

The Standard Forms of the Base Indonesian Words

Sudarno Sinduwiryo¹

Jakarta, Indonesia

1. Introduction

The Indonesian language, which is currently used as the official language of the Republic of Indonesia, originated from the Malayan language. The Malayan language has been the 'lingua franca' for the south-east Asia region since time immemorial. The native speakers of the Malayan language were originally the inhabitants of the coastal area of the Malaka Peninsula and Sumatra, and the surrounding islands. Due to its strategic location, many ships of different nationalities pass by that area, from South Asia heading for East Asia and vice versa. In that condition the Malayan language was influenced by many foreign languages, such as Sanskrit, Arabic, Persian, Chinese, Portuguese, Dutch, and English, besides by local Indonesian languages, such as Javanese, Sundanese, Minangkabau.

The Malayan or the Indonesian language was firstly written in about the first century. The character used at that time was that of Palawa, a kind of script originated from South India. After the coming of Islam, Indonesian was written in a slightly modified Arabic script known as 'Huruf Arab Melayu'. From the beginning of this century the language has been written in Roman script. The Indonesian language has so far developed its writing system. Firstly, the spelling of Ch. van Ophuysen version, secondly, the spelling of Suwandi version, and finally 'Ejaan Yang Disempurnakan' or the Revised Spelling System, the system of spelling that is currently used. The Indonesian spelling system is to some degree more phonetic than, say, the English one.

Due to its historical development, many words in the Indonesian language have more than one form, both in the written and in the spoken forms. The cognate Indonesian words as **lubang** 'hole', **pihak** 'side', **ubah** 'change', sometimes appear as **lobang**, **fihak**, and **obah** respectively. The Indonesian loan words like **tentara** 'arm forces' (Sanskrit), **majelis** 'council' (Arabic), **metode** 'method' (Dutch), **manajer** 'manager' (English), may appear as **tentera**, **majlis**, **metoda**, and **menejer** respectively. Even words of different spelling as **khawatir**, **kuatir**, **hawatir**, and **kawatir** refer to one meaning, namely 'anxious'.

It is true that some dictionaries do indicate which one among the many forms is the standard form. However, they often differ from one dictionary to another in choosing the standard forms of the words. For example, for the words which means 'shark', 'Thursday', and 'victim', one Indonesian dictionary uses **yu**, **Kemis**, and **korban** respectively, while the other uses **hiu**, **Kamis**, and **kurban**.

I really do not know the reason underlying the choosing of the standard forms. That is why, in this paper, an alternative of choosing the standard forms of both cognate and loan words of the Indonesian base words is offered. The standard form offered is based on the nature of

¹ Sudarno Sinduwiryo is a lecturer in State Institute for Islamic Studies (IAIN) Jakarta and Muhammadiyah Institute for Teacher Training and Paedagogy.

the cognate Indonesian base words, their phonemes, their syllables, and the arrangement of the syllables in forming the base words.

2. The Indonesian Phonemes, Syllables, and Base Words

I have a document containing Indonesian base words taken from some Indonesian dictionaries. I have separated cognate words from loan words. Indonesian cognate words have the following characteristics :

2.1 The Phonemes

The Indonesian phonemes consist of vowels and consonants, each followed by an example of a word containing the phoneme concerned :

/a/	aku	'I'	/c/	cukup	'enough'
/i/	itu	'that'	/g/	guntur	'thunder'
/u/	ulang	'repeat'	/k/	kulit	'skin'
/e/	emas	'gold'	/w/	sawah	'rice field'
/e/	enak	'nice'	/y/	kayu	'wood'
/o/	obat	'medicine'	/r/	rumah	'house'
/ai/	pantai	'coast'	/l/	lari	'run'
/au/	pulau	'island'	/s/	sapu	'brooms'
/b/	batu	'stone'	/h/	leher	'neck'
/p/	padi	'rice'	/m/	makan	'eat'
/d/	daun	'leaf'	/n/	tangan	'hand'
/t/	tahun	'year'	/ny/	nyamuk	'mosquito'
/j/	jauh	'far'	/ng/	siang	'noon'

There are still other phonemes or letters indicated by some Indonesian dictionaries. However, the 'phonemes/letters' are found only in loan words, such as :

/ie/	diesel	'diese' (Dutch)
/eu/	eufemisme	'euphemism' (Dutch)
/ei/	apartheid	'apartheid' (Dutch/English)
/bh/	bhineka	'diversity' (Sanskrit)
/dh/	adhesi	'adhesion' (Dutch)
/dj/	adjektif	'adjective' (Dutch)
/f/	faktor	'factor' (Dutch)
/kh/	khusus	'special' (Arabic)
/q/	qari	'Qur'an reader' (Arabic)
/sh/	squash	'squash' (English),
	shalat	'prayer' (Arabic)
/sy/	syarat	'condition' (Arabic)
/ts/	tsunami	'Wave' (Japanese)

/v/	revolusi	'revolution' (Dutch)
/x/	marxisme	'marxism' (Dutch)

I do not know whether the combination of letters such as <ie> or <eu> each represents a diphthong or a vowel sequence, and the combination of consonants <dh> or <dj> each represents a consonant cluster or two consonants.

Some loan words containing such phonemes have two forms, for examples :

Februari	- Pebruari	'February' (Dutch)
khalayak	- kalayak	'public' (Arabic)
masyhur	- mashur	'famous' (Arabic)
November	- Nopember	'November' (Dutch).

Some have one form only, such as :

faktor	'factor' (Dutch/English)
khas	'specific' (Arabic)
syarat	'condition' (Arabic)

2.2 The Syllables

Most Indonesian cognate base words consist of two syllables, such as **a-ku** 'I', **a-ir** 'water', **i-kan** 'fish', **pa-di** 'rice', **tem-pat** 'place'. The syllable consists of one, two, or three phonemes. The patterns are (v stands for vowel and c stands for consonant) :

V	as i	in	i-kan	'fish'
VC	as ir	in	a-ir	'water'
CV	as ku	in	a-ku	'I'
CVC	as pan	in	pan-tai	'coast'

So there are basically only four Indonesian syllable patterns. In the Indonesian language currently used, there are still other patterns which are different from the above mentioned patterns. But, such patterns are always found in loan words. For examples :

VCC	as eks	in	eks-por	'export' (Dutch)
CVCC	as teks	in	teks	'text' (Dutch)
CCVCC	as pleks	in	kom-pleks	'complex' (Dutch)
CCV	as pro	in	pro-sa	'prose' (Dutch)
CCVC	as prak	in	prak-tek	'practice' (Dutch)
CCCV	as stra	in	stra-tegi	'strategy' (Dutch)
CCCVC	as struk	in	struk-tur	'structure' (Dutch)

Many loan words of such patterns have two forms, namely the forms which still reserve the loan patterns while the others have been adjusted to the cognate patterns, such as :

test, tes	'test' (Dutch)
klas, kelas	'class' (Dutch)
ons, on	'ounce' (Dutch)

In the cognate syllable patterns there are some rules. Among others are :

1. In a closed syllables, i.e. that of the patterns VC and CVC, the (second) consonant must be /p/, /t/, /k/, /r/, /l/, /s/, /h/, /m/, /n/, /ny/, (always symbolised with /n/), or /ng/. So words containing other than those consonants at the end of each syllable can easily be recognized as loan words. For example:

se-bab	'cause' (Arabic)
ad-mi-ni-stra-si	'administration' (Dutch)
mik-raj	'Prophet Muhammad's journey' (Arabic)
bi-o-log	'biologist' (Dutch)

However, such deviation, i.e. syllables ending in voiced stops, is permissible in written Indonesian, but not in the oral form. Thus the pronunciation of words as **se-bab** is **se-bap** etc.

2. The vowel of a closed syllable can only be a monophthong. For example:

am-pun	'pardon'
tam-pak	'seem'
ker-dil	'dwarf'
kan-dung	'bear'
pang-gang	'bake'

If in such position the vowel is a diphthong, the word containing it must be a loan word, such as :

syaikh	'sheikh' (Arabic)
haid	'menses' (Arabic)

A loan word with a diphthong in its closed syllable is usually adjusted to the cognate one by replacing the diphthong by a monophthong or by considering the diphthong as sequence of vowels. For examples the Arabic word **syaikh** 'sheikh' becomes **syekh**, **haid** becomes **hed** and **bait** becomes **ba-it**.

3. In the CV and CVC patterns the first consonant and vowel cannot be **wu** or **yi**. If it is, the word must be a loan word. For example:

nah-wu	'grammar' (Arabic)
ta-sa-wuf	'Islamic school' (Arabic)
ma-yit	'corpse' (Arabic)
ba-yi	'baby' (Javanese)

Some loan words containing /wu/ and /yi/ have been adjusted to the Indonesian rules by omitting the consonant, i.e. /w/ or /y/. For example (the syllabification follows the cognate one, here):

na-hu	instead of	nah-wu
ta-sa-uf	instead of	ta-sa-wuf
ma-it	instead of	ma-yit
ba-i	instead of	ba-yi

2.3 The Base Word

As stated above, an Indonesian base word consists of one or more syllables. If a word consists of more than one syllable, the arrangement of the syllables follows certain rules.

To explain the arrangement of syllables in base words, I will first highlight the probabilities of the arrangements of syllables. Since there are four syllable patterns, there will theoretically be sixteen probabilities of arrangements. Those are :

- | | | | |
|----------|-----------|------------|-------------|
| 1. V-V | 5. CV-V | 9. VC-V | 13. CVC-V |
| 2. V-VC | 6. CV-VC | 10. VC-VC | 14. CVC-VC |
| 3. V-CV | 7. CV-CV | 11. VC-CV | 15. CVC-CV |
| 4. V-CVC | 8. CV-CVC | 12. VC-CVC | 16. CVC-CVC |

Practically, there are merely twelve kinds of arrangement. Those are (with the base words as examples) :

- | | | |
|-------------|---------|--------------|
| 1. V-V | i-i | 'he/she' |
| 2. V-VC | a-ir | 'water' |
| 3. V-CV | a-ku | 'I' |
| 4. V-CVC | i-kan | 'fish' |
| 5. CV-V | du-a | 'two' |
| 6. CV-VC | bu-ah | 'fruit' |
| 7. CV-CV | te-bu | 'sugar cane' |
| 8. CV-CVC | pu-tih | 'white' |
| 9. VC-CV | un-tai | 'set' |
| 10. VC-CVC | em-pat | 'four' |
| 11. CVC-CV | tung-gu | 'wait' |
| 12. CVC-CVC | tem-pat | 'place' |

The other four kinds of arrangements are found only in loan base words, such as :

- | | |
|------------|------------------------------|
| 13. VC-V | (not found) |
| 14. VC-VC | af-al 'deed' (Arabic) |
| 15. CVC-V | mas-a-lah 'problem' (Arabic) |
| 16. CVC-VC | Jum-at 'Friday' (Arabic) |

However, the last three base words have other forms which do not contradict to the rules since they have been adjusted to the cognate arrangement :

- | | | |
|-----------|---|-----------|
| af-al | = | a-fal |
| mas-a-lah | = | ma-sa-lah |
| Jum-at | = | Ju-mat |

The cognate arrangements still have other conditions which must be fulfilled. The conditions are, among others :

1. In the sequence (C)VC - CV(C) the (last) consonant of the first syllable is only a nasal consonant, i.e. either /m/, /n/, /ny/, /ng/, or a rolled /r/. If the last consonant of the first syllable is nasal, the first consonant of the second syllable is only a homogenic plosive consonant or /s/, or /l/. If the last consonant of the first syllable is /r/, the first consonant of the second syllable will never be /w/, /y/, or /h/. For example:

tum-buh	'grow'	ang-kat	'lift'
jum-pa	'meet'	lang-sung	'direct'
un-dang	'call'	lung-lai	'weak'
pan-tai	'coast'	ker-bau	'buffalo'
jen-jang	'rank'	ter-ka	'guess'
pan-cung	'cut'	ser-bu	'attract'
tung-gu	'wait'		

In loan words the last consonant of the first syllable and the first consonant of the second syllable are not always in the patterns explained above. For example:

in-san	'person' (Arabic)
tan-pa	'without' (Javanese)
ab-jad	'alphabet' (Arabic)
ak-ta	'Act' (Dutch)
jad-wal	'schedule' (Arabic)
kar-ya	'work' (Sanskrit)
mar-hum	'the late ...' (Arabic)

2. In the sequence (C)V - V(C) the vowels cannot be the same vowels, and in pronunciation there is a semivowel as a connector. For example:

i-a	'he/she' pronounced as iya
di-am	'stay' pronounced as diyam
lu-ar	'outside' pronounced as luwar
tu-a	'old' pronounced as tuwa

In some loan words the two vowels can be the same vowel and in the pronunciation there is no semivowel but rather a glottal stop /ʔ/ as a connector. For example:

sa-at	'time' pronounced as saʔat (Arabic)
ma-af	'pardon' pronounced as maʔaf (Arabic)
fi-il	'deed' pronounced as fiʔil (Arabic)

Some loan words from Dutch still have semivowels. For example:

di-it	'diet' pronounced as diyit
ri-il	'real' pronounced as riyil

3. In the sequence (C)V - CV(C) if the vowel of the first syllable is /u/, the first consonant of the second syllable is not /w/; if the vowel is /i/, the first consonant is not /y/. For example:

ru-mah	'house'
ti-ba	'arrive'

In some loan words though the two vowels are different; there are no semivowels. For example:

sya-ir	'poem' pronounced as sya^ʔir (Arabic)
so-al	'question' pronounced as so^ʔal (Arabic)
do-a	'prayer' pronounced as do^ʔa (Arabic)
syi-ar	'preach' pronounced as syi^ʔar (Arabic)

Some have two pronunciation, one with semivowel due to the adjustment to the cognate rules, while the other without semivowel. For example:

ta-bi-at	'conduct' pronounced as tabi^ʔat or tabiyat (Arabic)
da-e-rah	'region' pronounced as da^ʔerah or daherah (Arabic)

Some loan words from Dutch still have semivowels. For example:

di-it	'diet', pronounced as diyit
ri-il	'real', pronounced as riyil

3. In the sequence (C)V - CV(C) if the vowel of the first syllable is /u/, the first consonant of the second syllable is not /w/, and if the vowel is /i/, the first consonant is not /y/. For example:

ru-mah	'house'
ti-ba	'arrive'
bu-nga	'flower'
si-kat	'brush'

Some loan words with such a syllable arrangement the vowel /u/ is followed by /w/ and the vowel /i/ is followed by /y/. For example:

mu-wa-fa-kat	'agree' (Arabic)
u-khu-wah	'brotherhood' (Arabic)
wi-ya-nga	'gamelan player' (Javanese)
wi-ya-ta	'lesson' (Javanese)

3. The Standard Forms

Some Indonesian dictionaries indicate which one among the many forms of the Indonesian base words is the standard form. Three of the dictionaries here are taken as examples. Those are :

1. Sutan Muhammad Zain, *Kamus Moderen Bahasa Indonesia*, Grafica, Jakarta, (no year of publication indicated, but it was about 1964).

2. W.J.S.Poerwadarminta, *Kamus Umum Bahasa Indonesia*, PN.Balai Pustaka, Jakarta, 1966.
3. Anton M. Moeliono et. al., *Kamus Besar Bahasa Indonesia*, PN.Balai Pustaka, Jakarta, 1988.

The *Kamus Moderen* has 12,841 entries, the *Kamus Umum* has 22,412 entries, while the *Kamus Besar* has 29,937 entries. The first two dictionaries use the old spelling system, while the last uses the new spelling system.

3.1 The *Kamus Moderen*

In the *Kamus Moderen*, letters <q>, <v> and <x> are not used in its spelling system, even in spelling of loan words (<c> and <y> are not used in consequence of the old spelling system). Loan words can be recognized only by the use of the letters <ch> (in the current spelling system it is <kh>), <f>, <sj> (in the current spelling systems it is <sy>) and <z>. But strange enough, the diacritical marks <"> and <'> are used in the spelling of loan words from Arabic, such as:

Kurän	'Koran'
soäl	'question'
do´a	'prayer'
fa´al	'deed'

There are no consonant clusters used in the spelling of a syllable but few, especially in the middle of a word. So the loan words from Dutch such as **modern**, **protes**, **setrum**, **sekretaris** in this dictionary are respectively written **moderen**, **perotes**, **seterum**, and **sekertaris**.

Consonant /w/ is never followed by vowel /u/; neither is /y/ by /i/. So instead of **wujud**, **nahwu**, **bayi**, **priyayi** they are spelled as **ujud**, **nahu**, **bai**, and **periari** respectively.

The loan (C)VC - V(C) arrangement of syllables in this dictionary is recognized and an umlaut is used to refer the syllable boundary or even a morpheme boundary, such as:

Jumät	'Friday'	pronounced as Jum[?]at	(Arabic)
Korän	'Koran'	pronounced as Kur[?]an	(Arabic)

3.2 The *Kamus Umum*

In the *Kamus Umum*, letters <q>, <v> and <x> are also not used in its spelling system. As in the case of the *Kamus Moderen*, loan words can be recognized only by the use of the letters <ch>, <f>, <sj> and <z>. A stroke is used to spell some of the loan words from Arabic, such as:

Jum´at	'Friday'
mu´tamar	'conference'

As it is different from the *Kamus Moderen*, the spelling of loan words containing consonant clusters are found in this dictionary. For example:

praktek	'practice' (Dutch)
status	'status' (Dutch)
strategi	'strategy' (Dutch)

But we rarely find consonant clusters at the end of syllables in this dictionary; the case is not the same as in the *Kamus Besar*.

The loan arrangement of syllable (C)VC - V(C) is found also in this dictionary. But, unlike in the *Kamus Moderen*, it is written with a stroke <´> such as:

Jum´at	‘Friday’ (Arabic, also Jumaat)
af´al	‘Deed’ (Arabic)

3.3 The *Kamus Besar*

The *Kamus Besar* is the most “modern” dictionary viewed from the Malayan point of view, in phoneme, syllable, and the arrangement of syllables of the base words. All letters of the Roman alphabet are applied in this dictionary. The letters <q> and <x> which are not used in the two other dictionaries, are used in the *Kamus Besar* to spell loan words. In many letter, sequences (including loan words) it is not clear to me, whether the two letters represent one phoneme or two — are also found, such as:

diesel	‘diesel’ (Dutch)
eufemisme	‘euphemism’ (Dutch)
adhesi	‘adhesion’ (Dutch)
tsunami	‘tsunami’ (Japanese)

In syllabification, it seems to follow loan patterns, such as (with their respective examples in words):

VCC	as in	ons	‘ounce’ (Dutch)
CVCC	as in	teks	‘text’ (Dutch)
CVCCC	as in	hertz	‘hertz’ (Dutch)
CCV	as in	pro - tes	‘protest’ (Dutch)
CCVC	as in	prak - tek	‘practice’ (Dutch)
CCVCC	as in	kom - pleks	‘complex’ (Dutch)
CCCV	as in	stra - te - gi	‘strategy’ (Dutch)
CCCVC	as in	struk - tur	‘structure’ (Dutch)

Strange enough, in the syllable arrangement, the *Kamus Besar* almost always follows the Malayan rule. Loan syllable arrangement (C)VC - V(C) is not found in this dictionary. So loan words from Arabic such as **Jum-at**, **Qur-an**, **mas-a-lah** for example, are written in the *Kamus Besar* as **Jumat**, **Quran**, and **masalah** respectively.

Comparing the three dictionaries in dealing with the base Indonesian words, I would like to highlight some words which have different standard forms. Those are (among others):

<i>Kamus Moderen</i>	<i>Kamus Umum</i>	<i>Kamus Besar</i>	
bai	bai	bayi	‘baby’
do´a	doa	doa	‘prayer’
Pebruari	Pebruari	Februari	‘February’
Jumat	Jum´at, Jumaat	Jumat	‘Friday’
kwartal	kwartal, kuartal	kuartal	‘quarter’

<i>Kamus Moderen</i>	<i>Kamus Umum</i>	<i>Kamus Besar</i>	
ruh	ruh, roh	roh	'soul'
salaf	salap, salep	salep	'ointment'
senteri	senteri, santri	santri	'disciple'
ujud	ujud	wujud	'appearance'

The different standard forms of the Indonesian base words inevitably causes difficulties for Indonesian language teachers.

4. An alternative proposed

As it is clear that the standard forms of the Indonesian base words vary from one dictionary to another, an alternative is offered below to solve the problem. The arrangement follows the above treatment.

4.1 The Phonemes

The Indonesian phonemes and/or letters to symbolize them should be (each followed by an example of word containing the phoneme concerned) :

/a/	- aku	'I'	/k/	- kulit	'skin'
/i/	- itu	'that'	/w/	- sawah	'rice field'
/u/	- ulang	'repeat'	/y/	- kayu	'wood'
/e/	- emas	'gold'	/r/	- rumah	'house'
/e/	- enak	'nice'	/l/	- lari	'run'
/o/	- obat	'medicine'	/s/	- sapu	'broom'
/ai/	- pantai	'coast'	/h/	- leher	'neck'
/au/	- pulau	'island'	/m/	- makan	'eat'
/b/	- batu	'stone'	/n/	- tangan	'hand'
/p/	- padi	'rice'	/ny/	- nyamuk	'mosquito'
/d/	- daun	'leaf'	/ng/	- siang	'noon'
/t/	- tahun	'year'	/f/	- faktor	'factor'
/j/	- jauh	'far'	/kh/	- khusus	'special'
/c/	- cukup	'enough'	/sy/	- syarat	'condition'
/g/	- guntur	'thunder'	/z/	- zaman	'era'

So any other sequence of letters/phonemes other than those mentioned above, either vowels or consonants, must be considered as a sequence of letters/phonemes, not as a diphthong or as a consonant cluster.

4.2 The Syllable

The syllable patterns of Indonesian base words are :

V	as in	i-kan	'fish'
VC	as in	a-ir	'water'
CV	as in	ta-nah	'land'

CVC	as in	em-pat	'four'
CCV	as in	pro-tes	'protest'
CCVC	as in	prak-tis	'practical'

The last two patterns are for loan words only.

However, there are still some regulations concerning the syllables. Those are :

1. The consonants concluding the syllable are merely /p/, /t/, /k/, /r/, /l/, /s/, /h/, /m/, /n/, /ny/ (written with <n>), /b/, /d/ and /g/. The last three are for loan words only.
2. The vowel of patterns VC, CVC and CCVC must not be a diphthong.
3. The second member of consonant clusters are only consonants /k/, /p/, /t/, /w/, /y/, /r/, /l/. If the second member of consonant clusters are /k/, /p/ or /t/, the first member must be /s/.

4.3 The Arrangement of Syllables

The arrangement of syllables should follow the patterns below :

- | | |
|-----------------|-------------------|
| 1. V-V | 2. V-VC |
| 3. V-(C)CV | 4. V-(C)CVC |
| 5. VC-(C)CV | 6. V-(C)CVC |
| 7. (C)CV-V | 8. (C)CV-VC |
| 9. (C)CV-(C)CV | 10. (C)CV-(C)CVC |
| 11. (C)CV-(C)CV | 12. (C)CVC-(C)CVC |

The regulations relating to those arrangements are among others :

1. There should be no diphthong in the middle or initial position of a base word.
2. There should be no double consonants in a base word.

Exceptions of the above mentioned conditions should only be for religious and scientific terminologies.