

Numeral Classifier *Buah* in Malay: A Corpus-based Study^{*}

Siaw-Fong Chung
National Chengchi University

This paper examines the classifier *buah* in Standard Malay in Malaysia. Many studies (e.g. Adams & Conklin 1973, Allan 1977, Saalbach & Imai 2005) have agreed that classifiers are closely related to categorization. In Malay, it has been suggested that the classifier *buah* is able to classify things that do not have definite types and shapes. Its classification of varied and seemingly dissimilar nouns raises the question of whether *buah* is a general classifier. By analyzing nouns classified by *buah* in Standard Malay from a collection of 5,009 news articles in Malaysia, consisting of over a million words, this paper argues that *buah* not only has a semantic function but also serves a cultural role. This paper also investigates the connection between the different senses of *buah* and shows that its senses are often metaphorically linked to 'products' (e.g. products of artifacts, products of telecommunication, products of nature, products of thoughts, etc.). Through a corpus-based analysis, we discern the meanings of *buah* and further postulate two possible mechanisms for its selection, a result that eliminates the status of *buah* as a general classifier.

Key words: numeral classifier, sense, Malay, *buah*, corpus

1. Introduction

Classifiers in general can be used to depict how objects are conceptualized by humans. Tai & Wang (1990:38), in their investigation of Mandarin classifiers, state clearly that there are similarities in the nouns classified by the same classifier:

A classifier categorizes a class of nouns by picking out some salient perceptual properties, either physically or functionally based, which are permanently associated with entities named by the class of nouns; a measure word does not categorize but denotes the quantity of the entity named by a noun.

^{*} The author would like to thank I-Wen Su and Louis Weilun Lu, National Taiwan University, for commenting on the previous version of this work. Comments from the reviewers are also highly appreciated. The research herein is supported partly by NSC (project number NSC 97-2410-H-004-001).

In a different study, Allan (1977:285) also states that classifiers “have meaning, in the sense that a classifier denotes some salient perceived or imputed characteristic of the entity to which an associated noun refers.” These observations regarding the feature properties of entities under the same classifier can also be seen in some cross-linguistic work, such as Saalbach & Imai’s (2005) study, which examines the similarities of concept classification for Chinese, Japanese, and German, and in Schmitt & Zhang’s (1998) investigation of product classification for objects such as hair spray, soft drinks, lipstick, pianos, and fax machines in Mandarin, Cantonese, and Japanese, based on the features of size and shape. Through comparisons of the different languages, Schmitt & Zhang (1998:120) emphasize how linguistic forms can display “representations of grammatical structures” that “operate in the mind.”

In addition to cross-linguistic comparisons, there have been studies that try to create a typology of classifiers. Aikhenvald (2003) is one such study that provides a distinction of all types of classifiers. As a type of ‘classifier,’ noun classifiers can be seen in the use of 人 in Chinese in which noun phrases such as 投資人 invest-person ‘investor’ and 合夥人 partnership-person ‘partner’ are formulated. Unlike noun classifiers, numeral classifiers usually quantify nouns by grouping those with similar features under the same classifier. In addition to these two types of classifiers, there are others, such as verbal classifiers, relational classifiers, and deictic classifiers, but these are not within the scope of the current research.

For numeral classifiers specifically, their rich grammatical systems have been found in many languages, such as Japanese, Chinese, Vietnamese, the Mon-Khmer, and the Austroasiatic languages. Adams & Conklin (1973) and Jones (1970) represent two earlier studies that have compared the features of numeral classifiers in several languages. Adams & Conklin’s (1973:1) study investigates 37 Asian languages, “including representatives of the Malayo-Polynesian, Austro-Asiatic, Mon-Khmer, Sino-Tibetan, Altaic, Dravidian, and Indo-Aryan families,” while Jones’ (1970) study examines the typological differences between languages in Southeast Asia. In his investigation, Jones was able to draw some similarities between the Malay numeral classifier system and those of Chinese, Cebuano, Vietnamese, and two Mon-Khmer languages (Brôu and Sedang).

In addition to the above-mentioned typological research, studies on numeral classifiers have also focused on the comparisons of semantic functions. In (1) below, some of the semantic functions of numeral classifiers taken from Aikhenvald (2003:306) are given.

- (1) a. Animacy
- b. Social status and kinship relation
- c. Directionality and orientation

- d. Physical properties
- e. Nature
- f. Quanta
- g. Arrangement
- h. Functional properties

Among the dimensions in (1), Adams (1986:248) suggests that most Austroasiatic languages reckon the animate function (1a) “to be primary.” Many languages, such as Malay and Mandarin, also separate animate nouns from inanimate nouns. For instance, animals are classified using *ekor* in Malay and 隻 in Mandarin. In contrast, inanimate nouns usually appear within the parameters of (2) below.

- (2) a. Round
- b. Long and rigid
- c. Long and slender or flexible
- d. Flat

The parameters in (2) reflect the prototypical features of objects, such as round for ‘ball,’ long and rigid for ‘stick,’ long and slender or flexible for ‘hair,’ and flat for ‘land.’ These parameters, when manipulated, as was found in Tien et al. (2002) using psycholinguistic rating and judgment tasks, can affect the selection of numeral classifiers for less prototypical objects, such as a shorter-than-usual bone. Compared to the prototypical long bone, the less prototypical one might result in less agreement among the subjects when naming a suitable numeral classifier for it. This probably explains the name for numeral classifiers, *penjodoh bilangan*, in Malay, which means ‘match-maker for number.’¹ As ‘match-maker,’ numeral classifiers share certain typical semantic functions with their corresponding nouns. Examples of numeral classifiers in Malay can be seen in (3) below, with the construction of [number-classifier-noun]. In Malay, only the number ‘one’ *se-* appears as a clitic to the classifier, as in (3b).

- (3) a. *dua **keping** kuih*
 two Class. cake
 ‘two slices of cake’
- b. *se**keping** duit kertas*
 one.Class. money paper
 ‘a bank note’

¹ Hereafter, unless specified, all mentions of Malay refer to Malaysian Malay.

In example (3) above, *keping* is a numeral classifier for thin, flat, and wide objects (thus, two slices of cake and a bank note). Other numeral classifiers for inanimate objects in Malay include *biji* for eggs and stones (2a); *batang* for sticks (2b); *helai* for hair (2c); and *bidang* for land (2d). In addition to these commonly seen semantic features (i.e., shapes and sizes), Salehuddin & Winskel (2008:65) suggest that “the subcategorisation of the Malay classifiers has [further] characteristics that are specific to Malay.” For instance, *kuntum* is used only with flowers while *laras* is used only with firearms (p.71).

In contrast to the specific classifiers, *buah* seems to appear at the other continuum of generality as a numeral classifier that can categorize ‘small round objects, in general’ in Minangkabau (Marnita 1996:104) and one that can group ‘things with indefinite shapes and types’ in Malay (the latter is a definition translated from a prestigious Malay dictionary, *Kamus Dewan* (2002), cf. Table 2). Salehuddin and Winskel further state that *buah* is used to classify three-dimensional objects that are inanimate count nouns with large shapes and sizes (e.g. books, computers, and cupboards). In an attempt to find corresponding numeral classifiers for *buah*, the following excerpt by Adams (1986:249) relates the word *trái*, a word in Vietnamese that possesses a similar original meaning as *buah* (i.e., ‘fruit’), to the counterparts of *trái* in different languages:

In those Austroasiatic languages which do use a classifier meaning ‘fruit’ to count round objects, these classes normally include fruits, vegetables, eggs, along with many other naturally occurring items as well as various artifacts. For instance in Vietnamese, the word *trái* can be used to count fruits and vegetables; body organs such as hearts and livers; pieces of earth and hills; and bombs, balls and globes. Other languages with such classes include large bulky items like mountains and houses. However, sometimes, numeral classifiers of this type are used only to count such things as fruits, vegetables and nuts. This is true in Khmu and in Vietnamese where there is a second classifier meaning ‘fruit’ which is borrowed from Chinese. [underlining added]

From the excerpt above, it seems clear that even though two different languages share a classifier with similar function, differences can still be found. Therefore, it is not surprising that in Malay, *buah* is used with bulky objects of both solid (e.g. house, car, big rock, etc.) and abstract (e.g. idea, country, etc.) nature yet in Minangkabau, it is used with small round objects.

Another issue related to *buah* is its status as a general classifier. In Hopper’s (1986) study based on a classic 1923 manuscript entitled *Autobiography of Abdullah bin Abdul Qadir “Munshi”*, *buah* was found to be highly frequent, co-appearing with a variety of dissimilar nouns, including ship, hole, and abstract nouns such as idea. Hopper (1986:323;

also in 1991) says the following regarding *buah*:

The most general classifier, BUAH, has almost lost any semantic, conceptual content and is clearly going the way of WATU ‘stone’, which now exists only as an appendage to *sa-* ‘one, a’ in the frozen form S(U)ATU ‘one, a’.

In this paper, we intend to respond to this observation concerning *buah*—that is, whether it is a general classifier. We shall also examine the uses of *buah* in modern Malay written texts, especially in newspapers. Using a self-collected corpus, the following research question is asked:

- (4) How can a corpus-based approach help identify the semantic functions of *buah*?

This question will be answered using news articles from the Malaysian online newspaper *Utusan Malaysia* (<http://www.utusan.com.my/>) collected between 2005 and 2010. A majority (93%) of the data come from 2010, while the remaining 7% are constituted by data spreading between 2005 and 2007. This difference in the years of collection is partly due to the restriction of the online newspaper archive, which allows only searches of recent articles. The large amount of data from 2010 was collected automatically for a continuous period of time in which 200 articles were recorded each day.²

Regarding the research question in (4), the hypothesis is that linguistic predictions based on a particular genre can be made via a corpus. In order to shed light on the issue of whether corpora data can help clarify the (non)general-classifier’s status of *buah* in Malay, we inspected the noun collocates following *buah* in the corpus. We also examined the origins of the nouns that *buah* classifies. By doing so, we were able to postulate that the origins of these nouns may contribute to the determination of their corresponding classifiers, in addition to the influence of semantic features (such as bulky in size, big in shape, etc.).

Based on our corpora data, we can predict the possible mechanisms that may govern the selection of *buah* as a numeral classifier to a noun by hypothesizing what would happen if a newly invented, classifier-needing noun entered the Malay vocabulary. The following section will discuss further the different uses of *buah*, including its relationship with the ‘fruit’ reading.

² The difference between periods, however, did not hinder the overall results because similar patterns were found for the smaller sample of data from 2005 to 2007 and the larger, but more representative, set of data in 2010.

2. *Buah* as a classifier and its semantic functions

Dewan Bahasa lists nine senses for *buah* (2002:175). Their translations (by the author) and parts-of-speech are provided in Table 1 below.

Table 1: Senses of *buah* in *Dewan Bahasa* (2002:175) and their translations

Sense	Parts-of-Speech	Definitions	Translations	Example Sentences/Phrases
1	Noun	<i>bahagian tumbuhan yang terjadi daripada bunga atau putik (biasanya berbiji)</i>	Part of a plant that grows from a flower or pistil (usually those that produce seeds)	<i>Pokok ini sudah ada buahnya</i> 'This tree has borne fruit.'
2	Noun	<i>biji yang agak besar dan keras</i>	Fruits that are quite big and hard	<i>Buah kemiri</i> 'fruit of <i>kemiri</i> tree'
3	Noun	<i>benda yang menyerupai buah tertentu</i>	Things that are fruit-like	<i>Buah bidara</i> 'bead that looks like a type of fruit named <i>bidara</i> '
4	Noun	<i>bahagian (badan manusia dan lain-lain) yang menyerupai buah</i>	Parts (body parts and others) that are fruit-like	<i>Buah lengan</i> BUAH upper.arm 'biceps'
5	Noun	<i>bermacam-macam benda bulat</i>	All kinds of round objects	<i>Buah guli</i> 'marble'
6	Nominal classifier	<i>penjodoh bilangan untuk benda (bentuk atau jenisnya tidak tertentu)</i>	Nominal classifiers for things with indefinite shapes and types	<i>Sebuah rumah</i> 'one.BUAH house'
7	Noun	<i>(Kiasan) hasil sesuatu perbuatan (usaha dan lain-lain)</i>	(Figurative meaning) product of an action (effort, etc.)	<i>Segala usahanya tidak memberi buah</i> 'All his/her effort did not bring any result.'
8	Noun	<i>(Kiasan) bahan, pokok, pangkal</i>	(Figurative meaning) issue, important thing, or base of discussion	<i>Buah cakap</i> BUAH speak 'product of speech'
9	Noun	<i>pergerakan atau teknik dalam seni silat until mematahkan serangan lawan</i>	Movement or technique in self-defense arts to stop the attack of an opponent	<i>Buah sendeng sudah dipelajarinya</i> 'The technique of slanting has been learnt by him/her.'

These nine senses clearly show that the term *buah* in Malay displays the following meanings: literal fruit (sense 1) or a type of fruit (sense 2), fruit-like objects (senses 3 and 4), round objects (sense 5), a nominal classifier (sense 6), products (senses 7 and 8), and a type of movement in martial arts (sense 9). Among these senses, sense 9 seems to be a

different lemma, as its meaning cannot be directly related to the others. Senses 2 to 5 are parts of the noun classifiers and an example is given in (5) below for sense 2. (All translations are provided by the author. The date of the news articles from which the instance is taken is given at the end of the translation.)

- (5) *kemasukan **buah mangga** dari negara jiran...*
 entrance BUAH mango from country neighbor
 ‘the import of mangos from a neighboring country...’ (2005.5.3)

The meaning in (5), a type of fruit, is distinguished from the literal *buah* of sense 1, which means ‘the product of plant growth’ (the basic meaning of ‘fruit’ in English from Merriam-Webster online).³ In addition, many uses of *buah* are metaphorically motivated (‘fruits of the body’) and they appear as compound nouns (e.g. *buah lengan* ‘biceps,’ and *buah pinggang* ‘kidneys’).⁴ An example taken from the corpora is given in (6) below.

- (6) *subsidi rawatan dialisis kepada pesakit **buah pinggang**...*
 subsidy treatment diagnosis to patient BUAH waist
 ‘subsidy for diagnosis treatment for patients with kidney problem...’ (2005.2.15)

Buah in this example constitutes one complete meaning to the words underlined (i.e., kidney). These uses of *buah* (i.e., fruits and body part terms) correspond to Adams’ (1986) previous excerpt regarding the uses of *trái* for ‘body organs,’ such as ‘hearts’ and ‘livers.’ Nonetheless, these compound nouns are not the target expressions we shall investigate in the current work and they will be removed from the corpora data. In contrast, the target of this research is that from sense 6 (i.e., the numeral classifier meaning), exemplified in (7) below.

- (7) a. *dua **buah** meja disediakan untuk kaunter pendaftaran.*
 two BUAH table be.prepared for counter registration
 ‘Two tables are prepared as registration counters.’ (2005.2.14)
 b. *syarikat itu sudah menerbitkan tujuh **buah** drama.*
 company that already publish seven BUAH drama
 ‘That company has published seven dramas (on CDs).’ (2010.2.21)

³ <http://www.merriam-webster.com/>

⁴ One of the reviewers also pointed out that unless the body parts come in pairs (e.g. *sepasang mata* ‘a pair of eyes’ and *sepasang tangan* ‘a pair of hands’), body parts usually do not need a numeral classifier (thus, **sebiju kepala* ‘one.BIJU head’ is incorrect) (cf. Asmah binte Haji Omar 1972).

- c. *sebuah* *idea yang lahir daripada kepala Dr. Mahathir...*
 one.BUAH idea Rel. born from head Dr. Mahathir
 ‘an idea that comes from (the head of) Dr. Mahathir...’ (2005.11.28)

From the examples in the corpus, we can see that *buah* classifies concrete objects (7a), published objects (7b), and abstract ideas (7c). Other nouns classified by *buah* include boat, house, table, chair, artwork, country, state, etc. Considering these various types of nouns, the questions, therefore, are whether the definition ‘to group things with indefinite shapes and types’ is good enough to represent the uses of *buah* as a numeral classifier, and whether we should accept *buah* as a general classifier—a numeral classifier that can take objects of all shapes and types.

In this paper, we propose that there are cognitive mechanisms governing the selection of *buah* as a numeral classifier instead of the collapsed definition as given in Table 1 previously. In order to examine the uses of *buah* in modern Malay texts, we searched for instances in Malay newspapers, as illustrated in the section below. This study also tries to provide an explanation for the relatedness of the different uses of *buah* as a numeral classifier via quantitative data.

3. Methodology

In order to examine the different nouns classified by different Malay classifiers, we first referred to online teaching resources for grammar lessons, the results of which can be seen in Table 2 below.⁵ The semantic functions of the numeral classifiers, translated based on the definitions from *Kamus Dewan*, are shown in column two, and the list of nouns assembled from the online resources is shown in column three.

⁵ See <http://tatabahasabm.tripod.com/tata/pjodoh.htm> and <http://dst.schoolnet.my/default.php?page=Senarai+Penjodoh+Bilangan> for a list of nouns used with these different classifiers (accessed April 9, 2010).

Table 2: Examples of Malay numeral classifiers and their semantic functions

Malay Numeral Classifiers	Semantic Functions of the Numeral Classifiers	Malay Nouns (Taken from online teaching resources)	
butir	<i>Untuk barang-barang kecil yang bulat</i> For things that are small and round	<i>telur</i> ‘egg’ <i>manik</i> ‘bead’	<i>beras</i> ‘rice’ <i>peluru</i> ‘bullet’
biji	<i>Bagi benda kecil dan lain-lain</i> For things that are small	<i>labu</i> ‘gourd’ <i>lobak</i> ‘carrot’ <i>kubis</i> ‘cabbage’ <i>terung</i> ‘brinjal’ <i>bola</i> ‘ball’ <i>cawan</i> ‘cup’	<i>guli</i> ‘marble’ <i>belon</i> ‘balloon’ <i>ketupat</i> ‘dumpling’ <i>telur</i> ‘egg’ <i>gelas</i> ‘glass’
ketul	<i>gumpal, kepal (nasi, tanah, dan lain-lain), kerat (tulang, kayu, dan lain-lain)</i> Clod, lump of (rice, soil, etc.) Incised (bone, wood, etc.)	<i>ais</i> ‘ice’ <i>daging</i> ‘meat’ <i>tulang</i> ‘bone’	<i>batu</i> ‘stone’ <i>kepala ikan</i> ‘head of a fish’
buah	<i>Untuk benda (bentuk dan jenisnya tidak tentu) (sense 6 in Table 1 previously)</i> For things with indefinite shapes and types	<i>rumah</i> ‘house’ <i>kereta</i> ‘car’ <i>komputer</i> ‘computer’ <i>meja</i> ‘table’ <i>kerusi</i> ‘chair’ <i>beg</i> ‘bag’ <i>katil</i> ‘bed’ <i>tilam</i> ‘mattress’ <i>buku</i> ‘book’ <i>bandar</i> ‘town’	<i>kotak</i> ‘box’ <i>radio</i> ‘radio’ <i>piano</i> ‘piano’ <i>kamera</i> ‘camera’ <i>negara</i> ‘country’ <i>negeri</i> ‘state’ <i>lagu</i> ‘song’ <i>karya</i> ‘artwork’ <i>catatan</i> ‘records’
keping	<i>Bagi sesuatu yang pipih dan nipis (seperti kad, kertas, papan, dan sebagainya)</i> For things that are flat and thin (a card, paper, board, etc.)	<i>CD</i> ‘CD’ <i>disket</i> ‘diskette’ <i>DVD</i> ‘DVD’ <i>roti</i> ‘bread’ <i>stem</i> ‘stamp’ <i>wang kertas</i> ‘bank note’	<i>syiling</i> ‘shilling’ <i>gambar</i> ‘photo’ <i>kek</i> ‘cake slice’ <i>kad</i> ‘card’ <i>tiket</i> ‘ticket’
helai	<i>Bagi benda yang tipis (seperti daun, rambut, kertas, kain, dan lain-lain)</i> For things that are thin (leaf, hair, paper, cloth, etc.)	<i>kertas</i> ‘paper’ <i>daun</i> ‘leaf’ <i>rambut</i> ‘hair’	<i>kain</i> ‘cloth’ <i>baju</i> ‘clothes’
bidang	<i>Bagi benda yang luas (seperti tanah dan sebagainya)</i> For things that are wide (land, etc.)	<i>tanah</i> ‘land’ <i>sawah</i> ‘farmland’	<i>kebun</i> ‘plantation’ <i>kain</i> ‘cloth’

Table 2 above shows some of the specific semantic properties classified by different classifiers in Malay. In fact, Salehuddin & Winskel (2008:66) claim that “[t]he system of Malay numeral classifiers is very complex and has to be formally taught in schools,” especially those that are used with one specific type of noun but not any other (e.g. *kuntum*). Most of the numeral classifiers in Table 2 above are defined using shapes and sizes (*butir*, *biji*, *keping*, *helai*, and *bidang*). Only *ketul* is defined using examples by the lexicographers. The nouns for *buah* (shaded), by contrast, seem to vary from big bulky objects (e.g. house, car, and mattress) to more abstract objects (e.g. song, artwork, and records). Based on these nouns, it is unclear what the relationship might be between ‘bag’ and ‘country.’ Should these dissimilar nouns lead to the conclusion regarding the general classifier status of *buah* is a question this paper aims to address.

In order to answer the research question in (4), we examined a list of noun collocates classified by *buah* based on 5,009 newspaper articles randomly collected via the search archive system of *Utusan Malaysia*. For all news articles, section headings, titles, and dates were kept. The results of this examination are shown below in Table 3.

Table 3: Types and tokens of the Malay corpus based on randomly collected news articles

Word Tokens (Total Word Count)	Word Types	Sentences	Mean Sentence Length (in Words)
1,671,570	45,751	84,859	20.02

Table 3 shows that the corpus collected comprises more than one million tokens (or number of words), with 45,751 word types, indicating an average ratio of 37 tokens per type. In addition, there are 84,859 sentences and each sentence is about 20 words in length. Based on this corpus, we searched for uses of *buah* in the data. Two concordancers, AntConc 3.2.1 (Anthony 2005) and WordSmith version 5 (Scott 2008), were then used to extract all instances that end with *buah* (**buah* with an asterisk serving as a wild card or just *buah*).

4. Results

The search returned 2,602 instances of [*buah*+NOUN], constituting 0.17% of the total word count. Some sentences with omitted nouns were also kept. An example of [*buah*+NOUN] is given in (8a) while (8b) is a case for an omitted noun.

- (8) a. *ia mendarat di sebuah padang di*
 it land at one.BUAH field at
Kalimantan Timur hari ini.
 Kalimantan East today this
 ‘It [flying jet] lands at a field in East Kalimantan today.’ (2010.2.11)
- b. *...dan menjangka untuk menjual 800,000 buah, tahun ini.*
 and estimate to sell 800,000 BUAH, year this
 ‘...and estimate to sell 800,000 [cars] this year.’ (2010.2.5)

Both examples in (8) were included in this study. Nonetheless, the 2,602 instances returned exclude one instance of *berbuah* ‘with fruit/outcome,’ 52 instances of *buah-buahan* ‘fruits,’ and 26 instances of *anak buah* ‘nephew/niece,’ as they are not the focus of this study. From these 2,602 instances, 107 (4%) were compound nouns (e.g. *buah limau* ‘lime,’ *buah rasberi* ‘raspberry,’ and metaphors such as *buah fikiran* ‘thoughts’) that are not related to the use of *buah* as a numeral classifier. After removing the compound nouns, we further analyzed the remaining 2,495 instances for the semantic meanings of *buah* and the types of nouns *buah* co-occurs with, which resulted in 395 types of nouns. The distribution of senses is shown in Table 4 below. The first column lists the senses; the second and third columns provide the instances and their percentages, respectively; and the last column provides the number of different types of nouns found within each sense.

Table 4: Distributions of senses for the nouns classified by *buah*

Senses	Instances	Percentages	Number of Noun Types
Organization (Country, company, team, state, government, organization, etc.)	546	21.88	20
Artifact (Subtypes: transportation, telecommunication, furniture, small tools and objects, etc.)	494	19.80	97
Place (Home, center, restaurant, room, university, city, etc.)	462	18.52	37
Building (Mosque, factory, condominium, complex, bungalow, warehouse, etc.)	439	17.60	25
Product of Arts/Writing (Subtypes: writing in print, arts, performance, etc.)	255	10.22	17
Area (Land, field, area, etc.)	106	4.25	59

System (Program, model, system, tradition, etc.)	58	2.32	40
Event (Unification, development, placement, application, etc.)	57	2.29	56
Product of Nature (Cave, pool, stone, mountain, hill, etc.)	51	2.04	24
Product of Thoughts/Intellect (Philosophy, hope, decision, etc.)	21	0.84	4
General Things (Thing, entity, asset, etc.)	6	0.24	16
Total	2,495	100.00	395

Table 4 shows that the most frequent sense for nouns classified by *buah* is ‘organization,’ as in *sebuah negara* ‘one.BUAH country,’ followed by three senses (‘artifact,’ ‘place,’ and ‘building’) with about 18% to 20% each. The author differentiated ‘place’ from ‘building’ (and from ‘area’) because ‘place’ refers to a non-definite building, as shown in (9a) below.

- (9) a. *Fara Nadia, 21, yang kini menuntut di sebuah pusat pengajian*
Fara Nadia 21 Rel. now study at one.BUAH center research
tinggi di Amerika Syarikat (AS)...
high at United States...
‘Fara Nadia, 21 years old, who is now studying in a research center (of higher education) at the United States...’ (2010.2.7)
- b. *...dalam satu perhimpunan haram anjuran Hindraf di depan.*
in one gathering illegal hosting Hindraf at front
sebuah kuil, tiga tahun lalu.
one.BUAH temple three year ago
‘...in an illegal gathering hosted by Hindraf in front of a temple three years ago.’ (2010.2.6)

In (9a), ‘a research center of higher education’ refers to a ‘place’ rather than a ‘building.’ In contrast, (9b) refers to a physical ‘building,’ a temple, rather than a ‘place.’ In addition to interpreting via context, the author used a prototypical criterion (cf. Ungerer & Schmid 2006) to decide whether a visual of a building could emerge when analyzing a building versus a place. In the example above, (9b) provides a more vivid visual of a building than (9a) does.

‘Building,’ which possesses a category of its own in Table 4, was also differentiated from ‘artifacts,’ which consist mainly of transportation (*kereta* ‘car,’ *lori* ‘lorry,’ etc.),

telecommunication (*telefon* ‘telephone,’ *satelit* ‘satellite,’ etc.), furniture (*meja* ‘table,’ *almari* ‘cupboard,’ etc.), tools (*alat* ‘tool,’ *jentera/mesin* ‘machine,’ etc.) and other small objects (*beg* ‘bag,’ *album* ‘album,’ etc.). ‘Area,’ with only 4.25% of the total nouns, is a category for objects that are neither ‘building’ nor ‘place.’ This group comprises nouns such as *kawasan* ‘area,’ *unit* ‘unit,’ *ruang* ‘space,’ and *kampus* ‘campus,’ which reflect a meaning of spatial area.

In addition, there are three majors types of products found—‘product of arts/writing’ (e.g. *filem* ‘film’ and *akhbar* ‘newspaper’), ‘product of nature’ (e.g. *gua* ‘cave’ and *batu* ‘stone’), and ‘product of thoughts/intellect’ (e.g. *falsafah* ‘philosophy,’ *forum* ‘forum,’ and *pilihan* ‘selection’). ‘Event’ (2.28%) is constituted mainly by deverbal nouns, such as *permohonan* ‘application’ and *penyatuan* ‘unification,’ which refer to a process dissimilar to other categories listed in Table 4. An example is given in (10) below.

- (10) ...*sebuah* letupan besar di *sebuah* loji tenaga di Connecticut di
 one.BUAH explosion big at one.BUAH site electric at Connecticut at
Amerika Syarikat (AS) hari ini menyebabkan beberapa orang cedera.
 United States today this cause several people hurt
 ‘...today a huge explosion at an electrical site/factory in Connecticut in the
 United States has caused several injuries.’ (2010.2.7)

The first *sebuah* in (10) denotes an ‘event’ (explosion) whereas the second *sebuah* denotes a ‘place’ (an electrical site/factory).

Finally, there are 0.24% of ‘general things’ that cannot be categorized into other groups because they are collective nouns, such as *benda* ‘thing,’ *entiti* ‘entity,’ or *aset* ‘asset.’ (Asset includes both natural products and artifacts, which causes it to be grouped under ‘general things.’)

Based on these senses, this study found that the reduction of all types of senses in Table 4 to ‘classify things with indefinite shapes and types’ is unsatisfactory (but the author was aware of the economic principle of lexicography). Furthermore, from this study, as shown in the last column of Table 4, we discovered that the types of nouns falling into each category did not similarly pattern with the number of instances in the second column. For instance, we found that only 20 nouns consistently re-appear in the 21.88% of instances for ‘organization.’ This means that a limited set of nouns is used frequently with *buah* when referring to ‘organizations.’ Therefore, they are quite predictable.

As for ‘artifact,’ 97 types of nouns were found, which is four times higher in number than that of ‘organization.’ This means that the types of artifacts found with *buah* are usually unpredictable—there can be as many types of artifacts as possible (demonstrated by the number of subtypes that were found). As for ‘area’ and ‘event,’ with a combined percentage of less than 7 in the entire data, they feature over 50 types of nouns. This,

again, indicates that nouns falling into these two senses are often non-recurring nouns. For instance, there could always be events created through deverbalization and this makes the nouns classified by *sebuah* less predictable in ‘event’ than in ‘organization.’

In Table 4, one can see the overall picture of the numeral classifier *buah*, including how it patterns in the corpora. In order to know further which types of nouns appear most frequently in the entire dataset, the noun collocates with frequency above 20 instances are provided in Table 5 below.

Table 5: Nouns (above twenty instances) classified by *buah*

Nouns Classified by <i>Buah</i>	Gloss	Frequency
<i>negara</i>	country	150
<i>rumah</i>	home	128
<i>syarikat</i>	company	123
<i>sekolah</i>	school	101
<i>kereta</i>	car	89
<i>filem</i>	film	68
<i>pusat</i>	center	54
<i>kawasan</i>	area	37
<i>bank</i>	bank	37
<i>hotel</i>	hotel	36
<i>restoran/en</i>	restaurant	36
<i>motosikal</i>	motorcycle	31
<i>stesen</i>	station	31
<i>lori</i>	lorry	30
<i>parti</i>	party	30
<i>lagu (-lagu)</i>	song(s)	30
<i>akhbar</i>	news	30
<i>bilik</i>	room	29
<i>kenderaan</i>	car	28
<i>bas</i>	bus	27
<i>pasukan</i>	team	27
<i>negeri</i>	state	27
<i>masjid</i>	mosque	25
<i>kilang</i>	factory	25
<i>kedai</i>	shop	25
<i>badan</i>	body	25
<i>kapal</i>	ship	24
<i>hospital</i>	hospital	21
<i>kerajaan</i>	government	20
<i>universiti</i>	university	20
<i>bandar</i>	town	20

In Table 5, it appears that only 31 nouns (7.8%) have a frequency of more than 20 in the total 395 types of nouns found with *buah*. These 31 nouns constitute 1,384 tokens, occupying half of the total instances (2,495). This means that the rest of the nouns below the frequency of 20 (i.e., the remaining 364) are low-frequency items. When we analyzed the type of senses found in these 31 nouns, ‘building’ appeared at the top of the list (25.81%), followed by ‘artifact’ and ‘place’ (respectively, 19.35%), and ‘organization’ (16.13%). The remaining senses were found less often (‘product’ (9.68%) and ‘area,’ ‘event,’ and ‘product of nature’ (3.23% each)).

The high frequency of ‘building’ reveals information in addition to the previous Table 4—even though ‘building’ is not the most frequently appearing sense for nouns of *buah*, nouns denoting ‘building’ are still ranked first in frequency. From Table 5, the top four nouns (*negara* ‘country,’ *rumah* ‘home,’ *syarikat* ‘company,’ and *sekolah* ‘school’) are also distinctively different in frequency than the remaining ones. This may be due to the genre we selected—newspaper articles—because these terms appear frequently in newspapers.

While going through the list of nouns in Table 5, we found many loanwords from other languages (e.g. *filem* from *film* in English). We therefore carried out another analysis whereby we tried to find the origins of these nouns. The results show that from 395 types of nouns, 195 (51.3%; 1,574 tokens) can be traced back to its language of origin (15 proper nouns, such as NGO (organization name), were removed).⁶ The results are presented in Table 6 below.

⁶ Part of the results in Table 6 is based on searches in dictionaries, encyclopedias, and resources online, among which are the following websites (accessed March 22, 2010):

http://www.101languages.net/malay/borrowed_words.html

http://en.wikipedia.org/wiki/List_of_Malay_loanwords

http://www.experiencefestival.com/a/Malay_language_-_Some_simple_phrases_in_Malay/id/1741729

More about loanwords can also be found in Asmah binte Haji Omar (1966), Wignesan (1995), and Kuiper (1962), while Maxwell (2008) has a comprehensive book on the Malay language, sketching the influence of different languages in animal, vegetable, kingdoms, etc.

Table 6: Origins of nouns classified by *buah*

Language of Origin	No. of Noun Types	Tokens
English (only the top ten are shown)	152	854
<i>filem</i>	film	(68)
<i>bank</i>	bank	(37)
<i>hotel</i>	hotel	(36)
<i>restoran/en</i>	restaurant	(36)
<i>motosikal</i>	motorcycle	(31)
<i>stesen</i>	station	(31)
<i>lori</i>	lorry	(30)
<i>parti</i>	party	(30)
<i>bas</i>	bus	(27)
<i>hospital</i>	hospital	(21)
Sanskrit	12	244
<i>negara</i>	country	(150)
<i>kenderaan</i>	car	(28)
<i>negeri</i>	state	(27)
<i>keluarga</i>	family	(16)
<i>karya</i>	artwork	(8)
<i>asrama</i>	hostel	(2)
<i>anugerah(an)</i>	award	(2)
<i>kementerian</i>	ministry	(4)
<i>cerita</i>	story	(4)
<i>peta</i>	map	(1)
<i>acara</i>	event	(1)
<i>angkatan</i>	troop	(1)
Portuguese	7	205
<i>sekolah</i>	school	(101)
<i>kereta</i>	car	(89)
<i>gereja</i>	church	(5)
<i>tangki</i>	tank	(3)
<i>meja</i>	table	(3)
<i>almari</i>	cupboard	(3)
<i>bangku</i>	stool	(1)
Arabic	14	190
<i>syarikat</i>	company	(123)
<i>akhbar</i>	news	(30)
<i>masjid</i>	mosque	(25)
<i>khemah</i>	camp	(2)
<i>kubah</i>	dome	(1)
<i>makmal</i>	laboratory	(1)
<i>musyawarah</i>	negotiation	(1)

<i>bahtera</i>	big ship	(1)
<i>wilayah</i>	province	(1)
<i>hadis</i>	Arabic Bible	(1)
<i>kiamat</i>	doomsday	(1)
<i>syurga</i>	heaven	(1)
<i>dunia</i>	world	(1)
<i>falsafah</i>	philosophy	(1)
Tamil	3	50
<i>kedai</i>	shop	(25)
<i>kapal</i>	ship	(24)
<i>katil</i>	bed	(1)
Persian	2	23
<i>bandar</i>	town	(20)
<i>dewan</i>	hall	(3)
Chinese	3	4
<i>longkang</i>	drain	(2)
<i>teko</i>	teapot	(1)
<i>cawan</i>	cup	(1)
French	1	3
<i>kafe</i>	café	(3)
Iban	1	1
<i>benua</i>	continent	(1)
Total	195	1,574

From Table 6, nine languages were found for 195 types of nouns classified by *buah*. The influence from English was found to be the greatest, with 152 (77.9%) from the total 195 noun types, indicating that many nouns classified by *buah* are of English origin. (However, this does not mean that all nouns with English origin are necessarily used with *buah*.) Sanskrit is another language of influence, with 12 types and 244 (15.5%) tokens from a total 1,574 tokens that could be traced to that language of origin. When the senses for these languages were analyzed, only English and Sanskrit showed identifiable patterns. Nouns originating from English are constituted mainly by ‘artifact’ (27%, or 41 from 152 tokens), while nouns from Sanskrit have a majority sense of ‘organization’ (42%, but with only 5 from 12 tokens). The other senses do not form a particular pattern in different languages.

5. Discussion and conclusion

Gathering all information from the previous discussion, the senses of *buah* can be summarized in Figure 1 below, detailing the possible metaphorical extensions between them.

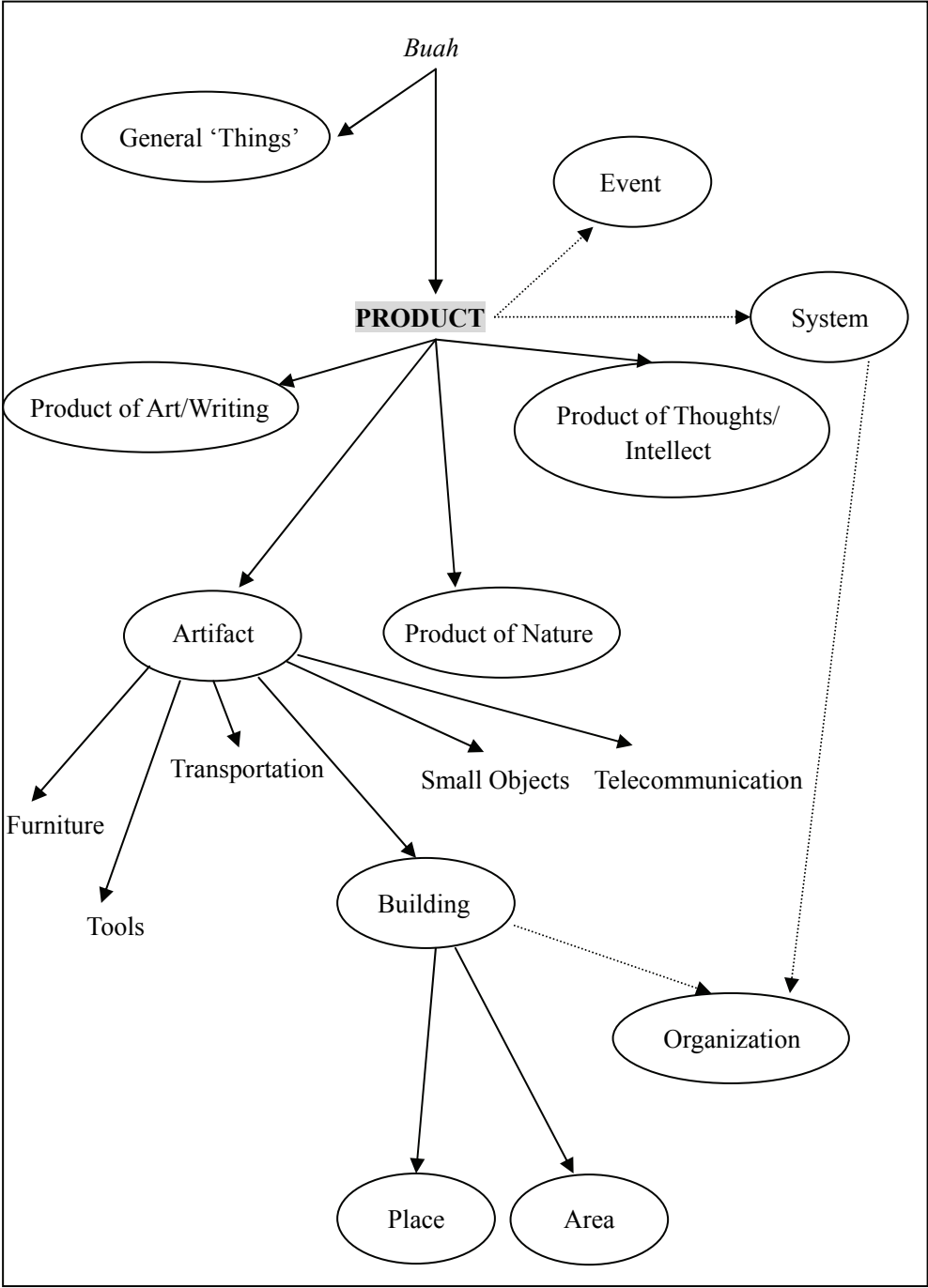


Figure 1: Uses of *buah* and their semantic extensions

In Figure 1, *buah* is suggested to be a classifier that categorizes nouns denoting a core meaning of ‘product.’ The types of ‘products’ include ‘product of arts/writing’ (article, newspaper, pamphlet), ‘product of thoughts/intellect’ (thoughts, philosophy, decision), ‘product of nature’ (cave, stone), ‘artifact’ and two other weakly linked (dotted arrows) ‘products’—‘event’ (as a product of an action: e.g. unification is a product of unifying effort) and ‘system’ (as a product of a set of rules and beliefs). In addition, ‘system’ is also linked to ‘organization,’ where rules and beliefs are put into practice. At the same time, ‘organization’ is a metonymic extension of ‘building,’ alongside ‘place’ and ‘area.’ ‘Building,’ which forms a separate category of meaning by itself, is also part of an ‘artifact.’ The other members of ‘artifacts’ include ‘furniture’ (cupboard, chair, stool), ‘transportation’ (car, train, lorry), ‘small object’ (photo album, bag), ‘telecommunication’ (computer, telephone), and ‘tool’ (gadget, tool). In addition to these categories, there is also a group of ‘general things,’ constituted by non-specific collective nouns such as entity, thing, and asset. (Note that further relationships can also be established between these senses—from ‘place’ to ‘area’ (metonymy), from ‘furniture’ to ‘building’ (partonymy or part-whole relation), etc.)

From the establishment of relationships, it is not hard to imagine how the compound nouns *buah dada* ‘breasts’ and *buah pinggang* ‘kidneys’ can be extended to *buah pemikiran* ‘fruit of thoughts,’ and later from ‘fruit of thoughts’ to an ‘event’ (e.g. forum, discussion, and negotiation).

From what has been studied so far, we discovered that not only is *buah* used with names of artifacts, such as buildings, machines, artwork, and organizations, it is also used with nouns that are loanwords (at least 50% of them were such nouns, excluding those that could not be found). A quick reference to a classic Malay manuscript dating from the 1370s from the Malay Concordance Project online shows that both the ‘fruit’ reading and the numeral classifier *buah* existed, but its primary use is to classify terms such as *kapal* ‘ship,’ *rumah* ‘house,’ *gunung* ‘mountain,’ *gua* ‘cave,’ *khemah* ‘camp,’ *kolam* ‘pond,’ and *perahu* ‘sampan.’⁷ These same nouns are still used with *buah* in modern Malay. However, an additional concern of this work is that when new technologies, such as computer, radio, and video camera (which are also now used with *buah*), appeared after the 1370s, what kinds of mechanisms decide whether they should be classified using *buah*? The results herein suggest that there must be a certain semantic function for *buah* so that things that did not exist in the 1370s can be used with *buah* in modern Malay.

Figure 1 provides a hint that the artifacts for telecommunication are likely to be grouped using *buah*. Furthermore, previous results suggest that a classifier-needing loanword is likely to be grouped using *buah* as well. These results are not without doubt

⁷ <http://mcp.anu.edu.au/>

because CDs and DVDs are not grouped with *buah* but, rather, with *keping* (for flat, thin objects; cf. Table 2). With the findings obtained so far, we postulate the following two mechanisms, which may help answer the above question. These mechanisms inform us about the conditions that might have determined the choice of *buah* as a numeral classifier for classifier-needing new terms when they were first used in modern Malay. These two mechanisms are given in (11) below.

- (11) a. First, when a loanword that is foreign to the culture of Malay enters its vocabulary and this noun needs a classifier, *buah* will serve as a default classifier if this foreign object does not have particular shapes that can be classified by other classifiers, such as *keping* (CD, DVD, paper, etc.). Therefore, the other classifiers emphasizing the shape (and/or size) of the entity will serve as the ‘filter’ before deciding whether *buah* can apply to these foreign products.
- b. Second, products of nature, artifact, and abstract creation of intellect, as well as organization or system, are generally used with *buah*. These products are often (but not necessarily) big.

Using the above two mechanisms, we are able to explain why there are seemingly dissimilar nouns classified using *buah*. An answer based on the definition of ‘classifying things with indefinite shapes and types’ is not sufficient to explain the rich cultural meaning of *buah*. The cultural part of *buah* may also be related to solidarity, whereby foreign objects are classified separately from others (if they do not have particular shapes as a precondition). This self-and-other metaphor (Lakoff & Johnson 1999) is seen even in classifying objects.

Referring to the comment that *buah* has no semantic content (cf. Hopper 1986, 1991 in his comment on *sa-buah*), we provide two responses.⁸ First, it might be that Hopper was referring to the grammaticalization of *sa-buah*, not *buah* itself. It could be that *buah* acts more like a general classifier when used with the clitic *se-* in *se-buah* (or *sa-buah* as its derived form in historical texts) but not when it appears in *dua buah* ‘two BUAH.’ The results from this study prove that *buah* is definitely used for a reason. Second, using the general classifier 個 in Chinese as a comparison, we compare the characteristics of 個 with *buah* and observe that for almost all nouns used with the general classifier of 個 in Chinese, they can take both the general classifier and one or more non-general classifiers, as shown in example (12) below.

⁸ The hyphen after *se-* appears only when we refer to Hopper’s study. In written Malay, *se-* is combined with *buah*.

- (12) a. 一個 房子 (general)
 one Class. house
 ‘a house’
- b. 一間 房子 (specific)
 one Class. house
 ‘a house (emphasis on the opening of the house)’
- c. 一棟 房子 (specific)
 one Class. house
 ‘a house (emphasis on the building of the house)’

For (12a), one can use the general classifier 個 to mean one house. (Note the necessity of adding a description when translating the phrases in example (12) because such specific meanings (as shown in (12b) and (12c)) are not expressed in the same way in English.) However, one can also substitute 個 with other classifiers, such as 間 in (12b) to emphasize the opening or 棟 in (12c) to emphasize the structure. The examples above show that 個 is indeed a general classifier that can serve as the “default-rule” (cf. Myers 2000) when someone does not know which classifier to use. For *buah*, however, nouns that are classified by it cannot take other classifiers to form more specific meanings. Furthermore, Malay has another indefinite quantifier of *suatu* ‘one’ that serves more like a general classifier than *buah* (cf. Hopper 1986, 1991). Therefore, *buah* does not seem to be a general classifier, according to the findings of this work.⁹

To conclude, based on our study and investigation, the results show that *buah* serves a function that is important to the culture. This paper argues for the non-general-classifier’s status of *buah* as it does display semantic functions and many of them can be used to classify various forms of products (literal and metaphorical). In addition, this study found that *buah* is not only salient in terms of its cultural function but also the nouns it classifies have some characteristics in common. The grouping of loanwords by a particular classifier may reflect the self-and-other metaphor, and this aspect of the culture is seldom emphasized in literature.

As future work, the differences between *buah* in *sebuah* ‘one.BUAH’ versus *dua buah* ‘two BUAH’ can be carried out. There is a possibility that *sebuah* (but not *buah*) can be used interchangeably with *suatu* as in *sebuah idea* ‘one.BUAH idea’ and *suatu idea* ‘one idea,’ yet it cannot be multiplied to mean **dua buah idea* ‘two BUAH idea.’ Whether this phenomenon is caused by the abstract nature of ‘idea’ or the grammaticization of *sebuah* will be an area for future research.

⁹ *Buah* is unlike *suatu* because they both have different constructions, where *suatu* is a number that means ‘one’ and *buah* needs another number to quantify it.

In addition, the presence and absence of *buah*, as Hopper (1991) has noticed, may also bring a different light to this study. Hopper found that when a noun (e.g. *rumah batu* ‘brick house’) is not the subject of discussion, its numeral classifier *sa-buah* can be dropped. He then analyzed the nouns of numeral classifiers when they appeared with or without classifiers using Givón’s measures of discourse persistence to NP (cf. Givón 1995). His results show that the appearance of the numeral classifiers “bears a striking resemblance to the emergence of an indefinite article in English.” His observation demonstrates that not only is the choice of classifier an important issue in research, but also the absence of supposedly-present classifiers can bring about a pragmatic impact on the language. Therefore, it is also worth investigating not only the presence of *buah* used in the newspaper genre but also their absence. These we reserve for future work.

In conclusion, arguing against the claim that *buah* is used to classify things that do not have definite shapes and types, this paper utilizes a corpus-based, quantitative approach through which the semantic functions of *buah* are presented. While most work in classifier systems (e.g. Adams & Conklin 1973, Allan 1977, Salehuddin & Winskel 2008) focuses on establishing a typology of classifiers or on discovering the semantic features of classifiers, this paper crystallizes the influence of culture in classification and explains how a grammatical system such as numeral classifier may reflect a community’s treatment of categorization. This study hopes to shed light on the cognitive importance of numeral classifiers. Study in this direction will also support a corpus-based approach where (re-)analysis through exemplars is emphasized.

References

- Adams, Karen L. 1986. Numeral classifiers in Austroasiatic. *Noun Classes and Categorization: Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*, ed. by Colette G. Craig, 241-262. Amsterdam & Philadelphia: John Benjamins.
- Adams, Karen, and Nancy F. Conklin. 1973. Toward a theory of natural classification. *Chicago Linguistic Society (CLS)* 9:1-10.
- Aikhenvald, Alexandra Y. 2003. *Classifier: A Typology of Noun Categorization Devices*. Oxford & New York: Oxford University Press.
- Allan, Keith. 1977. Classifier. *Language* 53.2:285-311.
- Anthony, Laurence. 2005. AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. *Proceedings of 2005 IEEE International Professional Communication Conference (IPCC 2005)*, 729-737. Limerick, Ireland.
- Asmah binte Haji Omar. 1966. The nature of Tamil loanwords in Malay. *Proceedings of the First International Conference of Tamil Studies*, Vol. 2, 534-558. Kuala Lumpur, Malaysia.
- Asmah binte Haji Omar. 1972. Numeral classifiers in Malay and Iban. *Anthropological Linguistics* 14.3:87-96.
- Dewan Bahasa and Pustaka. (eds.) 2002. *Kamus Dewan [The Institute Dictionary]* (3rd edition). Kuala Lumpur: Dewan Bahasa and Pustaka.
- Givón, Talmy. 1995. *Functionalism and Grammar*. Amsterdam & Philadelphia: John Benjamins.
- Hopper, Paul J. 1986. Some discourse functions of classifiers in Malay. *Noun Classes and Categorization: Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*, ed. by Colette G. Craig, 309-325. Amsterdam & Philadelphia: John Benjamins.
- Hopper, Paul J. 1991. Some remarks on the grammaticalization of classifiers in Western Austronesian, especially Malay. Paper presented at the Sixth International Conference on Austronesian Linguistics (6-ICAL). Honolulu, Hawai'i, USA.
- Jones, Robert B. 1970. Classifier constructions in Southeast Asia. *Journal of the American Oriental Society* 90.1:1-12.
- Kuiper, F. B. J. 1962. A Malay word in Tamil. *Indo-Iranian Journal* 5.3:237-241.
- Lakoff, George, and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.
- Marnita, Rina A. 1996. *Classifier in Minangkabau*. Canberra: The Australian National University MA thesis.

- Maxwell, William Edward, Sir. 2008. *A Manual of the Malay Language: With an Introductory Sketch of the Sanskrit Element in Malay*. E-book available from <http://www.gutenberg.org/etext/25604>.
- Myers, James. 2000. Rules vs. analogy in Mandarin classifier selection. *Language and Linguistics* 1.2:187-209.
- Saalbach, Henrik, and Mutsumi Imai. 2005. Do classifier categories structure our concepts? Paper presented at the 27th Annual Conference of the Cognitive Science Society (CogSci 2005). Stresa, Italy.
- Salehuddin, Khazriyati, and Heather Winskel. 2008. Malay numeral classifiers: sketching conceptual representation from a native speaker's perspective. *SEALS XVI: Papers from the 16th Annual Meeting of the Southeast Asian Linguistics Society 2006*, ed. by Paul Sidwell & Uri Tadmor, 65-74. Canberra: The Australian National University.
- Schmitt, Bernd H., and Shi Zhang. 1998. Language structure and categorization: a study of classifiers in consumer, cognition, judgment and choice. *Journal of Consumer Research* 25.2:108-122.
- Scott, Michael. 2008. *WordSmith Tools*, Version 5. Liverpool: Lexical Analysis Software.
- Tai, James H-Y. 1994. Chinese classifier systems and human categorization. In *Honor of William S-Y. Wang: Interdisciplinary Studies on Language and Language Change*, ed. by Matthew Y. Chen & Ovid J. L. Tzeng, 479-494. Taipei: Pyramid.
- Tai, James H-Y., and Lianqing Wang. 1990. A semantic study of the classifier *tiao* (條). *Journal of the Chinese Language Teachers Association* 25.1:35-56.
- Tien, Yi-Min, Ovid J. L. Tzeng, and Daisy L. Hung. 2002. Semantic and cognitive basis of Chinese classifiers: a functional approach. *Language and Linguistics* 3.1:101-132. (In Chinese)
- Ungerer, Friedrich, and Hans-Jörg Schmid. 2006. *An Introduction to Cognitive Linguistics*. Harlow: Pearson/Longman.
- Wignesan, T. 1995. The extent of the influence of Tamil on the Malay language: a comparative study. Paper presented at the Eighth World Tamil Studies Congress. Tanjavur, India: Tamil University. Also available at <http://stateless.freehosting.net/TAMILINFLUENCEMALAY.htm>.
- Wilkinson, Richard J. 1952. *An Abridged Malay-English Dictionary*. London: Macmillan.

[Received 3 December 2009; revised 31 March 2010; accepted 8 April 2010]

Department of English
National Chengchi University
64, Sec. 2, Zhinan Road
Taipei 116, Taiwan
sfchung@nccu.edu.tw

馬來語量詞「Buah」之語料庫研究

鍾曉芳

國立政治大學

本論文針對標準馬來語之量詞「buah」進行研究。「buah」可區分無特定型態或形體，且具差異性之名詞，因此「buah」是否為「general classifier」開始引起討論。本研究分析 5,009 篇，字數超過 100 萬字之馬來西亞新聞語料中，以「buah」分類之名詞，研究認為「buah」除區分語意外，亦賦有文化意涵；進一步探討「buah」各種含義之相關性，發現「buah」的含義與各類「產物」（工藝、通訊、自然或思想等）之衍生義有隱喻關連；由語料庫分析「buah」之語意，並提出兩種篩選方式，得到「buah」並非「general classifier」之結論。

關鍵詞：量詞，語意，馬來語，buah，語料庫