

PROVIDING ETYMOLOGICAL INFORMATION FOR SINITIC LOANWORDS IN THE KBBI INDONESIAN DICTIONARY

David Moeljadi

Kanda University of International Studies, Japan

davidmoeljadi@gmail.com

Abstract

This paper documents the process of adding the etymological information of loanwords from Sinitic languages in Indonesian language into the KBBI Indonesian dictionary fifth edition, the most comprehensive and authoritative Indonesian monolingual dictionary, published by The Language Development and Cultivation Agency, under the Ministry of Education and Culture. It is a part of the etymology project which involves experts from universities in Indonesia (Moeljadi et al. 2019). Data of Sinitic loanwords from various sources such as Schlegel (1891), Hamilton (1924), Png (1967), Leo (1976), Kong (1994), and Jones (2009) were compiled. Data selection is based on the dictionary headwords, thus words which are listed in the KBBI dictionary were chosen and further analyzed. Finally, a database of Sinitic loanwords for the KBBI dictionary was built.

Historically, there are four major periods associated with external cultural and linguistic influence in the Indonesian archipelago: (1) Indian, (2) Chinese, (3) Islamic, and (4) European (Blust 2009). As of February 2021, the KBBI dictionary has etymological information of loanwords from Semitic languages (especially Arabic) and Indic languages (especially Sanskrit). Since languages in southern part of China were the early donor languages, it is worth adding the etymological information of loanwords from those languages into the KBBI dictionary.

The earliest instance of a Sinitic loanword is *tahu* ‘bean curd’ which is attested in an Old Javanese inscription from the tenth century (Jones 2009). Some tools such as *gunting* ‘scissors’ which is also found in Old Javanese texts might be borrowed from a southern Chinese language (Blust 2009). In the early Ming period after 1368, various Sinitic loanwords were borrowed through trade such as *opau* ‘money belt, small wallet’ and *honcoe* ‘smoking pipe’ (Blust 2009).

I found that there are more than 350 Sinitic loanwords in the KBBI dictionary. Regarding semantic domains, many of them are related to food, tradition and customs, and commerce, the rests are related to tools, clothes, kinship terms, martial arts, opium, prostitution, medicine, etc. Regarding donor languages, most of them are from Hokkien, others are from Cantonese, Hakka, and Mandarin.

Keywords: Sinitic loanwords, KBBI dictionary, lexical borrowing

1 Introduction

Kamus Besar Bahasa Indonesia (KBBI) is the official dictionary of the Indonesian language, published by Badan Pengembangan dan Pembinaan Bahasa (The Language Development and Cultivation Agency) or Badan Bahasa, under the Ministry of Education and Culture, Republic of Indonesia. Up until present, KBBI is the most comprehensive and the most authoritative reference for the Indonesian language. Etymological information was added in October 2019 for Semitic (especially Arabic) loanwords and in October 2020 for Indic (especially Sanskrit) loanwords. This paper discusses the inclusion of etymological information from Sinitic languages into the KBBI database which is planned in October 2021. It is a part of the KBBI etymology project (Moeljadi et al. 2019).

1.1 Scope of research

The term “Sinitic loanwords” in this paper refers to those loanwords from various languages in China, especially Hokkien, Hakka, Cantonese, and Mandarin, which are completely borrowed and thus listed as words in Kamus Besar Bahasa Indonesia (KBBI). Sinitic loanwords appear to be more widely used in the late colonial period than in more recent times, both by Chinese and non-Chinese speakers (Hoogervorst 2017). For example, the personal pronouns *gua* (我) ‘I’ and *lu* (汝) ‘you’ have been taken over by non-Chinese speakers of Betawi Malay and several other varieties. These pronouns are still used nowadays. The use of pronouns *bwansing* (晚生) ‘I’ and *owe* (喂) ‘I’ is restricted to the ethnic Chinese (Nio 1955: 43-44). To the best of my knowledge, these pronouns can be considered as archaic.

It is important to make a distinction between Sinitic lexical influence that has entered the mainstream Malay/Indonesian language and loanwords only understood by ethnic Chinese (Leo 1975). Thus, there are two types of Sinitic loanwords. The first one is those which are used by both ethnic Chinese and non-Chinese speakers, standardized, and thus listed in KBBI.¹ The second one is those which are used only by ethnic Chinese, not standardized (there are variations in spellings) and thus not listed in KBBI. Regarding the first type, we can divide into two groups: the first one is those which are used until present-time or those having attested status as words in the present language, whether people are aware that they are borrowed or not, for example *mi* ‘noodles’ and *tahu* ‘tofu, bean curd’. The second one is archaic words, such as *kimantu* ‘derogatory term for a Chinese newcomer’. Such words are labelled “ark” (a short form of *arkais* ‘archaic’) in KBBI. Similarly, regarding the second type, we can divide into two groups, i.e. those which are used until present-time (widely or narrowly in some communities) such as *Cungkuo* ‘China’ and those which were used in the past time or archaic, such as *owe* ‘I’. The present paper only deals with the first type, i.e. those which are listed in KBBI. See Figure 1 for the types of Sinitic loanwords.

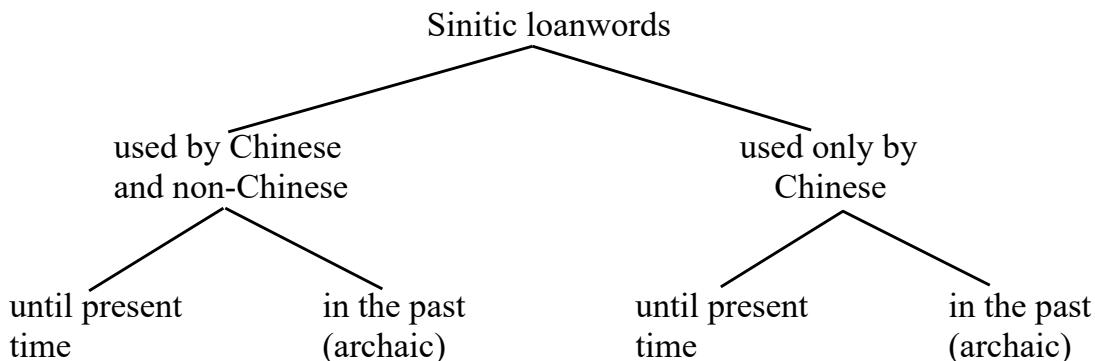


Figure 1. Types of Sinitic loanwords in Indonesian

In addition to Sinitic loanwords, there are Sinitic loan translations, as well as hybrid forms and ad hoc creations. Constructions such as *nasi pagi* ‘breakfast’ or *makan pagi* ‘breakfast’, *moeloet pintoe* ‘doorway’ and *keloear pintoe* ‘to go out’ are literal translations of Chinese 早飯, 早膳, 門口, dan 出門 respectively (Salmon 1974). The present paper does not deal with these loan translations although some of them are listed in KBBI.

1.2 Historical background and Sinitic influence in Malay lexical sources

Contact with speakers from China had happened from the seventh to the tenth century A.D., when Chinese merchants traded to Riau Islands, West Kalimantan, and East Kalimantan, even until North Maluku, long

¹ KBBI has language labels. For Chinese or Sinitic languages, the language label is “Cn” (a short form of *Cina* ‘Chinese’). Lexical entries which are used particularly by ethnic Chinese are given the label “Cn”.

time before the arrival of Portuguese and Dutch people. When Sriwijaya kingdom appeared and became strong, China also opened a diplomatic relation with Sriwijaya to secure its trade and shipping business. In the year 922, Chinese travelers visited Kahuripan kingdom in East Java. Since the 11th century, hundreds of thousands of Chinese migrants left their ancestral land and settled in many parts of the Archipelago.² During the Dutch colonial period, more Chinese migrants who were contracted by the Dutch came to the archipelago. The Chinese population increased. In 20th century during the revolutionary movement in China, more and more Chinese people came to Indonesia.

The influence of Sinitic languages in Malay can be seen from the lexical sources. Before paper dictionaries, there are word-lists or lists of words in a foreign language with the equivalent meaning in Malay. The earliest extant word-list of Malay is a Chinese-Malay vocabulary, dated to the 15th century, containing 482 entries which is written wholly in Chinese characters and employed both to give the Chinese word and for a transcription of the sound of the Malay word (Edwards and Blagden 1931). It already contains a number of Sinitic loanwords. The next word-list is the one compiled by Antonio Pigafetta, an Italian, from materials collected in about 1521, probably from the eastern islands of Indonesia. This contains some 426 items, the Italian word being given first, followed by the Malay equivalent (Marsden 1984). It has at least one Sinitic loanword. From colonial times, all documented varieties of Malay seem to have undergone some degree of Chinese influence. During the 19th century and the beginning of 20th century of colonial period, some word-lists related to Sinitic loanwords were published in the Archipelago, such as Schlegel (1891) and Hamilton (1924).

In the 20th century, there are a number of publications related to Sinitic loanwords which are mentioned in the following chapter.

2 Method

Data of Sinitic loanwords in Malay/Indonesian from various sources were gathered and compiled from 2019 to 2020. Table 1 summarizes the data sources.

Table 1. Data sources

No.	Source	Number of Sinitic loanwords	Comments
1	Blust and Trussel (2010)	8	The Austronesian Comparative Dictionary (web edition) with information on loanwords ³
2	Chow (2010)	515	M.A. thesis on Sinitic loanwords in Standard Malay
3	Hamilton (1924)	189	Sinitic loanwords in Malay Peninsula
4	Jones (2009)	1,469	Sinitic loanwords in Indonesian and Standard Malay
5	Kong (1994)	1,046	Sinitic loanwords in Indonesian and Standard Malay
6	Leo (1976)	288	Sinitic loanwords spoken by the inhabitants of Jakarta
7	Png (1967)	416	Sinitic loanwords in Malay
8	Schlegel (1891)	92	Sinitic loanwords in Malay
9	Sutami (2016)	407	Sinitic loanwords in Indonesian

In addition, eleven words which are not mentioned in any of those sources but are listed in KBBI were

² The Archipelago or *Nusantara* refers to Malay-related cultural and linguistic lands, such as the present Indonesia, Singapore, Brunei, and Malaysia. Standard Malay spoken in Malaysia and Indonesian spoken in Indonesia are two standardized varieties of the Malay language. There are other Malay varieties such as Singapore Malay and Brunei Malay.

³ https://www.trussel2.com/acd/acd-lo_a.htm

manually added, i.e. *butongpai* ‘a kind of martial arts’, *micin* ‘MSG, vetsin’, *laucu* ‘Laozi’, *tokwi* ‘tablecloth’, *shou sui* ‘a tradition on the night of Chinese New Year’, *syantung* ‘finely woven cloth from Shandong’, *takoah* ‘bean curd skin’, *hoisem* ‘sea cucumber’, *saucu* ‘grilled pork’, *ako* ‘elder brother’, and *dizi* ‘Chinese flute’.

Loanwords from each source, together with its details and explanations, were gathered and summarized in a table which contains the following information: part-of-speech, ID, Indonesian word in KBBI, Indonesian word in the source, Indonesian word meaning, source language, original word, Chinese character, original meaning, semantic domain, and earliest in corpus. The Indonesian words and parts-of-speech are taken from KBBI. The Indonesian/Malay words, word meanings, source languages, original words, Chinese characters, and original meanings are taken from the source directly. Thus, the Indonesian/Malay words in the sources and Indonesian words in KBBI sometimes differ in orthography. The ones in KBBI are the standard ones. The data sources also differ in the information on donor languages; some sources have “Chinese”, while some have “Hokkien” or “Hokkian”. The semantic domains were decided by myself. The information on “earliest appearance in corpus” was based on the Malay Concordance Project (MCP)⁴ and the Austronesian Comparative Dictionary (ACD). Table 2 contains three example items from the Sinitic loanwords data I have compiled.

Table 2. Some examples of Sinitic loanwords data

ID	SCHLEGEL_83	KONG_364	LEO_2_12
Indonesian word in KBBI	<i>cuki</i>	<i>honcoe</i>	<i>ceki</i>
part-of-speech	noun	noun	noun
Indonesian/Malay word	<i>tjuki</i>	<i>huncue</i>	<i>ceki</i>
Indonesian/Malay word meaning	a kind of draughts played with white and black beans	<i>paip penghisap tembakau</i> ‘pipe for smoking tobacco’	a Chinese card game the same as <i>capjiki</i>
Source language	Chinese	Fujian/Hokkian	Hokkien
Original word	<i>tioh kî</i>	<i>hun ts'ui</i>	<i>tsit-kî</i>
Chinese character	着棋	熏喙	一枝
Original meaning	to play at draughts or chess	-	<i>tsit</i> , one; <i>kî</i> , piece; literally, one piece
Semantic domain	gambling	tobacco	gambling
Earliest in corpus	Hikayat Bayan Budiman, 1371 (MCP)	probably during the Ming dynasty (ACD)	Warkah Buton, 1790s? (MCP)

During this compilation process, I found challenges regarding data sources, as follows.

1. Different sources have different information on donor or source languages, as noted in Table 2. Most of the sources only mention Hokkien or Hokkian or Fujian as the donor language while Jones (2009) writes the dialect names of Hokkien, e.g. Amoy (Xiamen) dialect, Changchiu (Zhangzhou) dialect, Tsoanchiu (Quanzhou) dialect etc.
2. Different sources have different romanization systems.
3. Different sources have different Chinese characters, for example the Chinese characters for *cincau* ‘grassjelly’ are 清草 in Jones (2009) and Sutami (2016), 青草 in Kong (1994) and Leo (1976), and 薦草 in Schlegel (1891). There is no standardization in writing Chinese characters for Hokkien because Hokkien is mainly a spoken language.
4. Authors have different opinions on which words can be considered as Sinitic loanwords, i.e. loanwords which are used in the present modern Indonesian or those which are used in pre-independence or during late-colonialism (Sino-Malay literature).

5. There is lack of documentation regarding when a word is recorded for the first time and thus we do not know the first appearance of most of the words in documents and published materials as well as their meaning development.

After the data compilation, I did a manual entry selection process i.e. I chose one entry from the candidate entries based on the amount of information. I chose the most informative and specific one, as illustrated in Table 3. If there is only one entry from one source, I examined and decided if it is a suitable entry and can be selected or it is a doubtful entry and marked it as “doubtful”.

Table 3. Entry selection

ID	JONES_544	LEO_1_27	SUTAMI_187
Indonesian word in KBBI	<i>jok</i>	<i>jok</i>	<i>jok</i>
part-of-speech	noun	noun	noun
Indonesian/Malay word	<i>jok</i>	<i>jok</i>	<i>jok</i>
Indonesian/Malay word meaning	seat of car or pedicab*; cf <i>loanjok</i>	mattress, seat of car or pedicab	<i>alas tempat duduk</i> ‘seat cover’
Source language	Chiangchiu (Zhangzhou)	Hokkien	Chinese
Original word	<i>jiòk</i>	<i>dziók</i>	<i>rì</i>
Chinese character	褥	褥	褥
Original meaning	-	<i>dziók</i> , mattress	-
Semantic domain	vehicle	vehicle	vehicle
Notes	selected	-	-

Afterwards, I summarized all the chosen entries in a table for KBBI database. There is a column for data sources. All sources which support the chosen entries should be mentioned in this column. I follow Carstairs Douglas' romanization system for Hokkien (Douglas 1899). The format of the table is shown in Table 4.

Table 4. The chosen entry for *jok* ‘seat of car or pedicab’ in KBBI database format

node_ID	entry_ID	entry	language	orig_word	translit	orig_meaning	source_ID
10199	34790	<i>jok</i>	Hokkien Zhangzhou dialect	褥	<i>jiòk</i>	<i>matras, kasur</i> ‘mattress’	JONES, LEO, SUTAMI

3 Result

After the selection process, there are 357 Sinitic loanwords and 40 doubtful entries. The 357 loanwords belong to six parts-of-speech: 307 nouns, 22 adjectives, 12 numerals, 10 verbs, 4 pronouns, and 2 particles (interjection and expression). It is interesting to know that there are more numerals borrowed than verbs. The borrowed numerals are, for example, *ceban* ‘ten thousand’, *ceceng* ‘one thousand’, *cepek* ‘one hundred’, *goban* ‘fifty thousand’, *goceng* ‘five thousand’, and *gopek* ‘five hundred’ which I think correspond to the Indonesian currency denominations and they are usually used in business or commerce.

Regarding the archaic words, 18 out of 357 Sinitic loanwords (5%) are labelled as archaic in KBBI. Regarding the words mainly used by ethnic Chinese, 88 of them (24.6%) have Chinese as the language label, i.e. people still recognize them as Sinitic loanwords. It means that more than half or most of the Sinitic loanwords are used by both ethnic Chinese and non-Chinese. 3.4% or 12 of them have Jakarta Malay as the language label, i.e. the Sinitic words were borrowed via Jakarta Malay speakers in Jakarta.

In addition, there are two words labelled “Mal” (Standard Malay), one word labelled “Jw” (Javanese), one word labelled “Sd” (Sundanese), and one word labelled “Jb” (Jambi Malay). Through these language labels, we know that some Sinitic words were borrowed via Malay, Javanese, Sundanese, and Jambi Malay.

I classified the Sinitic loanwords into 37 semantic categories as follows: 85 of them (23.8%) are related to food, 48 of them (13.4%) tradition and customs, and 36 of them (10.1%) commerce, the rests are related to clothes (5.9%), kinship terms (3.9%), gambling (3.9%), medicine (3.9%), house (3.9%), tools (3.6%), martial arts (3.6%), seafaring (2.5%), body parts (2.5%), prostitution (2.2%), opium (2%), place (2%), vehicle (1.4%), government (1.1%), tobacco (1.1%), plants (0.8%), animals (0.8%), and others (6.2%). Regarding donor languages, most of them are from Hokkien (89.4%), others are from Cantonese (5.3%), Mandarin (3.1%), and Hakka (1.7%).

4 Analysis and Discussion

This paper has reached its goal in terms of adding etymological information for Sinitic loanwords into the KBBI. However, in order to make a proper, comprehensive etymological information for Sinitic loanwords, etymological information for words particularly used by ethnic Chinese and Sino-Malay words should be added and extensive etymological research based on corpora should be conducted. In order to get more reliable data with years of occurrences, digital corpora of Malay/Indonesian should be made. The following data sources can be employed for that purpose.

1. Malay Concordance Project (MCP). I employed this source to check the occurrences of Sinitic loanwords in Classical Malay texts between 1302 and 1953. This source contains 165 texts and 5.8 million words, including 140,000 verses. The corpora have been digitized so we can do various kinds of searches.
2. Monash University’s Sin Po newspaper collection.⁵ This collection contains newspaper articles in Sino-Malay language between April 1923 and December 1941. It only has PDF and microfilms.
3. University of Washington collection which contains two kinds of Sino-Malay literature, written between 1886 and 2000, owned by two temple libraries in Java.⁶ The first one is 5,000 scanned pages from a collection of religious books and magazines in Chinese and Malay languages. The second one is 12,500 pages of popular Sino-Malay novels and magazines.
4. National Library of Indonesia collection of Sino-Malay literature between 1804 and 1950.⁷ It contains 14,162 pages from 122 titles of novels, magazines, story books, poems, and picture books. All of them are in PDF.
5. Leiden University Library collection of Sino-Malay literature between 1870s and 1950s.⁸ It consists of two newspapers (Sin Po and Hoakiao) and circa 1,400 books (dime novels, educational works, translations, works on religion, poems, and literary works). However, it is not open to people outside Leiden University.

Sources 2 to 5 have the data saved in PDF and/or microfilms. The data should be digitized first before being able to be employed for various kinds of searches. In the end, we should answer the following etymological questions which we cannot answer all of them only by data gathering, compilation, and selection described in this paper:

1. What are the source or donor languages?
2. When were the words borrowed or when did they appear for the first time in which text and what

5 <https://repository.monash.edu/collections/show/117>

6 <https://digital.lib.washington.edu/researchworks/handle/1773/21474>

7 <http://e-resources.perpusnas.go.id/library.php?id=00031>

8 <https://digitalcollections.universiteitleiden.nl/sinomalaytexts>

did it mean?

3. Were they directly or indirectly borrowed? What were the intermediate stages of the borrowing process?
4. How did they reflect the social history at that time?
5. Were they borrowed together with many other (semantically related) words?
6. Did they replace the (previous) Malay word?
7. Did they undergo meaning shifts and sound changes?

In addition, Old Javanese and Sinitic sources should be referred to if possible. Regarding Sinitic sources, some of the challenges are to make corpora of Hokkien/Min Nan, Cantonese, and Hakka and to conduct research on etymological information of Sinitic words.

Acknowledgments

I thank Anna Šlajová who helped me input most of the data into an Excel spreadsheet. I gratefully acknowledge the support of the European Regional Development Fund Project “Sinophone Borderlands – Interaction at the Edges” CZ.02.1.01/0.0/0.0/16_019/0000791.

5 References

- Blust, Robert A. (2009). *The Austronesian languages*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Blust, Robert and Stephen Trussel. (2010). Austronesian Comparative Dictionary, web edition (www.trussel2.com/ACD). Revision 21/06/2020. Accessed on 22/05/2021.
- Chow, Chai Khim. (2010). *The Study of Loanwords between Chinese Language and Malay Language in Malaysia*. M.A. thesis. National University of Singapore and Peking University.
- Douglas, Carstairs. (1899). *Chinese-English Dictionary of the Vernacular or Spoken Language of Amoy, with the Principal Variations of the Chang-chew and Chin-chew Dialects*. London: Publishing Office of the Presbyterian Church of England.
- Edwards, E. D. and Blagden, C. O. (1931). A Chinese vocabulary of Malacca Malay words and phrases collected between AD 1403 and 1511 (?). *Bulletin of the School of Oriental Studies*, University of London, 6(3), 715-749.
- Hamilton, A. W. (1924). Chinese loan-words in Malay. *Journal of the Malayan Branch of the Royal Asiatic Society*, 2(1 (90), 48-56.
- Hoogervorst, T. G. (2017). What kind of language was ‘Chinese Malay’ in late colonial Java?. *Indonesia and the Malay World*, 45(133), 294-314.
- Jones, Russell. (2009). *Chinese Loan-words in Malay and Indonesian: A Background Study*. Kuala Lumpur: University of Malaya.
- Kong Yuanzhi. (1994). Kata pinjaman Bahasa Cina dalam Bahasa Melayu. In *Jurnal Dewan Bahasa*, Aug., pp. 676-702; Sept., pp. 772-795.
- Leo, Philip. (1975). *Chinese Loanwords Spoken by the Inhabitants of the City of Jakarta*. Jakarta: Lembaga Research Kebudayaan Nasional L.I.P.I.
- Marsden, William. (1984). *A dictionary and grammar of the Malay language*, [Facsimile reprint of the 1812 ed.], Singapore: Oxford University Press, 2 vols.

- Moeljadi, David, Ian Kamajaya, and Azhari Dasman Darnis. (2019). Considerations for Providing Etymological Information in the KBBI Indonesian Dictionary. In Mehmet Gürlek, Ahmet Naim Çiçekler, and Yasin Taşdemir (Eds.), *Proceedings of the 13th International Conference of the Asian Association for Lexicography*, Istanbul University, pp. 161–178. Istanbul: Asos Publisher.
- Nio, Joe Lan. (1955). Kata-kata ‘goea’ dan ‘loe’ di Indonesia [The words ‘goea’ and ‘loe’ in Indonesia]. *Bahasa dan Budaja* 3 (3):41–44.
- Png, Poh-Seng. (1967). A preliminary survey of Chinese Loan-words in the Malay Language. The Journal of Southeast Asian History. Singapore: The Island Society.
- Salmon, Claudine. (1974). Les traductions de romans chinois en malais (1880–1930) [Translations of Chinese novels into Malay]. In P-B. Lafont and D. Lombard (eds), *Littératures contemporaines de l'Asie du sud-est* [Contemporary literatures of Southeast Asia]. Paris: l'Asiathèque, pp. 183–201.
- Schlegel, G. (1891). Chinese Loan-words in the Malay Language. In *T'oung Pao*, I, pp. 391-405.
- Sutami, Hermina. (2016). Menelusuri Penggunaan Sumbangan Kosakata Bahasa Cina dalam Bahasa Indonesia. Lampiran tulisan untuk memperingati hari jadi ke-85 Prof. Dr. Muhamajir, Guru Besar