



JALT Hokkaido Journal Vol. 9 pp. 1-16

journal@jalthokkaido.org

## Common *Gairaigo* Corresponding to High-frequency and Academic English -- Are Japanese Students Ready for Foreign Study?

Frank Daulton

[daulton@econ.ryukoku.ac.jp](mailto:daulton@econ.ryukoku.ac.jp)

Ryukoku University

---

For Japanese students considering university study in an English-speaking country, high-frequency vocabulary for everyday communication, and academic vocabulary for lectures and homework are essential. The first part of this study indicates that the average Japanese university student has deficiencies in both types of vocabulary; however, English-based loanwords in Japanese are already helping them. A second part of this study reveals that a quarter of the academic words correspond to loanwords.

英語圏の国における大学留学を考慮している日本人学生は、日常のコミュニケーション、講義を受けるための大学用語、そして宿題のために高頻出語彙は必要不可欠である。スタデイ 1 では平均的日本人大学生の場合両方の語彙を学習する必要があり英語ベース外来語はすでに役立っている。2 では高頻出英語の半数近くと 4 分の 1 の大学用語が日本語における一般外来語と一致しているという事を語ります。

### Introduction

How adequate are the English lexicons of Japanese students for everyday conversation and university study? Students first need to master the basic lexicon: the *high-frequency vocabulary*. Thereafter, those planning to enter a university in an English-speaking country should additionally master the words most often met in textbooks and lectures: the *academic vocabulary*. First in this paper, these two important frequency types of English words will be introduced as well as two relevant word lists. In Study One,

freshman university students' knowledge of high-frequency and academic vocabulary will be examined. Next, research revealing how English-based loanwords in Japanese help the learning of the original borrowed words will be briefly summarized. Daulton (2004) has shown that nearly half of the high-frequency words of English correspond to common loanwords in Japanese; Study Two will determine how many of the academic words similarly correspond to common loanwords.

### **High-frequency and Academic Vocabulary**

A careful choice of vocabulary can allow teachers and material makers to more aptly meet learners' lexical needs. Word lists are a first step to ensuring that learners spend time on the words most valuable to them (see Nation, 2001). One prominent use of word lists is in graded readers, such as the Oxford Bookworms series. In creating lists, one way to distinguish words is by their percentage of coverage within certain types of texts. For students with a variety of goals for tertiary study in English speaking countries, the most relevant types of vocabulary are the high-frequency and academic ones.

High-frequency words are those most frequently used (in a given language). For example, "the" is the most frequently appearing word in English. In order to achieve good comprehension of informal conversation and reading, as well as to have a well-rounded arsenal for production, at least 2000 and preferably 3000 word families (head words plus inflections and common derivations) are required (see Nation, 2001). If one takes proper nouns as a given (i.e. understood from context), this level provides nearly 95 percent coverage of most non-academic texts.

Meanwhile, academic words are quite useful for university students. In a sense,

academic words, e.g., “lecture,” are high-frequency words in tertiary education. Academic words are shared by several fields of study and are supportive of, but not central to, the topics (see Coxhead 1998).

The high-frequency words of English appear in Nation’s BNC 3000 (forthcoming). It contains 3000 word families derived from the British National Corpus (BNC), a collection of more than 100 million words sampled from a wide range of written and spoken sources. The academic words of English appear in Coxhead’s (1998) *An Academic Word List (AWL)*. The *AWL* contains 570 word families that are frequent and of wide range in a wide variety of academic texts. It is based on a corpus of 3.5 million running words from the Arts, Sciences, Law, and Commerce and was designed to help students prepare for tertiary education (Coxhead, 1998).

### **Study One: Student Knowledge of High-frequency and Academic Words**

The first study measured both the high-frequency as well as the academic word lexicons of a group of freshmen students in Japan in order to assess their lexical readiness for everyday English communication and tertiary study in an English-speaking country, respectively.

#### *Participants*

The participants were 72 first-year Business and Law students in a required English class at a Japanese university in Kyoto. As first-year students, their ages were 18 and 19, with a four-to-one male/female ratio. The university is commonly referred to as an upper-middle-level university. The level of English ability was typical of many non-majors (i.e., “false beginner”) with widely varying levels of motivation. No TOEIC or TOEFL information was available. Participants were informed their test data were

intended for self-awareness and research purposes.

### *Instruments*

Sections from the Vocabulary Levels Test -- also known as the “Levels Test” -- were used to measure the students’ knowledge of high-frequency and academic vocabulary. The Levels Test samples words at given levels and is quick to take and easy to mark and interpret; as a multiple choice test, it gives credit for partial knowledge of a word. The “Revised 2,000 Word Level Test Form A” was used to measure the participants’ high-frequency lexicons, and “Test A: Academic Vocabulary” was used to measure knowledge of academic words.

The original Levels Test first appeared in 1983 and was later republished in Nation’s (1990) book. Since then, various versions have been employed around the world in both assessment and research (e.g., Schmitt and Meara, 1997). Schmitt, et al. (2001) found that the Levels Test provided accurate estimates of the vocabulary size of students at the targeted frequency levels.

In the “Revised 2,000 Word Level Test Form A,” target items are given in clusters of three, with six choices (three correct answers and three distracters) for each cluster. In “Test A: Academic Vocabulary,” six academic words appear of which three can be matched to single-word or phrasal answers. In contrast to subsequent levels of the Levels Test, the 2000 level contains two levels of words; 1-to-1000-level and 1001-to-2000-level words are sampled at a 1:2 ratio.

### *The Data*

Participants’ performance in the two tests is summarized in Table 1:

**Table 1: Participant Knowledge of High-frequency and Academic Vocabulary**

	<b>High-frequency: % (# correct out of 27)</b>	<b>Academic: % (# correct out of 30)</b>
<b>Average</b>	82.0% (22.1)	69.2% (20.8)
<b>Median</b>	66.7% (18)	58.4% (17.5)
<b>High</b>	100% (27)	100% (30)
<b>Low</b>	33.3% (9)	16.7% (5)

Regarding high-frequency vocabulary, which includes words such as “debt”, “roar” and “pride”, the participants’ average score was 82 percent. That means 82 percent of 3000 word families were known by participants, or 2460 families. The median score was 66.7, the high score 100 and the low 33.3. The average score in particular indicates that most participants’ have not mastered the basic vocabulary for communication.

Regarding academic vocabulary, which includes words such as “labour”, “percent” and “principle”, the participants’ average score fell to 69.2 percent, which means 69.2 percent of 570 word families were known, or only about 394 families. The median was 58.4, the high score 100 and the low 16.7.

Considering the different natures of high-frequency and academic vocabulary, it is remarkable that the average scores were so close; English-based loanwords in Japanese may help explain this. Roughly half of the items in the “Revised 2,000 Word Level Test Form A” happened to be common loanwords in Japanese and roughly a quarter of the “Test A: Academic Vocabulary” items were also. Taking the first five non-borrowed and borrowed words occurring in each test, participants’ accuracy on individual items

was surveyed. The results are presented in Table 2:

**Table 2. Participants' Accuracy on Non-borrowed Versus Borrowed Test Items**

<u>h.f. word</u> <b>(borrowed)</b>	accuracy % <b>(borrowed)</b>	<u>academic word</u> <b>(borrowed)</b>	accuracy % <b>(borrowed)</b>
roar	56.9%	labour	83.3%
debt	81.9%	principle	37.5%
temperature	86.1%	philosophy	86.1%
flesh	79.2%	sum	83.3%
wage	93.1%	investigation	56.9%
<b>pride</b>	<b>91.7%</b>	<b>percent</b>	<b>97.2%</b>
<b>salary</b>	<b>95.8%</b>	<b>technique</b>	<b>94.4%</b>
<b>justice</b>	<b>94.4%</b>	<b>topic</b>	<b>94.4%</b>
<b>cream</b>	<b>95.8%</b>	<b>link</b>	<b>95.8%</b>
<b>motor</b>	<b>98.6%</b>	<b>publish</b>	<b>81.9%</b>
	+120.0%		+133.6%

In each column, words and figures appearing in bold refer to borrowed words (e.g., “pride”), and those not in bold refer to non-borrowed words (e.g., “roar”). We see that participants were 120 percent (+120.0%) more likely to respond correctly for a borrowed high-frequency word than a non-borrowed one. Moreover, participants were 133.6 percent (+133.6%) more likely to respond correctly for a borrowed academic word (e.g., “percent”) than a non-borrowed one (e.g., “labour”). Remarkably, as far as borrowed words were concerned, participants did almost as well for high-frequency words as for academic words, supporting Ringbom’s (1987) assertion that the borrowed word effect is robust.

## **The Positive Effect of Loanwords on the Learning of English Borrowed Words**

The familiar argument that loanwords in Japanese obstruct the learning of English has been largely anecdotal -- based on observations of errors with little empirical evidence presented. Researchers such as Ringbom (1987) believe that a focus on errors is not an appropriate way to approach cognates. Indeed, regarding the effect of loanwords in Japanese, empirical studies of university students have shown an overall positive effect on various learners and various aspects of word knowledge.

Loanwords in Japanese (e.g., *takushi*) assist the learning of their related borrowed words (e.g., “taxi”). Borrowed words are easier for Japanese to discern (Hashimoto, 1992), easier to comprehend (Brown & Williams, 1985; Kimura, 1989; Daulton, 1998), easier to spell (Hashimoto, 1993; Daulton, 1998) and preferred in multiple choice tests (Brown, 1995) as well as in production (Daulton, 2003).

Assuming learners are able to process the inflections and common derivations of English (see Schmitt & Meara, 1997), knowledge of even a single word family member (e.g., “win”) will give access to its related members (e.g., “wins”, won, “winner”, “winners”, and “winning”). Thus loanwords can be regarded as a latent vocabulary base, or a “built-in lexicon.”

### **Study Two: The Number of Correspondences between Academic English and Japanese**

For Japanese learners of English, Daulton (2004) showed that nearly half of the 3000 most-common word families of the Nation’s (forthcoming) BNC 3000 correspond to common loanwords in Japanese. His data is summarized in Table 3:

**Table 3. Common Loanwords in Japanese Corresponding to the BNC 3000**

	# of words corresponding to loanwords	# of word families corresponding to loanwords	% of families corresponding to loans
first 1000	803 words	550 families	55.5 percent
second 1000	632 words	493 families	49.3 percent
third 1000	371 words	317 families	31.7 percent
overall	1806 words	1360 families	45.3 percent

For the first 1000 word families of the BNC 3000, 803 words within 550 word families corresponded to loanwords – 55.5 percent of families. This percentage was 49.3 percent for the second 1000 words and 31.7 percent for the third 1000. Overall, 45.3 percent of BNC 3000 word families had at least one member corresponding to a common loanword.

Each correspondence resulted when one or more of the following was borrowed from English: a word family headword; an inflected word; or a derived word. For instance, the “act” word family had correspondences in Japanese with both its headword (*akuto*) and the derived form “action” (*akushon*). Roughly speaking, each word-family correspondence involved about 1.5 family members.

As Daulton (2004) regarded formal relatedness as primary to cognateness, he examined the resemblance between borrowed words and loanwords by first considering rephonologization. The adaptation of English words to the Japanese sound system creates obvious disharmony. However Daulton argued that by virtue of their “*katakana* filter,” when Japanese hear an English words such as “taxi,” they are likely to perceive it as “*takushii*” -- thus allowing access to their native-language knowledge. This observation is supported by research such as Hashimoto (1992), which found that loanwords in Japanese aid word recognition in spoken English as well as written. However, this mechanism is likely to break down during radical shortening {e.g.,

“apartment (house)” and *apaato* }.

Daulton (2004) proceeded to employ an integrated evaluation of shortening and semantic relatedness. In a sampling of every 20<sup>th</sup> word family among the corresponding BNC 3000 word families, he found no instances of shortening, which he attributed to the fact that his list of correspondences had excluded the longest loanwords, e.g., compounds such as *salariman*. Semantically, he found that while many cognates are convergent (having one meaning while the borrowed word has multiple), it was common for their single meaning to be the most common in English (e.g., “leisure” and *rejaa*; “motor” and *mootaa*).

Based on Daulton (2004), the similarity between English borrowed words and Japanese loanwords is sufficient to allow recognition of interlingual relatedness and helpful transfer of L2 concepts. Study Two will determine how many correspondences exist between common loanwords and academic English.

#### *Sources of Words*

Coxhead’s (1998) *AWL* is divided into sub-lists based on the frequency of the words in the academic corpus. Word family members from the *AWL* were first sought in a Japanese/English dictionary of common loanwords -- Motwani’s (1991) *A Dictionary of Loanwords Usage*, which contains a convenient index. There are 3019 headwords in Motwani’s dictionary, with many other words embedded under particularly productive headwords. One hundred eighty-two *AWL* words were found to correspond to words in the Motwani dictionary. A 28-year old native informant located 82 additional correspondences, bringing the total to 264 correspondences.

### *Testing Corresponding Loanwords*

For loanword cognates to help in learning foreign words, they must themselves be known. In Japanese, there are tens of thousands of loanwords with varying degrees of integration, most clustering at the obscure end of the spectrum. Therefore, as in Daulton (2004), self appraisal was used to confirm university students' comprehension of the corresponding loanwords -- a yes/no test instrument with a neutral "I'm not sure" option to discourage false positives. In the yes/no format, participants simply indicated whether or not they knew the meaning of a word, or to what degree they knew it. By including nonsense words, Daulton (2004) demonstrated that Japanese university students' responses are highly reliable. Therefore, it was deemed unnecessary to include nonsense words in the present study.

The 264 loanwords corresponding to the *AWL* were distributed among seven test versions. Each of the versions was seen by 10 participants -- the same students as in Study One. The words to be tested were further grouped by the *AWL* sub-list. A given word had to receive seven "yes" responses (out of a possible 10) to remain in the list of correspondences. This criteria is supported by Chall & Dale (1950: 206).

### **Results and Discussion**

Eighty seven poorly understood loanwords (33 percent of the original 264) that had corresponded to 58 *AWL* word families were eliminated, including *akademikku* ("academic") itself. The remaining correspondences between common loanwords and the *AWL* are summarized in Table 4:

**Table 4: Details of verified loanword correspondences to the AWL**

	# of words corresponding to loanwords	# of word families corresponding to loanwords	% of families corresponding to loanwords
Sublist 1	29 words	23 families	38 percent
Sublist 2	23 words	19 families	32 percent
Sublist 3	30 words	22 families	37 percent
Sublist 4	20 words	19 families	32 percent
Sublist 5	19 words	16 families	27 percent
Sublist 6	12 words	12 families	20 percent
Sublist 7	19 words	18 families	30 percent
Sublist 8	14 words	13 families	22 percent
Sublist 9	9 words	9 families	15 percent
Sublist 10*	2 words	2 families	7 percent
overall	177 words	153 families	27.0 percent

For instance, regarding the academic words in Sub-list 1 of the AWL, 29 words from 23 word-families corresponded to common Japanese loanwords, which is 38 percent of Sub-list 1. In all, 177 words from 153 word families in the AWL corresponded to loanwords -- 27 percent of word families. At each sub-list of the AWL, the percent of correspondences generally decreased with frequency of vocabulary; by Sub-list 10 it falls to seven percent. Typically each word family correspondence involved roughly 1.2 family members.

Study Two revealed that about one quarter of the academic words of English have entered the everyday Japanese lexicon. This is a conservative count, since as in Daulton (2004), it ignores the abundant compound forms and deals only with free-standing

loanwords. For the same reason, shortening is unlikely to have occurred. Moreover, due to internationally defined and accepted standards, academic words in English and Japanese are likely to share meanings, as with technical and scientific terms (see Nishiyama, 1995, p. 33).

### **Limitations and Implications**

The present study of one intact class cannot be generalized to wider populations before further studies confirm the findings. Indeed, it may also be useful to test certain assumptions made by this study with regard to the use of the Levels Tests. Having said that, some tentative remarks can be made. Were the students in this study *lexically* prepared for academic study in an English-speaking country? It appears most students have considerable work ahead in preparing for university study in an English environment. Fortunately, by virtue of their native language, Japanese have a considerable built-in lexicon of high-frequency and academic English. To what degree this resource can be further developed is a topic for future study.

A second limitation to this study is to acknowledge the varieties of research approaches to vocabulary learning that differ from the one taken in this study, such as studies based on collocation and other lexical approaches. Instead, this study is based on the assumption that identifying high-frequency words is most useful for teachers designing a content syllabus. In particular, I recommend the list of BNC 3000 words corresponding to common Japanese loanwords which is online at <http://www.angelfire.com/wa/yakineko/gairai.html>. The AWL words corresponding to common loanwords can be found in the appendix, while the complete AWL is also available online (Coxhead, 1998). These lists can be applied in various ways. Basically, teachers should introduce learners early on to these borrowed words in order to make

rapid progress. For more advanced learners, as borrowed words are already “known” to a certain degree, a better use of time might be to focus on the non-borrowed high-frequency and academic words. Meanwhile, fluency-practice with all high-frequency and academic words is suggested for most learners.

## References

- Brown, J. B. and Williams, C. J. (1985). Gairaigo: a latent English vocabulary base? *Tohoku Gakuin University Review: Essays and Studies in English Eibungaku* 76. Sendai, Japan: 129-146.
- Brown, J. B. (1995). Is gairaigo English? *The Internet TESL Journal*, 1, 12. (December). Retrieved Nov. 28, 2005 from <http://iteslj.org/Articles/Brown-Gairaigo/index.html>
- Chall, J. S. and Dale, E. (1950). Familiarity of selected health terms. *Educational Research Bulletin*, 39, 197-206.
- Coxhead, A. (1998). *An academic word list*. Occasional Publication Number 18, LALS, Victoria University of Wellington, New Zealand. (Also: <http://www.vuw.ac.nz/lals/research/awl/>)
- Daulton, F. E. (1998). Japanese loanword cognates and the acquisition of English vocabulary. *The Language Teacher* 20, 1: 17-25.
- Daulton, F. E. (2003). The effect of Japanese loanwords on written English production - A pilot study. *JALT Hokkaido Journal*, 7, 4-14. <http://journal.jalthokkaido.org/index.htm>
- Daulton, F. E. (2004). Gairaigo – The built-in lexicon: The common loanwords in Japanese based on high-frequency English vocabulary and their effect on language acquisition. Ph.D. dissertation, Victoria University of Wellington.
- Hashimoto, R. (1992). English loanword interference for Japanese students of English. Unpublished M.Ed. thesis. Mankoto State University.
- Kimura, M. (1989). The effect of Japanese loanwords on the acquisition of the correct range of meanings of English words. Masters Thesis, Brigham Young University, Department of Linguistics.
- Motwani, P. (1991). *A Dictionary of loanword usage*. Tokyo: Maruzen Co., Ltd.

Nation, I. S. P. (2001). How many high frequency words are there in English? In M. Gill, A. W. Johnson, L. M. Koski, R. D. Sell and B. Warvik (eds.) *Language, learning and literature: Studies presented to Hakan Ringbom. English Department Publications 4*, Abo Akademi University, Abo: 167-181.

Nation, I. S. P. (forthcoming). A study of the most frequent word families in the British National Corpus.

Nishiyama, S. (1995). Speaking English with a Japanese mind. *World Englishes*, 14, 55-66.

Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon & Philadelphia: Multilingual Matters.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition*, 9, 17-36.

**Appendix: The List of AWL Words Corresponding to Common Loanwords**

List explanation: Borrowed AWL words are represented by sub-list in alphabetical order. Underline words are headwords. Headwords appearing in italics do not correspond to loanwords but are only given for reference. Word family members (the inflections and common derivations) appear next to their headwords.

**Sub-list 1**

<u>approach</u>	<u>area</u>	<u>benefit</u>	<u>concept</u>	<u>contract</u>	<u>create</u>	creative
creator	<u>data</u>	<u>economy</u>	economic	economist	<i>finance</i>	
financial	<i>identify</i>	identity	<u>income</u>	<u>legal</u>	<u>major</u>	majority
<u>method</u>	<u>percent</u>	percentage	<u>policy</u>	<u>process</u>	<i>respond</i>	
response	<u>section</u>	<u>source</u>	<i>vary</i>	variation		

**Sub-list 2**

assist      assistant      category      chapter      community      complex  
compute      computer      credit      culture      design      designer      final  
focus      impact      item      normal      resource      select      selection  
site      text      tradition      traditional

**Sub-list 3**

comment      commentator      consent      coordinate      coordinator  
demonstrate      demonstration      demonstrator      document      illustrate  
illustration      initial      link      locate      location      maximize      max  
maximum      minor      minority      partner      proportion      publish  
react      reaction      sex      sexual      shift      task      technical  
technique      technology      volume

**Sub-list 4**

access      code      communicate      communication      contrast      cycle  
cycling      debate      domestic      error      ethnic      goal      job  
label      mechanism      option      professional      project      promote  
promotion      series      stress

**Sub-list 5**

challenge      challenger      contact      energy      generation      image  
license      medical      mental      mentality      monitor      network  
orient      orientating      style      stylish      symbol      target      trend  
version

**Sub-list 6**

cite      cooperate      cooperation      display      domain      edit      editor  
expert      gender      input      instruct      instructor      lecture      motive  
motivation      tape

**Sub-list 7**

adult      channel      chemical      classic      classical      couple  
dynamic      extract      file      globe      global      grade      media  
mode      release      reverse      reversible      survive      survival      topic  
unique      voluntary      volunteer

**Sub-list 8**

automate      automatic      chart      drama      dramatic      guideline  
paragraph      plus      random      schedule      tense      tension  
terminate      terminal      theme      uniform      visual

**Sub-list 9**

manual      medium      relax      revolution      route      scenario  
supplement      team      vision

**Sub-list 10**

panel      pose