

## **Diachronic Corpora as a Tool for Tracing Etymological Information of Indonesian-Malay Lexicon**

**\*Kamal Yusuf<sup>1</sup>, Dewi Puspita<sup>2</sup>**

UIN Sunan Ampel Surabaya, Indonesia<sup>1</sup>; Universitas Indonesia, Indonesia<sup>2</sup>

**E-mail:** [kamalinev@gmail.com](mailto:kamalinev@gmail.com)<sup>1</sup>, [dewi\\_puspita18@yahoo.com](mailto:dewi_puspita18@yahoo.com)<sup>2</sup>

Corresponding Author\*

DOI: <https://doi.org/10.18326/rqt.v13i1.153-182>

### **Submission Track:**

Received: 07-02-2020

Final Revision: 23-05-2020

Available online: 01-06-2020

### **ABSTRACT**

*Indonesian lexicon comprises numerous loanwords which some of them already exist since the 7<sup>th</sup> century. The large number of loanwords is the reason why many dictionaries of Indonesian etymology available today contain merely the origin of the words. Meanwhile, there are several aspects in a word etymology that can be studied and presented in a*

*dictionary, such as the change in a word form and in its meaning. This article seeks to demonstrate the use of corpora in identifying the etymological information of Malay words from diachronic corpora and to figure out the semantic change of the Malay words undergo from time to time until they turn out to be Indonesian lexicon. More specifically, two selected Malay words were examined: bersiram and peraduan. By exploring data resources from the corpus of Malay Concordance Project and Leipzig Corpora, this study attempts to collect etymological information of Indonesian lexicon originated from Malay by employing a corpus based research. The findings show that the examined words have changed in meaning through generalization and metaphor. However, unlike the word bersiram, the change that the word peraduan happened only occurs in semantic level. This information, ultimately, can be used as informative data for a more comprehensive Indonesian etymology dictionary. Drawing on corpus analysis, this paper addresses the importance use of diachronic corpora in tracing words origin.*

**Keywords:** *diachronic corpora, etymology, corpus analysis, semantic change, Malay-Indonesian*

## **INTRODUCTION**

For most language users, etymological information is perceived just evidence of which a word originated from, especially when a language absorbs many loanwords; Indonesian take as an example (Russel et al., 2007; Tadmor, 2009). Kridalaksana (2001) pointed out that the content of Indonesian etymology dictionaries which have been compiled and available today is merely an inventory of words origin which needs to be continued with research and interpretation from various aspects. This is in line with the opinion of Durkin (2009) and

Liberman (2009) who stated that the study of etymology is related to the history of a word, the history of meaning, formal history, or the history of its spread from one language to another, or from one group to another. In accordance with that, it is important to point out that at least there are six etymological information that can be applied to trace a word: (1) the year of usage, (2) the initial form (morphology) and the initial sound (phonology), (3) the language of the donor (for loan word), (4) the person who coined the word for the first time, (5) the initial meaning, and (6) the change of meaning. Therefore, an etymological dictionary should not only contain information of the word's origin but also be given more clear-cut description of a word.

Other things from Indonesian etymology dictionaries that are available until today is the scarcity of etymology information of words originated from Malay. Malay is the root of Indonesian (Teeuw, 1967; Andaya, 2001; Hoogervorst, 2015). In the early centuries, the language spoken in some part of the Indonesian archipelago and the Malay Peninsula might be the same. Over time, there are many things and events, socially and politically that affects the regions and causes the language to change and to be different. Information of changes that occur in Malay words--that now become the vocabulary of Indonesia, phonologically, morphologically, semantically, or syntactically-- are parts of the etymology information (Mohamed & Yusoff, 2014).

A number of studies have previously been done concerning the etymology and semantic change of words in various languages, to mention some are Wijaya & Yeniterzi, 2011; Yurivna, 2014; Jatowt &

Duh, 2014; Hasan, 2015; and Altakhaineh, 2018. However, to date, there is not ample works that pay their attention to scrutinize how the etymological information can be approached using corpora, especially in the relation of Malay and Indonesian language.

Wijaya & Yeniterzi (2011) identified semantic change of words over centuries using computational linguistics method. They used Topics-Over-Time (TOT) and k-means clustering on Google Books N-gram dataset. Through their methods, they show how clustering words that co-occur with an entity of interest in 5-grams can shed some lights to the nature of change that occurs to the entity and identify the period for which the change occurs. Yurivna (2014) only classified changes in meaning that occur in English medical terms. Classification of changes in meaning in question is specialization, generalization, pejoration or amelioration, also metaphor and metonymy.

Jatowt & Duh (2014) explored digitized historical texts, which were also carried out in our study. The difference is, Jatowt & Duh uses the NLP (Natural Language Processing) method while we used the corpus-based method. Another study was conducted by Hasan (2015) which dealt with semantic change of borrowing words, especially Arabic words in Bengali. This kind of research in Indonesian is plenty. In fact, most of etymological research in Indonesian is about borrowing words.

Altakhaineh (2018) examined the semantic change of positive vs. negative adjectives in Modern English. He compared the meaning of those adjectives in dictionary than look up their frequency of use in the corpus. He wanted to see whether the adjective had been negative or

positive from the beginning or whether the adjective turns negative or positive because there are things that cause it. The research object of the those researchs are English vocabulary.

Until now there has been no theory that specifically addresses the search for etymological information through diachronic corpora. Existing theories still separate theory of etymology and theories about corpus linguistics. Of the two theories, the theory referred to in this study is the theory of Collins (2003) and the theory of collocation via concordance from McEnery & Hardie (2012). The search for etymological information through diachronic corpus is nothing new for Indo-European languages, especially English. However, for the Indonesian language, especially those from Malay, no one has ever discussed this matter. Our current research deals with original words, i.e. the Indonesian words that are rooted from Malay, not from any foreign loan words. Most importantly, none of the previous studies used a corpus-based method, in particular comparing data through diachronic corpora.

This paper offers a tool that can be used to trace etymological information, especially to trace changes in their meaning. The tool that can provide a large collection of text from past centuries to be examined is diachronic corpora (de Melo, 2014). According to Allan & Robinson (2012), the use of corpus is the state of the art in the study of historical semantics, which is part of etymology study. Malay is fortunate to have Malay Concordance Project (MCP) developed by Australian National University (Proudfoot, 1991; Gallop, 2013). It consists of old classical

Malay manuscripts from 14<sup>th</sup> to 20<sup>th</sup> centuries that can be used to examine the usage of a Malay word during that time (Johary & Rahim, 2014). This present study, therefore, seeks to explore any etymological information of Malay words that become part of Indonesian lexicon which are still used until today by employing the MCP compared with a more recent potential corpus from the 21<sup>st</sup> century.

There are thousands of Malay-Indonesian original vocabularies. It would take a very long time to be able to analyze the entire original vocabulary. For this reason, as a preliminary study, the current research was conducted using data samples. Two samples were chosen to be presented in this paper; they are *bersiram* and *peraduan*. The sample selection process is explained in the research method section.

Thus, the aims of the current study have three folds: (1) to identify the etymological information of Malay words *bersiram* and *peraduan* from diachronic corpora, (2) to investigate what kind of changes those Malay words undergo from time to time until they turn out to be Indonesian lexicon, and at the end (3) to demonstrate the use of diachronic corpora as a tool in examining etymological Malay-Indonesia lexicon.

## **RESEARCH METHOD**

This research is a corpus-based research. To prove that etymological information can be collected from diachronic corpora, this study employed two corpora that were set in chronological order. We started with the methodological issue by selecting the proper corpora

collection available online. We found two major salient collections regarding Malay and Indonesian corpus. The first corpus is MCP, which comprises 5.8 million words (including 140,000 verses) from more than 165 sources of pre-modern Malay written text. The oldest script is from the year 1302 and the most up-to-date is from 1950 (Gallop, 2013; Bakar, 2020). However, the dates of some old scripts are somewhat hypothetical. The second corpus as a source for identifying the etymological information of this study is taken from the Indonesian corpus from Leipzig Corpora. This corpus is based on online material from 2012 to 2014 that consists of 74,329,815 sentences, 7,964,109 types, and 1,206,281,985 tokens (Richter et al., 2006; Biemann et al., 2007). The two corpora are available online and they demonstrate the context uses of Malay lexicon from the 14<sup>th</sup> to 21<sup>st</sup> century.

The search results of the words investigated from the two corpora then were analyzed qualitatively. The changes that each word undergoes were examined from the concordance lines and the word's collocations. Collocation analysis usually involved statistical measurement. Yet, McEnery & Hardie (2012) proposed a non-statistical method called collocation-via-concordance technique. In this technique, researchers must use their intuitive to check the concordance lines that yield up notable examples and patterns, and then examine each line individually. Therefore, with certain considerations, we undertook this technique for the recent study.

Regarding the data, we selected two samples from a number of Malay-Indonesian words to be further investigated as a model study in

this paper, i.e. *bersiram* and *peraduan*. Those words are taken from the list of honored words in Kamus Besar Bahasa Indonesia (KBBI). Honored words mean words that are used in formal situation and only for selected and respected people. There are 26 words in that list (Table 1). However, not all of them are originated from the Malay. Some of the words listed are originated from Sanskrit and Old Javanese. Most importantly, not all of them experience changes in their meaning. From that not so many Malay words that undergo changes in meaning, we found the word *bersiram* and *peraduan*.

**Table 1.** List of words labeled *hor* in KBBI

Nr.	Word	Nr.	Word
1	abangda	14	bersiram
2	adimas	15	dalem
3	adinda	16	eyang
4	ananda	17	jenazah
5	apa	18	mangkat
6	asma	19	meninggal
7	ayahanda	20	nenenda
8	ayunda	21	pakanira
9	baginda	22	pamanda
10	beliau	23	peraduan
11	beradu	24	pesiraman
12	bersantap	25	suaminda
13	bersemayam	26	surai

## RESULTS & DISCUSSION

Given the above description, we primarily present an analysis model of utilizing diachronic corpora to discover the etymological

information of Malay-Indonesia lexicon. We selected *bersiram* and *peraduan* and traced their use in the sentences deposited from the two corpora as presented below.

### ***Diachronic Use of ‘bersiram’***

The word *bersiram* is a high classical Malay word. The word has been recorded in the dictionaries of Malay (Kamus Dewan, 2015; Pusat Rujukan Persuratan Melayu, 2020) and Indonesian (Kamus Besar Bahasa Indonesia, 2020) with the meaning of ‘to take a bath’. The word can only be used for the royal family. In MCP, this word appeared 157 times in 24 old manuscripts dated from the year 1370s to 1930s (as can be seen in Table 1 and 2). *Bersiram* occurred 32 times in *Salasilah Melayu dan Bugis* (1865) and only one time in the manuscripts *Syair Ken Tambuhan* (~1750), *Hikayat Raja Bikrama Sakti* (1800), *Syair Nyamuk dan Lalat* (<1873), *Hikayat Purasara* (>1890), *Syair almarhum Sultan Abubakar* (1896), *Syair Raja Johor* (1899), and *Cerita Jenaka* (<1908).

**Table 2.** Distribution of results of *bersiram* in MCP  
~bersiram~ (1300-2000)

1370s	4	Bayan 4
1700s	2	Tuah 2
1750s	1	KT 1
1770s	5	ARM 5
1780s	5	Misa 5
1800s	1	Bik 1
1810s	17	PNB 17
1820s	8	Zub 8

1870s	5	Nymk 1, NAsik 2, KS 2
1880s	28	1880s 28 Pah 2, Mpt 26
1890s	5	SSiti 2, Puras 1, SAB 1, RJoh 1
1900s	1	CJen 1
1910s	6	J&P 6
1930s	13	TZA 13
157		

All those 157 tokens of *bersiram* demonstrate the same meaning in the contexts with those recorded in dictionaries. Below are some examples of *bersiram* in some contexts:

### 1370s

- (1) *sudah Élah kembali itu, maka bagindapun pergilah bersiram ke kolam itu. Setelah sudah baginda bersiram itu,*  
‘Elah came back, so the king takes a bath in the pool. After taking a bath,’

### 1770s

- (2) *Setelah selesailah daripada bercukur dan bersiram putera Baginda itu, maka datanglah bidan menjunjung duli ...*  
‘After the prince has shaved and took a bath, the midwife took a bow’

### 1810s

- (3) ...*anéka jenis daripada bungaan. Setelah sudah mandi bersiram maka naiklah segala puteri-puteri itu mengentas bunga2an ada yang ...*

‘various of flowers. After taking a bath then the girls get up to remove all those flowers...’

### 1890s

- (4) ... *sama elok parasnya. | Setelah genap tujuh hari, Bersiramlah baginda laki isteri, Dikerjakan oleh perdana menteri,*

‘Look as pretty. After seven days, the king and his queen took a bath, Done by the prime minister’

### 1910s

- (5) *Pada suatu hari Sultan Mahmud hendak berangkat bersiram, duduk di atas julangan, ditikam oleh Megat Sri Rama dengan ...*

‘Once upon a time Sultan Mahmud is going to take a bath, while sitting, stabbed by Megat Sri Rama using...’

### 1930s

- (6) ... *bestari, manakala siang keluar matahari, selesai bersiram mahkota negeri. | Berangkat keluar ia bertakhta, tersenyum ...*

‘...smart, when the sun rises in the afternoon, the crowned head took a bath. He comes out to enthrone, smiles...’

The above concordance lines show that the word *bersiram* collocated with the word *baginda* (king), *puteri-puteri* (princesses), *perdana menteri* (prime minister), Sultan Mahmud (King Mahmud), and

*mahkota negeri* (crowned head). The other concordance lines which are not presented here also show the same collocates. Those collocates indicate that the word *bersiram* is only used for the royal family. The line from the 1890s (as in the sentence 4) even shows that the bath was not just a usual bath, it was a kind of ceremony.

(4) ... *Setelah genap tujuh hari, **Bersiramlah** baginda laki istri, Dikerjakan oleh perdana menteri, ...*

‘... After seven days, The king and his queen **took a bath**, Done by the prime minister, ...’

**Table 3.** Frequency of occurrences *bersiram* in the MCP manuscripts

M&B Salasilah Melayu dan Bugis	1865	32
Mpt Hikayat Merpati Mas dan Merpati Perak	>1883 (MS 1887)	26
PNB Hikayat Perintah Negeri	1811 (MS 1811)	17
Dmsy.S Syair Raja Damsyik	1864	16
TZA Syair Tawarikh Zainal Abidin yang Ketiga	1936	13
Zub Syair Siti Zubaidah Perang Cina?	~1800 (MS 1840)	8
TN Tuhfat alNafis	1866 (MS 1890)	6
J&P Hikayat Johor serta Pahang	1917	6
ARM Adat Raja Melayu	1779, ~1850 (MSS 1817, 1873)	5
Misa Misa Melayu	~1780 (MS 1836)	5
Bayan Hikayat Bayan Budiman	1371 (MS 1852)	4
Tuah Hikayat Hang Tuah	~1700 (MS 1849)	2
NurP Syair Sultan Nur Peri	< ~1865	2
NAsik Hikayat Nakhoda Asik	~1870 (MS 1890)	2
KS Kitab Suci PL	1879, PB: 1935	2
Pah Hikayat Pahang	1883 (MS 1883, 1932)	2
SSiti Syair Seratus Siti?	~1890	2

KT Syair Ken Tambuhan	~1750 (MS 1791-1872)	1
Bik Hikayat Raja Bikrama Sakti	~1800 (MS 1830)	1
Nymk Syair Nyamuk dan Lalat	<1873	1
Puras Hikayat Purasara	>1890	1
SAB Syair almarhum Sultan Abubakar	1896	1
RJoh Syair Raja Johor	1899	1
CJen Cerita Jenaka	<1908	1

### ***The Semantic Change of ‘bersiram’***

After its independence in 1945, Indonesia has become a republic. The royal system is no longer used. For that reason, the frequency of use of the word *bersiram* might also be decreased. However, in a more recent corpus like Indonesian corpus in *Leipzig Corpora*, we can still find the use of the word *bersiram* in many different contexts (see graphic in Figure 1). The figure is an automatic generated collocation graphic from Leipzig Corpora. As can be seen, the graphic contains words related to the word *bersiram* found in the corpus. The graphic depicts that in their corpus, the word *bersiram* collocates or related to all those words. The collocates found in the graphic are different with those found in MCP corpus. In corpus linguistics, collocation means words which occur within the neighbourhood of another word. Those words don't have to be located right next to the main word (Baker et al, 2006: 36-37). The search of the word *bersiram* in Leipzig Corpora are turned in 55 lines. There are some interesting findings from the lines, such as:

- a. From 55 occurrences, only 15 of them have the literal meaning of ‘to take a bath’ or ‘to shower’. Ten lines, which come from Malaysian website, use the word *bersiram* as the equivalent of to take a bath or shower in the daily activity of common people, while the other five lines, which come from Indonesian website, still use the word only for a respected person.
- b. Six lines contain the word *bersiram* in a figurative meaning. In those lines, the word *bersiram* is used in the form of a metaphor. *Bersiram* is juxtaposed with *darah* (blood) and *cahaya* (light) as in the sentence (7):

(7) *Di kejauhan tampak gedung-gedung jangkung yang **bersiram** cahaya lampu.* (www.mayasanti.blogspot.com, crawled on 08/05/2012)

‘In the distance, tall buildings are seen *bathed* in light.’

- c. The most interesting finding is, 34 lines of them appeared in the contexts of food and carry a figurative meaning. In those lines, *bersiram* mostly collocates with *saus* (sauce), *jamur* (mushroom), *keju* (cheese), *cokelat* (chocolate). One example of the word usage in the context of food is as in the sentence (8):

(8) *Dari deretan menu terbaru, ada BBQ Beef Ribs & Alice Springs Chicken **bersiram** saus keju Monterey Jack-Cheddar.*

(www.femina.co.id, crawled on 06/02/2014)

‘From the latest menu, there are a BBQ Beef Ribs & Alice Springs Chicken *covered* with Monterey Jack-Cheddar cheese sauce.’



‘... then the King went to the pool *to take a bath.*’

The phrase *ke kolam itu* in above sentence (1) is not an object, but it is an adverb of place. An object is not needed after the word *bersiram* in that sentence.

However, in its figurative meaning, the verb *bersiram* has become transitive. Below is a concordance line of the verb *bersiram* in figurative meaning followed by its objects (in upright letters).

(10) *Tempat orang berniaga dikepalai seorang batin bijaksana yang mengharamkan negeri **bersiram darah.***

‘A place for trading is lead by a wise man who forbid the land from **bloodshed.**’

(11) *Di kejauhan tampak gedung-gedung jangkung yang **bersiram cahaya lampu.***

‘In the distance, tall buildings are seen **bathed** in light.’

(12) *Sejumput mi **bersiram saus** dengan potongan udang gemuk di atasnya.*

‘a pinch of noodles doused in sauce with a oiece of fat shrimp on its top.’

(13) *Versi Michel’s disebut Marble Mud Cake, **bersiram ganache** coklat putih dan coklat pekat.*

‘*Michel’s version* is called *Marble Mud Cake*, **covered** by ganache chocalate’

(14) *Dan, sebagai penutup pesanlah Roti Cane Gula atau Roti Cane Susu, **bersiram susu kental manis.***

‘And, as a desserts please order Roti Cane Gula or Roti Cane Susu, **watered** with condensed milk’

Objects in the above sentences are mandatory because without objects the sentences would be incomplete and meaningless.

### *Diachronic Use of ‘peraduan’*

Another example that we would like to present for tracing the semantic change and the etymological information utilizing MCP is the word *peraduan*. This word is a classical-high Malay as well, that is used strictly for the royal family. It has the meaning of ‘bed’ or ‘bedroom’. Compared to *bersiram*, the frequency of *peradaun*’s appearance in MCP was found higher. It appeared 357 times in 31 old manuscripts dated from the 1370s to 1950s. The word can be found in the manuscript *Syair Siti Zubaidah Perang Cina* (32 times) and mostly occurred one time in sixteen manuscripts (see Table 3). The sentences below are some examples from the concordance lines from the manuscripts showing the uses of *peraduan* in the contexts whereby all with the meaning of ‘bed’ and ‘bedroom’.

(15) .. *ketiganya itu pun masing-masing mendapatkan biliknya peraduan, lalu beradulah sekaliannya itu.*

‘...all of three get their **bedrooms**, then they sleep.’

(16) ... *beri rawan, sendu rupa kelakuan, buka ranjang peraduan. | Lalu makai Sinyor Gilang, baju lakan hitam gilang*

‘...be vulnerable, sentiment behavior, go to the **bed**. Then Sinyor Gilang wearing, a black trouser’

(17) *Maka Kuda Nestapa pun masuk ke dalam **peraduan** lalu menyingkap tirai kelambu itu. Maka dilihatnya Raden...*

‘...Then the Sorrow Horse gets into the **bedroom**, opens the curtain. So, he looks at Raden...‘

(18) *Sambil memakai bau-bauan. Adinda disambut masuk **peraduan**. / Lalulah duduk menanggalkan jubah,*

‘...wearing parfumes. Adinda welcomed into his **bedroom**. Then he took off his robe,‘

(19) *... biliknya dan pada tiap-tiap bilik itu ditaruhnya geta **peraduan** lengkap dengan kasur, tilam dan tirai ...*

‘...his room and in every **bedroom** equipped with mattress and curtain‘

(20) *... bilik yang indah. / Istana besar apa gunanya, **Peraduan** lengkap dengan perhiasannya, Asingnya tidak ...*

‘beautiful **bedroom**, what a point of a big palace, a full furnished bedroom with decorations, ...‘

**Table 4.** Frequency of occurrences *peraduan* in the MCP manuscripts

Zub Syair Siti Zubaidah Perang Cina	? ~1800 (MS 1840)	32
Panji Hikayat Panji Kuda Semirang	? ~1750 (MS 1832)	28
Mar Hikayat Maharaja Marakarma	1844 or 1848	27
SSiti Syair Seratus Siti	? ~1890	22
PandL Hikayat Pandawa Lima	? ~1525	15
Bayan Hikayat Bayan Budiman	1371 (MS 1852)	14
Boma Hikayat Sang Boma	~1600 (MS 1850)	12
Ind Hikayat Inderaputera	<1600 (MS 1700)	10
Bid Syair Bidasari	~1750 (MS 1825)	10

---

AHmz Hikayat Amir Hamzah	<1380	9
JPati Hikayat Putera Jaya Pati	? 17.. (MS 1819)	9
KT Syair Ken Tambuhan	~1750 (MS 1791- 1872)	9
Dmsy.H Hikayat Raja Damsyik	1863	9
Kosta Syair Sinyor Kosta	<1821 (MS 1862)	8
Dmsy.S Syair Raja Damsyik	1864	8
S Saudara	19301935	8
Anb Surat alAnbiya'	~1750 (MS 1877)	7
Sikka Hikayat Kerajaan Sikka	19251953 (MS 1979)	7
Aceh Hikayat Aceh	~1625 (MS ~1675)	6
Tuah Hikayat Hang Tuah	~1700 (MS 1849)	6
PiagM Piagam Muara Mendras	~16601880	6
SMrdn Hikayat Syah Mardan	~1720 (MS <1871)	5
ARM Adat Raja Melayu	1779, ~1850 (MSS 1817, 1873)	5
ABS Hikayat Ali Bad Syah	? 1800 (MS 1826)	5
Bik Hikayat Raja Bikrama Sakti	~1800 (MS 1830)	5
Nuri Syair Nuri dengan Simbangan	~1860	5
Nymk Syair Nyamuk dan Lalat	<1873	5
Kumb Syair Kumbang Mengindera	<1859	4
NurP Syair Sultan Nur Peri	< ~1865	4
Puras Hikayat Purasara	1625 (MS 1849)	4
Kutai Cerita Kutai	1625 (MS 1849)	3
SM Sejarah Melayu	~1612 (MS 1808)	3
Hemop Syair Hemop	1750s (MS 1817)	3
MW Hikayat Merong Mahawangsa	~1821 (edition 1898)	3
Mpt Hikayat Merpati Mas dan Merpati Perak	>1883 (MS 1887)	3
Pasai Hikayat Raja Pasai	~1390 (MS 1815)	2
Rama Hikayat Seri Rama	15.. (MS <1633)	2
Bkht Hikayat Bakhtiar	~1650 (text 1881)	2
IbrA Hikayat Ibrahim ibn Adham (long version)	~1650 (MS 1775)	2
KBima Syair Kerajaan Bima	~1830 (MS 1857)	2

---

Awai Syair Awai	1868	2
INata Hikayat Indera Nata	~1870 (MS <1874)	2
SAB Syair almarhum Sultan Abubakar	1896	2
CJen Cerita Jenaka	<1908	2
TZA Syair Tawarikh Zainal Abidin yang Ketiga	1936	2
PK PuisiPuisi Kebangsaan	19131942	2
BS Bustan alSalatin	~1640	1
ABJD Asal Bangsa Jin & DewaDewa	~1700 (MS <1851)	1
Pat Hikayat Patani	1730 (MS 1839)	1
Misa Misa Melayu	~1780 (MS 1836)	1
Perb Syair Tengku Perbu	1835	1
Siak Hikayat Siak	1855 (MS 1893)	1
M&B Salasilah Melayu dan Bugis	1865	1
TN Tuhfat alNafis	1866 (MS 1890)	1
PPRiau Peringatan Pelayaran ke Riau	1868 (MS ? 1917)	1
NAsik Hikayat Nakhoda Asik	~1870 (MS 1890)	1
KS Kitab Suci	PL: 1879, PB: 1935	1
Pah Hikayat Pahang	1883 (MS 1883, 1932)	1
Guna Perhimpunan Gunawan bagi Laki2 dan Perempuan	1911	1
SKel Hikayat Seri Kelantan	1783-1914 (MS 1914)	1
M Majlis	1932-1935	1
VN Vernacular Newspapers Editorials	1887-1940	1
		357

In a more recent corpus such as Leipzig Corpora, the frequency of occurrence of the word *peraduan* is also high. There are 650 occurrences from websites dated from 2012 to 2014. However, the meaning that the word carries in this corpus is rather different from those in MCP. From about 100 lines examined from the concordance lines, there are three types of usage of the word *peraduan*.

### ***The Semantic Change of ‘peraduan’***

The first type has the same meaning and usage as those in previous corpus, which is bed or bedroom of the royal family. The word *peraduan* in the first type, as shown in sentences number (21), (22), and (23), are collocated with *raja* (king) and *kerajaan* (royal).

(21) *Sementara itu, sang raja telah tidur di **peraduan** kerajaan.*

‘Meanwhile, the King had slept in the royal **bed**.’

(22) *Jika nanti sudah berada dalam **peraduan** raja, cincin itu harus dilepas, dan ditaruh didekat Pusaka Keraton karena dirinya sudah berada di dalam cincin itu.*

‘When already in the king's **bedroom**, the ring must be taken off and placed near the heritage of the palace because he is already in the ring.’

(23) *Sebelum mencabut tombak, ia kembali keluar dari **peraduan** raja yang kesakitan itu.*

‘Before pulling the spear, he came back out of the afflicted king's **bedroom**.’

In the second type of usage, the word *peraduan*, as found in the Leipzig Corpora, carries the same meaning but it is then used by common people.

(24) *Membaca buku, majalah, atau sekadar mendengarkan musik, sebelum Anda beranjak ke **peraduan** untuk tidur.*

‘Read book, magazine, or simply listen to the music before you go to **bed**.’



*matahari* (as shown in Figure 2), such as in the sentence (27), (28) and (29); and *sang surya* which also means ‘sun’ in (30). In those sentences, the sun is depicted as if it goes to bed to rest so the day turns into night, or gets out of the bed and starts to shine.

(27) *Matahari beranjak ke **peraduan** dan malam mulai menggeliat ke atas bumi.*

‘The sun goes down to its *resting place* and the night begins to climb the earth.’

(28) *Ketika matahari telah kembali ke **peraduan**, malam pun tiba.*

‘When the sun has gone to *bed*, the night has come.’

(29) *Matahari sudah beranjak ke **peraduan**, tetapi langit biru masih tersisa.*

‘The sun has gone to *bed*, but there is still some blue sky.’

(30) *Salah satunya adalah untuk melihat secara langsung, Sang Surya keluar dari **peraduan** di ufuk timur.*

‘One of the reasons is to see directly the sun out of its *bed* in the eastern horizon.’

In Indonesian, there is a metaphor that equates the sun as the king of the day and the moon as the night goddess (\_matahari=raja siang; bulan=dewi malam\_). Without the sun there will be no daylight. In some cultures, there are also tribes who regard the sun as a god or as the giver of life just like a king. Because of this metaphor and belief, some of the vocabulary reserved only for kings is also applied to the sun.

Finally, those different types of usage of *peraduan* found in Leipzig Corpora show that the word has changed in meaning through

generalization and metaphor. However, unlike the word *bersiram*, the change that the word *peraduan* experienced only occurs in semantic level. The other linguistic aspects of the word are not affected.

## CONCLUSION

Using corpora, this paper identified the etymological information particularly of the exemplary words *bersiram* and *peraduan* to determine to what extent these words diachronically changed through time. Drawing on data obtained from the analysis, the findings showed information as follows.

Entry	: <i>bersiram</i>
Initial meaning	: ‘to take a bath (intransitive), used for the royal family’
Additional meaning (in 21 <sup>th</sup> century)	1. To bathe (transitive, figurative meaning), 2. To cover (transitive, figurative meaning).
Entry	: <i>peraduan</i>
Initial meaning	: ‘bed or bedroom, used for the royal family’
Additional meaning (in 21 <sup>th</sup> century)	1. Bed or bedroom (general), 2. Resting place (figurative meaning).

The presentation of the etymological information in the dictionary can also be made in the narrative form, so the reader could get a clearer picture of the semantic change (Bochkarev et al., 2020).

This paper has demonstrated that diachronic corpora can be a useful tool in the investigation of etymological information, especially

to find changes in meaning. The corpora that are set chronologically can also tell us the approximate time of change. Although the precise year of change remains unknown, it is able to at least reveal in which era the change happen. The activity of collecting etymological information from diachronic corpora, however, can only be done to the lexicon in written texts. Furthermore, information about the usage of the words in spoken forms, whether or not they are used in the same register with the same meaning, is undisclosed. We found that, it does not lessen the effectiveness of diachronic corpora as a tool in collecting etymological information. Finally, this paper could strategically contribute to the model of development for a more comprehensive Indonesian etymology dictionary.

## REFERENCES

- Allan, K. & Robinson, J.A. (eds.). (2012). *Current Methods in Historical Semantics*. Berlin: Walter de Gruyter.
- Altakhaineh, A.R.M. (2018). The Semantic Change of Positive vs. Negative Adjectives in Modern English. *Lingua Posnaniensis*. 60(2), 25-37. DOI: <https://doi.org.10.2478/linpo-2018-00010>.
- Andaya, L.Y. (2001). The Search for the 'Origin' of Melayu. *Journal of Southeast Asian Studies*. 32(3), 315-330. DOI: <https://doi.org/10.1017/S0022463401000169>.
- Bakar, N.S.A.A. (2020). The Development of an Integrated Corpus for Malay Language. In: Alfred R., Lim Y., Havaluddin H., On C.

- (eds) *Computational Science and Technology. Lecture Notes in Electrical Engineering*, vol 603. Springer, Singapore. DOI: [https://doi.org/10.1007/978-981-15-0058-9\\_41](https://doi.org/10.1007/978-981-15-0058-9_41).
- Baker, Paul, et al. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Bieman, Chris, et al. (2007). The Leipzig Corpora Collection: Monolingual Corpora of Standard Size. Accessed from [https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/190paper.pdf&ved=2ahUKEwi6i\\_j4ubTnAhVSB30KHdBfADIQFjAAegQUBRAB&usg=AOvVaw1Fr6CtQ1ChDmmSbkH0fzZ](https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/190paper.pdf&ved=2ahUKEwi6i_j4ubTnAhVSB30KHdBfADIQFjAAegQUBRAB&usg=AOvVaw1Fr6CtQ1ChDmmSbkH0fzZ)
- Bochkarev, V., et al. (2020). A Method of Semantic Change Detection Using Diachronic Corpora Data. In: van der Aalst W., et al. (eds.). *Analysis of Images, Social Networks and Text*. AIST 2019. *Communications in Computer and Information Science*, vol 1086, 94-106. Springer, Cham. DOI: [https://doi.org/10.1007/978-3-030-39575-9\\_10](https://doi.org/10.1007/978-3-030-39575-9_10).
- de Melo, Gerard. (2014). Etymological Wordnet: Tracing the History of Words. In: *Proceeding of the Ninth International Conference on Language Resources and Evaluation. European Language Resources Association*: Reykjavil, 1148-1154.
- Collins, J.T. (1998/2018). *Malay, World Language: A Short History*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

- Collins, J.T. (2003). *Mukadimah Ilmu Etimologi*. Kuala Lumpur: Dewan Bahasa dan Pustaka
- Comrie, Bernard (ed.). (2009). *The World's Major Languages* (3<sup>rd</sup> Edition). London: Routledge.
- Durkin, Philip. (2009). *The Oxford Guide to Etymology*. New York: Oxford University Press
- Gallop, Annabel Teh. (2013). The Language of Malay Manuscript Art: A Tribute to Ian Proudfoot and the Malay Concordance Project. *International Journal of the Malay World and Civilisation*. 1(3), 11-27.
- Hasan, Mahade. (2015). Semantic Change of Words Entered into another Language through the Process of Language Borrowing: A Case of Arabic Words in Bengali. *People International Journal of Social Sciences*, Special Issue 1 (1), 1375-1390.
- Hoogervorst, Tom G. (2015). Tracing the Linguistic Crossroads between Malay and Tamil. *Wacana*. 16(2), 249-283. DOI: <https://doi.org/10.17510/wacana.v16i2.378>.
- Jatowt, A & Duh, K. (2014). A Framework for Analyzing Semantic Change of Words across Time. In *IEEE/ACM Joint Conference on Digital Libraries*, London, 229-238. DOI: <https://doi.org/10.1109/JCDL.2014.6970173>.
- Joharry, Siti Aeisha & Rahim, Hajar Abdul. (2014). Corpus Research in Malaysia: a Bibliographic Analysis. *Kajian Malaysia*. 32(1), 17-43.

- Kamus Besar Bahasa Indonesia (KBBI). (2020). Accessed from <https://kbbi.kemdikbud.go.id/>
- Kamus Dewan. (2015). Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Kridalaksana, H. (2001). Arah Pengembangan Kajian Etimologi Indonesia. *Kata*. April, 2001.
- Leipzig Corpora. (2019). Retrieved from [http://corpora.uni-leipzig.de/en?corpusId=ind\\_mixed\\_2013](http://corpora.uni-leipzig.de/en?corpusId=ind_mixed_2013)
- Liberman, A. (2009). *Word Origins and How We Know Them: Etymology for Everyone*. New York: Oxford University Press.
- Malay Concordance Project. (2020). Accessed from <http://mcp.anu.edu.au/>
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Melebek, Abdul Rashid & Moain, Amat Juhari. (2006). *Sejarah Bahasa Melayu*. Kuala Lumpur: Utusan Publications.
- Moeljadi, David, et al. (2019). Considerations for Providing Etymological Information in the KBBI Indonesian Dictionary. In *Proceeding of the 13<sup>th</sup> International Conference of the Asian Association for Lexicography*. Istanbul, 223-240.
- Mohamed, Noriah & Yusof, Radiah. (2014). *Pendekatan Kontranstif dan Komparatif Bahasa-Bahasa di Malaysia*. Penang: Universiti Sains Malaysia Press.
- Omar, Asmah Haji. (2005). *The Encyclopedia of Malaysia: Languages and Literature*. Singapore: Editions Didier Millet Pte Ltd.

- Proudfoot, I. (1991). Concordances and Classical Malay. *Bijdragen tot de Taal-, Land- en Volkenkunde* 147(1), 74-95. DOI: <https://doi.org/10.1163/22134379-90003200>.
- Pusat Rujukan Persuratan Melayu. (2020). Accessed from <https://prpm.dbp.gov.my/>
- Richter, Matthias, et al. (2006). Exploiting the Leipzig Corpora Collection. Accessed from [https://.researchgate.net/publication/228517118\\_C\\_Exploiting\\_the\\_leipzig\\_corpora\\_collection](https://.researchgate.net/publication/228517118_C_Exploiting_the_leipzig_corpora_collection)
- Ricklefs, M.C. (2008). *A History of Modern Indonesia since C.1200* (revised edition). London: Macmillan.
- Russel, Jones, et al. (eds). (2007). *Loanwords in Indonesian and Malay*. Leiden: KITLV.
- Sneddon, J.N. (2003). *The Indonesian Language: Its History and Role in Modern Society*. Sydney: UNSW Press.
- Tadmor, Uri. (2009). Loanwords in Indonesian. In: Haspelmath, Martin and Tadmor, Uri (eds.). *Loanwords in the Wolrld's Languages: A Comparative Handbook*. Berlin: Walter de Gruyter.
- Teeuw, A. (1967). The History of the Malay Language. In: *Modern Indonesian Literature*. KITLV. Springer: Dordrecht. DOI: [https://doi.org/10.1007/978-94-015-0768-4\\_2](https://doi.org/10.1007/978-94-015-0768-4_2).
- Wijaya, Derry Tanti & Yeniterzi, Reyyan. (2011). Understanding Semantic Change of Words Over Centuries. In the *Proceeding of the 2011 International Workshop on DETecting and Exploiting*

Kamal Yusuf, Dewi Puspita

*Cultural Diversity on the Social Web*, 35-40. DOI:  
<https://doi.org/10.1145/2064448.2064475>.

Yurivna, Shumylo Myroslava. (2014). Etymological and Semantic Changes of the English Medical Terms. *Science and Education a New Dimension Philology*, 2(6), Issue 29, 29-31.