Instruction. 2.37% includes proper nouns and defined words. Totals of percentages are not 100 because of rounding.

Pedagogical Implications

The first pedagogical implication is that students should be able to recognize the most frequent 3,000 word families aurally. As pointed out by Nishino and Watanabe (2008), most of the classroom time

at high schools in Japan tends to be devoted to improving reading proficiency. Instead, classroom practitioners should compensate for the lack of aural input by having students engage in extensive listening. Published graded readers often include audio CDs and using such resources can be a good option. Teachers can also provide aural input using websites. For instance, by creating an account at Extensive Reading Central (Waring & Browne, 2012), audio recordings of over 1,000 graded readers are available for free. Students can increase the amount of aural input by listening to those simplified listening materials.

Once students are used to listening to simplified listening materials, authentic listening materials should be incorporated because lack of exposure to common spoken language features such as contractions and connected speech may lead to poor listening comprehension. As Webb and Rodgers (2009b) suggest, knowing the written forms of individual word items does not necessarily ensure that contractions or connected speech are known. In fact, Bonk's (2000) study showed that nearly a quarter of the Japanese EFL subjects in his study were unable to comprehend connected speech appearing in the study although all the separate components were familiar to the participants. This suggests that contractions and connected speech may impair L2 listening comprehension. Thus, classroom practitioners should create opportunities for students to learn and strengthen the knowledge of contractions and connected speech. Concerning connected speech appearing in the listening passages from the five past TOEFL iBTs, *y'know*, *gonna*, and *kinda* appeared in three tests out of the five (range 3), *wanna*, *dunno*, and *gotta* in two tests (range 2), and *how'd*, *sorta*, and *outta* in one test (range 1).

The other pedagogical implication is that test takers of the TOEFL need to listen to much longer texts than those on the National Center Test. As mentioned earlier, the average number of tokens for the listening section of the five past TOEFL iBTs was 4,190 while the mean running words in the listening section from the 2006–2011 National Center Tests were 1,719 (Kaneko, 2012). TOEFL test takers are required to listen to passages whose tokens are more than twice as long as those in the Center Test. This clearly indicates the importance of ensuring much more aural input to prepare for the TOEFL test.

Limitations and Suggestions for Further Research

One limitation with the present study was that it was not able to calculate the coverage figures by multi-word units such as phrasal verbs and idioms because, as Nation acknowledges, the RANGE program cannot recognize multi-word units (2006).

The other limitation was that the corpus used for compiling Nation's word-family lists, the BNC, may not be the most appropriate for exploring the vocabulary of the TOEFL, considering the fact that some common spoken American English words such as *goof*, may not be included on Nation's BNC lists. As Nation (2004) acknowledges, the BNC is mainly written and in British English. Using Nation's (2012) newly created word-family lists based on the BNC and the COCA might lead to better coverage figures although use of the COCA/BNC lists would not ensure a meaningful comparison between results derived using the COCA/BNC lists and those in past TOEFL text-coverage studies.

Finally, it should be noted that the findings of the present study should be considered tentative because research investigating the effects of text coverage on comprehension in spoken discourse is in its infancy. Only a few studies are available so far (Bonk, 2000; Stahr, 2009; Van Zeeland & Schmitt, 2013) and findings are not consistent. We need more studies to be conducted before the findings of the present study can be properly evaluated. Yet, the findings seem to provide the answer to research questions 2 and 3: A vocabulary beyond the 10,000-word frequency level is not necessary for the reading and listening sections of the TOEFL iBT, and the vocabulary required for the listening section is within the level of high school students unless 98% coverage or more is desired. If future research investigating how text coverage influences listening comprehension supports that a lower text coverage figure, say 90%, is sufficient in spoken discourse as Van Zeeland and Schmitt (2013) suggested, then the most frequent 2,000 word families may suffice to comprehend listening passages in the TOEFL iBT.

References

ALC. (n.d.). Standard Vocabulary List 12000. Retrieved from <http://www.alc.co.jp/eng/vocab/svl/>

Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14*, 14–31. doi: 10.1080/10904018.2000.10499033

Chujo, K., & Oghigian, K. (2009). How many words do you need to know to understand TOEIC, TOEFL and Eiken? An examination of text coverage and high frequency vocabulary. *The Journal of Asia TEFL, 6*(2), 121–148.

Davies, M. (2008). The corpus of contemporary American English. Retrieved from <http://www.americancorpus.org/>

Educational Testing Service. (2013). *Official TOEFL iBT tests with audio* (Vol. 1). New York: McGraw-Hill.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programs. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Hirai, A. (2000). An analysis of vocabulary used in STEP 1st and pre-1st grades, TOEFL, and TOEIC. *言語文化論集, 54*, 1–16.

Ishida, M. (2004, Summer). *英語教員が備えておくべき英語力―英検準1級、ＴＯＥＦＬ５５０点、ＴＯＥＩＣ７３０点の目標値を中心に* [Recommended English proficiency levels for English teachers: With the focus on Eiken Grade Pre-1st, TOEFL 550 points, and TOEIC 730 points]. *英語展望, 111*, 10–17.

Ishikawa, S., Uemura, T., Kaneda, M., Shimizu, S., Sugimori, N., Tono, Y., ... Mochizuki, M. (2003). *大学英語教育学会基本語リスト* [JACET list of 8000 basic words]. Tokyo: JACET.

Kaneko, M. (2012). How large a vocabulary is required for the National Center Test? *Vocabulary Education and Research Bulletin*, *1*(2), 4–5.

Kaneko, M. (2014). Is the vocabulary level of the reading section of the TOEFL internet-based test beyond the lexical level of Japanese senior high school students? *Vocabulary Learning and Instruction*. Advance online publication. doi: 10.7820/vli.v03.1.kaneko

McCarthy, M., & Carter, R. (1997). Written and spoken vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 20–39). Cambridge, UK: Cambridge University Press.

Ministry of Education, Culture, Sports, Science and Technology. (2009). *Kotogakko gakushu shido yoryo* [The course of study guidelines for upper secondary school]. Tokyo: Author.

Mizumoto, A. (2006). *ＴＯＥＦＬ語彙の語彙レベル、真正性、特徴の検証* [Word levels, authenticity, and characteristics verification of the TOEFL vocabulary]. 言語文化学会論集, *27*, 3–15.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Amsterdam: John Benjamins.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59–82. doi: 10.3138/cmlr.63.1.59

Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly, 42*(1), 133–138.

Okada, K. (2014, February 17). 横浜市立高2年生でTOEFL受験 [Grade 11 students at public high schools in Yokohama to take the TOEFL test][Evening edition]. *The Asahi Shimbun*, p. 2.

Sato, M. (2013, May 31). 大学入試TOEFL導入への疑義 [Dispute over the adoption of the TOEFL as part of university entrance examinations]. *内外教育, 6250*, 1.

Schonell, F. J., Meddleton, I. G., Shaw, B .A., Routh, M. R., Popham, D. F., Gill, J. D., ... Stephens, C. (1956). *A study of the oral vocabulary of adults* (Report No. 1). Brisbane: The University of Queensland Press.

Stahr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*, 577–607. doi: 10.1017/S0272263109990039

Tonedachi, M. (2013, May 1). 争論　大学入試にTOEFL [Debate: Adoption of the TOEFL as part of university entrance examinations]. *The Asahi Shimbun*, p.15.

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: the same or different from reading comprehension? *Applied Linguistics, 34*(4), 457–479. doi: 10.1093/applin/ams074

Waring, R., & Browne, C. (2012). Extensive Reading Central. Retrieved from <http://er-central.com/>

Webb, S., & Rodgers, M. P. H. (2009a). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366.

Webb, S., & Rodgers, M. P. H. (2009b). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427. doi: 10.1093/applin/amp010

West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.

Xue, G., & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215–229.

Yoshida, R. (2013, April 9). LDP panel binds TOEFL to degrees. *The Japan Times*, p. 1.

Masaya Kaneko earned his M. A. in TESOL from Temple University Japan, and has taught English at various educational levels. He is currently a full-time instructor at Tokyo Denki University in Saitama, Japan. His research interests include L2 vocabulary acquisition, media corpus studies, and text coverage studies on standardized tests. He can be contacted at <m-kaneko@mail.dendai.ac.jp>.