

Explorations of the Lexical Interface between Japanese and English

Charles Adamson

Miyagi University School of Nursing

キーワード

辞書、日本語、英語、書き換えの方法、順番

Lexicon, Japanese, English, transcription method, sequence

要 旨

この研究は日本語と英語との間の関係をレキシカルインターフェースにより分析した初めての研究である。同一単語を日本語のひらがな、カタカナ、漢字および英語によって表記したものを提示し、その順位関係をリカルトスケールを用いて評価した。その結果、次のような首尾一貫した6つの順位が認められた。その順位とは、1) 英語を英語表記、2) 英語をカタカナ表記、3) 英語をひらがなで表記したものと、日本語をローマ字表記、4) 日本語をカタカナ表記、5) 日本語をひらがな表記、6) 日本語を漢字表記である。以上の結果は今後の研究のための示唆に富むものであった。

Abstract

This pilot study investigates aspects of the lexical interface between Japanese and English. The transcription of a single concept using the English and Japanese roots presented in English letters, hiragana, katakana, and kanji were evaluated by the subjects using a Likert scale. The results suggest that there is a consistent perceived order in the relative Japaneseness of the tokens: 1) English transcribed in English, 2) English in katakana, 3) English in hiragana and Japanese in romaji, 4) Japanese in katakana, 5) Japanese in hiragana, and 6) Japanese in kanji. Suggestions are made for further research.

Introduction

A quick glance at any Japanese language newspaper will reveal large numbers of lexical tokens that are apparently English or at least derived from English. The obvious question is then "How does the Japanese reader categorize these tokens?" Are the tokens English or Japanese? Or something in between? The answers to these questions can have a large influence on the methods used to teach these tokens in the Japanese classroom. This paper describes a first investigation into these questions. The most direct approach is used, subjects are asked to rate each token on a Likert scale which has pure English at one extreme and pure Japanese at the other.

Method

Through consultations with native speakers of Japanese, the author determined three families of concepts which might yield appropriate results: two are nouns in English (*book* and *milk*) and the other is an adjective (*big*). Verbs and adverbs were not selected since these do not usually appear as loan words in Japanese. The concepts were presented to the subjects as English written in English letters, katakana, and hiragana and as Japanese written in romaji, katakana, hiragana, and kanji. Two versions of the romaji concept for *MILK* (one indicating and one omitting the long vowel) were included. See Figure 1 for a complete listing.

The subjects were 79 first year nursing students

at Miyagi University who completed the questionnaires as an activity during a required English class. The subjects are mostly 18 and 19 year old females who had been streamed into two classes at the beginning of the semester by use of an in-house placement test. The students were required to write their name and student number on their questionnaire. This information was used to cross-reference the student's questionnaire with their class placement.

Two versions of the questionnaire were developed: (1) all the tokens arranged in a random order; (2) three sections, one for each concept family, with the tokens randomly placed within the section. A Likert scale with nine levels (1 to 9) was selected so that subjects could theoretically place each token on a separate level. The questionnaires were distributed so that every other student received the same questionnaire. The students were instructed to work quickly and to use their first impressions.

Data

The data used in the following analyses consisted of the responses to the questionnaire and the students' section. Coding was used to protect the students' privacy. Statistica '99 for Windows was employed for all calculations and figures.

Figure 2 shows the total responses by token and Figure 3 shows the means, standard deviations and standard errors of the means. There was no effect for class placement or questionnaire.

BIG	BOOK	MILK
big	book	milk
ビッグ	ブック	ミルク
びっぐ	ぶっく	みるく
ookii	hon	gyunyu gyuunyuu
オオキイ	ホン	ギユウニユウ
おおきい	ほん	ぎゅうにゅう
大きい	本	牛乳

Figure 1. The families of tokens used to present the concepts.

Token	Number of Responses/Likert category								
	1	2	3	4	5	6	7	8	9
big	0	1	0	1	2	0	4	5	66
book	1	1	0	2	1	4	2	3	65
milk	0	1	0	1	2	1	6	6	62
ビッグ	0	0	4	6	9	9	13	6	31
ブック	4	1	11	7	14	9	11	9	13
ミルク	5	3	12	7	17	7	12	6	10
びっく	12	11	11	9	12	12	7	2	3
ぶっく	13	7	11	13	11	15	3	2	4
みるく	22	12	16	9	8	7	3	1	1
ookii	16	5	11	11	14	10	6	1	5
hon	11	12	12	18	11	9	4	3	3
gyunyu	13	8	8	10	18	9	4	2	7
gyuunyuu	10	8	13	13	11	5	9	3	7
オオキイ	27	11	18	10	9	1	3	0	0
ホン	27	13	12	16	5	2	2	1	0
ギユウニユウ	29	13	19	7	7	2	1	1	0
おおきい	71	8	0	0	0	0	0	0	0
ほん	62	11	5	1	0	0	0	0	0
みるく	55	16	3	1	1	1	0	1	0
大きい	79	0	0	0	0	0	0	0	0
本	78	0	0	0	1	0	0	0	0
牛乳	78	0	0	0	0	0	0	0	1

Figure 2: Responses to tokens by Likert scale levels. Level 1 represents Japanese and Level 9 represents English.

The tokens fall into seven groups according to their method of transcription: ENG, EK, EH, JR, JK, JH, and KAN (see Fig. 3 for the meanings of these codes). Inspection of Figure 3 shows that, except for the EH and JR groups and JH_BIG/KAN_BIG, the groups are visually identifiable. The correlations verify this: in group correlations range from 0.44 to 0.90 ($p=0.001$).

The EH and JR tokens appear to form a single group with correlations ranging from 0.47 to 0.73 ($p=0.001$).

There are also highly significant correlations in the range of 0.37 to 0.62 ($p=0.001$) between the members of each group and the preceding and following groups when arranged in the order presented in Figure 3. All other correlations are generally slightly negative and not significant, even at $p=0.1$.

The EH token for /MILK/ differs from the other

tokens in its group in that it is substantially more Japanese, while the JH and KAN tokens have substantially larger standard deviations than the other members of the group, indicating less agreement between subjects. Finally, the EK token for /BIG/ is more English than the other tokens in its group.

Discussion

The correlations and visual inspection indicate that there is a sequencing by transcription type and language. This sequence is English transcribed in English, English in katakana, English in hiragana and Japanese in romaji, Japanese in katakana, Japanese in hiragana, and finally Japanese in kanji. The correlations are reasonably high but at the same time are low enough to indicate that other factors are involved.

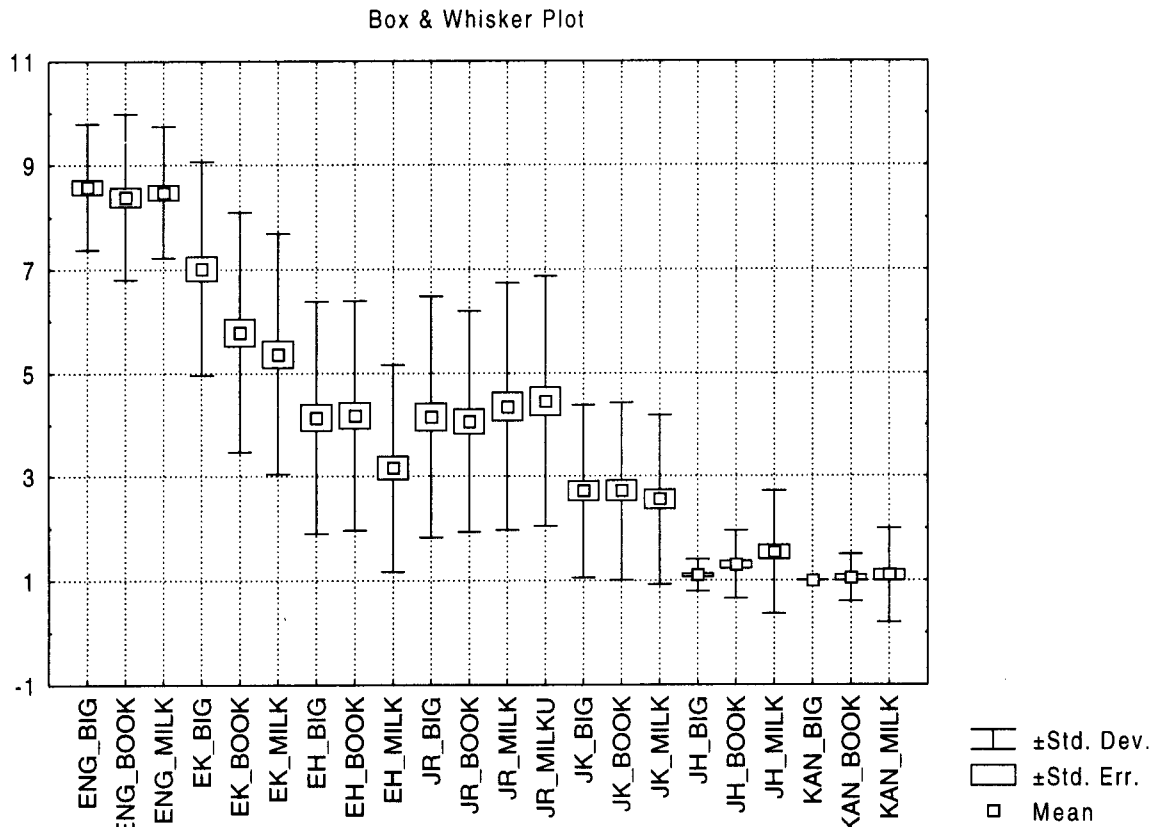


Figure 3: Box and whisker plots of means, standard deviations, and standard error of means. [Eng = English word transcribed in English, EK = English word in katakana, EH = English word in hiragana, JR = Japanese word in romaji, JK = Japanese word in katakana, JH = Japanese word in hiragana, and KAN = Japanese word in kanji; MILK indicates the transcription *gyunyu* and MILKU indicates *gyuunyuu*]

Only the means for ENG group and EK group fall in the English portion of the chart, but even the ENG group has a Japanese factor. This may be because these tokens can be used unchanged in some Japanese contexts, for example, advertisements. All other means are more Japanese than English even though they may be derived from English.

Katakana is always more English-like than the same token in hiragana. This supports the intuition that katakana is used for writing foreign words and hiragana for Japanese.

There are basically three anomalies in the data: in comparison to the other groups members, EK_BIG is more strongly English, EH_MILK is more strongly Japanese, and JH_MILK is less Japanese. My native-speaker Japanese sources suggest the following reasons for this. Big is a very easy, well-known

word that is learned as English in junior high school and seen frequently. Milk on the other hand is seldom seen in hiragana transcriptions of either the English or Japanese tokens.

The standard deviations indicate that there is more agreement at the ends of the scale and for Japanese except for JR. Inspection of the raw data reveals that much of the data is two peaked, possibly indicating that the data is made up from two different populations. However, additional data should be collected before trying to separate the possible populations.

Conclusions

This pilot study demonstrates that there is a recognizable and consistent sequence within at least portions of the English/Japanese lexical

interface. It also confirms to a certain extent the intuition that for many Japanese English is just an alternate way of writing Japanese, since all of the tokens were at least partly Japanese whereas the Japanese tokens were Japanese. While the details are beyond the scope of this paper, this has strong implications for the teaching of both English and Japanese as second languages.

The findings of this pilot study could be verified and generalized by the following additional research:

- (1) Increasing the number of subjects so the substructure of the responses for each token could be analyzed. This would also involve determining if individuals tended to follow specific patterns. If a substructure is found, it could suggest that there are two or more groups of students and that each group would benefit from different methods of vocabulary presentation.
- (2) Additional tokens could be found in Japanese sources and their results compared to the results obtained here. These results could suggest that there should be different methods of presentation for tokens which follow the normal pattern and those which are exceptions.
- (3) The possible effects of context could be explored by embedding the tokens in a context for presentation. The understanding gained from these results could lead to improved methods of presenting vocabulary items to students.
- (4) Comparing the responses for differing age groups. These results suggest differing methods of vocabulary presentation based on the students' age.
- (5) Comparing the results of Japanese to the results for native speakers of English who are studying Japanese. These results would give the teacher a better understanding of how his or her perceptions differ from the students'.