



ISSN 2187-4972

**Working Papers  
in Language Education and Research**

<http://eliworkingpapers.org/>

**A PRELIMINARY ANALYSIS OF CROSS-  
LINGUISTIC LEXICAL SIMILARITIES  
BETWEEN ENGLISH AND JAPANESE**

Keith Barrs

Corresponding author:

[keith-b@kanda.kuis.ac.jp](mailto:keith-b@kanda.kuis.ac.jp)

Publication date: February, 2013

**To cite this article**

Barrs, K. (2013). A preliminary analysis of cross-linguistic lexical similarities between English and Japanese. *Working Papers in Language Education and Research*, 1 (1), 11-20.

Scroll down for article

**A PRELIMINARY ANALAYIS OF CROSS-LINGUISTIC LEXICAL  
SIMILARITIES BETWEEN ENGLISH AND JAPANESE.**

Keith Barrs

*Kanda University of International Studies*

**ABSTRACT**

*Cross-linguistic similarities found between any two languages can have positive pedagogical implications for all participants in the language learning process. This article reports on a small-scale investigation into the quantity of English words existing in the Japanese lexicon through a study of English loanword occurrences in a week of Japanese-language newspaper articles. The results show that there is clear evidence of an integrated English sub-lexicon in the Japanese language, which can be conceptualised as a useful linguistic resource for Japanese learners of English.*

**1. INTRODUCTION**

An examination of the Japanese language which looks at English words encoded in the script of katakana, the script primarily employed in modern Japanese to represent words of foreign origin, reveals a vast quantity of English-based lexical items that have been integrated into the workings of the Japanese vocabulary. It is at this level of analysis, with words that exist as a sub-lexicon of the Japanese language, that any cross-linguistic lexical similarities could benefit the Japanese learner of English; one such benefit being the reduction of the 'learning burden' involved in the acquisition of vocabulary (Nation, 1990). This article reports on a small-scale research investigation

into the extent to which English vocabulary can be said to be integrated into the Japanese lexicon and, as such, be potentially available as a sub-lexicon to Japanese learners of English. Despite the small-scale nature of this research, it is hoped that the findings can contribute to a further understanding of to what extent English words form a sub-lexicon of the Japanese language and are potentially available as a cross-linguistic knowledge resource to Japanese learners of English (see Daulton, 2008 for a related but much larger-scale study).

## **2. METHODOLOGY**

This study reports on inventory research into the extent of English language lexical item incorporation and integration into a complete week of Japanese-language newspaper front pages. The research involved (1) a quantitative analysis of instances of English language items encoded in the katakana script and (2) an examination of whether these items exist with or without a translation into native or Sino-Japanese equivalents. The purpose of the first part of the investigation is to shed light on one specific example of English lexical item incorporation in the Japanese language. Subsequently, the purpose of the second part of the investigation is to offer insights into the extent to which these English lexical items are fully integrated into the syntactical and semantic workings of the Japanese sentence. This is done by analysing whether or not the incorporated English lexical items occur with a translation into native-Japanese or Sino-Japanese vocabulary, a practice that would suggest the incorporated item presents difficulties in its understanding and therefore has not been fully integrated into Japanese.

For this study it was decided to randomly select a full week of front pages, the start day and month were chosen by generating two numbers randomly, one number being generated from the range 1 to 12 to select the month, and the second number then being randomly generated from the number of days in that month to select the start day of the week. This number generation led to May 13<sup>th</sup> being the start day with seven front pages collected for analysis, ending on May 19<sup>th</sup>. The entire front page (except the bottom section containing a regular block of adverts and any other variously placed adverts) was selected for analysis. The exclusion of adverts was because of the different nature of foreign word usage in this genre, where foreign words and especially those of English appear in a variety of scripts and often for a more decorative (symbolic) purpose (see: Haarmann, 1984; Takashi, 1990).

The articles were scanned for katakana items, which were listed in an Excel file, and then categorised by language of origin into the groups: (1) English Foreign Language (EFL), representing any individual or hybrid item corresponding to words in the English language; (2) Other Foreign Language (OFL), representing any language other than Japanese and English; and (3) Japanese Language (JL), representing native and Sino-Japanese items.

The next stage involved reanalysing each front page both to catch any missed katakana lexical items, but more importantly to analyse each instance of katakana English for the occurrence of a native-Japanese and/or Sino-Japanese translation. Any translation occurrence was recorded alongside the entry of the lexical item. The entries were then summed by (1) total of types, (2) total of tokens, (3) total of items by language category, and (4) total of English-derived items with/without a translation.

### **3. DATA ANALYSIS AND DISCUSSION**

The total number of lexical item types found on the front pages of one week of the Asahi Shinbun newspaper, in the katakana script, totalled 222. The number of tokens totalled 313. This figure excludes proper nouns such as names and places which totalled 105 occurrences. Table 1 shows details of the research findings for each front page and shows that there were on average 31.7 katakana item types and 44.7 tokens found across the seven pages. A complete count of all characters/words in all scripts on each front page was beyond the scope and practical means of this research, but each front page can be divided up into, on average, 6 sections (4 main articles, an opinion piece and a side bar which acts as a newspaper index) which means that each section had an average of 4.5 katakana word types and 6.4 tokens incorporated into the text. Such an important but relatively infrequent occurrence of katakana is typical of the regular make up of writing involving the three main Japanese scripts, where there is a mix of kanji, as the most frequent, hiragana as the second most frequent, and katakana as the least frequent script (Shibatani, 1990; Gottlieb, 2005). In this respect, there existed several articles with zero occurrences of katakana, but no articles were found that only consisted of kanji or hiragana.

Table 1.

*A Frequency Count of Katakana Script Types and Tokens on the 7 Front Pages*

<b>Front Page Date</b>	<b>Katakana Script Items: Type</b>	<b>Katakana Script Items: Token</b>
<b>May 13<sup>th</sup></b>	21	32
<b>May 14<sup>th</sup></b>	35	47
<b>May 15<sup>th</sup></b>	28	38
<b>May 16<sup>th</sup></b>	25	30
<b>May 17<sup>th</sup></b>	43	60
<b>May 18<sup>th</sup></b>	33	59
<b>May 19<sup>th</sup></b>	37	47
<b>Total:</b>	222	313
<b>Average:</b>	31.7	44.7

The next stage of analysis, shown in Table 2, categorises all of the 222 katakana item types into the two lexical strata of: (1) Japanese Language (JL) (including Native Japanese and Sino-Japanese), and (2) Foreign Language (FL) (including all foreign languages except for Chinese which is a foreign language traditionally grouped into the Japanese Language strata). It is clear that the vast majority (92.3%) of katakana items are those within the foreign words strata of the Japanese lexicon. This finding is in line with the principal purpose of this script, to be used for the representation of foreign items used in the Japanese language (Shibatani, 1990). The languages within this FL strata as found on the seven front pages are (1) English (e.g. モニタリング

/monitoring/monitoring), (2) German (e.g. テーマ/teema/thema), (3) Portuguese (パン/pan/pão), (4) Dutch (コーヒー/koohii/koffie), (5) French (ピーマン/piiman/piment) and (6) Korean (ユッケ/yukke/yukhoe).

Table 2.

*Number and Percentage of Types in the Japanese (JL) and Foreign (FL) Lexical Strata*

Lexical Strata	Number of Types	Percentage of Types
Japanese (JL)	17	7.7%
Foreign (FL)	205	92.3%
<b>Total:</b>	<b>222</b>	<b>100%</b>

Table 3 unpacks the FL strata items into: (1) English Foreign Language (EFL), and (2) Other Foreign Language (OFL). Here it is clear that the overwhelming majority (95%) of FL items are of English origin. Some of these items have been morphologically modified when absorbed into Japanese, such as the shortened アパート/aparto/apart being used in place of アパートメント/apartomento/apartment and スーパー/suupaa/super being used in place of the longer スーパーマーケット/suupaamaaketto/supermarket, but they are still derived from English language lexical items. Even if the items are now considered to be words of the Japanese language more than the English language, due to their modifications within the host language, the extent to which the items appear in the newspapers, in any modified manner, clearly shows the extensive incorporation of the English language into Japanese. But in order to

claim that the words are integrated into Japanese, in a linguistic sense, whereby the words can be seen as a fully functional semantic and syntactical part of the sentence, it is necessary to show whether or not the EFL items have been translated. The existence of a translation occurring with the English-derived lexical items would suggest that the word is probably beyond the comprehension of the intended reader, and therefore that the item has not been fully linguistically integrated. It could be suggested in this sense that words appearing with a translation are employed more for a stylistic (symbolic) than linguistic purpose.

Table 3.

*Number and Percentage of English (EFL) and Other Foreign Language (OFL) Types*

<b>FL Strata</b>	<b>Number of Types</b>	<b>Percentage of Types</b>
<b>Divisions</b>		
<b>English (EFL)</b>	195	95%
<b>Other (OFL)</b>	10	5%
<b>Total:</b>	205	100%

Table 4 shows the occurrence of EFL items with and without a translation. As is clear, the vast majority of items (96.4%) are not associated with a translation and only 7 English-derived words found in the data set, of a total of 195, had a translation occurring either before or after the placement of the lexical item. This clearly shows that the vast majority of English-derived items found in the front pages have been fully integrated into the linguistic workings of this particular print medium.



Table 4.

*Number and Percentage of Types of EFL Items Occurring With/Without a Translation*

<b>Translation</b>	<b>Number of Types</b>	<b>Percentage of Types</b>
<b>Occurrence</b>		
<b>With Translation</b>	7	3.6%
<b>Without Translation</b>	188	96.4%
<b>Total:</b>	195	100%

The lexical items appearing with a translation, and their associated frequencies, are listed in Table 5. One of the items, *メルトダウン/meltdown*, occurred 4 times and is associated with the nuclear crisis following the 2011 earthquake and tsunami. It is significant that this item also appeared 3 times across the 7 front pages with no associated translation, suggesting that the use of a translation was more a personal choice of the article writer than a particular necessity connected with the difficulty of the word.

Table 5.

*The Katakana English-Derived Lexical Items Appearing With a Translation*

<b>Katakana Lexical Item</b>	<b>English Equivalent</b>	<b>Frequency</b>
メルtdown	Meltdown	4
エコタウン	Eco Town	1
シビアアクシデント	Severe Accident	1
ロードマップ	Road Map	1

**4. CONCLUSION**

The preceding analysis and discussion shows that there is not only a large number of English-derived lexical items incorporated into this particular print medium (195 types), but also that there is a very high level of integration of these items into the linguistic workings of the newspaper front pages (96.4% having no translation). This clearly indicates that English is playing more than just a symbolic role in this context: the incorporated and integrated lexical items are acting as fully functioning linguistic building blocks in the syntactical and semantic construction of the front pages.

The findings show that even in this small-scale research investigation, focusing on one specific literature medium, a vast amount of integrated English vocabulary could be found embedded in the Japanese language. This raises the issue of implications for Japanese learners of English who could draw on this sub-lexicon when engaged in English-language studies (see Daulton, 2008; Ringbom, 2007 for similar discussions),

and such implications can form the basis of further research investigations into the area of the extent of similarities found between the items in the English sub-lexicon of Japanese and the English-language equivalents.

## 5. ACKNOWLEDGEMENTS

Thank you to RILS&LE for assistance with the funding to present this activity at a research seminar at the University of Leicester, 2012.

## 6. REFERENCES

- Daulton, F.E. (2008). *Japan's built in lexicon of English-based loanwords*. Clevedon: Multilingual Matters Ltd.
- Gottlieb, N. (2005). *Language and society in Japan*. Cambridge: CUP.
- Haarman, H. (1984). The role of ethnocultural stereotypes and foreign languages in Japanese commercials. *International Journal of Sociology of Language*, 50, 101-121.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle Publishers.
- Ringbom, H. (2007). The importance of cross-linguistic similarities. *The Language Teacher*, 31(9), 3-5.
- Shibatani, M. (1990). *The languages of Japan*. Cambridge: CUP.
- Takashi, K. (1990). A socio-linguistic analysis of English borrowings in Japanese advertising texts. *World Englishes*, 9(3), 327-341.