

How many high frequency words of English do Japanese university freshmen`know'?

著者(英)	James Martin Rogers
journal or publication title	Journal of Inquiry and Research
volume	97
page range	237-252
year	2013-03
URL	http://doi.org/10.18956/00006092

How many high frequency words of English do Japanese university freshmen 'know' ?

James Martin Rogers

Abstract

Knowledge of high-frequency vocabulary is essential to language fluency. However, there is more to knowing a word than simply knowing its meaning. Full vocabulary depth knowledge includes not only semantics, but also knowledge of a word's phonology, orthography, collocations, word parts, grammar, constraints on use, concepts and referents, and associations.

This research examined the extent of knowledge that a group Japanese university freshmen have of high-frequency English vocabulary. First, students were made to give judgments on how well they knew a list of 3,000 high-frequency English lemma. Then, items deemed as being known were examined to determine whether they functioned as a cognate between English and Japanese. Finally, a sample of the items marked as being known was tested in regards to the full range of vocabulary depth knowledge.

The results of this study showed that while the students tested in this study are familiar with a majority of high-frequency English vocabulary, their vocabulary depth knowledge of these items is shallow. The results of this study should help guide teachers as to which aspects of vocabulary depth they need to concentrate on to help make language learning more efficient.

Keywords: vocabulary depth, high-frequency vocabulary, Japanese learners

Introduction

By the time Japanese students enter university, they have studied English for six years or more. In addition, Japanese shares many cognates with English, with some estimating that upwards of 10% of the Japanese language consists of cognates (Orita, 1999). With these two factors in mind, one might then assume that this would result in these learners having a strong base knowledge of the English language by the time they enter university. A pertinent question to ask is to what extent do Japanese students truly 'know' the most common vocabulary? Specifically, what is their vocabulary depth knowledge of high frequency English vocabulary? It remains to be seen just how many of the high-frequency

words of English Japanese university students truly know, and examining their depth of vocabulary knowledge is essential to answering this question.

High Frequency Vocabulary

High frequency vocabulary can cover 80% or more of the words in most texts (Nation, 2008). Therefore, knowing these words enables learners to access a wide variety of input and understand a large proportion of them.

Nation (2001) states:

The distinction between high frequency and low frequency words is critical in language teaching and course design, because teachers should deal with high frequency words in different ways from low frequency words. High frequency words deserve direct teaching and classroom time. Low frequency words do not. (p. 167)

So, what exactly should be considered as high-frequency vocabulary? What words are considered as such will vary according to the corpus used and the arbitrary cutoff point one sets. The goals of the course or learners, the intent of the research, time constraints, etc. all affect where this line is drawn. Despite the subjective nature of this factor, teachers and researchers must distinguish what is to be considered high frequency vocabulary and what is to be considered low frequency vocabulary.

Nation (2001:180) himself states that this cutoff point is ultimately “an arbitrary one.” He cites various research that point to a cutoff point at 3,000 for lemma, and also a cutoff point at 2,000 word families. Regardless of where one chooses to limit their list, the cost to benefit ratio is what ultimately divides low and high frequency vocabulary. Nation remarks that the “critical factor in making a decision about the high frequency/low frequency boundary will be the amount of coverage provided” (p. 172). Simply stated, if a word occurs often enough, then that is a word worth learning.

Loanwords

There is a notion that if “two languages share a lot of cognate vocabulary, then some of the learning burden will be lighter” (Nation, 2006a:448). Japanese certainly shares many cognates with English, and research has shown that the majority of these items are semantically direct or nearly direct borrowings. Ishikawa and Rubrecht’s (2008:314)

study of English loanwords used on Japanese television found that “the most common type of loanword found in the sample as a whole, and indeed, within each genre, was...direct borrowings” at 63%. Higa (1973) also found a large percentage (80.8%) of borrowed words in Japanese to be similar or the same in regards to form and semantics.

Oshima (2002:52) asks the following pertinent question: “Could *gairaigo* be a help rather than a hindrance to English learning?” Brown (1995), Daulton (1998, 1999, 2003, 2005, 2008), Underwood (1999), Rebuck (2002, 2007), Oshima (2002), Uchiwa (2007), Ishikawa and Rubrecht (2008), and Van Benthuyzen (2004) believe loanwords can facilitate language learning.

Daulton's (2008) study on how many loanwords there were in Nation's top 3,000 word families of English and in Coxhead's *Academic Word List* resulted in 45.2% and 26.8% corresponding to loanwords, respectively. With nearly half of the high frequency words of English and a good portion of academic words corresponding to loanwords, and being that such words have been said to have a low learning burden, one would assume that Japanese students would have good depth of knowledge of such items. For example, Van Benthuyzen (2004:169) states that “it seems almost self-evident that L2 vocabulary acquisition will be easier for an item which has a corresponding cognate in the L1 or which has been taken into the L1 as a loanword.” However, Daulton (2008:120) believes it is “imprudent to assume that Japanese learners can extend word knowledge within word families – even those assembled on the criteria of transparency and minimum learning burden.”

This paper will thus highlight how, beyond semantics and visual recognition, the students tested in this study have very little vocabulary depth knowledge of such items. The existence of a plenitude of direct or nearly direct borrowings may lead to the misconception that students 'know' such items, when the reality is that their knowledge of such items is quite shallow.

Vocabulary Depth Introduction

Vocabulary knowledge is central to mastery of a language. At its core, language learners must master tens of thousands of vocabulary items before they can be considered to be fluent in that language. An inventory of semantically known vocabulary items is commonly referred to as *breadth* of vocabulary knowledge (Qian, 1999:284). Qian states that breadth of vocabulary knowledge is one of the primary dimensions of vocabulary knowledge, while the other is *depth* of vocabulary knowledge. Many researchers have

created conceptual frameworks to describe depth of vocabulary knowledge, with many complementing each other. An example of such would be Nation's (2001) vocabulary depth framework, which includes both receptive and productive knowledge of a word's semantics, phonology, orthography, collocations, word parts, grammar, constraints on use, concepts and referents, and associations.

One of the main goals for language teachers should be to help students achieve a strong depth of vocabulary knowledge, in addition to breadth. So, by not only knowing a wide variety of words, but also knowing them well, students are able to access a wide variety of input and understand a large proportion of them.

Semantics

Semantics is an area of vocabulary depth that most researchers investigate. Schmitt (2010) points out that most studies of vocabulary knowledge only measure this form-meaning link. He states that "good vocabulary research is advantaged by multiple measures of vocabulary, to better capture a wide range of lexical knowledge" (152). Thus, this study aims to specifically look beyond mere semantics, and judge what other aspects of vocabulary depth students possess.

Phonology and Orthography

Fundamentally, before any other aspects of vocabulary knowledge can be known, students must be familiar with a word's form. To be able to produce a word's phonological and orthographic features accurately is an important first step in mastering a word's knowledge. But for students of English, grappling with its deep orthography may be a daunting task. The orthographic depth hypothesis (Katz & Frost, 1992) states that some languages have more variability between their orthography and phonology than others. English has a deep orthography. For instance, the phoneme /ε/ is represented in 6 different ways in the following words: *guest*, *revenge*, *friend*, *measure*, *said*, and *message*, while in Japanese /ε/ is only represented by the single grapheme ɛ. Therefore, accurate production of orthographic representations can be problematic. Furthermore, phonological mediation (Van Orden, 1987), or phonology's influence on visual word recognition can affect Japanese students' ability to accurately produce a word's phonological features. This is especially true

for loanwords, which undergo rephonalization (Kay, 1995), or an adaptation of pronunciation in Japanese. For example, the letter 'a' in *mat*, *go-cart*, and *America* is represented by the grapheme ア in Japanese. In contrast, the three words are represented by /æ/, /ä/, and /ə/ respectively in English.

Collocations

The extent to which Japanese students are familiar with collocations for high frequency English loanwords is another essential question, perhaps one of the most pertinent questions due to their sheer numbers. Defining the term 'collocation' is complex and there is much variety in what different researchers consider to be collocations, but Nation (2001) simply defines a collocation as a word and the "words that commonly occur with it" (p. 27). Bahns and Eldaw (1993), Lewis (2000), and Nesselhauf (2003) all acknowledge the significance of collocation knowledge for second language learners. However, many studies have found collocations to be highly problematic for L2 learners, such as in Gabrys-Biskup (1992), Hussein (1990), and Ozaki (2011).

Word Parts

Another question that needs to be addressed is whether or not students are able to extend their knowledge of words throughout entire word families through affix knowledge? Schmitt and Meara (1997) found that Japanese high school and university students had very poor English affix knowledge. Daulton (2008:120) agrees, stating that it is "imprudent to assume that Japanese learners can extend word knowledge within word families – even those assembled on the criteria of transparency and minimum learning burden". However, Daulton (2004) did find that Japanese students in general could extend their knowledge of one headword which is a loanword to comprehend other words within that word family better than with non-loanwords, but this depended on the level of difficulty of the affix, which Mochizuki & Aizawa (2000) also point out as a factor. Therefore, Japanese students may lack knowledge of specific morphological derivations of high frequency headwords. For instance, Japanese students may be aware of *support* and its derivation *supporter* since they both function as loanwords in Japanese, but they may not be able to produce other derivations, such as *supportive*, *supportable*, and *unsupportable*.

Grammar

Additionally, we must inquire as to whether students are aware of other parts of speech that a word can function as. For instance, a student may be aware of the noun *house*, but may not be aware of the verb *house*. Nation (2001:55) states that “in order to use a word it is necessary to know what part of speech it is and what grammatical patterns it can fit into.” A misconception as to the part of speech a word represents has potential to hinder comprehension, and teachers should be aware of which parts of speech are unknown to their students.

Constraints on Use

Finally, it is worthwhile to also investigate whether Japanese students are aware of constraints of use for high frequency vocabulary. Are they aware that certain words are only used in specific situations or in specific regions? Are they aware of formal or informal registers that certain words comprise? Nation (2001:57) states that “failure to observe these can result in inappropriate use.” So, for example, the Corpus of Contemporary American English’s (COCA) (Davies, 2008) second highest frequency collocation for the high frequency loanword *opening* is *opening statement*. Are students aware of the constraint of use on that collocation which limits its use in general to the courtroom? This study will help to shed light onto just what constraints students are actually aware of for high frequency vocabulary.

Concepts and Referents

Being familiar with an underlying concept behind the semantics of a word is yet another aspect of vocabulary depth knowledge. Nation (2001:50) asks the pertinent question: “Should teachers be trying to show the meaning underlying different uses of a word or should teachers treat the different uses as different items to learn?” It is clear to see the benefit of how being aware of an underlying concept in which a word, with two seemingly separate meanings, can make language learning more efficient. Ruhl (1989) shows that a singular abstract meaning can be seen as a thread running through all senses of a word. Parent (2009) agreed, and went as far as to question whether true polysemy really even existed.

Associations

It also remains to be seen whether Japanese students can produce associations of high frequency vocabulary anywhere near native levels. Schmitt (2010:58) states that the network connections between lexical items can lead “not only to appropriate lexical usage...but also more automaticity in using this knowledge.” Aitchison (2003) found three reoccurring trends when examining the associations made by native speakers. First, responses were almost always from the same semantic field (*needle* would elicit *sewing*). Second, items with clear pairs (*husband/wife*) or clear antonyms (*tall/short*) were typical. Third, response part of speech usually matched stimuli (nouns elicited nouns). Schmitt (2010) also notes that another reoccurring trend was that as a language learner becomes more fluent, their responses shift from being syntagmatic (responses with a sequential relationship, such as *black magic*) to more paradigmatic (responses which share part of speech).

Materials

In choosing an appropriate high-frequency word list to test students with, the choice between a lemmatized list and a word family list is a crucial one in regards to Japanese students. A lemma is a “set of related words that consists of the stem form and inflected forms that are all the same part of speech” (Nation & Meara, 2002:36). So, the verb *run* would be the lemma that represents the forms *runs*, *running*, *ran*. The noun *run* would be listed as a separate entry.

To clarify the difference between these two different but equally valid ways of counting words, one needs to think of word families as “a base word and all its derived and inflected forms” (Bauer and Nation, 1993:11). It is easy to see how a list which combines items in such a way can be problematic in estimating vocabulary knowledge when we look at the headword *govern* in a word family list, which also includes *governs*, *governing*, *government*, *governments*, *governor*, *governors*, *governable*, and *ungovernable*. Daulton's (2008) study makes it clear that extending knowledge of one item throughout a word family is difficult for Japanese students. Furthermore, the potential for confusion that heteronyms create, such as how a learner could easily mistake the verb *house* for the noun, must also be considered.

These issues contributed greatly to the decision to work with a lemmatized list, and thus Kilgarriff's (1995) lemmatized list was used in this study.

To test for word associations, high frequency associations from Moss and Older's (1996) study on native association norms were utilized as a reference.

Procedure

The lemma list utilized in this study was presented to 52 Japanese university freshmen. The proficiency of these students was level one (the highest) in a five-band proficiency system at the university. Students are broken up into these five bands based on their grades on Bennessé's GTEC (2004) test. Level one students scored from 246 to 352 points on this test.

These students were presented with Kilgarriff's (1995) list and were asked to rate their knowledge of a word using the following 5 criteria:

1. 100% I know it and can use it.
2. 75% I know it and can use it, but not perfectly
3. 50% I know it if I hear it, but can't use it
4. 25% If I hear it, I can guess the meaning
5. 0% I don't know the meaning at all.

Students were informed that to give a rating of 1, they would have to be very confident in their ability to produce an example sentence using the word correctly. If they gave a rating of 2, they should still be able to produce an example sentence easily, but they may not be 100% confident that the sentence is perfect. Five non-words were included every 250 words to ensure that students were paying attention.

Next, the checked lists were combined and if any word had an average rating of 2 or more, then that word was considered to be thought of as 'known' to students.

To determine the extent of how the existence of cognates played a role in students' ratings, each item that was marked as being known was examined to determine whether it exists as a cognate in Japanese.

Then, this study utilized Nation's (2001) vocabulary depth framework to determine whether students are also familiar with a word's phonology, orthography, word parts, semantics, concepts and referents, associations, grammar, collocations, and constraints on use. Every n^{th} item (which filled the criteria to be tested) from the lemma list recorded as being known by students was selected, and questions were made to determine the extent

of students' productive depth knowledge of these words. This test was administered to 77 level one university freshmen.

Semantic knowledge was judged simply by showing students an L1 translation of 20 items, and having them try to produce the appropriate L2 translation. Minor spelling mistakes were ignored.

To judge phonological knowledge, students' pronunciation of 20 items was recorded and rated on a scale of 1-10 by three native judges. If a student's pronunciation was native-like, with no discernible accent whatsoever, then that student received a 10 for that particular item. If the student's pronunciation was unmistakably heavily accented, then such a student will be given a score of 1 for that item.

To rate orthographic knowledge, 20 target items recorded by a native speaker were played for students, and they had to try to spell the items properly. To judge their answers, two scores were given for 20 items: strict and sensitive. First, to obtain a strict score, an item was deemed as either simply correct or incorrect. Then, a more sensitive marking system was employed. Levenshtein edit distance (1965), which has been used to measure similarity quantifiably (Mulloni & Pekar, 2006), was utilized to obtain sensitive scores. This study utilized Garcia's (2005) Levenshtein edit distance tool to obtain edit distance scores.

Collocational knowledge was judged with a 20 question cloze test. Collocations for the target items were chosen from high-frequency examples from the Corpus of Contemporary American English (Davies, 2008). Single letter prompts were utilized for the target item to prevent students from providing alternative answers. An effort was also made to ensure that all words in the example sentences were also high-frequency, and within the top 3,000 word families of English by using Cobb's (n.d.) *Vocabprofile*. Minor spelling mistakes were ignored.

Word part knowledge was judged by providing students with 20 headwords, and students had to choose from 20 affixes to choose two affixes which can be used with the headword. Out of the 20 affixes, only two were possibilities. Students were given partial credit if they could guess one of the two possibilities.

Grammatical knowledge was judged with questions which listed target items which functioned as more than one part of speech. Students had to choose whether a word functioned as a noun, verb, adjective, or adverb. For each target item (20 questions), a filler item was added which only functioned as one part of speech to avoid students seeing a pattern and guessing. Thus, this section had 40 questions in total.

Knowledge of constraints on word usage was tested with cloze questions which described the appropriateness of the target items' usage, and students had to produce that word. Single letter prompts were utilized for the target item to prevent students from providing alternative answers. Minor spelling mistakes were ignored.

Students' knowledge of word concepts and referents was tested with 20 questions which described two separate usages of a word, both of which had a common underlying concept. Single letter prompts were utilized for the target item to prevent students from providing alternative answers. Minor spelling mistakes were ignored.

Association knowledge was judged orally by presenting 20 recorded stimuli and giving students 12 seconds to produce an association with the item. Students received points if they chose an association that more than 20% of participants in Moss and Older's (1996) native association norms study also gave. Small spelling mistakes were ignored.

All test questions were piloted with native speakers to ensure validity.

Results

Out of the 3,000 lemma that were tested, students rated themselves as 'knowing' 1,903 items. In addition, all non-words were identified and marked as unknown by students. When these 1,903 items were examined, it was found that 715 of them function as cognates between Japanese and English. At 23.8%, the percentage of loanwords was significantly less than Daulton's (2008) 45.2%, but comparing the two results is problematic. First, this study only examined items that students felt they knew. In addition, since Daulton's study utilized a word family list and this study utilized a lemma list, comparing the results is difficult because of the nature of difference between the two. So, while Daulton's 45.2% refers to 1,356 instances of a loanword being present within one of 3,000 word families, these 3,000 word families themselves contain 16,458 separate types. So, if we look at it from this perspective, we would divide his 1,808 types found by 16,458, leaving us with a much lesser percentage of 10.9%. But this figure may also be unreliable since word families will include a separate entry for simple grammatical form changes, such as pluralizing a word or verb tense. If we are to make the realistic assumption that students can make the jump between the cognate *door* and *doors*, or between *stop* and *stopped*, then the total types would be significantly less than 16,458.

Next, the results of the vocabulary depth test were tallied. The results of what

How many high frequency words of English do Japanese university freshmen 'know' ?

percentage of questions students got correct on average can be seen in Table 1 below. By far, students were mostly familiar with semantic and orthographic knowledge. Students got on average 75.5% of the items correct when asked to produce the meaning of a word in the L2 when prompted with an L1 translation. Students received a similar score on average for orthographic knowledge (71.46%) when sensitive marking was employed, but somewhat less (57.97%) with a strict marking method. The next highest score tallied was for phonological knowledge (46.73%). Significantly lower than that was knowledge of affixes (24.71%) and grammar (12.9%). Aspects of vocabulary depth that students had little or no knowledge of were collocations (5.36%), concepts and referents (4.98%), associations (4.35%), and constraints on use (4.14%).

Table 1: Results of the vocabulary depth test

Semantics	Phonology	Orthography (strict)	Orthography (sensitive)	Collocations
75.5%	46.73%	57.97%	71.46%	5.36%
Word Parts	Grammar	Constraints on Use	Concepts and Referents	Associations
24.71%	12.9%	4.14%	4.98%	4.35%

Limitations

It should be noted that due to time constraints, this study only examined the highest level of vocabulary depth knowledge: productive knowledge. While the results are important, it remains to be seen what students' receptive depth knowledge of these words is. Thus, more research is needed to give a full picture of Japanese students' depth of knowledge of high-frequency vocabulary. In addition, the amount of questions each aspect of vocabulary depth was tested with were limited due to time constraints. Further testing would be necessary to truly judge students' knowledge since each one of these aspects would be worthy of a full research paper alone.

Discussion

Looking at the data, one may assume that knowing 1,903 of the 3,000 high-frequency

lemma of English constitutes a strong base knowledge of the language. But when we examine the data from the semantic section of the depth test, we see that students may be overestimating their knowledge to a degree. For example, one of the items that was marked as being 'known' by students was *county*. Generally speaking, this is not a word that is familiar to most university students in Japan, and it is easy to see how many students may mistakenly read it as *country*. However, semantic knowledge was still the students' strongest aspect of vocabulary depth.

The fact that 715 of the 1,903 items marked as being 'known' by students are loanwords derived from English is something that teachers and researchers should consider. Brown (1995), Oshima (2002), and Rebeck (2007) all believe that English loanwords in Japanese have the potential to aid the acquisition of English, but that this potential is not being fully exploited and that teachers should train students to get the most out of this built-in resource.

It is easy to understand why students had better scores on the orthography section versus the phonology section. With only approximately 90,000 native English speakers living in Japan (Ministry of Internal Affairs and Communications, 2010), opportunities to listen and speak with foreigners is far less than that of opportunities to read. Thus, Japanese students may not be getting the necessary input or chances to produce language orally in comparison with reading/writing. However, being that Japanese students study English for up to 6 years before they enter university, to get a score of only 57.97% in strict orthographic scoring and only 46.73% for phonological accuracy points to an issue that may need to be addressed in secondary schooling.

By far, one of the most problematic areas of knowledge for Japanese students is collocational knowledge. This aspect is especially problematic in that not only was students' scores very low (5.36%), but the sheer number of collocations that are needed to be learned for a student to sound native-like is staggering. Pawley and Syder (1983) believe native speakers have hundreds of thousands at their disposal. While such a number is significantly higher than the 20,000 high-frequency word families that Nation (2006b) believes that native speakers possess, other aspects of vocabulary depth knowledge pale in comparison with collocational knowledge quantifiably. Learning the sound system or spelling system of a language consists of much less information that needs to be mastered in comparison. The same can be said for a language's affix system, and the lesser known parts of speech words function as. In comparison to collocational knowledge, concepts and referents knowledge

and knowledge of constraints on use is much less numerous. Also, collocational knowledge is viewed by many researchers as central to language, and has been referred to as a “decisive factor in developing fluency” (Almela & Sanchez, 2007:37). The same cannot exactly be said for being able to produce native-like associations.

It was not surprising to see that students struggled with lesser known parts of speech, concepts and referents, associations, and constraints on use knowledge because these are high-order skills which will be developed as one approaches fluency. However, in the same vein as mentioned above regarding orthographic and phonological knowledge, word part knowledge may be yet another aspect of vocabulary depth that is not being fully addressing during secondary schooling.

Conclusion

This research project has shown that it is quite clear that many of the high frequency words of English are known to the students tested in this study, but that their depth of knowledge of these words is limited. What is seemingly a strong knowledge base (1,903/3,000 items) is actually a base with good breadth but not depth. This shallowness of vocabulary depth knowledge points to issues that need to be address in secondary schooling, but also issues that students will continue to need help developing throughout their college years and beyond.

The findings of this study should prove useful to help teachers in secondary schools in Japan focus more on areas which students struggle to obtain fluency in. The findings also give insight into areas which deserve additional emphasis in university, such as collocational knowledge. In addition, this study highlighted how teachers should consider ways in which they can better exploit the low learning burden of the many high frequency cognates Japanese shares with English, since a large proportion of them make up the high-frequency vocabulary of English.

References

- Aitchison, J. (2003). *Words in the Mind (3rd Edition)*. Oxford: Blackwell Publishing
- Almela, M. & Sanchez, A. (2007). Words as "lexical units" in learning/teaching vocabulary. *International Journal of English Studies* 7(2), 21-40.
- Bahns, J. & Eldaw, M. (1993). Should we teach ESL students collocations? *System*, 21(1), 101-114.
- Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279.
- Benesse Corporation (2004). *GTEC (Global Test of English Communication)*. Retrieved from <http://www.benesse.co.jp/gtec/english/index.html>.
- Brown, J. (1995). Is *Gairaigo* English?. *The Internet TESL Journal*, 1(2). Retrieved from <http://iteslj.org/Articles/Brown-Gairaigo/index.html>.
- Cobb, T. (n.d.). *Vocabprofile*. Retrieved from <http://www.lex tutor.ca/vp/bnc/>.
- Daulton, F. (2008). *Japan's built in lexicon of English-based loanwords*. Clevedon: Multilingual Matters Ltd.
- Daulton, F. (2005). Common *Gairaigo* corresponding to high-frequency and academic English – Are Japanese students ready for foreign study?. *JALT Hokkaido Journal*, 9, 1-16.
- Daulton, F.E. (2004). *Gairaigo – The built-in lexicon?* (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington.
- Daulton, F. (2003). The effect of Japanese loanwords on written English production - A pilot study. *JALT Hokkaido Journal*, 7, 3-12.
- Daulton, F. (1999). English words in Japanese: The built-in lexicon. *The Internet TESL Journal*, January, 5(1), 1-2.
- Daulton, F. (1998). Japanese loanword cognates and the acquisition of English vocabulary. *The Language Teacher*, January, 20(1), 17-25.
- Davies, M. (2008). *The Corpus of Contemporary American English: 425 million words, 1990-Present*. Retrieved from at <http://corpus.byu.edu/coca/>.
- Gabrys-Biskup, D. (1992). L1 influence on learners' renderings of English collocations. A Polish/German empirical study. In P.J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 85-93). London: Macmillan.
- Garcia, E. (2005). Levenshtein Edit Distance. Retrieved from <http://www.miislita.com/searchito/levenshtein-edit-distance.html>.
- Higa, M. (1973). Sociolinguistic aspects of word borrowing. *Topics in Culture Learning*, 1, 75-85.
- Hussein, R.F. (1990). Collocations: The missing link in vocabulary acquisition amongst EFL learners. In J. Fisiak (Ed.), *Papers and studies in contrastive linguistics, Volume 26. The Polish-English contrastive*

How many high frequency words of English do Japanese university freshmen 'know' ?

- project* (pp. 123-136). Poznan, Poland: Adam Mickiewicz University & Washington, DC: Center for Applied Linguistics.
- Ishikawa, K. & Rubrecht, B. (2008). English loanword use on Japanese television. In K. Bradford Watts, T. Muller, & M. Swanson (Eds.), *JALT conference proceedings*. Tokyo: JALT.
- Katz, L. & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Haskins Laboratories Status Report on Speech Research*, 111/112, 147-160.
- Kay, G. (1995). English loanwords in Japanese. *World Englishes*, 14(1), 67-76.
- Kilgarriff, A. (1995). *Lemmatized frequency list*. Retrieved from <http://www.kilgarriff.co.uk/bnc-readme.html>.
- Levenshtein, V. (1965). Binary codes capable of correcting deletions, insertions and reversals. *Doklady Akademii Nauk SSSR*, 163(4), 845-848.
- Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 8-10). Hove, England: Language Teaching Publications.
- Ministry of Internal Affairs and Communications (2010). Ministry of Justice's Annual Report of Statistics on Legal Migrants. Retrieved from <http://www.stat.go.jp/data/sekai/02.htm>.
- Mochizuki, M. & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 23(1), 291-304.
- Moss, H., & Older, L. (1996). *Birkbeck word association norms*. Hove: Psychology Press.
- Mulloni, A. & Pekar, V. (2006, May). Automatic detection of orthographic cues for cognate recognition. Paper presented at the fifth international conference on language resources and evaluation, Genoa, Italy. Retrieved from http://clg.wlv.ac.uk/~andrea/data/mulloni_pekar_06_OrthographicCues.pdf
- 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, B. Warvik, (Eds.), *Language, Learning, Literature: Studies Presented to Hakan Ringbom* (pp. 167-181). Abo Akademi University, Turku: *English Department Publications*.
- Nation, I.S.P. (2006a). Second language vocabulary. In K. Brown (Ed.) *Encyclopaedia of Language and Linguistics*, 2nd Ed. Oxford: Elsevier, 13, 448-454.
- Nation, I.S.P. (2006b). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I.S.P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, Heinle.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223-242.
- Orita, M. (1999). Word association patterns of Japanese novice EFL learners: a preliminary study. *Annual Review of English Learning and Teaching*, 4, 79-94.
- Oshima, K. (2002). Semantic and structural shift patterns of gairaigo in Japan. *Intercultural Communication Studies*, 11(4), 51-65.

- Ozaki, S. (2011). Teaching collocations effectively with the aid of L1. *The Language Teacher*, 35(3), 37-40.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.
- Parent, K. (2009). *Polysemy: A second language pedagogical concern* (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Quin, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56(2), 282-308
- Rebuck, M. (2007). Raising awareness of English loanwords in Japanese. *The Language Teacher, The Japan Association for Language Teaching*, 31, 24-26.
- Rebuck, M. (2002). The function of English loanwords in Japanese. *NUCB JLCC*, 4(1), 53-64.
- Ruhl, C. (1989). *On Monosemy*. New York: State University of New York Press.
- Schmitt, N. (2010). *Researching Vocabulary*. New York: Palgrave MacMillan.
- Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19, 17-36.
- Uchiwa, E. (2007). Oral and written identification of L2 loanword cognates by initial Japanese learners of English. *The Language Teacher*, 31(9), 19-22.
- Underwood, W. (1999). *Katakana Eigo-Real world list: A loanword approach to vocabulary building. Kurume kougyou daigaku, Kenkyuu Houkoku*, 23, 91-108.
- Van Benthuisen, R. (2004). Japanese efl students' awareness of English loanword origins. *Bunkyo gakuin daigaku, Gaigogakubu, Bunkyou gakuin tanki daigaku kiyou*, 4, 169-174.
- Van Orden, G.C. (1987). A rows is a rose: Spelling, sound, and reading. *Memory and Cognition*, 15(3), 181-198.

(James Martin Rogers 国際言語学部講師)