Hiroki Nomoto and Hooi Ling Soh
# Malay

## Introduction

Malay (*bahasa Melayu*) is a Western Malayo-Polynesian language in the Austronesian family. The standard varieties of Malay known as *bahasa Malaysia* and *bahasa Indonesia* are the national languages of Malaysia and Indonesia respectively. These two Malay varieties share a core grammar, with differences mainly in the phonology and lexicon. Besides Malaysia and Indonesia, Malay also enjoys official status in Brunei and Singapore, with different varieties being used in Thailand (mainly in the southern states where the ethnic Malays constitute the majority of the population), Timor-Leste (East Timor), Sri Lanka and Australia (Cocos and Christmas Islands).

In this article, we will focus on the standard variety of Malay spoken in Peninsular Malaysia, i.e. Standard Malay (see Appendix for a sample text). By Standard Malay we mean the variety used when conversing with people of various dialect backgrounds, but not the prescriptive language imposed by Dewan Bahasa dan Pustaka (DBP), an institution established by the Malaysian government to stardardize Malay and promote its use. Within Standard Malay, there exist two registers with considerable differences, viz. Formal Malay as the high variety and Colloquial Malay as the low variety in the sense of Ferguson (1959). As the high variety, Formal Malay enjoys high prestige, unlike Colloquial Malay which is sometimes described as 'broken' and 'lacking grammar'.[1] We will explicitly state which variety is in question when the relevant trait is characteristic of only one variety, but not both.

Peninsular Malaysia has been a multilingual society since at least when Malacca thrived as a centre of international trade in the 16th century. Malay was used in trade as a *lingua franca* for people from different parts of the world. It is

---

[1] See Nomoto and Shoho (2007) for other differences between Formal Malay and Colloquial Malay.

**Hiroki Nomoto:** Tokyo University of Foreign Studies
E-Mail: nomoto@tufs.ac.jp
**Hooi Ling Soh:** University of Minnesota
E-Mail: sohxx001@umn.edu

still a *lingua franca* for different ethnic groups within Malaysia, though nowadays English also serves the same function. Among the languages that have been influential as sources of new vocabulary are Sanskrit (before the arrival of Islam in the region in the 12th century), Arabic (after the arrival of Islam) and English (since the British colonization of Malacca in 1824). While contact between speakers of Malay and these languages occurred in limited settings and over limited periods of time (e.g. during trade), not all language contact situations share this characteristic. In particular, there was extended and consistent contact between Chinese and Malay in the context of mixed marriages between Chinese men and Malay women during the 15th and 16th centuries, which resulted in the emergence of a Malay-based creole Baba Malay (Lim 1988; Thurgood 1998). Direct encounters among speakers of different languages are not at all uncommon in the present-day Peninsular Malaysia, especially in medium and large-sized cities.

According to the Malaysia Department of Statistics, the total population of the country as of 2017 is approximately 32 million. It is estimated that about 65 per cent of the population speaks Malay as a first language whilst the rest of the population speaks it as a second language. Major languages spoken in Malaysia other than Malay include Chinese languages (especially Hokkien, Cantonese and Mandarin), Tamil and English.

# 1 Phonology

## 1.1 Suprasegmental phonology: tone and register

Distinct tones and registers (phonation types) do not exist in Malay. Lexical contrasts are made only segmentally. It is generally agreed that stress is not lexically distinctive in Malay in that it does not differentiate word meaning. However, there are discussions on whether the concept of 'word stress', namely that one or more syllable within a word is associated with a higher degree of prominence than the other syllables, is relevant for Malay. Some researchers assume the existence of word stress; a common characterization is that stress falls on the penultimate syllable of a word in the citation form, except when the penultimate syllable contains a schwa.[2] This observation dates back at least to

---

**2** It should be noted that there are other descriptions of word stress in Malay/Indonesian. In addition to the pattern described above, stress has also been described as falling on the final syllable, or sometimes on the penultimate syllable and sometimes on the final syllable. These

Winstedt (1920: 4). However, there are also authors who question the relevance of word stress in Malay/Indonesian. For example, Halim (1981: 86–88) claims that there is no word stress in Indonesian, and that the distribution of stress in words is of exactly the same pattern as the distribution of stress in units larger than words, which he refers to as 'pause-units'.[3] A recent acoustic study by Zuraidah et al. (2008), in our evaluation of it, confirms the validity of Halim's observation in Malay, though the authors attempt to draw rather different conclusions. Along a slightly different line, though also questioning the relevance of word stress, Gil (2007) argues that stress falls on a foot rather than on a syllable in Riau Indonesian.

## 1.2 Segmental phonemes: consonants and vowels

The consonant phonemes in Malay can be classified into two types: those native to Malay (summarized in Figure 1) and those borrowed from other languages, especially Arabic and English (summarized in Figure 2). Some borrowed consonants are often substituted by phonetically similar native consonants, which are given in parentheses.
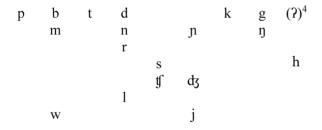
p    b    t    d                    k    g    (ʔ)[4]
     m         n         ɲ               ŋ
          r
               s                              h
               ʧ    ʤ
          l
     w                   j

**Figure 1:** Native consonants in Malay

f (p)    v    z    ʃ    x (k, h)    ɣ (g)    q (k)

**Figure 2:** Borrowed consonants and their native substitutes in Malay

---

varying descriptions may be due in part to authors' describing different regional or social varieties of Malay/Indonesian (see Tadmor 2000).

**3** In Halim's original generalization, the term 'accent' is used instead of 'stress'.

**4** The glottal stop (ʔ) is in parentheses, as its phonemic status is unclear.

All consonants can occur in the word-initial position. The stops are usually realized without aspiration, though the presence or absence of aspiration does not lead to lexical contrast. The presence of phonemic prenasalized voiced obstruents has been reported in many Austronesian languages (Klamer 2002; Donohue 2004). However, where Malay is concerned, we are not aware of any variety in which prenasalized obstruents are phonemically distinguished from plain ones, except for Sri Lanka Malay, which acquired phonemic prenasalized stops /ᵐb, ⁿd, ᶮɟ, ᵑg/ as a result of contact with Sinhala (Nordhoff 2009: 96–99). Prenasalized obstruents appear to only occur as allophones in some varieties of Malay such as Sarolangun Malay spoken in Jambi, Indonesia (Cole et al. 2008). Most native roots in Malay are disyllabic, of the CV.CV(C) pattern. Initial consonant clusters are only found in loan words. Examples of loan words from Sanskrit and English are listed in (1).

(1)  a.  Loan words from Sanskrit
         _sw_asta 'private', _dw_ibahasa 'bilingual', _pr_asangka 'prejudice'
     b.  Loan words from English
         _dr_ama 'drama', _kl_ausa 'clause', _kr_isis 'crisis', _sp_ek 'spectacles', _st_esen 'station', _tr_adisi 'tradition', _skr_ip 'script', _str_uktur 'structure'

Not all consonant phonemes can occur word-finally. Native Malay phonology allows only the following consonants in the word-final position: /p, t, k, m, n, ŋ, r, s, h, l/. Notice that the palatal nasal /ɲ/ is not included. In Peninsular Malaysia, there is a tendency to reduce final syllables in some way, either through the complete elimination of the coda or the elimination of the place feature of the final coda consonant.[5] This tendency is particularly evident in northern dialects. For example, in Kelantan Malay, final nasal consonants are dropped and the preceding vowel is nasalized as in (2a), and liquids are dropped as in (2b). Final stops lose their place features and are realized as a glottal stop as in (2c); /s/ is reduced to a glottal fricative as in (2d).

(2)  Kelantan Malay
     a.  /dalam/   →   [dalɛ̃]    'inside'
     b.  /mahal/   →   [maha]    'expensive'
     c.  /gəlap/   →   [gəlaʔ]   'dark'
     d.  /atas/    →   [atah]    'top'

---

**5** Sneddon (1993) points out a similar tendency towards final open syllables in languages in Sulawesi, Indonesia.

Some authors have noted that final voiced stops in loan words undergo devoicing to become voiceless (Farid 1980: 8–9; Teoh 1994: 53). However, Husni et al. (2006) claim that the devoicing is incomplete in that a final voiceless stop and a "devoiced" final voiced stop are not acoustically identical. Final consonant clusters are much more restricted than initial ones. No complex codas are allowed except in some loan words (e.g. se*ks* 'sex').

Standard Malay has six simple vowels /i, e, a, o, u, ə/ and three diphthongs /ai, au, oi/ (see Figure 3 for the inventory of simple vowels in Standard Malay). Vowel length is not distinctive.

| i | | u |
|---|---|---|
| e | ə | o |
| | a | |

**Figure 3:** The vowel system of Standard Malay

The vowel system in Standard Malay is thus considered simple compared to many Mainland Southeast Asian (MSEA) languages, which, according to Enfield (2005), commonly have nine simple vowel contrasts, often including a high central unrounded vowel (/ɨ/) and a number of complex vowel combinations. The Malay vowel system is perhaps more like Arakanese, which diverges from typical MSEA languages in having five simple vowel phonemes (the same ones as Standard Malay) and a diphthong /ai/ (Vittrant 2008). However, the number of vowel phonemes varies from dialect to dialect. It is common to find a vowel system with eight simple vowels consisting of the six vowels in Figure 3 and two additional half-open vowels /ɛ/ and /ɔ/ in northern dialects such as Kedah and Kelantan dialects (see Figure 4).

| i | | u |
|---|---|---|
| e | ə | o |
| ɛ | | ɔ |
| | a | |

**Figure 4:** The vowel system of Kedah and Kelantan Malay

The vowel systems of these northern dialects are thus more in line with most MSEA languages such as Thai (Iwasaki and Ingkaphirom 2005), Lao, Vietnamese and Burmese, which have eight or nine simple vowels (Vittrant 2008). In fact, northern dialects of Malay share with Burmese the same eight simple vowels (Vittrant 2008). While the northern dialects generally have a vowel system that is more complex than Standard Malay, their consonant systems are usually reduced compared to Standard Malay, as mentioned above.

## 1.3 Syllable structure

Syllables in Malay consist of 0–3 onsets, a nucleus and 0–2 codas, i.e. (C)(C)(C)V(C)(C). However, as mentioned above, complex onsets and codas only occur in loan words. Thus, syllables in native words have one of the following structures: V, CV, VC, CVC. A sequence of two consonants at the syllable boundary, i.e. $C_2$ and $C_3$ in $C_1VC_2.C_3VC_4$, consist of either a /r/ ($C_2$) and another consonant ($C_3$) as in /pər.lu/ 'need', or a homorganic nasal + obstruent sequence, as in /sim.pan/ 'to save'. Sesquisyllables (Matisoff 1973), i.e. disyllabic words comprising a minor/atonal syllable followed by a full major syllable typically found in languages of MESA, have not been reported in Malay.

# 2 Morphology

## 2.1 Word structure

Malay is an agglutinating language, in which a word may consist of multiple morphemes. Thus, there is no particular tendency towards monosyllables in Malay, unlike many MSEA languages. Although there is no such tendency, many common words in Colloquial Malay have a reduced number of syllables in comparison with their Formal Malay counterparts. Some examples are given in Table 1.

**Table 1:** Reduced forms in Colloquial Malay

| Formal Malay | Colloquial Malay | Meaning |
| --- | --- | --- |
| *tidak* | *tak* | 'not' |
| *sudah* | *dah* | 'already' |
| *hendak* | *nak* | 'will, to want' |
| *sahaja, saja* | *aja, ja* | 'only' |
| *ini* | *ni* | 'this' |

| Formal Malay | Colloquial Malay | Meaning |
|---|---|---|
| *itu* | *tu* | 'that' |
| *begitu* | *gitu* | 'like that' |
| *sedikit* | *sikit* | 'a little, a few' |
| *pergi* | *gi, pi* | 'to go' |
| *perlahan* | *pelan* | 'slow' |

Words in Malay may be formed through the attachment of affixes to roots. The number of productive affixes is relatively small in Malay. They can be categorized as (i) verbal, (ii) nominal, (iii) adjectival, and (iv) adverbial on the basis of the syntactic category of the affixed forms (Nik Safiah et al. 2008). A list of affixes and their functions (not exhaustive) are given in Table 2.

**Table 2:** Malay affixes and their functions

| Category | Affix | Function(s) | Example(s) |
|---|---|---|---|
| (i) Verbal | *meN-* | active voice[6] | *mem-baca* 'to read' [*baca* 'to read'] |
| | | inchoative | *mem-besar* 'to grow' [*besar* 'big'] |
| | *di-* | passive voice | *di-besar-kan* 'to be raised/enlarged' [*besar* 'big'] |
| | *ber-* | possession | *ber-makna* 'to mean' [*makna* 'meaning'] |
| | | action | *ber-doa* 'to pray' [*doa* 'prayer'] |
| | *ter-* | result state | *ter-tutup* 'to be closed' [*tutup* 'to close'] |
| | | accidental | *ter-tidur* 'to fall asleep' [*tidur* 'to sleep'] |
| | | ability | *ter-kawal* 'to be under control' [*kawal* 'to control'] |
| | *per-* | to make more ... | *(mem-)per-besar(-kan)* 'to enlarge' [*besar* 'big'] |
| | *-kan* | causative | *(me-)naik-kan* 'to raise (e.g. price)' [*naik* 'to rise'] |
| | *-i* | locative | *(me-)naik-i* 'to climb (e.g. a mountain)' [*naik* 'to rise'] |
| | *ke- -an* | adversative | *ke-habis-an* 'to run out of' [*habis* 'to finish'] |
| (ii) Nominal | *peN-* | agent | *peng-ajar* 'teacher' [*meng-ajar* 'to teach'] |
| | *pe-* | patient | *pel-ajar* 'student' [*bel-ajar* 'to learn'] |
| | *-an* | result | *ajar-an* 'teaching, lesson' [*di-ajar* 'to be taught'] |
| | *peN- -an* | action | *peng-ajar-an* 'teaching' [*meng-ajar* 'to teach'] |

---

**6** The prefix *meN-* is standardly associated with active voice and is glossed as ACT throughout. However, this analysis of *meN-* is not uncontroversial, considering the aspectual effects of *meN-* discussed in recent works (Soh and Nomoto 2009, 2011; Soh 2013). See the discussion below and in section 3.2.1.

| Category | Affix | Function(s) | Example(s) |
|---|---|---|---|
| | *per- -an* | action | *pel-ajar-an* 'learning, lesson' [*bel-ajar* 'to learn'] |
| | *ke- -an* | abstract entity | *ke-naik-an* 'increase' [*naik* 'to rise'] |
| (iii) Adjectival | *se-* | equative | *se-besar* 'as big as …' [*besar* 'big'] |
| | *ter-* | superlative | *ter-besar* 'biggest' [*besar* 'big'] |
| (iv) Adverbial | *se- -nya* | as … as possible, etc. | *se-cepat-nya* 'as quickly as possible' [*cepat* 'quick(ly)'] |

The verbal prefix *per-* is often preceded by the prefix *meN-* or *di-*. In Colloquial Malay, there is a tendency not to use the verbal prefix *meN-* (Benjamin 1993; Koh 1990: 153).

(3)  a.  *Dia    sudah    mem-baca    novel    itu.*                    Formal Malay
      3SG    PFV      ACT-read    novel    that
      'S/he read the novel.'

    b.  *Dia    dah    baca    novel    tu.*                    Colloquial Malay
       3SG    PFV    read    novel    that
       'S/he read the novel.'

The reason for this tendency is arguably due to the fact that the denotation of a verb without *meN-* subsumes that of the corresponding *meN-* form (Soh and Nomoto 2011), and that colloquial speech may allow for more context-dependency. Degree achievement sentences like (4) illustrate this point most clearly. While a sentence with a verb without *meN-* can be used as either a telic or an atelic predicate, one with a *meN-* verb can be used only as an atelic one (Soh and Nomoto 2011, 2015).

(4)  a.  *Harga    minyak    turun    selama/dalam    tiga    hari.*
      price    oil    fall    for/in    three    day
      'The oil price fell for/in three days.'

    b.  *Harga    minyak    men-[t]urun    selama/*dalam    tiga    hari.*
       price    oil    ACT-fall    for/in    three    day
       'The oil price was falling for three days.'

Besides affixation, compounding is also a productive morphological process in Malay. Compounds may be formed by the combination of two words of either the same or different syntactic categories. Examples are given below, with the syntactic category of each compound given in parentheses.

(5)  N-N compounds
    a.  *kereta api* 'train' (N)                [*kereta* 'car' + *api* 'fire']
    b.  *matahari* 'sun' (N)                 [*mata* 'eye'+ *hari* 'day']

(6)  A-A compounds
    a.  *kurus kering* 'thin and haggard' (A)    [*kurus* 'thin' + *kering* 'dry']
    b.  *kaya-raya* 'prosperous' (A)         [*kaya* 'rich' + *raya* 'great']

(7)  V-V compounds
    a.  *jual beli* 'trade' (N)               [*jual* 'to sell' + *beli* 'to buy']
    b.  *temu bual* 'interview' (N)        [*temu* 'to meet' + *bual* 'to talk']

(8)  P-P compounds
    a.  *daripada* 'from (human, source)' (P)   [*dari* 'from' + *pada* 'at (human)']
    b.  *kepada* 'to (human, target)' (P)     [*ke* 'to' + *pada* 'at (human)']

(9)  N-A compounds
    a.  *kerjasama* 'cooperation' (N)      [*kerja* 'work' + *sama* 'together']
    b.  *kencing manis* 'diabetes' (N)     [*kencing* 'urine' + *manis* 'sweet']

(10)  N-V compounds
    a.  *bola sepak* 'football' (N)       [*bola* 'ball' + *sepak* 'to kick']
    b.  *meja makan* 'dining table' (N)   [*meja* 'table' + *makan* 'to eat']

(11)  A-N compounds
    a.  *setiausaha* 'secretary' (N)       [*setia* 'loyal'+ *usaha* 'effort']
    b.  *lebih masa* 'overtime (work)' (N)  [*lebih* 'more' + *masa* 'time']

(12)  A-V compounds
    a.  *salah guna* 'to misuse' (V)      [*salah* 'wrong' + *guna* 'to use']
    b.  *kurang ajar* 'insolent' (A)     [*kurang* 'less' + *ajar* 'to teach']

(13)  V-A compounds
    a.  *kenal pasti* 'to identify' (V)    [*kenal* 'to recognize' + *pasti* 'certain']
    b.  *(kerja) sukarela* 'voluntary (work)' (A)  [*suka* 'to like' + *rela* 'willing']

In standardized orthography, some compounds are written as two separate units demarcated by a hyphen or a space, while others are written as one unit.

As with other languages, Malay compounds may be categorized as endo-centric or exocentric depending on how the elements of a compound relate to

the compound with respect to their structure and meaning. Endocentric compounds are ones where one of the elements of the compound serves as the head of the compound. The meaning of an endocentric compound stands in a "type-of" relation with the meaning of the head of the compound. Examples of endocentric compounds are given in (14).

(14)  Endocentric compounds
    a.   *anak angkat* 'adoptee' (N)    [*anak* 'son/daughter' + *angkat* 'take']
    b.   *harga mati* 'fixed price' (N)    [*harga* 'price' + *mati* 'dead']
    c.   *ubi kayu* 'cassava' (N)    [*ubi* 'a kind of tuber' + *kayu* 'wood']
    d.   *ibu jari* 'thumb' (N)    [*ibu* 'mother' + *jari* 'finger']

As shown in (14), endocentric nominal compounds in Malay may be left-headed or right-headed. (14a)–(14c) are left-headed in the sense that the left element of the compound determines the overall meaning and syntactic category of the compound. Thus, *anak angkat* 'adoptee' is a type of son/daughter, and *harga mati* 'fixed price' is a type of price, and so on. (14d) is right-headed in that *ibu jari* 'thumb' is a type of finger. Exocentric compounds on the other hand are ones without such a head. Examples are given below:

(15)  Exocentric compounds
    a.   *keras kepala* 'obstinate' (A)    [*keras* 'hard' + *kepala* 'head']
    b.   *kecil hati* 'spiteful' (A)    [*kecil* 'small' + *hati* 'liver']
    c.   *ibu bapa* 'parents' (N)    [*ibu* 'mother' + *bapa* 'father']
    d.   *pergi balik* 'to travel back and forth' (V)    [*pergi* 'go' + *balik* 'return']

Compounds in Malay may also be categorized according to the semantic relationship that holds between the elements of the compound, in terms of whether the relation is one of coordination or subordination. Endocentric compounds involve a subordination relationship between the elements of the compound, as the elements of the compound do not have equal status. Some exocentric compounds such as (15c) and (15d) involve a coordinate relation between their constituent parts.

    Being words rather than phrases (e.g. serial verb constructions), compounds can undergo further morphological processes such as affixation and reduplication.

(16)  Affixation
    a.   *ber-* + *kerjasama* 'cooperation' (N) = *bekerjasama* 'to cooperate' (V)
    b.   *peN- -an* + *salah guna* 'to misuse' (V) = *penyalahgunaan* 'abuse' (N)
    c.   *suka rela* 'voluntary' (A) + *-wan* = *sukarelawan* 'volunteer' (N)

(17) Reduplication
    a.   *kakitangan-kakitangan* 'staffs' (N)
    b.   *meja makan-meja makan* 'dining tables' (N)
    c.   *sukarelawan-sukarelawan* 'volunteers' (N)

## 2.2 Psycho-collocations

Matisoff (1986: 9) defines a 'psycho-collocation' as "a polymorphemic expression referring as a whole to a mental process, quality, or state, one of whose constituents is a psycho-noun, i.e. a noun with explicit psychological reference (translatable by English words like HEART, MIND, SPIRIT, SOUL, TEMPER, NATURE, DISPOSITION, MOOD)." Many of the psycho-nouns are body-part terms. Psycho-collocations are common in Malay (Oey 1990). Examples are given below.

(18) *hati* 'heart, liver'[7]
| | |
|---|---|
| *baik hati* 'kind-hearted' | [*baik* 'good'] |
| *besar hati* 'glad, proud' | [*besar* 'big'] |
| *hancur hati* 'broken-hearted' | [*hancur* 'crushed'] |
| *kecil hati* 'narrow-minded' | [*kecil* 'small'] |
| *keras hati* 'persevering; obstinate' | [*keras* 'hard'] |

(19) *kepala* 'head'
| | |
|---|---|
| *kepala batu* 'obstinate' | [*batu* 'stone'] |
| *kepala berat*, *berat kepala* 'stupid, dull' | [*berat* 'heavy'] |
| *kepala besar* 'proud' | [*besar* 'big'] |
| *kepala dingin*, *dingin kepala* 'cool-headed' | [*dingin* 'cold'] |
| *ringan kepala* 'clever, smart' | [*ringan* 'light'] |
| *keras kepala* 'stubborn' | [*keras* 'hard'] |

(20) *hidung* 'nose'
| | |
|---|---|
| *hidung tinggi* 'proud' | [*tinggi* 'high'] |
| *keras hidung* 'narrow-minded' | [*keras* 'hard'] |

(21) *mulut* 'mouth'
| | |
|---|---|
| *berat mulut* 'seldom talk, quiet' | [*berat* 'heavy'] |
| *besar mulut* 'brag' | [*besar* 'big'] |

---

**7** See Goddard (2001) for other psycho-collocations with the noun *hati* 'heart, liver'.

(22) *muka* 'face'

| | |
|---|---|
| *muka batu* 'poker face' | [*batu* 'stone'] |
| *muka manis* 'sweetface' | [*manis* 'sweet'] |
| *cari muka* 'to brown-nose' | [*cari* 'to search'] |
| *hilang muka* 'to lose face' | [*hilang* 'to lose'] |
| *tebal muka, muka tebal* 'impudent' | [*tebal* 'thick'] |

(23) *telinga* 'ear'

| | |
|---|---|
| *telinga nipis* 'easily affronted' | [*nipis* 'thin'] |
| *tebal telinga* 'indifferent' | [*tebal* 'thick'] |

## 2.3 Elaborate expressions

Elaborate expressions usually have four syllables and are "rhyming/alliterative sound symbolic items" according to Enfield (2005). Their equivalents are referred to as 'pair-words' (in English) in Zainu'l-Abidin (1927).

Pair-words in Malay can be classified into three types. The first type is compounds comprising two distinct but phonologically similar words that convey related meanings. Examples are listed in (24).

(24) *kurus kering* 'thin and haggard' (= *kurus* 'thin' + *kering* 'dry')
*kaya-raya* 'prosperous' (= *kaya* 'rich' + *raya* 'great')
*pecah belah* 'disintegrated' (= *pecah* 'broken' + *belah* 'slit; side')

The second type can be analyzed as a type of rhythmic reduplication ('imitative reduplication' in Sneddon et al's (2010) term), whereby the base is repeated with changes in vowels and/or consonants. Examples of this type of pair-words are given in (25).

(25) a. Vowel alternation
*gunung-ganang* 'mountain range' [*gunung* 'mountain']
*lenggak-lenggok* 'to sway' [*lenggok* 'body movement when dancing']
*serba-serbi* 'all sorts of' [*serba* 'all']
b. Consonant alternation
*kuih-muih* 'an assortment of cakes' [*kuih* 'cake']
*kacau-bilau* 'chaotic' [*kacau* 'in disorder']
*saki-baki* 'remainder' [*baki* 'remainder']
c. Vowel and consonant alternation
*liang-liuk* 'swaying' [*liuk* 'to turn to the left and right']

> *bengk**ang**-bengk**ok*** 'winding' [*bengkok* 'crooked']
> ***pi**ndah-**ra**ndah* 'to move constantly' [*pindah* 'to move']

The third type of pair-words involve neither compounding nor reduplication, and are referred to as 'inherently reduplicated words' in Nomoto (2012). While these "pair-words" appear to consist of two phonologically similar elements, neither of the apparent component parts cannot be considered the base. Examples are given in (26).

(26)  *haru-biru* 'uproarious'
      *huru-hara* 'clamourous'
      *dolak-dalik* 'vacillating, uncertain'

Further information about Malay pair-words and their sound patterns can be found in Abdullah (1974: 71–72), Tham (1979), Mohd Yunus and Zaitul Azma (2011) and Nomoto (2012).

## 2.4 Reduplication

In addition to the rhythmic reduplication discussed in the last section, Malay also has full reduplication and partial reduplication. In full reduplication, the entire stem is repeated as in (27), while in partial reduplication only a part of the stem is repeated as in (28). Full reduplication of noun stems is associated with the explicit marking of plurality as will be elaborated further below.

(27)  *pokok* 'tree(s)'        *pokok-pokok* 'trees'
      *buku* 'book(s)'         *buku-buku* 'books'
      *pendapat* 'opinion(s)'  *pendapat-pendapat* 'opinions'

(28)  *budak* 'kid(s)'     *be-budak* 'kids'
      *langit* 'sky'       *le-langit* 'roof of the mouth'
      *berapa* 'how many'  *be-berapa* 'several'

Stems of various syntactic categories can undergo full reduplication. The reduplication of different lexical categories is associated with different productive meanings. While full reduplication is in general productive in Malay, the productivity of partial reduplication is unclear. See Zaharani (2007: 59–60) for restrictions on partial reduplication in Malay.

The function of reduplication for noun stems is to express explicit plurality as mentioned in relation to (27) above. The plurality expressed is considered explicit because a bare NP such as *pokok* 'tree(s)' is number-neutral, and may refer to one or more trees, while its reduplicated form *pokok-pokok* 'trees' refers only to multiple trees. This use of reduplication is possible for all count nouns. The plural marking by reduplication here is considered 'ostensibly optional' in the sense that the denotation of the marked form (the reduplicated form) is a proper subset of the unmarked form (the bare NP). This is in contrast to 'genuinely optional' plural marking, where the denotations of the forms with and without marking are identical (see Nomoto 2013a). Because both a bare noun and a reduplicated noun can refer to plural entities, a question arises as to what governs the choice between the two. There is no clear answer to this question at the moment, though several factors have been offered in the literature. For example, Mintz (2002: 283) writes that a bare noun is used when the (plural) referents are seen as a unit, while a reduplicated form is used when they are seen as many individual parts. On the other hand, Nomoto (2013a) suggests that since bare NPs are vague with respect to singular and plural reference, when only multiple entities are intended as the referent, reduplication is obligatory. On a similar line, Sneddon et al. (2010: 20) state that in Indonesian, a noun is reduplicated only if the plurality of the referent is important to what the speaker wishes to convey, while Shiohara (2011, 2012) notes that the use of reduplicated nouns in Indonesian tends to occur when the referent is present visually.

Wh-words may be reduplicated in Malay. When they are reduplicated, they become indeterminates 'any X', as shown in (29).

(29) *apa* 'what'     *apa-apa* 'anything'
     *siapa* 'who'     *se-siapa/siapa-siapa* 'anybody'
     *bila* 'when'     *bila-bila* 'anytime'

The reduplication of verbs has been associated with one of the following meanings: (i) repetitiveness, (ii) reciprocity and (iii) immediacy (see Mintz 2002: 295–297 and Asmah 1993: 203–210 for details). Examples are given in (30). Note that in V-V reduplication, only the stem is reduplicated.

(30) *muntah* 'to vomit'                    *muntah-muntah* 'to keep on vomiting'
     *ber-* + *teriak* 'to shout'            *berteriak-teriak* 'to shout and shout'
     *meN-* + *tolong* 'to help'             *tolong-men-[t]olong* 'to help one another'
     *meN-* + *hormat* + *-i* 'to respect'   *hormat-menghormati* 'to respect each other'
     *bangun* 'to wake up'                   *bangun-bangun* 'as soon as X wakes up'

The reduplication of adjectives and adverbs indicates intensity as shown below:

(31) *marah* 'angry'    *marah-marah* 'very angry'
    *kecil* 'small'     *kecil-kecil* 'very small'

(32) *jarang* 'rarely'      *jarang-jarang* 'very rarely'
    *sering* 'frequently'    *sering-sering* 'very frequently'

Negation and auxiliaries may also be reduplicated. Reduplication does not appear to have a productive function in these cases.

(33) *bukan* 'not'    *bukan-bukan* 'unfounded'

(34) *sudah* 'have (perfect), already'    *sudah-sudah* 'that's enough'
    *se-+boleh* 'can'+-*nya*          *seboleh-bolehnya* 'as much as possible'

# 3 Grammar and Syntax

## 3.1 Nominal domain

### 3.1.1 Basic structure of the NP

The linear order of the elements composing NPs in Malay is as follows:

(35) **Quant Num CLF N        Direct modifier    Poss    Relative clause Dem**
    *semua tiga   ekor kucing berwarna hitam   saya   yang   comel    itu*
    all    three  CLF  cat    coloured black   1SG    REL    cute     that
    'all those three cute black cats of mine'

In Colloquial Malay, the possessor can also occur prenominally, before N, followed by the possessive marker *punya*.

(36) *dia     punya    mak*                          Colloquial Malay
    3SG    POSS    mother
    'his/her mother'

The emergence of *punya* as a productive possessive marker can be traced to around the middle of the 19th century, when a huge wave of Chinese immi-

grants were brought to work in the tin mining industry, though it is possible that they entered the language earlier, perhaps around the 15th century (Yap 2007). The immigrants spoke southern Chinese languages such as Hokkien, Cantonese, Hakka and Teochew. These Chinese languages have a form corresponding to *punya*, namely *e* in Hokkien, *ge* in Cantonese, *ke* in Hakka and *kai* in Teochew, and the nominal structure indicating possession in these languages are possessor-possessive marker-possessed, as shown in the Cantonese example in (37) (Yap 2007).

(37) *nei      ge       dikjan*
     2SG    POSS    enemy
     'your enemy'

Malay does not use definite or indefinite articles. Bare NPs can be either definite or indefinite. However, there are devices to mark an NP as definite: the demonstrative *itu* 'that' and *ini* 'this', as well as determiners like *tersebut* 'mentioned', *berkenaan* 'relevant', *terbabit* 'involved', etc. These words all occur in the demonstrative position in (35). There are also words that bring about indefinite interpretations. They are the numeral *satu* 'one' and expressions containing its prefixal form *se-* such as *se-*CLF 'a', *(se-)suatu* 'some', *se-barang* 'any', etc. (see Shoho 2001 and Nomoto and Aznur Aisyah, to appear, for a description of in-definite expressions in Malay). For number marking in Malay, see section 2.4 on reduplication.

### 3.1.2  Classifier (CLF) device

Malay has numeral classifiers. Zainudin's (2000) classifier dictionary for primary school students lists 100 words that are considered classifiers. Among them, the most frequently used are *buah* (lit. 'fruit') for three-dimensional objects and abstract entities, *orang* (lit. 'human') for humans and *ekor* (lit. 'tail') for animals.

    Classifiers are used when a numeral modifies a noun. However, numerals can modify nouns without the presence of numeral classifiers in Malay as shown in (38).

(38) *tiga      (buah)     majalah*
     three    CLF    magazine
     'three magazines'

Nomoto (2013b) points out that an interpretive difference exists between expressions with and without a classifier: while both expressions can refer to particular instantiations of a kind, only expressions without a classifier can also refer to subclasses of a kind.[8] Thus, both the noun phrases with and without the classifier can be understood to refer to copies of a magazine (instantiations of the 'magazine' kind) as shown in (39a). Only the noun phrase without the classifier can be understood to refer to different titles (subkinds) of magazines as shown in (39b).

(39) a. *Masih    tinggal   {tiga    buah   majalah/   tiga    majalah}.   Semuanya*
      still     left      three  CLF    magazine  three  magazine  all

      *majalah    Mastika.*
      magazine   Mastika.
      'We still have three (copies of) magazines. All of them are Mastika.'

   b. *Masih    tinggal   {\*tiga  buah   majalah/   tiga    majalah},   iaitu*
      still     left      three  CLF    magazine  three  magazine  namely

      *majalah    Mastika,   Majalah    PC    dan   Nona.*
      magazine   Mastika   magazine   PC   and   Nona
      'We still have three (titles of) magazines, namely Mastika, Majalah PC and Nona.'
      (The sentence is acceptable with *tiga buah majalah* [three CLF magazine] if exactly one copy is left for each title.)

While numerals can modify nouns without classifiers, it is not possible to use classifiers without numerals as shown in (40).

(40) *\*Saya    ingin    mem-beli    buah    buku.*
     1SG     want    ACT-buy    CLF    book

Malay is thus unlike some Chinese languages such as Cantonese, and some MSEA languages such as Vietnamese, where the classifier + noun sequence can occur without a numeral.

---

**8** This interpretative difference between expressions with and without a classifier has also been reported in Thai for the sequences 'NP CLF Dem' versus 'NP Dem' (Piriyawiboon 2009).

(41) a. *ngo    seung    maai    bun    syu.*                                    Cantonese
       1SG      want     buy     CLF    book
       'I want to buy a book.'

   b. *Tôi    muốn    mua    cuốn    sách.*                                 Vietnamese
       1SG    want    buy    CLF     book
       'I want to buy a book.' (T. H. Nguyen, 2004: 17, cited in Tran 2011: 36)

Malay is also unlike some classifier languages (e.g. Mandarin) where the numeral *one* may be absent (or unpronounced) in the presence of a demonstrative (Tang 1990).

(42) *\*(se-/satu)    buah    buku    ini*                                        Malay
      one            CLF     book    this

(43) *zhe    (yi)    ben    shu*                                             Mandarin
      this    one    CLF    book
      'this book'

The NP in the 'Num CLF NP' structure can be null in Malay. For example, an utterance like (44) is common at a local market when the seller and the buyer are clear on the referent intended. The classifier *biji* (lit. 'seed') is used for nouns denoting small objects.

(44) *Bagi    dua    biji.*
      give    two    CLF
      'Give me two.'

Such structures can also be used anaphorically, when a referent has been established in the prior linguistic context. This use of the 'Num CLF' sequence is also found in Thai (Haas 1942: 204).

## 3.2 Verbal domain

### 3.2.1 Inflection

Like many MSEA languages, verbs in Malay do not inflect for the person, gender or number of the subject, as shown in (45).

(45) *Saya/kami/kamu/dia/mereka    mem-beli    surat khabar.*
    1SG/1PL/2SG/3SG/3PL            ACT-buy      newspaper
     'I/we/he/she/they/you bought/buy/buys the newspapers.'

Tense is not overtly expressed in Malay either through verbal inflection or temporal auxiliaries. The temporal location of an event/state is determined contextually, or established by a temporal adverbial as in (46).

(46) *Ali    berjumpa    Hassan    pada    pukul    sepuluh    semalam.*
    Ali    meet        Hassan    at      o'clock  ten        yesterday
    'Ali met Hassan at ten yesterday.'

Besides these strategies, the situation aspect (Akstionsart) and viewpoint aspect expressed by the sentence also contribute to temporal interpretation. According to Smith and Erbaugh (2005), unbounded situations receive a default present intepretation, while bounded events receive a default past interpretation in Mandarin Chinese. The same is found in Malay. (47), which describes a state (an unbounded situation), is interpreted as holding in the present, while (48), which describes an achievement (a bounded event), is interpreted as having occurred in the past.

(47) *Mereka    berdua    berada    di    stesen    Kajang.*
    3PL        both      be        at    station   Kajang
    'They are at the Kajang station (together).'

(48) *Mereka    berdua    tiba    di    stesen    Kajang.*
    3PL        both      arrive  at    station   Kajang
    'They arrived at the Kajang station (together).'

The default temporal interpretation can be overridden. For example, the presence of the temporal adverbial in (49) ensures a past interpretation, even though the sentence describes a state.

(49) *Mereka    berdua    berada    di    stesen    Kajang    semalam.*
    3PL        both      be        at    station   Kajang    yesterday
    'They were at the Kajang station (together) yesterday.'

Aspectual and modal information in Malay are not expressed through verbal inflection, but rather through auxiliaries. Aspectual/modal auxilairies typically occur between the subject and the predicate as in (50).

(50)  *Ali*  ***pernah/boleh/sudah***  *bermain*  *gitar.*
      Ali   have.experienced/can/have.completed   play   guitar
      'Ali has played guitar before/can play guitar/has played the guitar.'

While Malay verbs do not inflect for tense, aspect or mood information, they take affixes, with a variety of functions (see Table 2), some of which have been noted to have aspectual effects. For example, the prefix *ter-* has been noted to encode the completive aspect (e.g. Winstedt 1927; Asmah 2009: 146; Nik Safiah et al. 2008: 174; Soh 1994; Chung 2011). This is because the use of *ter-* indicates the completion of the event described and/or the result state as in (51).

(51)  Completive/result state
      *Pencuri*  *itu*  *ter-tangkap*  *oleh*  *polis.*
      thief   that   TER-catch   by   police
      'The thief has been caught by the police.'

However, the prefix *ter-* is also used to indicate that the event described by the sentence is accidental or not intended by the agent, or to express that the agent has the ability to carry out the activity/event described by the sentence.[9]

(52)  Accidental
      *Dia*  *ter-makan*  *racun.*
      3SG   TER-eat   poison
      'S/he drank poison accidentally.'

(53)  Abilitative
      *Orang*  *itu*  *tidak*  *ter-baca*  *tulisan*  *Cina.*
      person   that   not   TER-read   writing   China
      'That person could not read Chinese characters.'

Given that the same morpheme is associated with these different semantic effects, many authors have attempted to determine the basic meaning of *ter-* that explains its different semantic effects (see Soh 1994; Siraj 2010; Chung 2011; Nomoto 2011).

   Another prefix that has been associated with an aspectual effect is the prefix *meN-* (Benjamin 1993; Soh and Nomoto 2009, 2011). Soh and Nomoto (2009,

---

**9** The abilitative use is not as productive as the other two uses, which leads Nomoto (2011) to claim that the abilitative reading of *ter-* sentences is obtained by pragmatic inference.

2011) note that *meN-* has progressive-like aspectual effects; *meN-* does not occur in stative sentences like the progressive aspect.

(54) *Saya     suka/\*meny-[s]uka     masakan     Jepun.*
     1SG     like/ACT-like        cuisine      Japan
     'I like Japanese cuisine.'

### 3.2.2 Verbal categories

Several morphemes in Malay have both lexical and functional uses. The functional use arguably evolved as a consequence of grammaticalization, as is often the case with many MSEA languages. We discuss two such morphemes in this section: *dapat* and *kena*. The word *dapat* is used as a verb meaning 'to get' as well as an auxiliary expressing ability/possibility/permission.

(55) a. *Aminah* **dapat** *hadiah     daripada     kekasih-nya.*
        Aminah    get      present     from        lover-3SG
        'Aminah got a present from her lover.'

     b. *Aminah* **dapat** *mem-beli     hadiah     untuk     kekasih-nya.*
        Aminah    can      ACT-buy    present    for       lover-3SG
        'Aminah could buy a present for her lover.'

The set of meanings associated with *dapat* 'to get' is similar to those expressed by the verbs meaning 'to acquire' in MSEA languages discussed in Enfield (2001) (e.g. Lao *dâj*, Northern Zhuang *dáy*, Cantonese *dak* and Mandarin *de*).

     The word *kena* may be a verb expressing physical and abstract contact, meaning 'to touch, hit' (56a) or 'to incur' (56b). It may also be a modal verb that occurs in adversative passive and debitive ('have to') sentences, as shown in (56c) and (56d) respectively.

(56) a. *Tangan     saya* **kena** *pisau.*
        hand      1SG     touch    knife
        'My hand touched a knife.'

     b. *Gerai-gerai     kecil     pun     boleh* **kena** *cukai.*
        stall-PL        small    even    can     incur    tax
        'Even small stalls can get taxed.'

c. *Ali* ***kena*** *pukul* *oleh* *budak* *itu.*
   Ali    get    hit    by    kid    that
   'Ali got hit by that kid.'

d. *Kita* ***kena*** *mem-baca* *lebih* *banyak* *buku.*
   1PL    have.to    ACT-read    more    many    book
   'We have to read more books.'

Malay patterns like several MSEA languages such as Thai, Vietnamese (Prasith-rathsint 2004), Hokkien (Bodman 1955), and Khmer (Hiromi Ueda, p. c.) in using the same morpheme for adversative passive and debitive sentences: *thùuk* (Thai), *bị* (Vietnamese), *tioq* (Hokkien) and *trəw* (Khmer). According to Nomoto and Kar-tini (2011), *kena* started its life as a transitive verb of physical contact meaning 'to hit, touch'. Then, the semantic domain expanded to subsume abstract contact to mean 'to get, incur'. An adversative passive use emerged as a result of a change in the verb's subcategorization frame; *kena* started to take a clausal complement in addition to a noun phrase. The debitive use arose because the modal meaning of *kena* led the speakers to reanalyze the verb as a member of a special predicate class, which involves ambiguity in terms of thematic role assignment.

### 3.2.3 Serial verb constructions

Serial verb constructions (SVCs) are monoclausal structures containing two or more verbs. The verbs/verb phrases present aspects of a single event and they are typically characterized as involving the sharing of arguments between or among verbs (see Muysken and Veenstra 2006 and Bril 2004 for further criteria). SVC-like structures are common in Malay. In (57), the subject serves as an external argument of both verbs, while in (58), the object serves as an internal argument of the first verb but the external argument of the second verb.[10]

(57) *Ali* ***mem-[p]ekik*** ***me-minta*** *tolong.*
     Ali    ACT-shout    ACT-ask    help
     'Ali shouted calling for help.'

(58) *Ali* ***meng-hantar*** *Siti* ***pulang*** *ke* *rumah-nya.*
     Ali    ACT-send    Siti    return    to    house-3SG
     'Ali took Siti back to her house.'

---

**10** (57) and (58) are cited with modifications from Koh (1990).

As illustrated in (57) and (58), the verbs may be contiguous or non-continguous, exemplifying 'nuclear type SVC' and 'core type SVC' respectively (Foley and Olson 1985).

Both symmetrical and asymmetrical SVCs are attested in Malay. In symmetrical SVCs, the predicates involved are of equal rank, while in asymmetrical SVCs, the predicates are hierarchized and are headed, implying a modifier-head relation (Bril 2004). Thus, one of the participating verbs may be drawn from a grammatically or semantically restricted class in asymmetrical SVCs (e.g. verbs expressing direction, motion, posture, cause-effect, aspect, modality, etc.), while no such restriction exists for symmetrical SVCs (Bril 2004; Aikhenvald 2006). (59) is an example of a symmetrical SVC.

(59) *Mereka* **terkejut** **men-dengar** *berita kematian-nya.*
  3PL    be.surprised ACT-hear    news  death-3SG
  'They were surprised, hearing the news about his death.'

(60)–(62) are examples of asymmetrical SVCs, expressing direction and motion, causative and aspect/modality respectively.

(60) a. *Dia* **pergi** *meng-ambil songkok-nya segera.*
    3SG   go    ACT-take  Malay.cap-3SG immediately
    'He went to get his songkok immediately.' (Koh 1990: 292)

   b. *Norisah* **masuk** *men-jatuh-kan diri-nya di atas sofa …*
    Norisah  enter   ACT-fall-CAUS  self-3SG at top sofa
    'Norisah entered (and) dropped herself onto the sofa.' (Koh 1990: 293)

(61) a. *Saya* *mem-***buat*** adik      me-nyanyi lagu.*
    1SG   ACT-make younger.sibling ACT-sing  song
    'I made my younger brother/sister sing a song.'

   b. *Emak* **bagi/kasi** *anak keluar bermain.*  Colloquial Malay
    mum   give    child go.out play
    'Mum let her child go out and play.'

(62) *Ali* **mula/habis/cuba/mahu** *makan kari ayam.*
  Ali  begin/finish/try/want   eat  curry chicken
  'Ali began/finished/tried/wanted to eat chicken curry.'

## 3.3 Clausal/sentential organization

Malay is an SVO language. Other word orders are possible with information structure effects. (63a) is an SVO sentence used in unmarked contexts. (63b) and (63c) are respectively OSV and VOS sentences used in marked contexts, as reflected in their English translations.

(63) a. *Siti    makan    sayur    saja.*
Siti    eat    vegetable    only
'Siti only eats vegetables.'

b. *Sayur    saja    (yang)    Siti    makan.*
vegetable    only    REL    Siti    eat
'(It is) only vegetables (that) Siti eats.'

c. *Makan    sayur    saja,    Siti.*
eat    vegetable    only    Siti
'(She) only eats vegetables, Siti.'

### 3.3.1 Ellipsis of arguments

Argument ellipsis occurs in both Formal and Colloquial Malay. However, it is more frequent and less restricted in terms of possible ellipsis environments in Colloquial Malay than in Formal Malay (Koh 1990: 142). Eliding the subject of a subordinate or coordinate clause as in (64) is commonly found in both varieties.

(64) *Sebelum    ___    men-[t]inggalkan    pejabat    tadi,*    Formal Malay
before    ACT-leave    office    just.now

*Lokman    kelihatan    gamam ...*
Lokman    appear    anxious
'Just before (he) left the office, Lokman appeared anxious ...' (Koh 1990: 142)

By contrast, eliding a matrix subject as in (65) is virtually limited to Colloquial Malay.

(65) ... *ini-lah    hari    terakhir    aku    bersekolah.*    Colloquial Malay
this-PART    day    last    1SG    go.to.school

___   *sedih   le   jugak,   tapi   apa*   ___   *boleh   buat* ...
      sad   PART   also   but   what       can   do

'... this was my last day of school. (I) was sad, but what could (I) do ...' (Koh 1990: 143)

Colloquial Malay also allows an object argument to be elided, which makes the following sentence with a relative clause ambiguous.

(66) *Ada   yang   tak   suka.*                 Colloquial Malay
    exist   REL   not   like
    a.   'There are people who don't like <something>.'
    b.   'There are people/things that <I/we/they> don't like.'

In interpretation (a), the object argument is elided. In interpretation (b), it is the subject argument that is elided.

    While argument ellipsis is common, its occurrence is not without restrictions. In particular, prepositional objects cannot be elided in Malay as shown in (67) (cf. Fortin 2007; Sato 2007).

(67) a.   *Dia   juga   nak   pergi   ke   \*(sana).*     Colloquial Malay
        3SG   also   want   go   to   there
        'S/he doesn't want to go there either.'

    b.   *Ali   tidak   suka   pada   \*(mereka).*     Formal Malay
        Ali   not   like   at/on   3PL
        'Ali doesn't like them.'

### 3.3.2  Information structure

In a Malay sentence, the topic can appear at the beginning or the end, separated from the rest of the sentence by a potential pause, which is more frequently realized when topics appear at the end of the sentence (Philips 1970).

(68) *Lelaki   itu,   nama-nya   Hasan.*
    man   that   name-3SG   Hasan
    'That man, his name is Hasan.'

(69) *Nama-nya    Hasan,    lelaki    itu.*
name-3SG    Hasan    man    that
'His name is Hasan, that man.'

In the examples above, the topic NP anaphorically binds the possessor, which is the pronoun *-nya*. There are also topic constructions that do not involve a pronoun anaphorically bound by the topic (Philips 1970). In the examples below, the position in which the topic NP would occur in the corresponding non-topic construction is indicated by the underline.

(70) *Itu    semua    orang    sudah    tahu* ___.
that    all    person    PFV    know
'That, everybody already knows.' (Utopia, October 2005, p. 51)

(71) *Berita    itu    kita    sudah    dengar    ___.*[11]
news    that    1PL    PFV    hear
'We have already heard the news.'

Furthermore, some topic NPs do not appear to be an argument of the predicate. For such topics, it may not be straightforward to identify the positions in the corresponding non-topic constructions in which they would occur. Examples are given below.

(72) *Kereta-kereta    itu,    ada    yang    berwarna    hitam.*
car-PL    that    exist    REL    coloured    black
'(Among) those cars, there are (ones) which are (coloured) black.' (Philips 1970: 561)

---

**11** There is a similar construction known as bare passives, also referred to as object-preposing constructions, passive type 2 and object voice. In bare passive sentences, the agent follows the auxiliary, adverb or negation and precedes the "bare" verb form, i.e. a form without the prefix *meN-* or *di-*. The bare passive counterpart of (71) is (i) below.

(i)    Berita    itu    sudah    kita    dengar.
news    that    PFV    1PL    hear
'We have already heard the news.'

While bare passive sentences may appear to be an instance of topicalization, Chung (1976) shows that they are syntactically passive.

(73) *Aduan      mengenai  longkang  tersumbat  akan  di-ambil    tindakan*
     complaint  regarding  drain      clogged    will  PASS-take   action

   *pemulihan    dalam    tempoh    3    hari.*
   remedy       within   period    3    day
   '(Regarding) complaints on clogged drains, remedial actions will be taken
   within three days.' (Machang District Council's website)[12]

It may be possible to analyze these sentences as involving a prepositional
phrase, but with the preposition not pronounced or dropped, a process that is
common in Colloquial Malay. The relevant prepositions for (72) and (73) would
be *antara* 'among' and *terhadap* 'on' respectively. The non-topic constructions
for these sentences would thus be as follows:

(74) **Antara**  *kereta    itu,      ada     yang    berwarna    hitam.*
     among    car       that      exist   REL     coloured    black
     'Among those cars, there are (ones) which are (coloured) black.'

(75) **Terhadap** *aduan       mengenai  longkang  tersumbat  akan  di-ambil*
     on          complaint  regarding  drain      clogged    will  PASS-take

   *tindakan    pemulihan    dalam    tempoh    3    hari.*
   action      remedy       within   period    3    day
   'Regarding complaints on clogged drains, remedial actions will be taken
   within three days.'

Colloquial Malay has a topic marker, namely *kalau*, as shown in (76) (Nomoto
2009).

(76) *Kalau    tempat    tu,      ramai    pelancong    Cina.*      Colloquial Malay
     TOP      place     that     many     tourist      China
     'Speaking of the place, there are lots of Chinese tourists there.'

*Kalau* as a topic marker is not available in Formal Malay, where it may only be
used to mean 'if'.

---

**12**  Pencapaian Piagam pelanggan – PORTAL RASMI MAJLIS DAERAH MACHANG (http://www.
mdmachang.gov.my/146?p_p_id=56_INSTANCE_h6eX&p_p_lifecycle=0&p_p_state=normal&p
_p_mode=view&p_p_col_id=column-2&p_p_col_count=2&page=3, accessed 29/07/2011).

Focus constituents in Malay may be marked by focus morphemes such as *kah* and *lah*; the former is interrogative while the latter is not (Mashudi 1981: 303; Nomoto 2009; Cole et al., to appear; Fortin 2009; Nomoto and Aznur Aisyah 2016). Constituents in focus marked by *kah* and *lah* may be fronted as illustrated in (77) and (78) respectively.

(77) a. *Dia mem-baca **di perpustakaan-lah** kelmarin.*
　　　 3SG ACT-read in library-FOC yesterday
　　　 'It was in the library that he was reading yesterday.' (Mashudi 1981: 99)

　 b. ***Di perpustakaan-lah** dia mem-baca kelmarin.*
　　　 at library-FOC 3SG ACT-read yesterday
　　　 'It was in the library that he was reading yesterday.' (Mashudi 1981: 99)

(78) a. *Guru itu pergi **ke universiti-kah** tadi?*
　　　 teacher that go to university-FOC just.now
　　　 'Was it to the university that the teacher went just now?' (Mashudi 1981: 100)

　 b. ***Ke universiti-kah** guru itu pergi tadi?*
　　　 to university-FOC teacher that go just.now
　　　 'Was it to the university that the teacher went just now?' (Mashudi 1981: 100)

Besides *kah* and *lah*, *pun* 'also, too, even' may also be considered a focus morpheme (Mashudi 1981: 96–98; Goddard 2001).[13, 14] (79a) is a neutral sentence while (79b) is a sentence where the verb phrase is in focus.

---

**13** Mashudi (1981: 97) also includes *tah* as a focus morpheme. *Tah* together with a wh-word marks a rhetorical question.

(i) Aminah itu apa-tah kurang-nya?
　　Aminah that what-FOC less-3SG
　　'What is less pretty in Aminah?' ('Aminah is just as pretty') (Mashudi 1981: 97)

The extent to which *tah* patterns like other focus morphemes is not discussed in Mashudi (1981).

**14** Nomoto (2006: 85) suggests that particles *pula* 'too' and *juga* 'also' are also focus markers.

(79) a. *Dia*    *tidak*    *mahu*    *mem-[p]andang.*
      3SG      not       want      ACT-look
      'S/he does not want to look.'

    b. ***Mem-[p]andang***    ***pun***    *dia*    *tidak*    *mahu.*
      ACT-look            FOC     3SG     not     want
      'He does not even want to look.'

### 3.3.3 Questions

Malay has several ways to form questions. In wh-questions, the wh-word may either stay in situ as in (80a) or occur in a sentence-initial position as in (80b). In addition, the wh-word can also occur in the initial position of an embedded clause as in (80c), showing that Malay has partial wh-movement (Cole and Hermon 2000). In all these sentences, the word *bila* 'when' may be associated with the embedded clause or the matrix clause, and hence may be questions about the time of the coming event or the time of the thinking event.

(80) a. *Awak*    *fikir*    *mereka*    *datang*    *ke*    *sini*    ***bila?***
      2SG      think    3PL       come      to     here    when
      'When did you think they came here?'

    b. ***Bila***    *awak*    *fikir*    *mereka*    *datang*    *ke*    *sini?*
      when    2SG     think    3PL        come     to    here
      'When did you think they came here?'

    c. *Awak*    *fikir*    ***bila***    *mereka*    *datang*    *ke*    *sini?*
      2SG     think    when    3PL        come     to    here
      'When did you think they came here?'

Yes-no questions are marked by a question intonation contour as in (81a)[15] and/ or one of the following morphosyntactic devices: using the question particle *kah* (*ka/ke* in Colloquial Malay) at the end of the sentence (81b), adding the question marker *adakah* at the beginning of the sentence (81c), adding *ke t(id)ak/belum* 'or not/not yet' at the end of the sentence (81d), or adding *tak* 'not' after the aux-

---

[15] Not all intonation contours involve a clear rising contour. Further research is needed to provide an adequate description of question intonation contours in Malay.

iliary or the predicate (81e) or using the particle *kan* at the end of the sentence (81f).

(81) a.  Question intonation
    *Awak    boleh    datang?*
    2SG    can    come
    'Can you come?'

    b.  Yes/no question particle *kah (ka/ke)*
    *Awak    boleh    datang-kah?*
    2SG    can    come-Q
    'Can you come?'

    c.  *Ada-kah* 'be + *kah*'
    *Ada-kah    awak    boleh    datang?*
    be-Q    2SG    can    come
    'Is it that you can come?'

    d.  *Ke tak/belum* 'or not/not yet'                       Colloquial Malay
    *Awak    boleh    datang    ke    tak?*
    2SG    can    come    or    not
    'Can you come or not?'

    e.  *Tak* 'not'                                      Colloquial Malay
    *Boleh    tak    awak    datang?*
    can    not    2SG    come
    'Can you come?'

    f.  Tag question with *kan*                         Colloquial Malay
    *Ali    dah    tahu    kan?*
    Ali    already    know    right
    'Ali already knew it, didn't he?'

Strategies exemplified in (80d)–(80f) are only found in Colloquial Malay.

# 4 Semantics and pragmatics

## 4.1 Common semantic domains

### 4.1.1 Food terminology

According to Matisoff (1983: 79), languages in Thailand are "unanimous in dividing food into two basic categories: rice and that which one eats with rice (e.g. Thai *khâaw* 'rice' and *kabkhâaw* 'with rice')," and this may serve as a semantic feature characterizing languages of Southeast Asia. Malay does so too, with distinct words for rice (i.e. *nasi*) and things eaten with rice (i.e. *lauk*).

Like most languages spoken in Thailand, Lahu and Mandarin (Matisoff 1983: 25), Malay has separate words for rice in the fields (i.e. *padi*) versus cooked rice ready to be eaten (i.e. *nasi*). In fact, Malay also has a word referring to uncooked rice, i.e. *beras*.

### 4.1.2 Expressions of "cutting", "carrying", "drying", "directional" (upper ~ lower, from ~ toward ...)

According to Matisoff (2004), Southeast Asian languages are rich in verbs of manipulation (e.g. CARRY, CUT), which reflect the physical lifestyle of the region. These verbs are in common use and are not specialized in a particular domain. Malay seems to fit this description. For example, there are many verbs for CARRY depending on the parts of the body involved in the act of carrying. Examples are given with the prefix *meN-*.

(82) CARRY
    *membawa* 'to carry along'
    *mengangkat*, *mendukung* 'to carry astride the back or waist'
    *menggendong* 'to carry on the back or hip'
    *mengusung* 'to carry on a stretcher'
    *menjinjing* 'to carry in hand'
    *menanai* 'to carry in hands'
    *memikul* 'to carry on shoulder'
    *menanggung*, *menjulang* 'to carry on shoulders'
    *menjunjung* 'to carry on head'
    *membimbit/mengelek* 'to carry under arm'

There are also a number of verbs for CUT and DRY, depending on how the action is carried out.

(83)  CUT
    *memotong* 'to cut'
    *menggunting* 'to cut with scissors'
    *menggergaji* 'to cut with saw'
    *mengapak* 'to cut with axe'
    *mencukur* 'to cut with razor'
    *menghiris* 'to cut by slicing'
    *membelah* 'to cut by splitting/halving'
    *mengerat* 'to cut into pieces'
    *mencencang* 'to cut into very small pieces'
    *menetak* 'to chop'

(84)  DRY
    *mengeringkan* 'to dry'
    *menjemurkan* 'to dry in the sun'
    *mengelap* 'to dry using a towel, piece of cloth, etc.'

Different verbs exist for PUSH, depending on whether or not an object moves as a result of being pushed.

(85)  PUSH
    *menekan* 'to push down, press'
    *menolak* 'to push away'

## 4.2 Pragmatics & discourse

### 4.2.1 Final particles

Malay has a number of final particles, which may be broadly categorized into two types: (i) interrogative particles and (ii) discourse particles. Interogative particles are discussed in section 3.3.3 above. Discourse particles indicate the speaker's attitude or belief about the propositional content of the utterance or his attitude or belief about other participants' attitude or belief on the relevant proposition. They include, but are not limited to, *kot/kut*, *punya*, *dah* and *lah/la/le*. *Kot* is an epistemic particle used to express uncertainty with respect to the propositional content of the utterance (Koh 1990: 78).

(86) *Ali    dah    tahu    **kot**.*                              Colloquial Malay
    Ali    PFV    know    I.guess
    'Ali knew it (but I may be wrong).'

*Punya*, on the other hand, indicates that the speaker is certain about the propositional content of the utterance (Koh 1990: 78; Yap 2007). Soh (2016) shows that in addition to expressing the attitude holder's certainty about the truth of the propositional content of the utterance, the use of *punya* also signals the source of the information presented as of the inferential type. The attitude holder may be the speaker or the external argument of verbs of saying and believing.

(87) *Ali    dah    tahu    **punya**.*                          Colloquial Malay
    Ali    PFV    know    for.sure
    'Ali knew it (I can assure you).'

The use of final *dah* may indicate the speaker's belief that one or more of the discourse participants hold a belief that is contrary to what is being asserted. Thus, the use of final *dah* in (88) indicates that the speaker believes that one or more of the discourse participants believe that Ali does not have knowledge of the matter (Soh 2017). This use of the final *dah* is comparable to the use of *already* in English (Soh 2011).

(88) *Ali    dah    tahu    **dah**.*                            Colloquial Malay
    Ali    PFV    know    already
    'Ali already knew it.'

The discourse particle *lah* is to be distinguished from the focus *lah* (cf. section 3.3.2). Depending on the intonation, the particle *lah* may convey either a "light-heartedness" or an "ill-tempered" effect in declaratives, and it may either "soften" or "harden" a request (Goddard 1994). According to Goddard (1994: 154), the different uses of *lah* can be explained by the fact that *lah* "offers an explanation of the speaker's illocutionary purpose, which is roughly to correct or preempt a misapprehension or misunderstanding of some kind." An example of the use of the discourse particle *lah* is given in (89).

(89) *Ali    dah    tahu    **lah**.*                            Colloquial Malay
    Ali    PFV    know    LAH
    'Ali already knew it (though you might have thought otherwise).'

It is possible for multiple final particles to co-occur. The order of multiple particles is not free, though a comprehensive description of it awaits further research. The interrogative particles must follow discourse particles as in (90) (Soh 2011).

(90) *Kau   sampai   kat   KLCC   {**dah**     ke/   *ke   **dah**}?*     Colloquial Malay
     2SG   arrive   at   KLCC   already   Q   Q     already
     'Have you arrived at KLCC?'

Within the discourse particles, Aoki (2012) notes that *punya* must precede *la(h)* as in (91).

(91) *Aku   tak   lari   ke   stesen   {**punya la**/   ***la punya**}.*     Colloquial Malay
     1SG   not   run   to   station   for.sure   LA     LA   for.sure
     'I won't run to the station.'


### 4.2.2  Politeness

In Malay, different pronouns are used in settings of different formality which require different levels of 'politeness' or 'closeness'. Singular pronouns in Malay have forms that differ in their degrees of politeness or closeness, as shown below.

(92)       polite/distant   ⟵——————⟶ non-polite/close
    1SG   *saya*                                    *aku*
    2SG   *anda*          *awak*   *kamu*   *engkau/kau*
    3SG   *beliau*                                  *dia*

As Koh (1990: 111) notes, non-polite/close pronouns "do not indicate impoliteness when used correctly." In addition to these pronouns, titles (e.g. *Dato'*, *Haji*, *Prof.*, *Dr.*, *Tuan*, *Encik*) and kin terms (e.g. *mak cik* 'aunt', *kakak* 'older sister', *adik* 'younger brother/sister', *abang* 'older brother') are also used as address terms and to encode different levels of politeness or closeness. For instance, a waiter would address his male customer as *abang* at a neighbourhood restaurant, but he would switch to *tuan* 'Sir (lit. master)' if he worked at a restaurant in a five-star hotel.

(93) *Abang/Tuan          nak     makan    apa?*
     older.brother/Sir   want   eat       what
     'What would you like?'

See Mintz (2002: 86–94) for a detailed description of different pronouns as well as titles and kin terms as pronoun substitutes.

# 5 Conclusion/ summary

Vittrant (2008) and Enfield (2005: 186–187) have hypothesized that the languages of the MSEA linguistic area share specific features in different components of the grammar. For example, the phonological systems of the languages in the area are rich in vowel phonemes, with suprasegmental phenomena such as tones or registers (Vittrant 2008). The vowel system of Standard Malay is considered small compared to many MSEA languages, while the vowel systems of northern Malay dialects are more in line with most MSEA languages. Unlike many MSEA languages, Malay does not have distinct tones and registers, nor is there evidence of the development of sesquisyllables. Like many Southeast Asian languages, consonant phonemes that may appear in word-final position in Malay represent a reduced set compared to those that may appear in word-initial position. However, it should be noted that this feature is also shared by many of the world's languages, including languages genetically related to Malay that are not spoken in Mainland Southeast Asia such as Mantauran (Rukai) (Zeitoun 2007: 21), and other Austronesian languages in Indonesia and Borneo such as Toba Batak (Percival 1970: 31) and Iban (Asmah 1981: 17). In this connection, Prince and Smolensky (2004: 188) put forward the following generalization relating to the asymmetry of onset/coda licensing: "There are languages in which some possible onsets are not possible codas, but no language in which some possible codas are not possible onsets."

In terms of morphosyntactic features, languages in the area have been described as exhibiting a tendency towards monosyllabicity, with extensive use of compounding (as opposed to derivational morphology), and showing common use of numeral classifiers, psycho-collocations and elaboratives (Vittrant 2008). There is no particular tendency towards monosyllabicity in Malay, as it is an agglutinating language, with many words consisting of multiple morphemes. Malay makes extensive use of compounding and derivation, as well as reduplication. It employs numeral classifiers in numeral modification, but direct numeral modification is also possible, unlike many MSEA languages. The use of psycho-collocations and elaboratives is common. It is relevant to note that the use of psycho-collocations is also common in other Austronesian languages, such as Tsou (Formosan, Taiwan), Gayo (Western Malayo-Polynesian, northern Sumatra) and Kambera (Central Malayo-Polynesian, Sumba Island, eastern Indone-

sia). The use of psycho-collocations is so extensive that Musgrave (2006: 233) argues that "it could have already been present in Proto-Austronesian and inherited wherever it occurs in the family." In terms of their syntax, MSEA languages are considered mainly SVO, with argument ellipsis, serial verb constructions and the lack of grammatical marking of tense. Malay shares these features with MSEA languages. However, unlike some MSEA languages such as Vietnamese, where wh-question words may only stay in situ (Bruening and Tran 2006), Malay wh-question words may either stay in situ or occur in the initial position of a clause, whether matrix or embedded. Like many MSEA languages, Malay has many verbs of manipulation such as CARRY and CUT. It also has a number of final particles that may be broadly categorized as interrogative or discourse-related. In addition, Malay has pronouns that differ in terms of levels of politeness or closeness.

It is possible that some of these features that Malay shares with MSEA languages have resulted from contacts with the latter languages. Needless to say, however, there are other possible reasons for the resemblance, including contact with non-MSEA languages (notably Chinese languages, which happened to share the relevant features with MSEA languages), genetic inheritance (only relevant under the Austric superfamily hypothesis, which claims that Austronesian and Austro-Asiatic languages developed from a common proto-language and hence genetically related), parallel development, and universals (see Campbell 2006; Tosco 2008). All of these possibilities must be taken into account when evaluating the extent to which a particular common feature in Malay and MSEA languages is due to contact between the two aided by the geographic contiguity.

# Abbreviations

| 1 | first person |
|---|---|
| 2 | second person |
| 3 | third person |
| A | adjective |
| ACT | active |
| C | consonant |
| CAUS | causative |
| CLF | classifier |
| DEM | demonstrative |
| FOC | focus |
| N | noun |
| NP | noun phrase |
| P | preposition |

| PART | particle |
| PASS | passive |
| PFV | perfective |
| PL | plural |
| POSS | possessive |
| PROG | progressive |
| Q | question |
| REL | relativizer |
| SG | singular |
| TOP | topic |
| V | verb |
| V | vowel |

# References

Abdullah Hassan. 1974. *The morphology of Malay*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. in Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions: A cross-linguistic typology*, 1–68. Oxford: Oxford University Press.

Aoki, Rina. 2012. *Kougo Mareego no danwashouji punya no tougoronteki kenkyuu*. [A syntactic study of the discourse particle *punya* in Colloquial Malay] BA Thesis, Tokyo University of Foreign Studies.

Asmah Haji Omar. 1981. *The Iban language of Sarawak: A grammatical description*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Asmah Haji Omar. 1993. *Essays on Malay linguistics*, 2nd ed. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Asmah Haji Omar. 2009. *Nahu Melayu mutakhir (edisi kelima)*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Benjamin, Geoffrey. 1993. Grammar and polity: The cultural and political background to Standard Malay. In William A. Foley (ed.), *The role of theory in language description*, 341–392. Berlin & New York: Mouton de Gruyter.

Bodman, Nicholas C. 1955. *Spoken Amoy Hokkien*. Kuala Lumpur: Charles Grenier & Company.

Bril, Isabelle. 2004. Complex nuclei in Oceanic languages: Contribution to an areal typology. In Isabelle Bril & Françoise Ozanne-Rivierre (eds.), *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundedness*, 1–48. Berlin & New York: Mouton de Gruyter.

Bruening, Benjamin & Thuan Tran. 2006. Wh-questions in Vietnamese. *Journal of East Asian Linguistics* 15. 319–347.

Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 1–31. Hampshire: Palgrave Macmillan.

Chung, Sandra. 1976. On the subject of two passives in Indonesian. In Charles N. Li (ed.), *Subject and topic*, 57–99. New York: Academic Press.

Chung, Siaw-Fong. 2011. Uses of *ter-* in Malay: A corpus-based study. *Journal of Pragmatics* 43. 799–813.

Cole, Peter & Gabriella Hermon. 2000. Partial wh-movement: Evidence from Malay. In Uli Lutz, Gereon Müller & Arnim Stechow (eds.), *Wh-scope marking*, 85–114. Amsterdam: John Benjamins.

Cole, Peter, Gabriella Hermon & Norhaida Aman. To appear. Clefted questions in Malay. In David Gil & James Collins (eds.), *Malay/Indonesian linguistics*. London: Curzon Press.

Cole, Peter, Gabriella Hermon & Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118. 1500–1553.

Donohue, Mark. 2004. Typology and linguistic areas. *Oceanic Linguistics* 43. 221–239.

Enfield, Nick J. 2001. On genetic and areal linguistics in Mainland South-East Asia: Parallel polyfunctionality of 'acquire'. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 255–290. Oxford: Oxford University Press.

Enfield, Nick J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.

Ferguson, Charles A. 1959. Diglossia. *Word* 15. 325–340.

Foley, William A. & Mike Olson. 1985. Clausehood and verb serialization. In Johanna Nichols & Anthony C. Woodbury (eds.), *Grammar inside and outside the clause: Some approaches to theory from the field*, 17–60. Cambridge: Cambridge University Press.

Fortin, Catherine R. 2007. *Indonesian sluicing and verb phrase ellipsis: Description and explanation in a Minimalist framework*. PhD thesis, University of Michigan.

Fortin, Catherine. 2009. On the left periphery in Indonesian. In Sandy Chung, Daniel Finer, Ileana Paul & Eric Potsdam (eds.), *Proceedings of the Sixteenth Meeting of the Austronesian Formal Linguistics Association*, 29–43. Santa Cruz: University of California.

Gil, David. 2007. Intonation and thematic roles in Riau Indonesian. In Chungmin Lee, Matthew Gordon & Daniel Büring (eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, 41–68. Dordrecht: Springer.

Goddard, Cliff. 1994. The meaning of *lah*: Understanding 'emphasis' in Malay (Bahasa Melayu). *Oceanic Linguistics* 33. 145–165.

Goddard, Cliff. 2001. The polyfunctional Malay focus particle *pun*. *Multilingua* 20. 27–59.

Haas, Mary R. 1942. The use of numeral classifiers in Thai. *Language* 18. 201–205.

Halim, Amran. 1981. *Intonation in relation to syntax in Bahasa Indonesia*. Canberra: Pacific Linguistics.

Husni Abu Bakar, Karthik Durvasula, Nadya Pincus & Tim McKinnon. 2006. Incomplete laryngeal neutralization in KL Malay. Paper presented at the 11th International Symposium on Malay/Indonesian Linguistics (ISMIL).

Iwasaki, Shoichi & Preeya Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge: Cambridge University Press.

Klamer, Marian. 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41. 363–383.

Koh, Ann Sweesun. 1990. *Topics in colloquial Malay*. PhD thesis, University of Melbourne.

Lim, Sonny. 1988. Baba Malay: The language of the 'Straits-born' Chinese. In Hein Steinhauer (ed.), *Papers in Western Austronesian linguistics* 3, 1–61. Canberra: Pacific Linguistics.

Mashudi Kader. 1981. *The syntax of Malay interrogatives*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Matisoff, James. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman (ed.), *Consonant types & tones*, 71–95. Los Angeles, CA: The Linguistic Program, University of Southern California.

Matisoff, James. 1983. Linguistic diversity and language contact in Thailand. In John McKinnon & Wanat Bhruksasri (eds.), *Highlanders of Thailand*, 56–86. Kuala Lumpur: Oxford University Press.

Matisoff, James. 1986. Hearts and minds in Southeast Asian languages and English: An essay in the comparative lexical semantics of psycho-collocations. *Cahiers de Lingustique – Asie Orientale* 15. 5–57.

Matisoff, James. 2004. Areal semantics – Is there such a thing? In Anju Saxena (ed.), *Himalayan languages, past and present*, 347–395. Berlin & New York: Mouton de Gruyter.

Mintz, Malcolm W. 2002. *An Indonesian & Malay grammar for students*. Perth: Indonesian/Malay Texts and Resources.

Mohd Yunus Sharum & Zaitul Azma Zainon Hamzah. 2011. Golongan dan rumus kata gandaan berima. *Jurnal Bahasa* 11. 27–47.

Musgrave, Simon. 2006. Complex emotion predicates in Eastern Indonesia: Evidence for language contact? In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 227–243. Hampshire: Palgrave Macmillan.

Muysken, Pieter & Tonjes Veenstra. 2006. Serial verbs. In Martin Everaert, Henk van Riemsdijk, Rob Goedemans & Bart Hollebrandse (eds.), *The Blackwell companion to syntax*, 234–270. Oxford: Blackwell.

Nguyen, Tuong Hung. 2004. *The structure of the Vietnamese noun phrase*. PhD thesis, Boston University.

Nik Safiah Karim, Farid M. Onn, Hashim Haji Musa & Abdul Hamid Mahmood. 2008. *Tatabahasa Dewan (edisi ketiga)*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Nomoto, Hiroki. 2006. *A study on complex existential sentences in Malay*. Master's thesis, Tokyo University of Foreign Studies.

Nomoto, Hiroki. 2009. On indefinite subjects of pivot verbs in Malay. *International Journal of Corpus Linguistics* 14. 221–254.

Nomoto, Hiroki. 2011. Mareeshiago no modariti no gaiyou [Modality in Malay: An overview]. *Gogaku Kenkyuujo Ronshuu* 16. 130–150.

Nomoto, Hiroki. 2012. Consonant-changing rhythmic reduplication in Malay as identity avoidance. Paper presented at the 19th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA).

Nomoto, Hiroki. 2013a. On the optionality of grammatical markers: A case study of voice marking in Malay/Indonesian. In Alexander Adelaar (ed.), *Voice variation in Austronesian languages of Indonesia*, 123–145. Jakarta & Tokyo: Universitas Katolik Indonesia Atma Jaya & Tokyo University of Foreign Studies.

Nomoto, Hiroki. 2013b. *Number in classifier languages*. PhD thesis, University of Minnesota.

Nomoto, Hiroki & Aznur Aisyah Abdullah. 2016. Mareeshiago no shouten hyougen to meishijutsugobun [Focus expressions and nominal predicate sentences in Malay]. *Gogaku Kenkyuujo Ronshuu* 21. 171–189.

Nomoto, Hiroki & Aznur Aisyah Abdullah. To appear. Mareeshiago no toritate joshi to hutei hyougen [Focus-sensitive particles and indefinites in Malay]. *Gogaku Kenkyuujo Ronshuu* 22.

Nomoto, Hiroki & Kartini Abd. Wahab. 2011. Konstruksi *kena* dalam bahasa Indonesia: Per-
    bandingan dengan bahasa Melayu. *Linguistik Indonesia* 29. 111–131.
Nomoto, Hiroki & Isamu Shoho. 2007. Voice in relative clauses in Malay: A comparison of writ-
    ten and spoken language. In Yuji Kawaguchi, Toshihiro Takagaki, Nobuo Tomimori &
    Yoi-chiro Tsuruga (eds.), *Corpus-based perspectives in linguistics*, 353–370.
    Amsterdam: John Benjamins.
Nordhoff, Sebastian. 2009. *A grammar of upcountry Sri Lanka Malay*. PhD thesis, University of
    Amsterdam.
Oey, Eric. 1990. Psycho-collocations in Malay: A Southeast Asian areal feature. *Linguistics of
    the Tibeto-Burman Area* 13. 141–158.
Farid M. Onn. 1980. *Aspects of Malay phonology and morphology: A generative approach*.
    Bangi: Universiti Kebangsaan Malaysia.
Percival, W. Keith. 1964. *A grammar of the Toba Batak*. PhD thesis, University of Michigan.
Philips, N. G. 1970. Topic clauses in Malay. *Bulletin of the School of Oriental and African Stud-
    ies, University of London* 33. 560–572.
Piriyawiboon, Nattaya. 2009. The role of classifiers in N + Dem in Thai. Paper presented at the
    19th annual meeting of the Southeast Asian Linguistics Society (SEALS).
Prasithrathsint, Amara. 2004. The adversative passive markers as a prominent areal feature of
    Southeast Asian languages. In Somsonge Buruphat (ed.), *Paper from the 11th Annual Mee-
    ting of Southeast Asian Linguistics Society 2001*, 583–598.
Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative
    Grammar*. Malden, MA: Blackwell.
Shoho, Isamu. 2001. Mareeshiago no hutei hyougen. [A research into Malay indefinite expres-
    sions] *Gogaku Kenkyuujo Ronshuu* 6. 71–91.
Sato, Yosuke. 2007. P-stranding generalization and Bahasa Indonesia: A myth? *Snippets* 16.
    17–18.
Shiohara, Asako. 2011. Reduplication and indication of plurality in Indonesian. Paper
    presented at the 15th International Symposium on Malay/Indonesian Linguistics  (ISMIL).

Shiohara, Asako. 2012. Indication of plurality in Indonesian as observed in Indonesian school
    textbooks. Paper presented at the 15th International Symposium on Malay/Indonesian
    Linguistics (ISMIL).
Siraj, Pasha. 2010. 'Accidental' *ter-* in Malay as an anti-bouletic modifier. Paper presented at
    the 17th meeting of the Austronesian Formal Linguistics Association (AFLA).
Smith, Carlota S. & Mary S. Erbaugh. 2005. Temporal interpretation in Mandarin Chinese. *Lin-
    guistics* 43. 713–756.
Sneddon, James N. 1993. The drift towards final open syllables in Sulawesi languages. *Oceanic
    Linguistics* 32. 1–44.
Sneddon, James N. S., Alexander K. Adelaar, Dwi N. Djenar & Michael Ewing. 2010. *Indonesian:
    A comprehensive grammar*, 2nd ed. London: Routledge.
Soh, Hooi Ling. 1994. *Aspect and the organization of argument structure and phrase structure:
    Evidence from Malay*. Master's thesis, University of Calgary.
Soh, Hooi Ling. 2011. The syntax of *dah* in Colloquial Malay. Paper presented at the 15th inter-
    national symposium on Malay/Indonesian linguistics (ISMIL).
Soh, Hooi Ling. 2013. Voice and aspect: Some notes from Malay. In Alexander Adelaar (ed.),
    *Voice variation in Austronesian languages of Indonesia*, 159–173. Jakarta & Tokyo: Univer-
    sitas Katolik Indonesia Atma Jaya & Tokyo University of Foreign Studies.

Soh, Hooi Ling. 2016. Evidentiality, modality, focus and presupposition: The case of the discourse particle *punya* in Colloquial Malay. Paper presented at the 23rd annual meeting of the Austronesian Formal Linguistics Association (AFLA).

Soh, Hooi Ling. 2017. The syntax and semantics of the aspectual particle *dah* in colloquial Malay. Manuscript, University of Minnesota.

Soh, Hooi Ling & Hiroki Nomoto. 2009. Progressive aspect, the verbal prefix *meN-*, and the stative sentences in Malay. *Oceanic Linguistics* 48. 148–171.

Soh, Hooi Ling & Hiroki Nomoto. 2011. The Malay verbal prefix *meN-* and the unergative/unaccusative distinction. *Journal of East Asian Linguistics* 20. 77–106.

Soh, Hooi Ling & Hiroki Nomoto. 2015. Degree achievements, telicity and the verbal prefix *meN-* in Malay. *Journal of Linguistics* 51. 147–183.

Tadmor, Uri. 2000. Rekonstruksi aksen kata bahasa Melayu. In Yassir Nasanius & Bambang Kaswanti Purwo (eds.), *PELBBA* 13: *Pertemuan Linguistik (Pusat Kajian) Bahasa dan Budaya Atma Jaya: Ketiga Belas,* 153–167. Pusat Kajian Bahasa dan Budaya, Unika Atma. Jaya, Jakarta.

Tang, C. C. Jane. 1990. *Chinese phrase structure and the extended X-bar theory*. PhD thesis, Cornell University.

Teoh, Boon Seong. 1994. *The sound system of Malay revisited*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Tham, Seong Chee. 1979. Vowel patterning and meaning in Malay pair-words. In Nguyen Dang Liem (ed.), *South-East Asian Linguistic Studies*, 365–377. Canberra: Pacific Linguistics.

Thurgood, Elzbieta A. 1998. *A description of nineteenth century Baba Malay: A Malay variety influenced by language shift*. PhD thesis, University of Hawai'i.

Tosco, Mauro. 2008. What to do when you are unhappy with language areas but you do not want to quit. *Journal of Language Contact* THEMA 2. 112–123.

Tran, Jennie. 2011. *The acquisition of Vietnamese classifiers*. PhD thesis (pre-defence version), University of Hawai'i.

Vittrant, Alice. 2008. Burmese as a language of the Mainland South-East Asian *Sprachbund*. Paper presented at the 21st International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL).

Yap, Foong Ha. 2007. On native and contact-induced grammaticalization: The case of Malay *empunya*. Manuscript, Department of Linguistics and Modern Languages, Chinese University of Hong Kong.

Winstedt, Richard O. 1920. *Colloquial Malay: A simple grammar with conversations*, 2nd ed. Singapore: Kelly & Walsh.

Winstedt, Richard O. 1927. *Malay grammar*. Oxford: Clarendon Press.

Zaharani Ahmad. 2007. *Pembentukan kata ganda separa bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Zainudin Dirin. 2000. *Kamus penjodoh bilangan untuk KBSR*. Shah Alam: Fajar Bakti. Zainu'l-Abidin bin Ahmad (Za'ba). 1927. Pair-words in Malay. *Journal of the Malaysian Branch of the Royal Asiatic Society* 5. 324–338.

Zeitoun, Elizabeth. 2007. *A Grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.

Zuraidah Mohd Don, Gerry Knowles & Janet Yong. 2008. How words can be misleading: A study of syllable timing and 'stress' in Malay. *The Linguistics Journal* 3. 66–81.

# Appendix 1: Summary of linguistic features

Legend
+++ the feature is pervasive or used obligatorily in the language
++  the feature is normal but selectively distributed in the language
+   the feature is merely possible or observable in the language
–   the feature is impossible or absent in the language

|  | Feature | +++/++/+/– | page and § ref. in your article |
|---|---|---|---|
| Phonetics | Lexical tone or register | – | 1.1 |
| Phonetics | Back unrounded vowels | – | 1.2 |
| Phonetics | Initial velar nasal | + | 1.2 |
| Phonetics | Implosive consonants | – | 1.2 |
| Phonetics | Sesquisyllabic structures | – | 1.3 |
| Morphology | Tendency towards monosyl-labicity | ++ | 2.1; "selectively distributed" in terms of register and syntactic category |
| Morphology | Tendency to form compounds | +++ | 2.1 |
| Morphology | Tendency towards isolating (rather than affixation) | ++ | 2.1; "selectively distributed" in terms of register |
| Morphology | Psycho-collocations | +++ | 2.2 |
| Morphology | Elaborate expressions (e.g. four-syllable or other set patterns) | +++ | 2.3 |
| Morphology | Reduplication generally | +++ | 2.4 |
| Morphology | Reduplication of nouns | +++ | 2.4 |
| Morphology | Reduplication of verbs | +++ | 2.4 |
| Grammar | Use of classifiers | ++ | 3.1.2 |
| Grammar | Classifiers used in counting | ++ | 3.1.2 |
| Grammar | Classifiers used with demonstratives | – | 3.1.2 |
| Grammar | Adjectival verbs | + | not discussed |
| Grammar | Grammatical number | – | 3.2.1 |
| Grammar | Inflection of verbs | – | 3.2.1 |
| Grammar | Use of tense/aspect markers | +++ | 3.2.1 |
| Grammar | Use of verb plural markers | – | not discussed |
| Grammar | Grammaticalization of GET/OBTAIN (potential mod. resultative/perfect aspect) | +++ | 3.2.2 |

| | Feature | +++/++/+/− | page and § ref. in your article |
|---|---|---|---|
| Grammar | Grammaticalization of PUT, SET (completed/resultative aspect) | − | not discussed |
| Grammar | Grammaticalization of GIVE (causative, benefactive; preposition) | +++ | not discussed, though an example is given in (61b), 3.2.3 |
| Grammar | Grammaticalization of FINISH (perfective/ complete aspect; conjunction/temporal subordinator) | − | not discussed |
| Grammar | Grammaticalization of directional verbs e.g. GO / COME (allative, venitive) | − | not discussed |
| Grammar | Grammaticalization of SEE, WATCH (temptative) | − | not discussed |
| Grammar | Grammaticalization of STAY, REMAIN (progressive and continuous, durative aspects) | − | not discussed |
| Grammar | Serial verb constructions | +++ | 3.2.3 |
| Syntax | Verb precedes object (VO) | +++ | 3.3 |
| Syntax | Auxiliary precedes verb | +++ | 3.3 |
| Syntax | Preposition precedes noun | +++ | 3.3 |
| Syntax | Noun precedes adjective | +++ | 3.1.1 |
| Syntax | Noun precedes demonstrative | +++ | 3.1.1 |
| Syntax | Noun precedes genitive | +++ | 3.1.1 |
| Syntax | Noun precedes relative clause | +++ | 3.1.1 |
| Syntax | Use of topic-comment structures | +++ | 3.3.2 |
| Syntax | Ellipsis of arguments known from context | +++ | 3.3.1 |
| Lexical semantics | Specific terms for forms of rice | +++ | 4.1.1 |
| Pragmatics | Use of utterance-final pragmatic particles | +++ | 4.2.1 |
| Pragmatics | Encoding of politeness | +++ | 4.2.2 |
| Pragmatics | Encoding of honorifics | +++ | 4.2.2 |

# Appendix 2: Text interlinearized

*'Mujur     tak     kena     tikam'*     (Harian Metro, 11/08/2011)
fortunate     not     get     stab
'Fortunate not getting stabbed'

*KUALA LUMPUR: "Orang kata saya turut di-tikam, tapi mujur     tak kena.*
Kuala Lumpur     person say 1SG     too     PASS-stab but fortunate not get

*Saya tak nampak pisau, tapi orang     di flat berdekatan memberitahu*
1SG     not see     knife but person at flat nearby     tell

*kelompok     perusuh     itu     bersenjatakan     pisau," kata*
group     rioter     that     use.as.weapon     knife     say

*Muhammed Asyraf Haziq Rossli,     20.*
Muhammed Asyraf Haziq Rossli     20
'KUALA LUMPUR: "People have been saying that I was stabbed too, but for-
tunately I wasn't. I didn't see a knife, but people living in a nearby flat said that
the rioters used knives as their weapon," said Muhammed Asyraf Haziq Rossli,
20.'

*Dia yang cedera di mulut akibat di-serang     dan di-rompak perusuh*
3SG REL     injured at mouth result PASS-attack and PASS-rob     rioter

*ketika     melalui     satu     kawasan     rusuhan     di timur     London,     terpaksa*
when     pass     one     area     riot     at east     London     have.to

*menjalani     pembedahan     membetulkan     rahang-nya     yang patah     dan*
undergo     surgery     fix     jaw-3SG     REL break     and

*kini menerima     rawatan     di Royal London     Hospital.*
now receive     treatment     at Royal London     Hospital.
'He, who suffered an injury to his mouth as a result of having been attacked and
robbed by rioters while passing through a riot area in east London, had to un-
dergo surgery to fix his broken jaw and is currently receiving treatment at Royal
London Hospital.'

| *Dalam* | *temubual* | *di-rakam* | *wakil* | | *Kelab* | *Umno* | *London* |
|---|---|---|---|---|---|---|---|
| in | interview | PASS-record | representative | club | Umno | London |

| *kelmarin,* | | *penuntut* | *perakaunan* | *tajaan* | | *Mara* | *di* | *Kolej* |
|---|---|---|---|---|---|---|---|---|
| day.before.yesterday | student | accounting | sponsorship | Mara | at | college |

| *Kaplan,* | *London* | *itu* | *memberitahu,* | *kejadian* | *berlaku* | *pantas* | *dan* | *dia* |
|---|---|---|---|---|---|---|---|---|
| Kaplan | London | that | tell | incident | happen | quick | and | 3SG |

| *tidak* | *dapat* | *melarikan* | *diri* | *ketika* | *di-kerumuni* | | *kumpulan* | *berkenaan.* |
|---|---|---|---|---|---|---|---|---|
| not | can | run | self | when | PASS-surround | group | relevant |

'In an interview recorded by the representative of Kelab Umno London the day before yesterday, the accounting student sponsored by Mara at Kaplan College, London, said that the incident occurred so quickly that he could not run away when surrounded by the group of rioters.'

| *"Saya* | *sedang* | *mengayuh* | *basikal* | *ketika* | *itu* | *dan* | *memang* | *tak* |
|---|---|---|---|---|---|---|---|---|
| 1SG | PROG | pedal | bicycle | time | that | and | really | not |

| *boleh* | *buat* | *apa-apa* | *kerana* | *mereka* | *ramai,* | *(ketika* | *di-kerumuni)* |
|---|---|---|---|---|---|---|---|
| can | do | anything | because | 3PL | many | when | PASS-surround |

| *ada* | *yang* | *seluk* | *belakang* | *baju* | *sejuk* | *ber-hud* | *yang* | *saya* |
|---|---|---|---|---|---|---|---|---|
| exist | REL | grab | back | clothes | cold | with-hood | REL | 1SG |

| *pakai* | *dan* | *ada* | *ambil* | *telefon.* |
|---|---|---|---|---|
| wear | and | exist | take | telephone |

'"I was riding my bicycle then and really couldn't do anything because there were many of them, (when I was surrounded by them) one grabbed the back of the hooded sweater I was wearing and another took my phone.'

| *"Mereka* | *juga* | *tarik* | *basikal* | *yang* | *saya* | *tunggang* | *menyebabkan* | *saya* |
|---|---|---|---|---|---|---|---|---|
| 3PL | also | pull | bicycle | REL | 1SG | ride | cause | 1SG |

| *jatuh* | *dan* | *mulut* | *berdarah.* | *Masa* | *jatuh* | *itu-lah* | | *saya* | *rasa* | *rahang* |
|---|---|---|---|---|---|---|---|---|---|---|
| fall | and | mouth | bleed | time | fall | that-PART | 1SG | feel | jaw |

| *saya* | *sudah* | *patah* | *dan* | *mereka* | *semua* | *lari* | *apabila* | *nampak* | *saya* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1SG | PFV | break | and | 3PL | all | flee | when | see | 1SG |

*berdarah," kata-nya dalam nada berhati-hati menahan sakit.*
bleed    say-3SG  in    tone  careful      withstand  pain
'They also pulled the bicycle I was riding, which made me fall and my mouth bleed. It was when I fell that I felt my jaw was broken and they all fled seeing me bleeding," he said in a careful tone, withstanding the pain.'

*Muhammed Asyraf Haziq  berkata, selepas di-serang,  dia  di-sapa  pula*
Muhammed Asyraf Haziq  say    after   PASS-attack  3SG  PASS-greet then

*beberapa lelaki lain  yang menawarkan pertolongan, tapi dalam masa*
a.few    man   other REL  offer       help        but  in    time

*sama ada  orang  di  belakang-nya  menyelongkar beg  sandang*
same exist person at  back-3SG      rummage      bag  sling

*di-pakai.*
PASS-wear
'Muhammed Asyraf Haziq said, after he was attacked, he was then greeted by a few other men who offered help, but at the same time there were also people behind him rummaging in his backpack.'

*"Ketika rusuhan berlaku  semua warga  kulit hitam saja yang saya*
when   riot    happen   all   citizen skin  black only REL  1SG

*nampak, cuma ada  dua orang saja kulit putih. Orang  yang*
see     just exist two CLF   only skin  white person REL

*menyerang dan merompak itu  daripada kumpulan sama juga.*
attack    and rob      that from     group    same also
'"When the riot happened, the people that I could see were all black-skinned citizens, there were only two white-skinned ones. The people who attacked and robbed me were also from that group.'

*"Dia macam ada satu kumpulan yang pergi, kemudian datang orang*
3SG like  exist one  group    REL  go    then     come   person

*lain dari belakang ... jadi saya nak  lari pun  tak boleh*
other from back        so   1SG  want flee even not can

dah,"        kata-nya.
already    say-3SG
'"It was like there was a group going and another group of people coming from behind … so even though I wanted to get away, I couldn't do so already," he said.'

Dia      memberitahu,   ada      kanak-kanak   kecil    di-percayai    murid
3SG      tell              exist    child              small    PASS-believe    pupil

sekolah rendah  terbabit  dalam  kumpulan  merusuh  itu,  malah  kanak-kanak
primary.school  involved  in       group       riot       that  even   child

terbabit    yang    mengetuk    dan    menginginkan    basikal-nya.
involved    REL     knock        and    want              bicycle-3SG
'He said that there was a small child, who was believed to be a primary school student, involved in the riot group, and, in fact, it was this child who knocked his bicycle and asked for it.'

Namun,    kata-nya   dia    tidak   pasti    sama ada   kanak-kanak   terbabit
however   say-3SG    3SG    not     certain  whether    child             involved

mendapat   apa    yang   di-inginkan   kerana    dia    terus         beredar
get            what   REL    PASS-want      because   3SG    immediately   leave

menyeberangi  jalan   berdekatan  sebelum  di-bantu   dua  wanita   ber-kulit
cross             street  nearby         before   PASS-help  two  woman    with-skin

hitam    yang    memberi-nya    tisu    (bagi    mengelap    darah).
black    REL     give              tissue  for      wipe          blood
'However, he said, he was not certain whether the child got what he wanted because he immediately moved away to cross the nearby street, when he was given help by two black-skinned women, who gave him tissue paper (to wipe the blood).'

"Budak    kecil    itu    memang    ada.   Dia    yang    ketuk-ketuk    basikal
kid          small    that   really      exist  3SG    REL     knock            bicycle

saya  tanda  mahukan.  Tapi,  selepas  apa   yang  berlaku,  saya  tak  tahu
1SG    sign    want         but     after     what  REL    happen    1SG    not   know

| *dia* | *dapat* | *atau* | *tidak* | *(basikal),* | *rasanya* | *orang* | *lain* | *yang* | *ambil* |
|-------|---------|--------|---------|--------------|-----------|---------|--------|--------|---------|
| 3SG | get | or | not | bicycle | seemingly | person | other | REL | take |

| *sebab* | *dalam* | *video* | *pun* | *nampak* | *orang* | *lain* | *yang* | *kayuh,"* | *kata-nya.* |
|---------|---------|---------|-------|----------|---------|--------|--------|----------|------------|
| because | in | video | too | see | person | other | REL | pedal | say-3SG |

'"The small kid was really there. It is s/he who knocked my bicycle to show that s/he wants it. But, after the incident that happened to me, I don't know if s/he got (the bicycle) or not. It seems another person had taken it because in the video too another person was seen riding it," he said.'