

FOREIGN LEXICAL PRESERVATION: A CASE OF PORTUGUESE WORDS BORROWED INTO JAPANESE

Atsushi ASAI (Daido Institute of Technology, Nagoya, Japan)

ABSTRACT: The borrowing of Portuguese words into Japanese began in the 16th century, and some words were adopted in a short period of time. Although, for these 100 years, the ratio of Portuguese words to foreign words has decreased from 20% to only 2%, several words for clothing, food and amusements in total are still used at an occurrence rate of about 0.3 per mil. Without the elaborate ways of transcribing foreign terms with ideographic Chinese characters in pursuit of both phonetic and semantic appropriateness, such high assimilation could not have occurred in the lexicon.

1. Introduction

This study reports on lexical changes from a sociolinguistic point of view by means of specifically designed corpora of Japanese texts written one century ago and in the past several years, and shows how some Portuguese words have been used in Japanese.

Portuguese traders and missionaries reached the Far East in the 16th century. Their arrival made a great impact on Japan. The local people were astonished to know about the uniqueness or usefulness of new clothing, food, and religion. Despite the nation's seclusion and ban on foreign missions, some Portuguese words as well as Austroasian words for expressing the names of products brought by Portuguese traders were rapidly and widely borrowed into Japanese. A key to the adoption was the elaborate ways of transcribing. At that time, loanwords were transcribed with Chinese characters or phonographic Hiragana characters mainly used for expressing native words. Some words, such as “capa”, “carta”, “confeito”, “tabaco”, and “tempero”, were transcribed with ideographic Chinese characters in order to achieve the highest appropriateness both phonetically and semantically. The word “capa” is represented with two Chinese characters that mean ‘fitted’ and ‘feather’, and the word “carta” is with three characters that mean ‘song’, ‘recording’, and ‘many’. Chinese characters had been part of the Japanese writing system for about 1000 years. The transcriptions devised in Chinese characters helped the Japanese to avoid the sense of foreignness, stabilize the representations, and maintain the words in the reader's memory. The seclusion and ban on foreign missions strictly limited the importation of goods and ideas. This social condition blocked the further increase of foreign terms coming from outside. If people needed many loanwords for a particular reason, many might not have enough capacity to memorize the transcriptions exactly.

2. Overview of the history of the Japanese lexicon

Japanese has received an influx of four large waves of foreign word as sketched in Figure 1, and thus its lexicon consists of four strata -- native words, Chinese words, foreign words, and onomatopoeia (Inoue, F., 2002). Because the Japanese had been eager to import the advanced continental cultures, they started to borrow Chinese characters in the 3rd century or slightly later for the practical purpose of writing down Japanese words (Muller, 1964; Ohno, 2000). Japanese envoys to the Tang Dynasty brought advanced products and knowledge back to the archipelagos in the 7th and 8th centuries (Goto, 1988; Steinberg, 1996). For the following several centuries, Chinese vocabulary gradually increased in the Japanese lexicon. Ultimately, the Chinese words came to form one lexical stratum of the Japanese language, and are not any longer considered foreign words (Miyajima, 1977; 1989).

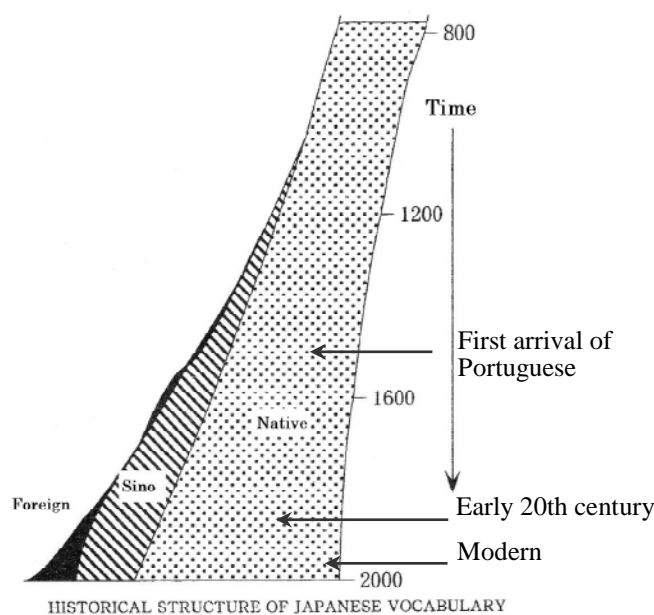


Figure 1. Historical structure of Japanese vocabulary.
Reproduced from Inoue, F. (2002).

The second wave came in the 16th century. Portuguese traders drifted ashore onto a small southwestern island of Japan in 1539. After that accident, Portuguese traders and missionaries arrived on the main islands and brought many European products and ideas. A small number of Dutch traders followed the Portuguese, and even during the years of seclusion brought some products whose names are also still used.

In the late 19th century, the third wave washed the isolated islands. The shogunate terminated its seclusion, and the Imperial Court took over the reins in 1867. American and British officers and visiting instructors introduced new products, which required the Japanese to learn new words in order to name those fruits of the industrial revolution. The efforts at creating semantic translations with Chinese characters could not catch up with the increased number of new foreign words. Many of those new words were thus transcribed simply with phonographic Katakana characters. Such convenient phonetic transcriptions inevitably produced variants.

The fourth wave began immediately after World War II. Foreign words from European languages have enriched the third lexical stratum (Inoue, H., 2002). Extensive research has been conducted to observe this wave, and several views from a couple of angles have measured the size successfully. For instance, Yamaguchi et al. (2004) revealed that about 5% of all the running words in articles of a national newspaper in Japan were foreign words. Note that the fourth stratum is onomatopoeia while some lexicologists set compound words that mix native words with Chinese words, called Sino-Japanese words, to this category.

3. 100-year-old text

Let us take a closer look at texts in the early 20th century. Portuguese words in total appeared at an occurrence rate of about 0.5 per mil on the token base. This period of time was chosen because nearly a half century had passed since the nearly 250-year seclusion had ended, and a social movement for the unification of the spoken and written language had made headway to the extent that we can now make a linguistic comparison between texts from that time and those from today (Kaganoi, 2002).

A mini corpus compiled for this study consists of novels in the early 20th century, and counts the characters included at about 1.29 million. The controlled conditions for this text collection were as follows. This special corpus contains 35 novels (Aozora Bunko, 2006) written by 32 well-known novelists who were born before 1900. The novels were written in the modern colloquial language from 1890's to 1940's. Two novels are so long that only single chapters were filed into this corpus to maintain a balance in the quantitative contributions of the 35 sources. A quasi-random choice extracted sentences from this corpus to make the number of words approximately 13,800. The parser Chasen Windows version 2.1 was used to divide the passages into words (Matsumoto et al., 2000). In this agglutinative language, morphemes can be defined as words, and the short beta unit scheme was applied in this syntactic parsing (NIJL, 1962). After the

automatic processing, the remaining errors were corrected manually. Some errors usually occur because almost every text has homonyms, polysemous words, proper nouns, and rare words not listed in the built-in dictionary.

Statistical analysis with homemade Perl codes showed that about 0.2 percent of the running words were European in origin. The proper nouns of Chinese or Korean origin in Chinese characters were not counted as foreign entries here. Japanese allows foreign words to be borrowed as nouns to fit into its syntactic system as observed in most cases of language contact. The descriptive statistics, therefore, include nouns and other indeclinable words, and exclude declinable words. Nominal adjectives are classified as the combinations of nouns and post particles (Kitahara, 1973; Tokieda, 1950). The word “tabaco” appeared 89 times in the texts, and indeed 82 of them were in Chinese characters representing semantic transcriptions that meant ‘smoking grass’. The other seven were in Hiragana characters. One example of the transcriptions in Chinese characters is shown in Figure 2.

先年三越で福井丸の船材で造った物を売り出した時に
巻煙草入を買って帰りました。その日に偶然ヘルンの
書いて置きました『廣瀬中佐の歌』が出ましたから私
は不思議に思いまして、それを丁度その箱に納めて置

Figure 2. One example of transcription of “tabaco”
(the underlined part).

4. Modern text

For these 100 years, our dramatic social changes are thought to have caused a large shift in the foreign lexical components of Japanese from Portuguese to English (Hoffer, 1990; Ozawa, 1976). In our modern lives, we must use English words to operate not only computers and other office equipments, but also digital TVs, video recorders, cell phones, car navigation systems, and many other multi-function home electronics appliances. This study created another corpus consisting of 24 novels written by 24 authors who are active today. This text counted its characters at about 1.25 million. In the same manner as the above, we made a subset of this special corpus of the modern novels, and analyzed the selected text. The validity of the genre selection was confirmed with Kabashima’s rule with the sample text. Kabashima’s rule shows that a ratio of the number of indeclinable words to the sum of declinable words – verbs and adjectives – and aspect words – conjunctions, interjections, and adverbs – is an indicator of genre. We made a linguistic comparison between the sample text of modern novels here and texts from the other four genres: (1) general newspaper articles, (2) sports newspaper articles, (3) cosmetics news, and (4) information technology news and manuals. This comparison allows us to know about the characteristics of a modern text (Asai & Ishikawa, 2006). The sizes of all the texts were controlled to around 13,800 words, whose size fulfills a requirement for obtaining minimum statistical reliability, 5,000 words (Maekawa, 1995). The dates of publication were also controlled. The source texts were issued between January 2001 and June 2006 in the novel section of The Asahi Shimbun, a general newspaper with a nationwide circulation, (Asahi Shimbun, 2006) and online novel sites (WEB Quilala et al., 2006). We obtained a parsed result that the ratio introduced above was about 55% for novels, and already knew that was about 77% for general news as shown in Figure 3, which plotted the percentage of each syntactical category to all of the four main categories: nouns, verbs, conjunctions and interjections, and adjectives, adverbs and attributes. The relative ratio of these two values is 0.71, which well agrees with a value reported in the previous research (Kabashima, 1954; 1955), 0.75, if we consider the fact that this kind of figure depends on sources and times and that half a century has passed since the previous research was conducted. In general, novels contain both description parts and conversation parts, and cover general topics, for example business, social conditions, arts, beliefs, adolescence, human relations, and general situations, for example streets, farms, workplaces, school campuses, and homes. This ratio for a novel is thus located between a ratio for a dialog and that for a newspaper article. In this sense, novels are a balanced source that can best reflect our language life.

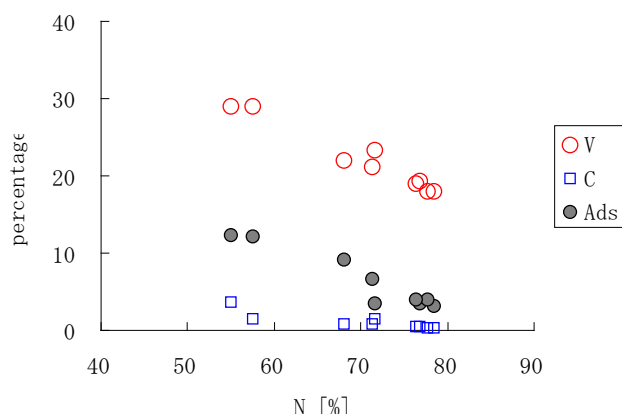


Figure 3. Ratios of nouns and other parts of speech.
N: Nouns; V: Verbs; C: Conjunctions and interjections;
Ads: Adjectives, adverbs and attributes.

Next, we extended the statistical data obtained from the selected sample back to the population which had 50 times as large a size as the sample did. This estimation showed that about 2% of all the running words were English in origin or from other European languages. Miyajima (1997) reviewed the data of the foreign component in texts of 90 kinds of magazines issued in 1956 (NIJL, 1962). According to the analyses, foreign words took 2.9% of the token. More than 85% of those words were from English, and only 1.2% from Portuguese. He listed 21 types of Portuguese words appearing in the texts. This shows the position of Portuguese words in the foreign lexicon has not changed significantly for this half-century.

5. Discussion

For these 100 years, some words derived from Portuguese have truly disappeared; for example, the ratio of Portuguese words to all the foreign words decreased from about 20% to only 2%. The details of statistic analysis can illustrate what kinds of categories have led to higher preservation than others. Words for clothing and food seem to maintain their popularity. Portuguese words for food appeared at 0.27 per mil in the early 20th century as shown in Table 1. That occurrence rate in this field decreased to 0.17 per mil in the modern texts. The frequent appearances of specific words, for instance “botão”, “veludo”, “pão” and “(pão de) Castella” (sponge cake) contribute significantly to this preservation. As a matter of fact, the words “pão”, and “tempero” are often combined with native or Chinese-origin words to form compound nouns, for example “an-pan” (bean-jam bun, or more concretely, round bread that has red bean paste inside), and furthermore to likely turn to be elements of the contractions, for example “ten-don” for ‘tempero donburi’ (a bowl of deep-fried seafoods and vegetables with soy-and-fish-based sauce over steamed rice). We can draw high schematicity for crisp and juicy in such construction. The part “ten” (heaven) is associated with a feeling of high temperature, lightening, bouncing, or rising. This formation suggests a high degree of assimilation into the Japanese lexicon. The word “capa” was also likely to be combined with native words. In the texts of the early 20th century, seven cases out of the eight occurrences were in Chinese characters, and four cases had native nouns or a quantifying prefix of Chinese origin, for example “ama-gappa” (rain coat), “abura-gappa” (oil-cloth coat), and “han-kappa” (half-size coat or short coat) as shown in Figure 4.

事を終わって出発するような伊勢参宮の講中もある。
黒の半合羽を着たまま奥の方に腰掛け、膳を前にして
、供の男を相手にしきりに箸を動かしている客もある

Figure 4. One case of transcription of “capa”
(the underlined part).

As summarized in Table 1, words in the amusement field seemed to gain in popularity, but the frequent use of a single word, “balanço”, in one novel resulted in this high value. In addition, we should note that this value included two Austroasian words, “khsier” (tobacco pipe) and “lao” (pipe tube), brought by

Portuguese traders. To improve this kind of large fluctuation in word occurrence, a larger variety of sources will be necessary. For reference, the Brown Corpus includes 500 kinds of sources, and the National Institute for Japanese Language gathered articles from 90 kinds of magazines. Asai and Ishikawa (2006) analyzed language styles in the information technology contexts by means of 101 kinds of text sources.

Table 1. Occurrence rates of Portuguese words.

categories	novels	
	modern	Meiji
clothing	0.08	0.11
religion	0.06	0.03
food	0.17	0.27
products	0.02	0.01
amusements	0.08	0.07

[per mil]

While this study investigated characteristics of foreign vocabulary by chronology and genre, we obtained some interesting findings. We should watch a trend that the terminology for soccer is coming from Brazilian Portuguese into Japanese. For example, “volante”, a mid-fielder who takes an important role of controlling the game in a key position between offensive and defensive players, and “seleção”, a player selected to a national team, are frequently observed in sports sections of newspapers. Words can be transferred from an advanced culture to a demanding culture. The team names of popular professional clubs are “Verdi” Tokyo named after “verde” and “Júbilo” Iwata for which Dunga once played. In these cases, foreign words are used for symbolizing professionalism. Of course English or Spanish names are popular because they are easier to understand than Portuguese words. English is a mandatory subject in junior high schools, senior high schools, and colleges. Spanish is a popular elective subject at many colleges, and several universities offer departments of Spanish and Latin cultures. On the other hand, Portuguese is an elective subject at a limited number of colleges. A couple of examples of English names are Kashima “Antlers”, at which Zico spent his last period as a player, Urawa “Red Diamonds”, and simply “Football Club” Tokyo. Apart from soccer, many baseball players wear “miçanga” for good luck. Although this word did not appear in the corpora of this study, we often hear the word on TV. Baseball is the most popular sport in Japan, and we can enjoy watching ballgames almost everyday during the pennant race. In brief, people can easily learn about new words when the topics are entertainment.

These observations also indicate that a cultural aspect is definitely an influencing factor in the lexicon (Fujimoto, 1988). We sometimes argue whether or not we can accept foreign words in Katakana characters with no semantic translation. This is largely because the general public is required to learn many technical terms today (Asai, Ishikawa, & Se, 2005; Asai & Sakai, 2005; Inoue, 2000). If new foreign words are to express interesting food and amusement, people can easily add those words to their vocabulary (Toyama, 1993). In fact, this shows a psychological aspect of acceptability: People are willing to use warm vocabulary but hardly learn cold vocabulary.

6. Conclusion

The borrowing of Portuguese words into Japanese began in the 16th century, and survived the 250-year seclusion. Certainly, for these 100 years, a ratio of Portuguese words to foreign words has decreased from 20% to only 2%. English has been taking a more significant role in the Japanese lexicon. However, several Portuguese words for clothing, food and amusements are still used at about 0.3 per mil in total on the token base. This can indicate that those items are part of Japanese life. A key to the adoption was the elaborate ways of transcribing with ideographic Chinese characters to achieve the highest appropriateness both phonetically and semantically. The Japanese people have thus been using some words of Portuguese origin without any sense of foreignness, and are adopting new words in a particular area of sports. This study reports on how lexical changes reflect concretely what kinds of culture have a major effect on communication life in the society.

Acknowledgments

The author would like to express his gratitude to Mr. Michael McCafferty of Indiana University at Bloomington for his useful comments on language styles. The author also is grateful for being allowed to download the novels in Aozora Bunko, which offers electronic plain texts whose copyrights have expired, and the other cited sites, and furthermore to use the parser Chasen and the Perl script engine. The author assumes all responsibility for any errors or flaws that may remain.

7. References

Aozora Bunko. (2006). <http://www.aozora.gr.jp>

Asahi Shimbun, (2006). Asahi Shimbun online database Kikuzo: <http://www.asahi.com>

Asai, A., & Ishikawa, Y. (2006). "Power of English seen in the new IT Japanese corpus", Paper presented at the 1st World Congress on the Power of Language, Bangkok, Thailand.

Asai, A., Ishikawa, Y., & Se, M. (2005). "A study on influences of loanwords in the IT field", Proceeding of the 15th Meeting of Japanese Association of Sociolinguistic Science, pp. 140-142, Tokyo.

Asai, A., & Sakai, M. (2005). "Joho senmon yogo to gakushu fuan tono kankei" [Relationship between terminology on information technology and anxiety in learning]. Proceedings of the 47th Annual Meeting of the Japan Association of Educational Psychology, p. 363, Sapporo.

Fujimoto, S. (1988). "Kodai Tyosen no gengo to moji bunka" [Languages and cultures in Old Korea]. In Toshio Kishi (ed.) "Languages and cultures" [Kotoba to bunka], pp. 175-240, Tokyo: Chuokoron.

Goto, A. (1988). "Kanbun no juyo" [Acceptance of Chinese literature]. In Toshio Kishi (ed.) "Languages and cultures" [Kotoba to bunka], pp. 241-261, Tokyo: Chuokoron.

Hoffer, B. (1990). "English loanwords in Japanese: Some Cultural Implications", Language Sciences, 12(1), 1-21.

Inoue, F (2000). "English as a Language of Science in Japan". In Ulrich Ammon (ed.) "The Dominance of English as a Language of Science on the Non-English Language Communities", Mouton de Gruyter.

Inoue, F. (2002). "Seiyo go no hatuon no eikyo" [Influences of pronunciations of European languages]. In Yoshifumi Hida and Takeyoshi Sato (Eds.) "Gendai nihon go koza 3 Hatsuen" [Modern Japanese Course: Pronunciation], pp. 59-84. Tokyo: Meiji Shoin.

Inoue, H. (2002). "Nihongo kansatsu noto" [A field note of Japanese], pp. 105-108, Tokyo: Chuokoron.

Kabashima, T. (1954). "Gendai bun niokeru hinshi no hiritsu to sono zogen no yoin nitsuite" [The ratios of parts of speech in modern literature and the factors of their increases and decreases], Kokugogaku, 18, 15-20.

Kabashima, T. (1955). "Ruibetsushita hinshi no hiritsu ni mirareru kisokusei" [The regularity of ratios of categorized parts of speech], Kokugo kokubun, 24(6), 385-387.

Kaganoi, S. (2002). "Nihongo wa shinka suru" [Japanese is developing], pp. 53-90, Tokyo: Nihon Hoso Shuppan Kyokai.

Kitahara, Y. (1973). "Hoju seibun to renjo shushoku seibun -- Watanabe Minoru Shi no renjo seibun nitsuiteno saikento" [Filling elements and adverbial modifying elements -- Revisiting Mr. Minoru Watanabe's study on adverbial elements], Kokugogaku, 95.

Maekawa, M. (1995). "1000-man nin no computer kagaku: Bungaku" [Computer science for ten million people: Literature], Tokyo: Iwanami.

- Matsumoto, Y., Kitauchi, A., Yamashita, T., Hirano, Y., Matsuda, H., Takaoka, K., & Asahara, M. (2000). "Japanese Morphological Analysis System ChaSen version 2.2.1", Nara Institute of Science and Technology.
- Miyajima, T. (1977). "Goi no taikai" [Lexical system]. In Susumu Ohno and Takeshi Shibata (eds.) "Goi to imi" [The lexicon and meanings], pp. 1-41, Tokyo: Iwanami.
- Miyajima, T. (1989). "Gendai nihongo: Goi" [Modern Japanese: Vocabulary]. In Takashi Kamei, Rokuro Kono, & Eiichi Chino (eds.) "The Sanseido Encyclopaedia of Linguistics", Vol.2, pp. 1750-1757, Tokyo: Sanseido.
- Miyajima, T. (1997). "Zasshi 90 shu hyoki hyo no tokei" [Statistics of representations in 90 kinds of magazines], Japanese Language Science, 1, 92-104.
- Muller, S. H. (1964). "The world's living languages", NY: Frederick Ungar.
- NIJL. (1962). "Gendai zasshi 90 shu no yoji jogo (1) soki oyobi goi hyo" [Words and their usage in 90 kinds of modern magazines (1) general remarks and vocabulary lists], edited by the National Institute for Japanese Language, Tokyo: Shuei Shuppan.
- Ohno, S. (2000). "Nihon-go no keisei" [Formation of Japanese language], p. 78, Tokyo: Iwanami.
- Ozawa, K. (1976). "An investigation of the influence of the English language on the Japanese language through lexical adaptation from 1955 to 1972", Ph.D. dissertation, Ohio University.
- Steinberg, J. D. (1996). "Lexical Borrowing and Modernization in China and Japan", Ph.D. dissertation, UCLA.
- Tokieda, M. (1950). "Nihon bunpo – Kogo" [Japanese Grammar – Colloquialism], pp. 130-131, Tokyo: Iwanami.
- Toyama, S. (1993). "Gairai-go wo ikeireru shinri" [Psychology of accepting foreign words]. In Heisei-no gairaigo hyakuninissu henshu-iinkai (ed), "Kotoba-dokuhon gairaigo", pp. 48-60, Tokyo: Kawadeshoboshinsha.
- Yamaguchi, M., Mogi, T., Kiryu, R., & Tanaka, M. (2004). "Goshu tonon kankei ni motozuita shimbun kiji niokeru goi no jikanteki henka bunseki" [An analysis of chronological changes of vocabulary concerning word types], Proceedings of the 13th Annual Meeting of the Japanese Association of Sociolinguistic Sciences, pp. 113-116, Tokyo.
- WEB Quilala et al. (2006). Online novel sites: <http://www.quilala.jp/>, <http://nnr.netnovel.org/>, <http://www.newvel.jp/>