Englishes in Practice 2022; 5(1): 107–132
DOI 10.2478/eip-2022-0005

A Corpus Analysis of Loanword Effects on Second Language Production

Jonathan Ferries

Lancaster University, Department of Educational Research, United Kingdom
j.ferries@lancaster.ac.uk

ABSTRACT

Research suggests that English-derived loanwords in Japanese can affect Japanese learners' acquisition and receptive knowledge of their English words of origin ('basewords'). This study adopts a corpus-based approach to expand on this research by exploring the effects of loanwords on learners' productive knowledge. It primarily uses a corpus of written English produced by Japanese learners of English, a corpus of written English produced by native English speakers, and samples from a corpus of written Japanese to compare quantitatively how basewords and loanwords are used in each. The results provide statistically non-significant evidence that basewords are used relatively more frequently by learners than by native speakers, and some significant evidence that learners' baseword usage exhibits features of loanword usage where loanwords have changed in meaning or part of speech from their words of origin. The corpora also provide weak evidence that loanwords' effects on baseword usage increase with length of study of English. The findings point the way to more targeted use of loanwords in the classroom, including through the exploration of corpora by learners themselves.

多くの外来語に関する研究は、語彙習得及び受容知識が中心となっている。それに対してこの研究ではコーパスを使って学習者の産出知識に対する外来語の影響を探る。日本人学習者が作成した英語コーパス、英語母語話者が作成した英語コーパス、および日本語コーパスのサンプルを使用して、外来語とその由来語がそれぞれどの

ように使用されているかを定量分析する。結果として由来語が英語母語話者よりも学習者によって多く使用されているが、有意差はなかった。また、外来語の意味や品詞が変化した場合、学習者の由来語の使用法が有意に影響を受ける。さらに、外来語は英語の学習期間が長いほど影響するという有意でない傾向を示す。この調査結果は、学習者自身によるコーパスを用いた学習など、教育現場での外来語のより効果的な使用への道を提案する。

KEYWORDS: Loanwords, Gairaigo, Corpus linguistics, Vocabulary acquisition, Language transfer

1. INTRODUCTION

Vocabulary development is crucial to language learning. Both learners and teachers therefore have a natural interest in resources that may accelerate vocabulary acquisition, and cognates are regarded as one such resource. Cognates are words in two or more languages that are similar in either spoken or written form and meaning due to their common origin (Richards & Schmidt, 2013), as in the case of the word *hospital* in English and Spanish. Nation (2013) explains that learners find second language (L2) cognates easier to learn because their similarity to known first language (L1) words allows them to apply their knowledge of the connection between form and meaning in L1 to the corresponding connection in L2.

Similarities can also exist between L1 loanwords and their L2 words of origin ('basewords'). Despite being borrowed from another language rather than sharing a common origin, loanwords have consequently long been treated as a subset of cognates (see, for example, Lado, 1955). This is of relevance to Japanese L1 learners of English due to the ubiquity of English-derived loanwords in Japanese. Loanwords accounted for 24.7% of the word types used in mass-market magazines in 1994 (Itou, 2007), and Stanlaw (2004) estimated that 94.1% of the loanwords used in Japanese were of English origin. Opinion of their impact on learners has not been universally positive, however: while some view loanwords as a resource that students can exploit to accelerate vocabulary acquisition (Allen, 2019; Daulton, 2008), others have argued that they are a liability that can confuse learners and cause miscommunication (Sheperd, 1996).

Studies of the influence of loanwords on language learning have largely relied on evidence generated by controlled elicitation of single words and definitions to test recognition and recall of basewords (see, for example, the studies by Brown, 1995; Daulton, 1998; Inagawa, 2007; and Kimura, 2004, discussed in the literature review below). In contrast, the focus of the present study is on the effects of loanwords on learner language production under less controlled conditions. Using learner and native English speaker (NS) corpora in conjunction with Japanese corpus samples, the study explores whether basewords are used relatively more frequently by Japanese L1 learners than by NSs, and whether their baseword usage more closely resembles loanword usage in Japanese than NS usage. It also looks at whether the effects of loanwords on learners' production of English decline with length of study. The implications of the findings for

use of loanwords in English language teaching in Japan are then considered.

2. LITERATURE REVIEW

2.1 Loanwords: Resource or Liability?

An early study to demonstrate the positive effects of Japanese loanwords on English vocabulary development was that by Yoshida (1978), who found that loanword knowledge accelerated receptive vocabulary acquisition by a Japanese L1 child in an English as second language environment. Subsequent studies have pointed to the positive effects of loanwords among older Japanese L1 learners too. In small-scale studies of college-level English majors performing decontextualized written tasks, Brown (1995) and Kimura (2004) detected greater meaning recognition of basewords, and Daulton (1998) detected greater meaning recognition and form recall of basewords compared to non-basewords.

Arguing that these findings underline loanwords' value as a learning resource, Daulton (2003) has compiled lists of high-frequency words in English that correspond to loanwords in Japanese for classroom use. He proposes that by raising learners' awareness of known L1 loanwords, L2 basewords can be more readily acquired and the focus of vocabulary instruction instead placed on non-basewords. In contrast, Sheperd (1996) refers to loanwords as a "pitfall for all students" in the title of his paper, and Simon-Maeda (1995, Introduction section) writes that "misuse" of loanwords in Japanese leads to misuse of basewords in English. These dangers, they argue, stem from the changes in form and meaning that loanwords undergo when they are adopted into Japanese, and this stance is evident in writings for non-academic audiences too (see, for example, Kahl, 2015).

2.2 Types of Loanword Change

When English words are assimilated into Japanese to create loanwords, they undergo changes in form and meaning. The changes described in the literature (e.g., Inagawa, 2007; Irwin, 2011; Kay, 1995; Millar, 2006; Scherling, 2009; Stanlaw, 2004) are summarized below.

(1)	Orthographical:	Representation in the native *katakana* syllabary, which obscures the original spelling of basewords in English. Written in *katakana*, the word プロダクト (*purodakuto*), for example, provides fewer clues to the spelling of its English word of origin ('product') than do cognate pairs in languages written in similar scripts.
(2)	Phonological:	Adaptation to the Japanese sound system by, for example, inserting vowels to convert the consonant clusters found in English into the consonant-vowel patterns more typical of Japanese, as in the example above.

(3)	Morphological:	Clipping or blending with native Japanese words. The word *akuseru*, for example, is back-clipped from the English word *accelerator*, while *haburashi* ('toothbrush') has been produced by blending the Japanese word *ha* ('tooth') with the English word *brush*.
(4)	Semantic:	Narrowing, broadening, and complete shifts in meaning. Unlike the English word *veteran*, for example, the loanword *beteran* that derives from it denotes only experience and not ex-service personnel. It has thus undergone a narrowing of meaning. *Sain*, on the other hand, which derives from the English word *sign*, differs from the English in denoting an autograph as well as a signature. Its meaning has thus broadened. *Manshon*, meanwhile, has undergone a complete shift in meaning from the English word *mansion* to denote an apartment building rather than an imposing residence.
(5)	Grammatical:	Changes in part of speech. For example, the English noun *memo* has been verbalized by the addition of the suffix *-ru* to produce *memoru* ('to take notes').
(6)	Coinages:	Entirely new words generated by combining English-derived elements. The word *sarariiman* ('office worker'), for example, has been coined from the English words *salary* and *man*.

English words often undergo a combination of these changes when borrowed into Japanese. *Haburashi* ('toothbrush'), for example, exhibits three changes: orthographical (it is written in *katakana*), phonological (vowels have been inserted), and morphological (it blends a native Japanese word with an English-derived word).

Recognizing that different loanwords undergo different changes when they enter Japanese, Inagawa (2007) and Millar (2006) have explored whether these differences affect learners' use and recognition of basewords in English. Using a sentence correction exercise administered to high school and university students, Inagawa detected negative effects in the form of 'erroneous' word choices and spellings, and found that grammatical changes created the most difficulties for learners. Inagawa also found that second-year high school students produced more errors than first-year high school students in all categories except minimally changed loanwords, suggesting that length of study of English does not necessarily correlate with fewer errors.

Millar (2006), on the other hand, used a learner corpus – the 86,000-word Japanese L1 learner subcorpus of the first edition of the International Corpus of Learner English (ICLE) – to compare Japanese L1 learners' usage of English basewords with their usage by NSs in a 1,230,000-word reference corpus of written British English from newspapers and other publications. His analysis produced some evidence that basewords are used relatively more frequently by learners than by NSs, and that learners' usage of basewords corresponding to loanwords that have undergone semantic or grammatical change exhibits features of L1 loanword usage. Millar's corpus-based study is

significant because it analyzes the language produced by learners under relatively natural conditions.

2.3 Baseword Knowledge Testing and Use of Corpora

Decontextualized tasks of the kind used by all of the above studies except Millar's are normally used to investigate learners' receptive knowledge. However, this represents only one aspect of the effects of loanwords on learners. For example, Yoshida (1978) found that learners could not always use basewords successfully in spoken output, even though they had been shown to have meaning-recognition knowledge of those basewords.

Learner corpora of the kind used by Millar (2006), on the other hand, consist of samples of running text produced by learners under less controlled conditions. By comparing them with corpora of texts produced by NSs, they can be used to identify distinctive features of learner usage (e.g., relative frequency of use of L2 basewords). They can thus be used to identity differences between English L1 and non-L1 usage without making prescriptive judgments as to their supposed 'erroneousness' from the point of view of NS norms.

For meaningful comparisons to be made, the learner and NS corpora should contain texts concerning a similar subject matter. As Millar noted, however, the learner corpus (argumentative essays produced by students on a narrow range of topics) and NS corpus (published materials on a wide range of topics) that he used differed in both genre and topic. He consequently could not exclude the possibility that it was these differences, rather than the effects of L1 loanword knowledge, that gave rise to the differences in baseword usage. Other limitations of Millar's study include the small size of the learner corpus and lack of a Japanese corpus to provide representative examples of L1 loanword.

2.4 Loanword Knowledge

The above studies, including Millar's, all assume that loanwords will be known to learners. Logically, L1 loanwords can only affect L2 baseword acquisition and use if those loanwords are known to learners. However, despite the large number of loanwords that appear in publications, it does not follow that most Japanese L1 speakers understand and use them. The difficulties posed by some loanwords for Japanese L1 speakers are confirmed by a series of surveys conducted by the National Institute for Japanese Language and Linguistics (2004b). To take as an example the loanwords derived from two of the words on a list of high-frequency loanwords compiled by Daulton (2003) based on the 3,000 most frequent word families in the British National Corpus (BNC) (Nation, 2017) – *stress* and *organizer* – these surveys show that while *sutoresu* ('stress') was recognized by 97.4% of the respondents, *ooganaizaa* ('organizer') was recognized by only 28.7%.

Loanword prevalence is also subject to change over time: some loanwords become fixtures of the language that are likely to be known to most Japanese L1 speakers, while others fall out of use after a brief period of popularity (Inoue, 2004). Thus, basewords may correspond to loanwords known to learners in one cohort, but not

another.

However, learners' familiarity with L1 loanwords is not reproducibly reflected in the sampling of words used in the studies above. For example, Brown's (1995) words were selected by an undisclosed method, Daulton (1998) selected words based on their level of difficulty for learners in L2, Kimura (2004) used a random sample of common loanwords from a dictionary, Millar's (2006) study was informed by "loanword dictionaries, consultation with native speakers and the Internet" (p. 484), and Inagawa (2007) selected words that "should be familiar" to learners (p. 89).

3. RESEARCH QUESTIONS

Still very little research into the effects of L1 loanword knowledge on L2 production has been undertaken. Adopting a corpus-based approach, this study aims to build upon the work of Inagawa (2007) and Millar (2006), and is guided by the following three research questions.

(1) Do the English basewords of Japanese loanwords occur relatively more frequently in the writing of Japanese L1 learners than that of NSs?
(2) Is Japanese L1 learners' usage of basewords in English affected by usage of the corresponding loanwords in Japanese?
(3) Do the effects of loanword knowledge on baseword usage decline with length of study of English?

4. METHOD

4.1 Materials

The study employed a learner corpus of written English produced by Japanese L1 learners, a NS corpus of written English produced by English L1 speakers, and samples from a corpus of written Japanese produced by Japanese L1 speakers. Supplementary use was also made of a corpus of written English produced by Chinese learners of English. The weaknesses identified by Millar regarding his own study were addressed by selecting learner and NS corpora from the same genre (argumentative essays on similar topics). The corpora were selected to be as contemporaneous as possible and produced by writers of similar ages and educational levels. This was to reduce the likelihood of differences between the corpora being due to diachronic change or writer profile.

In order to provide a more reliable means of selecting L2 basewords corresponding to known L1 loanwords than researcher intuition or consultation of dictionaries, baseword analysis was based on a survey of 3,000 Japanese L1 speakers' use and knowledge of 405 loanwords conducted by the National Institute for Japanese Language and Linguistics (2004b). Survey participants were interviewed individually by researchers and asked (a) whether they had ever heard or seen the words surveyed; (b) if so, whether they understood these words; and (c) if so, whether they had ever used these words. The words surveyed were presented in written form employing the *katakana* syllabary typically used to represent loanwords in Japanese, making it extremely likely

that even unknown words would be recognized by participants as loanwords (National Institute for Japanese Language and Linguistics, 2004a). To provide a manageable number of loanwords for analysis, the 15 most commonly used loanwords (used by at least 74% of the survey participants) were identified. One word (*rifuresshu*) was excluded because its baseword ('refresh') appeared in neither the learner nor the NS corpus, preventing comparative analysis of usage. This left a total of 14 highly commonly used basewords (see Appendix B).

4.1.1 Learner Corpus

The Japanese subcorpus (ICLEv2-jp) of version 2 of the International Corpus of Learner English (ICLEv2) was used as the learner corpus. This contains 198,241 words and consists of argumentative essays on topics including sport, crime, culture, technology, education, and the environment. The essays were written by 366 Japanese L1 students, aged between 18 and 26, at 21 universities in Japan, and were compiled in the 1990s and 2000s (Granger et al., 2009). The contributing universities consisted of both public and private universities of varying sizes from all over Japan. However, the majority (91%) of the essays were contributed by five institutions in Tokyo: two large, high-ranking private universities, Keio University (13.9%) and Waseda University (24.6%); two medium-sized, middle-ranking, humanities-focused private universities, Showa Women's University (33.3%) and Tamagawa University (7.7%); and one medium-sized, middle-ranking foreign languages-focused university, Tokyo University of Foreign Studies (11.7%). Most essays were produced under timed conditions (56.0%) without access to reference tools (82.%). All students were studying English as part of their courses of study at university. However, ICLEv2 provides no further information regarding students' majors, the courses for which essays were written, how many courses or hours of instruction students had received, whether and how much guidance was provided before completing the writing task, and whether feedback was provided during the writing process.

As essays can be selected from ICLEv2-jp based on selected learner variables, two subcorpora were extracted based on length of study of English: a 43,136-word '1-year' subcorpus of learners who had studied English for one year at university, and a 125,693-word '3-year' corpus of learners who had studied English for three years at university. It should also be noted that all learners would have studied English for at least six years in the course of their secondary education. The decision to use 1-year and 3-year subcorpora was made for two reasons: firstly, English produced by learners who had studied English for discrete lengths of time was sought to explore whether the effects of loanword knowledge on baseword usage decline with length of study of English; and secondly, use of subcorpora of English produced by learners who had studied English for two, four, and five years at university was ruled out because these subcorpora were too small (respectively 14,912 words, 12,845, and 1,655 words) to provide sufficient examples of usage for comparison with the NS corpus described below.

Supplementary use was made of the mainland Chinese learner subcorpus of ICLEv2. This contains 93,089 words and consists of 179 argumentative essays on a

range of topics written by students who were studying English as part of their university courses of study. The purpose of using this subcorpus was to test whether differences in frequency of use of basewords between the Japanese L1 learners and NSs could be observed among learners of an L1 that contains fewer English-derived loanwords than Japanese, the assumption being that these learners would use the basewords less frequently than the Japanese L1 learners if the Japanese L1 learners were using basewords due to their knowledge of loanwords. Chinese learners of English were selected because Mandarin Chinese contains the lowest percentage of loanwords of the 41 languages contained in the Max Planck Institute's World Loanword Database (Haspelmath & Tadmor, 2009).

4.1.2 NS Corpus

The American argumentative essays subcorpus (LOCNESS-usarg) of the Louvain Corpus of Native English Essays (LOCNESS) (Centre for English Corpus Linguistics, n.d.) was used as the NS corpus. This was chosen because of its similarity to the learner corpus in terms of genre, writer age, and contemporaneity. It contains 149,574 words and consists of argumentative essays on topics including sport, crime, and culture. The essays were written in 1995 by 176 English L1 students, aged mostly between 17 and 23, at seven universities in the United States.

4.1.3 Japanese Corpus Sample

The Japanese corpus sample was derived from the 10.3-million word Yahoo! Answers subcorpus and the 10.2-million Yahoo! Blog subcorpus of the Balanced Corpus of Contemporary Written Japanese (National Institute for Japanese Language and Linguistics, 2009), known as the BCCWJ, and consisted of the first 100 concordance lines (search items with co-text) produced by searches of these subcorpora for each of the loanwords corresponding to the basewords analyzed in the learner and NS corpora. The 100-line limitation was imposed to limit the sample to a size that could be manually analyzed given the research resources available, while still providing extensive non-researcher-selected examples of real-world usage.

These two subcorpora were chosen from among the several BCCWJ subcorpora available because they were considered to provide the most contemporaneous snapshot of L1 loanword use in non-fiction writing on a range of topics commonly engaged in by writers with similar attributes to the writers of the ICLEv2-jp. Contemporaneity is important because, as the National Institute for Japanese Language and Linguistics (2004b) and Inoue (2004) show, knowledge of loanwords can vary according to age group and over time. The texts in the Yahoo! Answers and Yahoo! Blog subcorpora respectively date from 2005 and 2008, and so are relatively contemporaneous with the learner and NS corpora. Writers were also judged likely to approximate the age profile learner corpus, because, according to Japan's Ministry of Internal Affairs and Communications (2005), the majority of Internet users in Japan were aged under 40 in 2004.

4.2 Selection of Words for Analysis

4.2.1 Basewords

The following 14 basewords were selected for analysis (see Appendix B for their corresponding loanwords and details of frequency of use).

stress**	privacy**	recycle	trouble*	volunteer***	rule*	cancel***
theme***	restructure	sample***	pride**	merit	message**	happy*

These correspond to the 14 loanwords selected as described in Section 4.1. Most of the basewords are also frequently used in English: 3 are in the 1,000 most frequent word families (indicated by one asterisk above), 4 are in the 1,001-2001 most frequent word families (indicated by two asterisks above), and 4 are in the 2,001-3,001 most frequent word families (indicated by three asterisks above) in the BNC (Nation, 2017).

4.2.2 Non-basewords

The following 14 high-frequency L2 non-basewords were selected for comparison with the basewords.

across	concern	further	glad	instead	into	laugh
ought	particular	quiet	rid	shall	usual	wrong

These were selected from among the 1,000 most frequent word families in the BNC (Nation, 2017) that could not be found to have loanword equivalents at the time that the corpora were compiled. They were selected after checking Ishiwata's (1998) dictionary of loanwords to ensure that they did not have loanword equivalents in Japanese (i.e., could be confirmed to be unknown as loanwords to learners) at the time that the learner corpus was created. Words were selected from the highest frequency BNC word list to address the possibility that learners might use basewords more because they were high-frequency words, rather than because they were basewords.

4.3 Procedure

4.3.1 Tagging and Lemmatization

The NS corpus was tagged with the CLAWS7 tagset using the CLAWS part-of-speech tagger for English (Lancaster University Centre for Computer Research on Language, n.d.). The part-of-speech functions of loanwords in the Japanese corpus sample were determined by visual inspection of the 100 concordance lines for each word. Word classes in Japanese were grouped by function; verbal nouns functioning as verbs, for example, were treated as verbs in this study. The learner corpus/subcorpora and NS corpus were in addition lemmatized by importing Someya's unhyphenated English lemma list (Anthony, 2018) into AntConc, a corpus analysis tool (Anthony, 2016). The purpose of this was to combine all inflected forms of the words studied in order to calculate single measures of frequency and keyness for the canonical forms (shown capitalized in this paper).

4.3.2 Keyness Analysis

The lemmatized corpora and subcorpora were imported into AntConc to determine the keyness of the 14 basewords and 14 non-basewords in the learner corpus/subcorpora relative to the NS corpus. 'Keyness' expresses the degree to which items occur significantly more or less frequently (i.e., are more or less salient) in one corpus compared to another, and is typically determined by using log-likelihood or chi-squared tests (Baker, 2010). Here, log-likelihood was used as research suggests that chi-squared is less reliable at the relatively low frequency levels found in the corpora selected for this study (e.g., McEnery & Hardie, 2011).

Log-likelihood was calculated automatically in AntConc by importing the learner corpus/subcorpora and assessing them against a lemmatized keyword list generated from the NS corpus. A log-likelihood score of 3.8 or higher would indicate that the difference between the corpora is significant at $p < 0.05$ (Rayson, 2015). To guard against the risk of family-wise error, however, a Bonferroni-adjusted significance threshold of $p < 0.004$ (0.05 / 14) was employed. For the keyness analysis, therefore, a log-likelihood score of 8.3 or higher is required to demonstrate statistical significance.

4.3.3 Usage Analysis

Usage of the four basewords found to occur most frequently in the learner corpus – STRESS, TROUBLE, RULE, and HAPPY (see Table C.1) – was then compared with usage in the NS corpus and usage of the corresponding loanwords in the Japanese corpus sample. The four most frequently occurring words were selected in order to provide the greatest possible number of examples of usage to compare with the NS corpus and explore whether learner usage in L2 was being affected by L1 loanword usage. Usage was analyzed from the angle of similarities and differences between the corpora in (1) part of speech, (2) meaning, and (3) the most frequently occurring groups of words occurring up to two words away from the baseword being searched for in the corpus (referred to below as 'co-occurrences).

Part-of-speech counts were performed automatically on the learner and NS corpora in AntConc, and manually on the Japanese corpus sample. Meaning-based counts were performed manually based on an analysis of concordances from each corpus. Co-occurrences were counted automatically in the learner and NS corpora using AntConc, and manually in the Japanese corpus sample.

For the meaning-based counts, the four most frequently occurring basewords were categorized into the following senses:

(1) STRESS: *strain* (mental or physical) and *emphasis* (prominence or focus)
(2) TROUBLE: *problem* (a particular difficulty or malfunction), *situation* (a general disagreement, conflict, or liability to be punished), *inconvenience* (caused to someone), and *effort* (made to achieve something)
(3) RULE: *regulation* (a convention or instruction governing what is allowed to be done), *decision* (an official finding), *control* (the exercise of power or authority), *norm* (a normal state of affairs), and *preclusion* (in the idiomatic usage of ruling something out)

(4) HAPPY: *content* (marked by pleasure or contentment)

The results of the part-of-speech and meaning-based counts were tested for significance as follows. Observed frequencies of each baseword in each category of part of speech or meaning and the total number of occurrences of each word were entered in the UCREL log-likelihood wizard (Rayson, 2015) to produce log-likelihood scores for each word. Bonferroni-corrected thresholds determined according to the number of pairs tested were adopted to judge the statistical significance of these scores (see Tables 3 to 8).

5. RESULTS

5.1 Keyness

As Table 1 shows, all 14 basewords exhibited positive keyness in the Japanese L1 learner corpus, but none were significantly key at the 0.05 level after Bonferroni correction. Thirteen of the 14 basewords exhibited positive keyness in the 1-year and 3-year subcorpora, with one baseword being absent from both, but none were significant after Bonferroni correction. There was also found to be no clear upward or downward trend in keyness according to length of study of English at university, with seven basewords being more key and six less key in the 3-year subcorpus than the 1-year subcorpus.

In the Chinese learner corpus, 11 of the 14 basewords were positively key and three were absent. As in the Japanese L1 learner corpus, none were found to be significantly key after Bonferroni correction.

Table 1

Lemmatized keyness values for basewords

Baseword lemma	Keyness (log-likelihood)			
	Learner corpus	1-year subcorpus	3-year subcorpus	Chinese corpus
STRESS	2.346	2.317	2.525	1.886
PRIVACY	1.354	0.772	1.823	0.189
RECYCLE	2.256	0.386	3.366	-
TROUBLE	4.061	4.635	4.067	2.452
VOLUNTEER	1.715	0.386	2.525	0.943
RULE	6.497	2.317	7.013	5.471
CANCEL	0.180	0.386	0.140	0.189
THEME	0.722	1.545	0.281	-
RESTRUCTURE	0.180	0.386	0.140	-
SAMPLE	0.090	-	-	0.189

PRIDE	0.722	0.772	0.561	0.377
MERIT	1.444	0.386	1.823	0.755
MESSAGE	1.895	2.704	0.982	0.943
HAPPY	6.407	2.704	6.872	3.018

An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 8.3 or higher is significant after Bonferroni correction.

As can be seen from Table 2, all 14 non-basewords exhibited positive keyness in the Japanese L1 learner corpus and both the 1-year and 3-year subcorpora. In all three, however, only one non-baseword (INTO) was found to be significant at the 0.05 level after Bonferroni correction. Keyness did not change markedly overall according to length of study of English, with seven non-basewords being more key and seven less key in the 3-year subcorpus than the 1-year subcorpus.

In the Chinese learner corpus, 13 of the 14 non-basewords were positively key and one was absent. Two of the non-basewords (INSTEAD and INTO) were found to be significantly key (respectively 8.300 and 21.694) after Bonferroni correction.

Table 2

Lemmatized keyness values for non-basewords

Non-baseword lemma	Keyness (log-likelihood)			
	Learner corpus	1-year subcorpus	3-year subcorpus	Chinese corpus
ACROSS	1.263	1.545	1.262	0.566
CONCERN	3.790	3.476	3.506	5.282
FURTHER	0.993	0.386	0.842	2.264
GLAD	0.902	1.159	0.982	-
INSTEAD	2.436	2.317	2.104	8.300
INTO	14.348	13.518	14.867	21.694
LAUGH	0.993	1.159	0.982	1.320
OUGHT	0.722	0.386	0.842	0.943
PARTICULAR	0.902	0.386	1.122	2.264
QUIET	0.722	0.772	0.842	1.698
RID	0.632	0.772	0.701	3.207
SHALL	0.271	0.386	0.281	0.566
USUAL	0.902	0.386	0.842	0.566
WRONG	6.407	7.725	5.049	5.848

An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 8.3 or higher is significant after Bonferroni correction.

5.2 Usage of STRESS, TROUBLE, RULE, and HAPPY

5.2.1 STRESS

Part-of-speech uses

While STRESS was used slightly more frequently as a verb than as a noun by the NSs, the learners used STRESS more as a noun (Table 3). STRESS was used significantly less frequently as a verb in the learner corpus (6.23) and the 3-year corpus (7.83) than in the NS corpus, and this tendency to use STRESS as a noun rather than as a verb closely resembled use of the corresponding loanword *sutoresu* in the Japanese corpus sample.

Table 3

Observed frequencies of STRESS by part of speech

Part of speech	NS corpus	Learner corpus	1-year subcorpus	3-year subcorpus	Japanese corpus sample
Noun	7 (41.2%)	15 (57.7%, 0.56)	4 (66.7%, 0.56)	17 (94.4%, 3.74)	100.0%
Verb	9 (52.9%)	3 (11.6%, 6.23)	2 (33.3%, 0.38)	1 (5.6%, 7.83)	0.0%

Percentage frequencies and LL scores are shown in parentheses. An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 5.0 or higher is significant after Bonferroni correction.

Semantic uses

STRESS was used approximately equally frequently by the NSs in the sense of strain and in the sense of emphasis (Table 4). In the learner corpus and both subcorpora, however, two thirds of occurrences signified strain. There was no difference according to length of study of English, and the differences between the NS corpus and the learner corpus/subcorpora were not significant (all LL scores were below 5.0). In the Japanese corpus sample, *sutoresu* was used solely in the sense of strain.

Table 4

Observed frequencies of STRESS by meaning

Category of meaning	NS corpus	Learner corpus	1-year subcorpus	3-year subcorpus	Japanese corpus sample
Strain	9 (52.9%)	18 (69.2%, 0.44)	4 (66.7%, 0.14)	12 (66.7%, 0.28)	100.0%
Emphasis	8 (47.1%)	8 (30.8%, 0.72)	2 (33.3%, 0.20)	6 (33.3%, 0.41)	0.0%

Percentage frequencies and LL scores are shown in parentheses. An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 5.0 or higher is significant after Bonferroni correction.

Co-occurrences

The most frequent co-occurrence containing STRESS in the NS corpus was *stress that*, where STRESS colligated either as a verb with a *that*-clause or as a noun introducing a relative clause. In the learner corpus, STRESS appeared most frequently as a noun with an *on*-prepositional phrase. Thus, even where STRESS was used in the sense of emphasis, the learners showed a preference for colligates that allowed STRESS to function as a noun.

Semantically, users in both the learner corpus and the Japanese corpus sample showed a preference for using STRESS/*sutoresu* with words that express the accumulation or release of a physical object (e.g., "she collect various stress" and "they seem to release their stress" in the learner corpus). Where STRESS was used to signify strain in the NS corpus, on the other hand, stress was represented as a state that arose or was entered (e.g., "the stress that comes about" in the NS corpus).

5.2.2 TROUBLE

Part-of-speech uses

TROUBLE was most commonly used as a noun in both the NS corpus and the learner corpus and subcorpora (Table 5). Learners differed from NSs in their more frequent use of TROUBLE as a verb (significantly so in the 1-year corpus) and less as an adjective (significantly so in the 3-year corpus). Within the learner corpus, the 3-year subcorpus most closely resembled the Japanese corpus sample in all three part-of-speech categories.

Table 5

Observed frequencies of TROUBLE by part of speech

Part of speech	NS corpus	Learner corpus	1-year subcorpus	3-year subcorpus	Japanese corpus sample
Noun	8 (72.7%)	33 (75.0%, 0.01)	6 (50.0%, 0.49)	24 (82.8%, 0.10)	100.0%
Verb	0 (0.0%)	10 (22.7%, 4.46)	5 (41.7%, 6.51)	5 (17.2%, 3.22)	0.0%
Adjective	3 (27.3%)	1 (2.3%, 5.60)	1 (8.3%, 1.23)	0 (0.0%, 7.75)	0.0

Percentage frequencies and LL scores are shown in parentheses. An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 5.7 or higher is significant after Bonferroni correction.

Semantic uses

Occurrences of TROUBLE/*toraburu* as problem were proportionately more frequent in both the learner corpus and the Japanese corpus sample than the NS corpus, with the 3-year subcorpus most closely resembling the Japanese sample in this regard (Table 6). Compared to the NS corpus, TROUBLE was used significantly less frequently to signify a situation in the learner corpus (6.26) and the 3-year corpus (6.55), and significantly more frequently to signify a problem in the 3-year corpus (6.32).

Table 6
Observed frequencies of TROUBLE by meaning

Category of meaning	NS corpus	Learner corpus	1-year subcorpus	3-year subcorpus	Japanese corpus sample
Problem	1 (9.1%)	22 (50.0%, 4.81)	3 (25.0%, 0.88)	18 (62.1%, 6.32)	67.0%
Situation	8 (72.7%)	9 (20.5%, 6.26)	2 (16.7%, 4.40)	5 (17.2%, 6.55)	33.3%
Inconvenience	1 (9.1%)	10 (22.7%, 0.98)	6 (50.0%, 3.54)	4 (13.8%, 0.15)	0.0%
Effort	1 (9.1%)	3 (6.8%, 0.06)	1 (8.3%, 0.00)	2 (6.9%, 0.05)	0.0%

Percentage frequencies and LL scores are shown in parentheses. An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 6.2 or higher is significant after Bonferroni correction.

Co-occurrences

No co-occurrences with TROUBLE occurred more than once in the NS sample. In the learner corpus, the commonest co-occurrence containing TROUBLE was the prepositional phrase *in trouble*, which occurred seven times. All seven occurrences were in the 3-year subcorpus. However, whereas *in trouble* is typically used by NSs to describe trouble with authority or being in a generally unfortunate situation (e.g., "how to recognize when somebody is in trouble and how to handle that situation" in the NS corpus), it was more frequently used by the learners in the 3-year subcorpus to describe a particular problem or difficulty (e.g., "foreign visitors, when they had in trouble and ask Japanese") in a manner that might be considered unusual by NSs.

5.2.3 RULE

Part-of-speech uses

Noun uses predominated by large margins in all the corpora (Table 7). The proportions of noun use and verb use were almost identical in the NS corpus, learner corpus, and 1-year and 3-year subcorpora, although instances of verb use were absent from the Japanese corpus sample. None of the differences between the NS corpus and the learner corpus/subcorpora were significant (all LL scores were below 5.7).

Table 7
Observed frequencies of RULE by part of speech

Part of speech	NS corpus	Learner corpus	1-year subcorpus	3-year subcorpus	Japanese corpus sample
Noun	62 (84.9%)	61 (84.7%, 0.00)	5 (83.3%, 0.00)	43 (86.0%, 0.00)	100.0%
Verb	10 (13.7%)	10 (13.9%, 0.00)	1 (16.7%, 0.03)	7 (14.0%, 0.00)	0.0%

Adjective	1 (1.4%)	1 (1.4%, 0.00)	0 (0.0%, 0.16)	0 (0.0%, 1.04)	0.0%

Percentage frequencies and LL scores are shown in parentheses. An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 5.7 or higher is significant after Bonferroni correction.

Semantic uses

RULE as regulation accounted for the majority of occurrences in all the corpora (Table 8). Despite accounting for almost one in six occurrences in the NS corpus, however, RULE as decision was absent from the learner corpus and both subcorpora, and the difference from the NS corpus was significant in the case of the learner corpus (15.10) and the 3-year subcorpus (11.48). Overall, RULE was used in a narrower range of meanings in the learner corpus and Japanese corpus sample than the NS corpus, and the 3-year subcorpus exhibited the smallest range of meanings and greatest resemblance to the Japanese corpus sample.

Table 8

Observed frequencies of RULE by meaning

Category of meaning	NS corpus	Learner corpus	1-year subcorpus	3-year subcorpus	Japanese corpus sample
Regulation	55 (75.3%)	58 (80.6%, 0.13)	4 (66.7%, 0.06)	43 (86.0%, 0.42)	100.0%
Decision	11 (15.1%)	0 (0.0%, 15.10)	0 (0.0%, 1.74%)	0 (0.0%, 11.48)	0.0%
Control	5 (6.8%)	12 (16.7%, 3.07)	1 (16.7%, 0.54)	7 (14.0%, 1.52)	0.0%
Norm	1 (1.3%)	2 (2.8%, 0.35)	1 (16.7%, 2.54)	0 (0.0%, 1.04)	0.0%
Preclusion	1 (1.3%)	0 (0.0%, 1.37)	0 (0.0%, 0.16)	0 (0.0%, 1.04)	0.0%

Percentage frequencies and LL scores are shown in parentheses. An LL of 3.8 or higher is significant at $p < 0.05$. An LL of 6.6 or higher is significant after Bonferroni correction.

Co-occurrences

The two commonest co-occurrences containing RULE in the NS corpus were *follow the rules* and *rules and regulations*, which both occurred seven times (9% of occurrences each). These were respectively absent from and occurred only once (1%) in the learner corpus. While three of the four commonest content word collocates in the NS corpus had strong authority-related or legal connotations (*court, supreme, regulations*), those in the learner corpus and both subcorpora had educational connotations (*grammar, children,* and *school*). However, this collocational tendency was not apparent in the Japanese corpus sample. Instead, the corresponding loanword *ruuru* was found to co-occur most frequently with words having sporting and driving connotations. (The wide differences between the corpora appear to be attributable in part to the higher proportion

of essays on educational themes in the learner corpus and the leisure interests of the writers in the Japanese corpus sample.)

5.2.4 Happy

Part-of-speech uses

All occurrences of HAPPY in both the NS corpus and all but one in the learner corpus were adjectival. (The one exception was a nominal occurrence in "this development brought the people many happy, on the other hand it brought misfortune," in the 1-year corpus, in a manner that NSs would typically consider erroneous.)

Predicate occurrences made up 82% of the total in the NS corpus, 76% in the learner corpus, 43% in the 1-year subcorpus, 78% in the 3-year subcorpus, and 6% in the Japanese corpus sample. Usage in both the NS and learner corpora thus differed markedly from the Japanese corpus sample, and the 3-year subcorpus most closely resembled the NS corpus.

Semantic uses

There were no clear semantic differences in usage between the corpora, and all occurrences signified contentment.

Co-occurrences

Colligationally, the NSs showed a preference for using HAPPY with the preposition *with* to signify contentment in the sense of satisfaction with a situation (e.g., "some women were happy with old traditions"). In the learner corpus and both subcorpora, on the other hand, HAPPY occurred most frequently with *to*, either as an infinitive marker or as a preposition in a manner suggestive of language transfer that NSs would typically consider erroneous (e.g., "school life was happy to me" in the learner corpus)

This colligational tendency was not apparent in the Japanese corpus sample; instead, the corresponding loanword *happii* frequently co-occurred attributively with other loanwords, often in fixed expressions borrowed or adapted from English.

6. DISCUSSION

6.1 Do English Basewords of Japanese Loanwords Occur Relatively More Frequently in Japanese L1 Learner Writing Than NS Writing?

Echoing Millar (2006), the keyness analysis provides some evidence that English basewords of Japanese loanwords are used relatively more frequently by learners than by NSs. However, although the choice of corpora for this study makes it less likely that this salience is attributable to genre or topic differences than in Millar, the results must be qualified in two regards: firstly, while all 14 basewords exhibited positive keyness in the learner corpus, none were found to be significant at the 0.05 level after Bonferroni correction (see Table 1); and secondly, the results for basewords in the Chinese learner corpus combined with the results for non-basewords in both the Japanese and Chinese learner corpora suggest that the basewords' greater salience cannot be attributed wholly

(or possibly even partly) to the effect of loanword knowledge.

Regarding the second qualification, the finding that 12 of the 14 basewords were positively key in the Chinese corpus and all 14 non-basewords exhibited positive keyness in both the learner and the Chinese corpora (significantly so in one case in the learner corpus and subcorpora and in two cases in the Chinese corpus after Bonferroni correction) indicates that both basewords and non-basewords were used relatively more frequently by learners than NSs. As all 14 of the non-basewords are among the 1,000 most frequent word families in the BNC and 11 of the 14 basewords are among the top 3,000 (Nation, 2017) (see Appendix B), this raises the possibility that learners were using basewords relatively more frequently not as a result of a loanword effect, but because they were overusing all high-frequency vocabulary items – basewords and non-basewords alike – compared to NSs. This may have occurred because, having smaller active vocabularies than the NSs and not having access to reference resources (300 of the 366 Japanese L1 learner texts were written without reference tools), learners drew on the words that they knew more frequently.

6.2 Is Japanese L1 learners' Usage of Basewords in English Affected by Usage of the Corresponding Loanwords in Japanese?

Usage of three of the four basewords – STRESS, TROUBLE, and RULE – exhibited one or more features associated with the Japanese corpus sample than the NS corpus. All three correspond to loanwords that have undergone semantic narrowing, as evidenced by the examples in the Japanese corpus sample, where *sutoresu* is used only in the sense of strain, *toraburu* largely in the sense of problem, and *ruuru* only in the sense of regulation.

Although these three basewords have also undergone a certain degree of grammatical narrowing in that their verbal and adjectival functions have been lost and only their noun functions retained in Japanese (Tables 3, 5, & 7), only STRESS was used markedly more as a noun by the learners than the NSs.

Usage of HAPPY, on the other hand, showed no signs of being affected by usage of *happii* in Japanese. Moreover, where colligations that deviated from NS norms did occur, they appeared to be influenced more by non-loanword language transfer. The construction "school life was happy to me" in the learner corpus, for example, shows signs of the learner choosing the English preposition *to* as a result of assuming an equivalence with the Japanese particle *ni*, which would be used in the corresponding L1 construction irrespective of loanword use, because *ni* functions as a dative of direction as well as a dative of reference in Japanese (Martin, 1988).

As the grammatical and semantic similarities noted in Section 5.2.4 suggest, HAPPY has undergone comparatively little change in the process of adoption into Japanese. HAPPY thus differs from the other three basewords in that its corresponding loanword is a straightforward borrowing, while the other three have all undergone semantic narrowing.

STRESS, TROUBLE, and RULE share a further similarity in that their corresponding loanwords – *sutoresu, toraburu, and ruuru* – are among the 2,500 most frequently used words in spoken and written Japanese (see Appendix B). *Happii*, on

the other hand, is not among the top 5,000, despite a very high proportion of Japanese speakers reporting that they had used the word themselves (National Institute for Japanese Language and Linguistics, 2004b). As the occurrences in the Japanese corpus sample show, however, it tends to occur in a limited number of familiar but fixed expressions, such as *happii endo* ('happy ending'), that possibly provide narrower scope for usage of HAPPY in English to be affected by how *happii* is used in Japanese.

The implication of these findings is that the effect of loanword knowledge on learners depends on the degree of loanword change and frequency of use: more modified (especially semantically narrowed) loanwords have a greater impact on learners' production of English than less modified (straightforward) loanwords; and high-frequency loanwords have a greater impact than low-frequency loanwords.

Although the literature has little to say on the relevance of L1 frequency of use of loanwords, the finding that semantically narrowed loanwords in particular affect learners' production of English is consistent with Millar (2006), who suggested that learners' use of basewords tends to be restricted, semantically, to their meanings in Japanese, and Inagawa (2007), who observed that learners produced more errors on tests of basewords corresponding to semantically modified loanwords (Inagawa, 2007).

This finding is also consistent with Ringbom (2007), who theorized that apparent similarities of form and meaning between L1 and L2 vocabulary items encourage learners to assume an oversimplified one-to-one relationship. Learners would therefore be expected to superimpose the semantically narrowed meanings of L1 loanwords directly onto their L2 baseword forms, and this expectation is borne out by the results.

As in the case of the keyness analysis, however, the results must be qualified by the observation that only a minority of the results were significant at the 0.05 level after Bonferroni correction.

6.3 Do the Effects of Loanword Knowledge on Baseword Usage Decline with Length of Study of English?

The results do not point to a decline in the effects of loanword knowledge. Instead, they provide weak evidence of an increase in the effects on baseword usage in that usage of three of the four basewords analyzed in detail – STRESS, TROUBLE, and RULE – was found to resemble usage of the corresponding Japanese loanwords more closely among learners who had studied English for longer. No difference in usage according to length of learning was observed in the case of the straightforward loanword HAPPY.

These results mirror to an extent those obtained by Inagawa (2007), who found that second-year high school students were more likely than first-year students to assume that English basewords behaved similarly to their corresponding loanwords in Japanese in all categories except straightforward loans. However, they contradict Ringbom (2007), who posited that the oversimplified assumption of a one-to-one relationship between L1 and L2 vocabulary items occurs more at the early stages of learning.

6.4 Pedagogical Implications

Proponents of loanwords as a resource advocate raising learners' general awareness of their "built-in lexicon" (Daulton, 2008) and encouraging them to use this resource

to intuit meaning and acquire new vocabulary (Barrs, 2013; Huynh, 2013; Kimura, 2004). However, the keyness results indicate that Japanese university students do not use basewords of commonly used loanwords significantly less than NSs, and they may even use them relatively more. There does not therefore appear to be a particular need to allocate valuable teaching time to raising loanword awareness further.

Research also shows that non-conformance to NS norms does not necessarily cause communication problems (Jenkins et al., 2011; Widdowson, 2012). Thus, where priority is placed on development of communicative ability rather than imitation of NS norms (such as in classrooms informed by an English as lingua franca approach), teaching time does not need to be allocated to countering the overuse of basewords and L1-influenced co-selections identified in this study, as these features of learner language usage would not affect intelligibility. Lack of vocabulary, on the other hand, can lead to communication problems, such as inability to explain an idea or understand an interlocutor if a vocabulary item is unknown and the speaker is unable to paraphrase (Lewandowska, 2019; Seidlhofer, 2004). Attention may therefore be usefully paid to raising learners' awareness of the broader semantic potential of the basewords that they are using in order to cover gaps in their vocabulary and improve their paraphrasing skills. This can be achieved by, for example, using corpora of texts produced by both English L1 and L2 users to explore with learners how usage of selected basewords resembles, and differs from, their own.

While Widdowson (1994) warns that classroom use of English L1 corpora may privilege NS expertise, this risk is here offset by the learner L1 knowledge that targeted teaching of basewords necessitates. Thus, whereas lists of loanwords and basewords can be compiled and used by NS teachers with relatively little L1 knowledge, interrogating NS corpora to compare L2 basewords' associations with those of L1 loanwords requires advanced L1 knowledge. This enhances the value of non-NS teachers' and learners' linguistic resources in the classroom in the manner advocated by proponents of English as a lingua franca and World Englishes such as Llurda (2018), and potentially mitigates what Kunert (2012) argues is the excessive deference shown toward NS teachers in English language teaching in Japan.

6.5 Limitations of Study

The implications of the study are limited by the design of the learner corpus, and in particular by the size of the corpus, the age of the learners who contributed to the corpus, the non-uniformity of the task conditions, and the lack of information provided regarding learner variables.

Regarding, firstly, corpus size, the 1-year subcorpus contained just 43,136 words, compared to 125,693 words in the 3-year corpus and 198,241 words in the learner corpus as a whole. The 1-year subcorpus consequently contains too few occurrences of the four basewords examined in depth (between 6 and 12 for each lemma, compared with between 18 and 50 in the 3-year corpus) to provide sufficient lines for part-of-speech, semantic, and co-occurring features to be considered representative of the English produced by first-year Japanese university students. This weakens the conclusions that can be drawn regarding changes in the effects of loanword knowledge

on baseword usage with length of study of English described in section 6.3.

Secondly, the learner corpus only contains texts written by Japanese L1 university students, making it difficult to extrapolate the results to other age groups with confidence.

Thirdly, the non-uniformity of the task conditions makes it difficult to completely isolate the effects of loanwords on the learners' production of written English. For example, while 53.7% of the texts in the 1-year corpus were written with access to reference tools, the corresponding figures for the 3-year subcorpus and the entire learner corpus were 0.9% and 13.1%. It consequently cannot be ruled out that the weak evidence for an increase in the effects of loanwords with length of study described in section 6.3 was due to the 3-year subcorpus writers having to rely more on their active vocabularies and their assumptions of baseword-loanword equivalences than the 1-year subcorpus writers.

Lastly, ICLEv2 provides no information on learner variables such as student majors, numbers of English courses taken, hours of instruction, or ability level. It is therefore not possible to draw conclusions about the impact of these factors on loanword use.

7. CONCLUSION

The findings provide non-significant evidence that Japanese L1 learners use basewords relatively more frequently in their writing than NSs, and some significant evidence that L2 baseword usage is affected by L1 usage of certain categories of loanwords, especially semantically narrowed loanwords, thereby pointing the way to more nuanced use of loanwords in the classroom. Corpus-based studies of loanwords' effects on learners remain scarce, however, and the weaknesses of the present study highlight three specific areas where further research is required.

Firstly, research is required using learner corpora covering other age groups and ability levels to demonstrate whether basewords' greater salience is characteristic of learners of all levels, and to corroborate, refine, or reject the finding that the effects of loanword knowledge increase with length of learning.

Secondly, learners' usage of similarly polysemous basewords and non-basewords needs to be compared to isolate the effects of semantically narrowed loanwords on L2 production from the challenge faced by learners of mastering the multiple meanings of polysemous words in general (Alnamer, 2017).

Lastly, further research is required to determine the extent to which frequency of use of basewords is attributable not to a loanword effect, but to the over-use of high-frequency words by learners. The supplementary use of the Chinese learner corpus in this study represented a first step in this direction. However, the keyness analysis of the Chinese learner corpus in this study did not investigate whether any of the basewords and non-basewords analyzed had loanword equivalents in Mandarin Chinese (or any other varieties of Chinese spoken by the learners). More granular research is therefore required using corpora of English produced by learners of other L1 backgrounds to isolate word frequency as a contributing factor.

REFERENCES

Allen, D. (2019). A procedure for determining Japanese loanword status for English words. *Vocabulary Learning and Instruction, 8*(2), 1-7. https://doi.org/10.7820/vli.v08.2.Allen

Alnamer, S. A. S. (2017). On the awareness of English polysemous words by Arabic-speaking EFL learners. *Advances in Language and Literary Studies, 8*(2), 112–121. http://dx.doi.org/10.7575/aiac.alls.v.8n.2p.112

Anthony, L. (2016). AntConc (Version 3.4.4) [Computer software]. Waseda University. http://www.laurenceanthony.net/software

Anthony, L. (2018). Someya's unhyphenated English lemma list. http://www.laurenceanthony.net/software/

Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh University Press.

Barker, F. (2010). How can corpora be used in language testing. In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 633–645). Routledge.

Barrs, K. (2013). L2 vocabulary in the L1 environment. *Working Papers in Language Education and Research, 1*(1), 3–10. https://eliworkingpapers.org/archives/ ophia y-2013/

Brown, J. (1995). Is gairaigo English? *The Internet TESL Journal, 1*(2). http://iteslj.org/Articles/Brown-Gairaigo/ (accessed 4 November 2017)

Centre for English Corpus Linguistics (CECL) (n.d.). *The Louvain corpus of native English essays*. Université catholique de Louvain. http://www.learnercorpusassociation.org/resources/tools/locness-corpus/

Daulton, F. E. (1998). Japanese loanword cognates and the acquisition of English vocabulary. *The Language Teacher, 22*(1). http://jalt-publications.org/tlt/articles/1916-japanese-loanword-cognates-and-acquisition-english-vocabulary

Daulton, F. E. (2003). List of high-frequency baseword vocabulary for Japanese EFL students #2. *The Internet TESL Journal, 9*(3). http://iteslj.org/lists/Daulton-BasewordVocab2.html

Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords*. Multilingual Matters.

Granger, S. 2002. A bird's-eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). John Benjamins Publishing.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *International corpus of learner English. Version 2*. Presses universitaires de Louvain.

Haspelmath, M. & Tadmor, U. (Eds.). (2009). *World Loanword Database*. Max Planck Institute for Evolutionary Anthropology. http://wold.clid.org

Huynh, C. (2013). English loanwords in Japanese as a bridge to vocabulary acquisition. *Konan Women's University Bulletin, 15*, 1–17. http://jairo.nii.ac.jp/0205/00003250

Inagawa, M. (2007). Influence of loanwords on English usage by Japanese learners of English. *The Bulletin of the Kanto-Koshin-Etsu English Language Education Society, 21*, 85–96. https://doi.org/10.20806/katejo.21.0_85

Inoue, M. (2004). Katakanago (gairaigo) kihon goi 550 go: Sono goi tokusei to sentei kijun [A basic vocabulary list of 550 katakana loanwords: Vocabulary characteristics and selection criteria]. *Journal of Kobe Yamate University, 6*, 65–79. https://ci.nii.ac.jp/naid/110004706763

Irwin, M. (2011). *Loanwords in Japanese*. John Benjamins Publishing.

Ishiwata, T. (Ed.). (1998). *Kihon gairaigo jiten* [A basic dictionary of loanwords]. Tokyodo Shuppan.

Itou, M. (2007). Zasshi ni mirareru gairaigo to gaikokugo [Loanwords and foreign words used in magazines]. *Dai 30 kai kotoba fooramu: Nihongo no naka no gairaigo to gaikokugo* [30th language forum: Loanwords and foreign words in Japanese], 6–13.

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching, 44*(3), 281-315.

Kahl, D. (2015) *Sono katakana eigo, gaikokujin ni ha tsuujimasen* [That katakana English won't be understood by foreigners]. Forest Publishing.

Kay, G. (1995). English loanwords in Japanese. *World Englishes, 14*(1), 67–76. https://doi.org/10.1111/j.1467-971X.1995.tb00340.x

Kimura, M. (2004). *The effect of Japanese loanwords on the acquisition of the correct range of meanings of English words*. Kyoto Women's University.

Kunert, H. (2012). Linguistic imperialism and English language teaching in Japan. *Sophia TESOL Forum/Working Papers in TESOL*. http://digital-archives.sophia.ac.jp/repository/view/repository/00000033032

Lado, R. (1955). Patterns of difficulty in vocabulary. *Language Learning: A Journal of Research in Language Studies, 6*(1-2), 23–41. https://doi.org/10.1111/j.1467-1770.1955.tb00829.x

Lancaster University Centre for Computer Research on Language (n.d.). CLAWS part-of-speech tagger for English. Lancaster University. http://ucrel.lancs.ac.uk/claws/

Lewandowska, E. (2019). English as a lingua franca: an overview of communicative strategies. In B. Loranc-Paszylk (Ed.), *Rethinking directions in language learning and teaching at university level* (pp. 27-52). Research-publishing.net.

Llurda, E. (2018). English language teachers and ELF. In J. Jenkins, W. Baker & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 518-528). Taylor & Francis Group.

Martin, S. (1988). *A reference grammar of Japanese*. Charles E. Tuttle Company.

McEnery, T. & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Millar, N. (2006). The use of Japanese loanword cognates in Japanese ESL learners' production: A corpus based study. *The Journal of Kanda University of International Studies, 18*, 479–496.

Ministry of Internal Affairs and Communications. (2005). *Jouhou tsuushin hakusho* [Information and communications white paper]. http://www.soumu.go.jp/johotsusintokei/whitepaper/h17.html

National Institute for Japanese Language and Linguistics. (2004a). *Gairaigo ni kansuru ishiki chousa (zenkoku chousa)* [A survey of attitudes towards loanwords (national survey)]. https://www.ninjal.ac.jp/archives/genzai/ishiki/

National Institute for Japanese Language and Linguistics. (2004b). *Gairaigo teichakudo chousa* [Surveys of familiarity with loanwords]. https://mmsrv.ninjal.ac.jp/gairaigo_yoron/

National Institute for Japanese Language and Linguistics. (2009). *Balanced corpus of contemporary written Japanese*. https://ccd.ninjal.ac.jp/bccwj/en/

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge

University Press.

Nation, I. S. P. (2017). *The BNC/COCA word family lists*. https://www.victoria.ac.nz/lals/about/staff/paul-nation

Rayson, P. (2015). *Log-likelihood calculator and effect size calculator*. University Centre for Computer Corpus Research on Language. http://ucrel.lancs.ac.uk/llwizard.html

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.

Richards, J. C. and Schmidt, R. (2013). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.

Scherling, J. (2009) *Embrassimilating the other: anglicisms in Japanese – a cultural pragmatic model for loanword integration* [Unpublished doctoral dissertation]. Karl-Franzens-Universität Graz. http://unipub.uni-graz.at/obvugrhs/content/titleinfo/206890

Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics, 24*, 209-239. https://doi.org/10.1017/S0267190504000145

Sheperd, J. (1996). Loanwords: A pitfall for all students. *The Internet TESL Journal, 2*(2). http://iteslj.org/Articles/Shepherd-Loanwords.html

Simon-Maeda, A. (1995). Language awareness: Use/misuse of loan-words in the English language in Japan. *The Internet TESL Journal, 1*(2). http://iteslj.org/Articles/Maeda-Loanwords.html

Stanlaw, J. (2004). *Japanese English: Language and culture contact*. Hong Kong University Press.

Tono, Y., Yamazaki, M. & Maekawa, K. (2013). *A frequency dictionary of Japanese*. Routledge.

Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly, 28*(2), 377–389. https://doi.org/10.2307/3587438

Widdowson, H. G. (2012). ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca, 1*(1), 5-26. https://doi.org/10.1515/jelf-2012-0002

Yoshida, M. (1978). The acquisition of English vocabulary by a Japanese-speaking

child. In Evelyn Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 91–100). Newbury House.