Quantifier and Reduplication in Malay: A Bridging Cross Reference Analysis

Nor Hashimah Jalaluddin
Zaharani Ahmad
School of Language and Linguistics
Faculty of Social Sciences and Humanities
National University of Malaysia

Abstract

Many Malay grammarians, namely Asmah (1993), Nik Safiah et.al (1997), Hashim Musa (1993), Liaw Yock Fang and Abdullah Hassan (1994) and Za'ba (2000) have discussed about reduplication and their focus primarily on the morphological aspect. There are only a few studies that explore meanings of the reduplicated forms. Nor Hashimah (2002) suggests that based on semantics perspective, Malay reduplication conveys specific and non-specific meanings. Numerical cardinals and quantifiers are also regarded as part of reduplication. The phrase *'tiga ekor ayam'* (three hens) is in plural forms – the numeral *'tiga'* marks the plurality and the nominal *'ayam'* is not reduplicated. Malay grammarians propose the same rule for quantifiers. The phrase *'semua guru'* (all teachers) is already in plural form and therefore the word *'guru'* does not need to be reduplicated. Nevertheless, the written and spoken data proved otherwise. The analysis on functions of numerical reduplications will be more exiting if corpus with big data is examined. Accordingly, semantic analysis of bridging cross-reference and reference assignment explains the differences between the two numerical reduplications adequately.

Introduction

The usage of numericals in the Malay language is quite a challenge to language users. A Malay numerical is often paired with a numerical classifier. Unlike English, the Malay language does not have any morphological changes to signify plurality. Therefore, a numerical is able to provide information whether a subject being discussed is singular or plural. *'Seekor kambing'* (a goat) is different from '*sepuluh ekor kambing*' (ten goats) based on its quantity. Clearly, there are no morphological changes to the word '*kambing'* (goat) itself. This type of numerical is called the cardinal numerical or the specific numerical. It is a different case for non-specific numericals/quantifiers, for instance '*para'* (a number), *'semua'* (the whole), *'segala'* (all), *'beberapa'* (a few), *'seluruh'* (entire) and others. According to the *Kamus Dewan* (1994), non-specific numericals carry the meaning of plurality. Does the plurality of this non-specific numerical/quantifier carry other meanings other than plurality? Shown here is the delegation of work in the form of complimentary distribution between the specific

numerical and the non-specific numerical. If singular or plural can be developed from specific numericals, the same case does not occur in non-specific numericals. Thus non-specific numericals have only plural connotations.

This research aims to help readers gain a deeper understanding on the numericals in the Malay language. The aspect of non-specific numerical seems to give more room for debates. The occurrence of reduplicated non-specific numericals in daily conversation seems to show that a non-specific numerical which is plural in nature can be reduplicated. Such occurrences of reduplicated non-specific numericals used by Malay native speakers found in spoken corpus can be used to challenge the definitions of numericals prescribed earlier by Malay grammarians. Digitalized corpus databased has certified the high occurrences of reduplication of non-specific numericals. Based on analysis of corpus (see below), a reassessment of the use of non-specific numericals must be carried out. A bridging cross-reference approach and reference assignment in semantics will be used to explain this emerging conflict.

Earlier Grammatical Description

Aspects of the numerical are discussed in most Malay grammar books. In *Tatabahasa Dewan* (1986), Nik Safiah et. al touched briefly on numericals. Nik Safiah et. al divided numericals into cardinal and non-specific numericals. According to Nik Safiah et. al, numericals which show amounts with specific attributes can be called cardinal numericals. Examples of cardinal numericals are '*dua'* (two), *'dua puluh'* (twenty), *'tiga ratus'* (three hundred), *'sepuluh ribu'* (ten thousand) and others. Whereas, non-specific numericals such as '*para'* (a number), *'beberapa'* (a few), *'semua'* (all), *'seluruh'* (whole), *'sekalian'* (everyone) and others indicate plurality and cannot be reduplicated again. Therefore, the following sentences are considered incorrect:

i) **Segala persiapan-persiapan telah dibuat.*
All preparations have been made
'All preparations have been made'

ii) **Semua pegawai-pegawai tidak dibenarkan bercuti esok.*
All officers not allowed take leave tomorrow
' All officers are not allowed to take leave tomorrow'

Other than the numericals above, Nik Safiah et. al also discuss collective numericals such as '*berabad-abad'* (for centuries)*, 'beribu-ribu'* (thousands)*, 'berguni-guni'* (sacks) and others. Numericals also provide a separate meaning such as '*masing-masing'* (each), *'tiap-tiap*' (every) and *'setiap'* (each). The aspect of fraction numericals is also discussed such as '*setengah'* (half)*, 'sepertiga'* (one-third)*, 'separuh'* (half).

Abdullah Hassan and Liaw Yock Fang in their book *Nahu Melayu Moden* (1994) also touch on this aspect of numericals. Their discussion is similar with Nik Safiah et. al (1986), except for some additional information from the functional aspects of the numerical itself. Abdullah and Liaw provide the main function of the numerical as a noun modifier. For example:

i) *Saya mempunyai **dua orang** sahabat karib.*
I have two class.. friend best.
'I have two best friends'

ii) *Orang itu ada **banyak** wang yang disimpan di dalam bank.*
Person that have a lot of money rel. saved in bank
'That person has a lot of money saved in the bank'.

iii) ***Seluruh** ruangan itu sunyi senyap.*
Whole room the silent
'The whole room was silent'.

Besides functioning as a noun modifier, numericals can also function as subjects, objects and adverbs. For example:

Subject:

i) *Semua pegawai sudah keluar dari bilik mesyuarat*
All officers have left from room meeting.
'All the officers have left the room'.

Object:

ii) *Umurnya baru **lima belas** tahun*
Age his only fifteen years.
'He is only fifteen years old'.

Adverb:

iii) *Mereka datang **bertiga***
They come three of them
'The three of them came together'.

In "*Nahu Melayu Mutakhir*" (1982), Asmah Hj. Omar explained the numericals in more detail. Asmah divided cardinal numericals into ordinal numbers and quantifiers. The difference between ordinal numbers and quantifiers is from the aspect of number, which can be the foundation of ordinal numbers but not for quantifiers. An interesting factor in Asmah's writing is that she does not state that a quantifier cannot be reduplicated. She states that the quantifier '*para*' (a number) is used to pluralise a noun at a higher social status, for instance '*para pegawai'* (a number of officers) and *'para mahasiswa'* (a number of graduates).

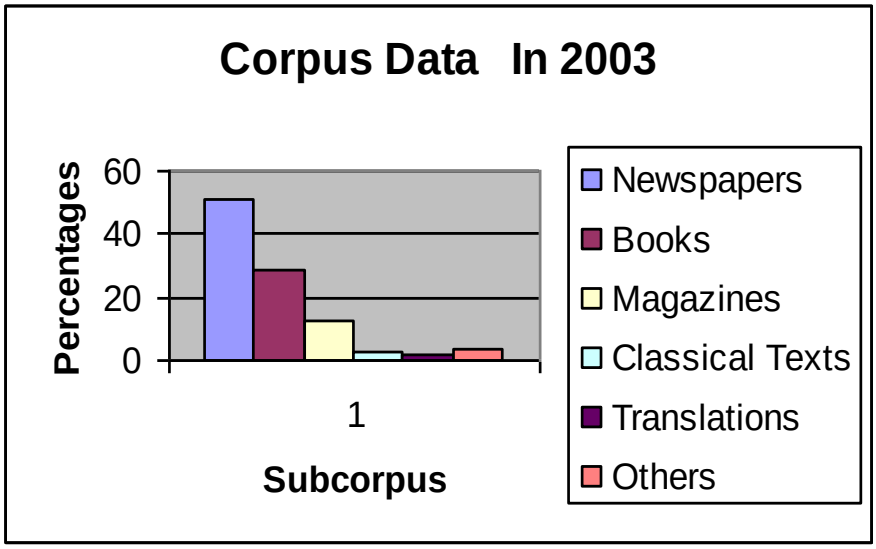
From the three views given by the grammarians discussed earlier, it can be summarized that Asmah's approach to explaning the numerical is more flexible. Asmah does not state that reduplicating a quantifier is grammatically wrong. However, Abdullah and Liaw and Nik Safiah et. al provide a clear division between specific numericals (cardinal) and non-specific numericals (quantifier). The descriptions provided by Abdullah and Liaw and Nik Safiah et. al are more practical and suitable for the purpose of language teaching at school level, which needs a simpler grammatical explanation. Nonetheless, for graduate studies level, Asmah's view with regard to the numerical leaves more room for future exploration. Asmah's discussion invites more academic debate.

Asmah's approach is more applicable when discussing corpus based numerical analysis. It has been shown that the use of non-specific numerical/quantifier being reduplicated occurs not only in digitalized corpus data but also in spoken corpus. Malay speakers produce phrases, clauses and sentences by reduplicating non-specific numericals with ease. The frequency of this utilisation is obvious especially in spoken corpus. It is

expressed not only in informal situations but also in formal contexts. In fact, many of us also tend to do the same without realising it. This is the conflict that occurs between what is stated in grammar books and what happens in the real world.

The reason this conflict happens can be further discussed. A sufficient amount of data was taken from 100 million samples from the corpus data bank compiled by Dewan Bahasa dan Pustaka. The compilation of corpus consists of data from newspaper articles (51.08%), books (28.34%), magazines (12.3%), classical texts (2.8%), translations (1.95%) and others (3.53%). This research uses only 5% of the sample data from the data bank which is adequate for the purpose of this paper. Table 1 shows the distribution of sub-corpus gathered in the data bank by the year 2003. From this data bank, we can explicate the problems of adjoining numericals and reduplication in Malay with the special focus on the non-specific numericals.

Table 1: ***Distribution of Sub Corpus***



Accepted and Non-Accepted Forms of Numericals

To explain what is known as grammatical and ungrammatical numericals, the following examples are taken from corpus data.

Accepted Forms of Specific and Non-specific Numericals

a. *"Tengok nasib?" Dia bertanya. ***Dua*** orang pemuda yang beruniform.* "

"Looking at your luck?" He asked. ***Two*** uniformed males.

b. *Yang ada cuma ***dua*** buah almari besi, dua buah meja.*

There were only two steel cupboards and *** two*** tables.

c. *kepada pihak berkenaan cuba fikirkan ***segala*** baik buruknya kursus vokasional di SMT sekiranya dihentikan.*

To those concerned, please think of ***all*** the advantages and disadvantages if the vocational courses at SMT are discontinued.

d. *"Oleh itu, kerajaan harus bertegas supaya ***semua*** urusan itu dibuat secara telus.*

Therefore, the government has to be firm so that ***all*** matters will be dealt with in a transparent manner.

e. *Saranan kerajaan untuk menggalakkan rakyat menumpukan semula kepada sektor pertanian sebagai ***sebahagian*** usaha menangani era kegawatan ekonomi sekarang,*

The government's suggestion to encourage citizens to focus again on the agricultural sector as ***part*** of the effort to counter the effect of the current economic downturn.

f. *kesan El Nino mula menurun dan ***kebanyakan*** iklim dan keadaan cuaca dunia mula kembali normal.*

The effect of El Nino has lessened and ***most*** of the world's climate and weather situations have gone back to normal.

Non Accepted Forms of Non-Specific Numericals

a. *Dia ingin menyatakan rasa kesalnya dengan sikap dan pemikiran ***sesetengah*** orang-orang kampung yang tidak berminat menghantar anak-anak mereka untuk mendapatkan pelajaran.*

He wants to express his disappointment with ***some*** of the villagers' attitudes and thinking in which they are not interested in sending their children to obtain education.

b. *Kita mungkin ada pengalaman melihat bagaimana ada ***sebilangan*** anak- anak muda yang menghabiskan malam-malam Ramadan dengan kerja sia-sia.*

We may have experienced how ***some*** youngsters spend the nights during Ramadhan involved in useless activities.

c. ****Sebahagian*** kilang-kilang yang akan didirikan di Taman Industri Getah itu adalah berkonsepkan inkubetor (incubator) dan bimbingan.*

Part of the factories to be built at the Rubber Industry Park is based on the incubator and guidance concept.

d. *peranan kontraktor (pemborong) sebagai orang tengah yang masih dikekalkan di ***kebanyakan*** ladang-ladang KETENGAH .*

The role of contractors as a middle persons is still used in ***most*** KETENGAH farms.

e. *Anda tidak boleh membantu ***semua*** anak-anak Iraq," katanya.*

"You cannot help ***all*** the Iraqi children", he said.

f. *Bolehkah kita mempercayai seseorang doktor itu jika ***segala*** butir-butir yang sepatutnya di rahsiakan menjadi bahan bualan?*

Can we trust a doctor if ***all*** details that must be confidential become topics of conversation?

The data examples above clearly show that there is a conflict between the views of grammarians and evidences found from the corpus of data. As mentioned earlier, the frequency of data using non-specific numericals with reduplicated noun words is quite high. Using evidence from the digitalized corpus of data to explicate Malay grammar is still new in grammatical analysis. Nonetheless, this method is one way of studying conflicts that may occur in the usage of a grammar. This is in line with a linguist's task and it is his or her responsibility to find out why such language patterns occur widely, contradicting the grammar perscribed by grammarians.

Explicating the Conflict

When a conflict occurs in a Malay grammatical description, it is necessary to find a solution to explain the language in a more scientific manner. Linguistics is a scientific discipline which enables research to be conducted. Linguistics is not static because language itself is a living entity. Nevertheless, all scientific studies need a solid academic description to ensure language users accept new findings to replace earlier perscribed grammar. With regard to conflicts in grammatical description and actual corpus data, specifically spoken corpus, everyday usage of the grammatical entity and spoken corpus of data can help towards finding a solution to these conflicts. However, a corpus of data alone cannot help towards finding a solution. A theoretical analysis must also accompany the corpus of data and in the case of numericals, a semantic analysis through a bridging cross reference approach and reference assignment could help to explain the problems that arise.

Corpus and Linguistic Evidence

McEnery (2001) states "corpus linguistics is perhaps best described for the moment in simple terms as the study of language based on examples of 'real life' language use". In fact, McEnery also adds, "corpus linguists study real language, other linguists just sit at their coffee table and think of wild and impossible sentences". Such statements are not well received by linguists from the Chomskyan school of thought. In fact, they regard corpus linguistics only as a method in language analysis.

This opinion can be disputed by observing how semanticists have solved conflicting problems with the help of corpus data. Mindt (1991) and McEnery (2001) proved that corpus data is able to: firstly, give an objective criteria in determining meaning in English and secondly, to solve fuzziness and gradience problems in English. With corpus assistance, Mindt successfully refuted earlier English grammarians' claims on tenses. Mindt discovered that personal subject occured frequently with present progressive, and rarely with simple future, *will* and *be going to*. It is clear that there is semantic relationship between the subject and the verb used in future time reference.

The same case can be seen in determining polysemy in a dictionary. Nor Hashimah Jalaluddin and Norsimah (2005) have shown how the meaning of polysemy

can be resolved in a dictionary. Using a cognitive semantics approach and corpus data, the meaning of *kuat* (strong) is described based on the gradience of meaning which is more systematic . Examples of the advantages of corpus in gradience is as follows:

The meaning of *kuat* /stong as given by Kamus Dewan (1994):

1. a lot of energy, not weak, energetic, able, strong and sturdy
2. not easily toppled, collapsible (as in damaged, broken etc), durable
3. does not give in easily, strong
4. does not change easily (pert. decision) determined (of faith)
5. solid (of weight), firm, loud and clear (of voice)
6. able or willing to do something
7. likes to eat (and drink etc.) a lot
8. having an advantage or skilful

The meaning listed above are no longer accurate when we refer to the corpus data taken from novels. *Strong* as seen above have changed from the core meaning, that is physical strength to emotion. From the data derived from popular novels, strong are now referring to:

i) the stand that one takes and it is about a person who does not give-up or change easily, determined and strong willed

ii) ability to go through something that is hurtful, willingness to sacrifice oneself for a person

iii) the raising of the voice to state one's position or stand or power

iv) a lot of energy, not weak, energetic, able, strong and sturdy

v) not easily toppled, collapsible (as in damaged, broken etc), durable

vi) likes to eat (and drink etc.) a lot

vii) having an advantage or skilful

From the meaning given above, it is clear that corpus data provide a semantic solution, especially in handling fuzziness and gradience. The meaning of *kuat* (strong) in *Kamus*

Dewan (1994) can be described with several meanings sistematically following the cognitive sequence. This assists users to see the extension of meaning of an entry in an orderly manner. The occurrences of *strong* in the corpus surely will help in meaning reordering of a dictionary. Here is the importance of corpus and linguistics in providing more authentic data with an elegant analysi

Going back to the question of numericals, a pertinent question can be asked: Can a semantic analysis using data base be done in the case of numericals? Further, is there a possibility that such analysis could be carried out? The answer is yes. According to earlier grammarians, the reduplication of a noun word following a non-specific numerical is incorrect based on Malay grammar. Asmah does not abide by this by excluding the word "*para*" (a number). The fact of the matter is that the supposed 'ungrammatical usage' occurs rampantly in language use.

Another role of corpus and linguistics can be proven from Zaharani Ahmad's study (2002). Often when there are language mistakes done by Malay speakers, redundancy will be blamed. The reason '*semua guru-guru*' (all teachers) is thought as ungrammatical is because of redundancy. *'Semua'* (all) carries the meaning of plurality and therefore it is considered ungrammatical to reduplicate *'guru'* (teacher). The same rules apply for the affixes '*memper-kan*'[1] which is considered wrong; '*yang paling cantik sekali*' (the most beautiful)[2] is also redundant. There are many morphological cases which are against the presciptive Malay grammar that can be solved with corpus linguistic evidence.

In the Malay language, the combination '*memper*' with adjectives ending in '*-kan*' is not allowed at all. This grammatical rule coincides with the grammatical description stated in *Tatabahasa Dewan* (Nik Safiah 1989). The presence of '*-kan*' is not needed as this will make the derivation redundant. Due to this, the grammatical derivative, which follows prescriptive grammar, is '*memperbesar*' (to cause to be big) and not '*memperbesarkan' (to cause to be big)*, '*memperdalam*' (to cause to be deep) and not '*memperdalamkan*' (to cause to be deep).

Although both derivatives are ungrammatical, Zaharani Ahmad's study (2006) shows that its usage among the Malay speakers are wide and very productive. Studies

concerning the utilization of the affix '*memper-'* and '*memper-kan*' are based on corpus data conducted by Abdul Jalil (2000) who studied corpus data contained in *Berita Harian* and *Berita Minggu* newspapers from April, May and June 2000. His findings show the utilization of the prefix '*memper-*' is not productive compared to the prefix '*memper-kan*'. This study also shows that the affix '*memper-kan*' can be combined with all Malay language words, including root words, which are adjectives. The table below shows the frequency of the usage of the affix '*memper-*' and '*memper-kan*', which is found in the studied derived word.

Table 2: ***Lists of Adjective with Suffixes From Corpus***

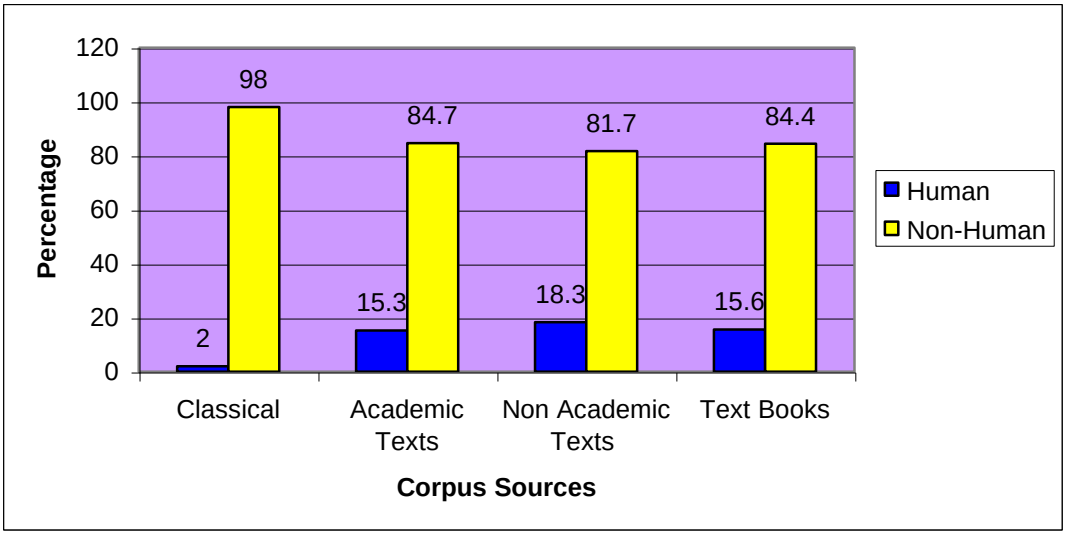
The frequency of the usage of the affix '*memper-*' and '*memper-kan*'

Root word	**Affix '*memper-*'**	**Affix '*memper-kan*'**
besar (big)	-	9
cepat (quick)	3	41
kecil (small)	2	12
kembang (expand)	2	15
kuat (strong)	1	11
kukuh (firm)	13	42
luas (wide)	8	33

Table 2 above shows that Malay language users are inclined to use the affix '*memper-kan*' rather than '*memper-*'. This inclination is not influenced by intuition alone, but is based on the speaker's efficiency in the language. This reflects the grammatical system, which is practiced by Malay users as native speakers.

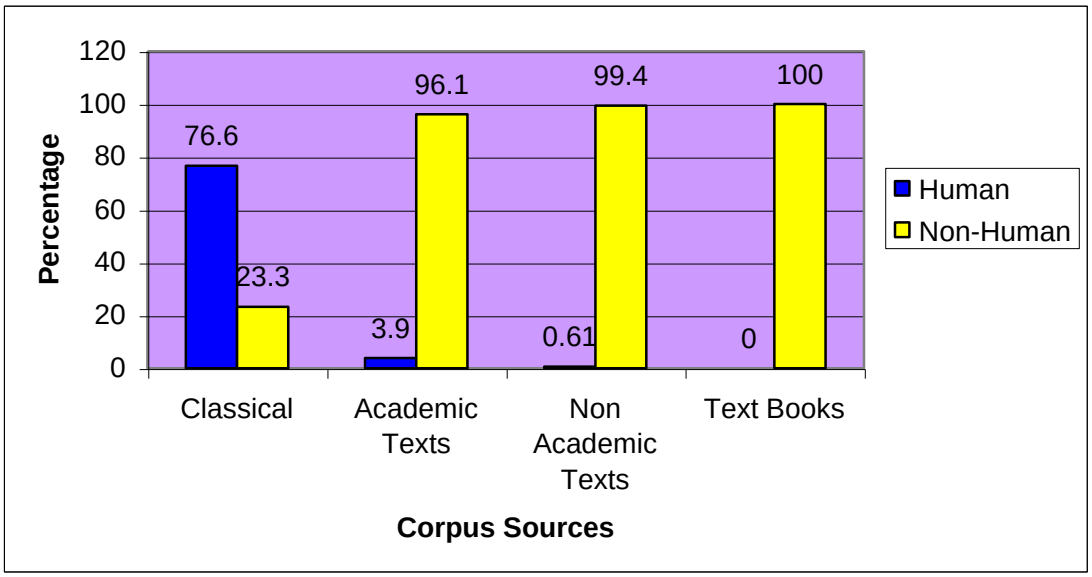
Another significance of corpus and linguistics in examining the conflicting problems between grammarians and findings of corpus based analysis is by looking at the study of pronouns. Nor Hashimah Jalaluddin (2006) has proven that the pronoun *ia* and '*ianya*' can be used to refer to both humans and non-humans. These findings conflicting with the prescribed grammar which stated that *ia* (it) can be used for human and *ianya* (it) is for abstract noun.

Table 3: Pronoun 'IA' From Corpus



From the table above (Table 3), the utilisation of the pronoun '*ia*' referring to non-living things is higher than the specific use for person. 98% of the corpus from classical data refer to non-living things, 84.7% from academic magazines, 81.7% from non-educational magazines and 84.4% from textbooks. These data show us that only the pronoun '*ia*' referring to the third person need to be corrected. Another instance is the pronoun *'ianya'* (it). It shows that the rules from prescribed grammar must be revised.

Table 4: *Pronoun "IANYA' from Corpus*



From Table 4 above, it can be seen that the pronoun '*ianya*' can refer to humans and non-humans. For classical data, there are 76.7% uses of '*ianya*' refering to persons, 3.9% for academic magazines and 0.6% for non-academic magazines. There is no data showing '*ianya*' referring to persons in textbooks. The data has proven that the pronoun '*ianya*' can refer to persons (from classical data) and to non-humans.

Linguistic evidence based on corpus, specifically using bridging cross reference semantic analysis is successful in convincing Malay speakers that the pronoun '*ia*' and '*ianya*' can be used with non-living things after being proven scientifically.

Specific and Non-SpecificThinking

Apart from corpus and linguistics evidence, the characteristics of Malay thinking can also be taken into consideration. In this case, it relates to specific and non-specific thinking. What is meant by specific and non-specific thinking is the presence of other meanings other than plurality in the Malay reduplication system of Malay. In this research, specific and non-specific thinking are supported by earlier studies and can now contribute to the discussion of numericals. Using non-specific numericals with reduplicated nouns is considered ungrammatical and this can be discussed with linguistic analysis, specifically through semantic analysis.

In her writing on '*Reduplication in Malay Language: Semantic dan Pragmatic Analysis*', Nor Hashimah (2002) has proven that a reduplicated noun can be specific and non-specific, other than being plural. When reduplicated, the word can have a non-specific meaning. Examples given are as follows:

Noun Word (Specific)	**Redundant Word (Non-Specific)**
Siapa (Who)	*siapa-siapa* (anyone)
Apa (What)	*apa-apa* (anything)
Mana (Where)	*mana-mana* (anywhere)
'*Untung* (Gain)	*untung-untung* (with luck)

Verb (Specific)	**Multipled Word (Non-Specific)**
Main (Play)	*main-main* (kidding)
Duduk (Sit)	*duduk-duduk* (relax)
Makan (Eat)	*makan-makan* (snacking)
Gila (Crazy)	*gila-gila* (over enthusiastic)
Jalan (Walk)	*jalan-jalan* (strolling)
Minum (Drink)	*minum-minum* (drink with leisure)
Rehat (Rest)	*rehat-rehat* (relax)
Adjective (Specific)	**Adjective (Non-Specific)**
Gemuk (Fat)	gemuk-gemuk (relatively fat)
Kurus (Thin)	kurus-kurus (relatively thin)
Lampai (Slender)	lampai-lampai (relatively slender)
Tinggi (Tall)	tinggi-tinggi (relatively tall)
Sakit (sick)	sakit-sakit (relatively sick)

'*Siapa-siapa*' has no specific reference for example, '*Siapa-siapa yang terambil barang Siti diharap dapat memulangkannya*' (Whoever has taken Siti's things, kindly return them). In this sentence, no one is focused specifically. The same goes with '*makan-makan*'. '*Makan-makan*' can refer to light snacks, for instance peanuts, pickles, sunflower seeds as compared to rice or noodles, which are heavy and specific. The next example is with regards to non-specific reference for e.g. '*gemuk-gemuk*', '*Budak Mat yang gemuk-gemuk itu yang menghantar surat ini*'. (Mat, the relatively fat boy was the one who sent this letter).

Other than the reduplicated words shown above, there is a type of verb that is always connected to activities in a Malay community that can accept the affix verb '*meN*' easily. However, there are cases where this verb cannot accept the affix verb '*meN-* '. Again, it can be explained that specific verbs that are very much associated with the Malays can accept the affix properly, whereas non-specific verbs which do not belong to the Malays cannot accept the affix. The examples below could be used as proof.

Specific	Non-specific
Memarang (to chop)	Memisau (to knife)
Merumput (to cut grass)	Melalang (to cut long grass)
Merokok (to smoke)	Memaip (to pipe)
Menari (to dance)	Menjoget (to dance)
menyusu/menetek (to feed)	Membotol (to bottle)

All activities in the left column are specific activities carried out by most Malays whereas those in the right column are unspecific activities in nature. Apparently, the verbs in the right column cannot receive the *'men-'*[3] affix. The characteristics of the above specific and non-specific verbs can relate to human cognitive processes while communicating. Sperber and Wilson (405:1999) stated that:

> "there is no reason to think that an interpretive expression for relevant thinking through optimum is the most literal. Speaker is supposed to achieve optimum relevance, and not literal correctness".

What Sperber and Wilson meant is that it is human nature in communication to give attention to those relevant to only oneself. Although the message conveyed is something that is non-specific, it is sufficient as long as the meaning is understood and the interpretation received is given proper attention. Sometimes the message to be conveyed is general and non-specific. It is the hearer's obligation to filter and process the information, where both speaker and hearer think is relevant.

To continue the discussion of numericals with reduplicated nouns, we will focus on specific and non-specific meanings. If someone wants to say what is being done or to describe someone's attribute specifically, the basic word is sufficient. The hearer can process the information and the correct input is received from the specification. However, using reduplicated nouns is more suitable if the message to be conveyed is something that is non-specific, not serious and does not require both the speaker's and the hearer's full concentration. Certainly, something non-specific does not require literal remarks. Man with heuristic thoughts can accept non-specific explanations in their

communication. They are able to process messages that need to be delivered without much problem. What is important, is that the context is clear and the words chosen are easy to process.

Discussion of the Problem with regard to Non-specific Numericals

The solution to the problematic pronouns of '*ia*' and '*ianya*' are based on corpus linguistic data and bridging cross reference (see the section below), thus numericals also follow the same procedure with the addition of the sub-concept of reference assignment which is discussed in Relevance Theory (1995). This research will again use the specific and non-specific approach in categorizing the meanings of non-specific numericals. Hence, any non-specific numerical using a singular noun will be regarded to mean **specified**. Whereas any non-specific numerical using a reduplicatied noun will be considered to mean **unspecified**. The example below will lessen the confusion:

Specified Non-specific Numerical	**Unspecified Non-Specific Numerical**
****sesetengah*** calon menggunakan khidmat bomoh Siam, bomoh Jawa dan bomoh Orang Asli untuk menang,* ***Some*** candidates use the services of Siamese, Javanese and Aborigine medicine-men to win,	*Dia ingin menyatakan rasa kesalnya dengan sikap dan pemikiran ***sesetengah*** orang-orang kampong* He wants to express his regret with ***some*** of the villagers' attitudes and thinking
*bilangan peniaga buah sawit sudah mencapai tahap maksimum kerana ***sebilangan*** besar mewujudkan persaingan tidak sihat sehingga menjadi penyumbang kejatuhan OER.* ***The number of*** oil palm traders has reached the maximum level because a big number has created an unhealthy competition and has contributed to the fall of OER.	*Kita mungkin ada pengalaman melihat bagaimana ada ***sebilangan*** anak-anak muda yang menghabiskan malam.* We may have experienced how ***some*** youngsters spend their nightlife.
*Ini kerana ***kebanyakan*** sumbangan karyawan seni yang sudah mati lebih bernilai berbanding rakan-rakan mereka yang masih hidup.* This is because ***most*** contributions	*Saya yakin keadaan kita lebih baik d aripada ***kebanyakan*** negara-negara lain* yang mana negara lain mereka memang pun tidak dapat memberi peluang kerja

from artists who have passed away are more valuable compared to those who are still alive.)	I am confident that our condition is better than ***most*** countries, where other countries are not giving job opportunities)
semua penghulu mukim, ketua kampung, Ahli Jawatankuasa Kemajuan dan Keselamatan Kampung (JKKK) dikehendaki memberi penerangan dan penjelasan ***All*** territory chief, village head, Village Safety and Development Committee Members are required to give explanations and justifications.	***Semua*** lesen-lesen perjudian dan tempat- tempat hiburan sebagai punca-punca maksiat dibekukan. ***All*** gambling and entertainment place licenses are the source of sinful acts and are therefore frozen.

Based on the examples above, a clear model can be found to enable both non-specific numericals to be analysed with their own individual roles. The solution to this problem is by using bridging cross reference and sub-task of reference assignment as proposed in the theory.

Bridging Cross Reference

The Bridging Cross Reference (BCR) approach based on Relevance Theory can assist grammarians to see how pronouns/noun(s) act as anaphora or cataphora. According to Kempson (1986), BCR is a phrase, which has no antecedent that can connect the second sentence with the first sentence in an article to help the hearer understand the said phrase. The antecedent does not exist at all whether it is from explicit content, or from visualized picture. Therefore, BCR must involve at least a two fold sentence. To help the listener understand the phrase, the speaker must give additional premises that can be part of the context. The hearer on the other hand will use the additional premise to detect the required antecedent to understand the real meaning intended by the speaker. For example,

i) *Maka dicampakkannya galahnya, katanya, "Lari kita, perahu ini **ia** tiada mahu mudik dipegang (oleh) hantu air*" .

(Hence the pole was thrown and he said, "Run, *it* does not want to be sailed by the water ghost".)

Here, '*ia*'(it) is a pronoun with a specific reference. '*Ia*' (it) is also known as anaphora. However, a clear reference can not be seen here. Using BCR, '*ia*' can be connected to the '*perahu*' (boat). The information '*galah*' (pole) and '*hantu air*' (water ghost) are additional information that strengthen our interpretation that '*ia*' refers to the boat. Since we use a pole to sail the boat and we seldom find water ghosts in the river, then it clearly shows that 'ia' must refer to the 'boat'. The information is based on the context given in the utterance itself. This extra information is what we call the reference assignment. The reference assignment can be dealt with by rules, which automatically integrate properties of the context with semantic properties of the utterance. Context information refers to non-linguistic information. From the semantic point of view, meaning will stop at the logical form. Meanwhile, pragmatics will further explicate the full intended meaning. The same procedure is used for specific and non-specific meaning. Reference assignment and bridging cross reference can explain why a noun can be categorized as specific and non-specific. The completed context information can assist in assigning the reference.

Non-specific Numerical and Bridging Cross Reference (BCR)

Based on the data above and the few additional data from the corpus, the problem of the reduplication of non-specific numericals can be explained from its usage. In this section, the non-specific numericals will be discussed. The numericals involved are '*semua'* (all), *'kebanyakan'* (most)*, 'sebilangan'* (a few), *'sebahagian'* (a part) and *'segala'* (whole).

***Semua* (All)**

1a) ****semua*** penghulu mukim, ketua kampung, Ahli Jawatankuasa Kemajuan dan Keselamatan Kampung (JKKK) dikehendaki memberi penerangan dan penjelasan* ***All*** territory chiefs, village heads, Village Safety and Development Committee Members are required to give explanations and justifications.	1b) *tahap penglibatan yang tinggi oleh ***semua*** pegawai-pegawai dan keseluruhan proses telah dipermudahkan.* the high level of involvement by ***all*** officers and the whole process has been simplified.
2a) ****semua*** pengurusan kilang perlu prihatin dan bersama membantu polis dengan menyediakan peralatan*	2b) *kerana Malaysia-Singapura perlu mencari jalan penyelesaian terbaik bagi meleraikan ***semua*** isu-isu yang*

keselamatan yang berkesan serta memastikan jaga benar-benar bertugas. ***all*** management of factories have to be observant and assist the police by preparing effective safety equipment as well as making sure that guards carry out their duties.	*pernah merenggangkan hubungan dua hala kedua-dua negara.* because Malaysia-Singapore have to find the best solution to settle ***all*** issues, which have separated the two way relationship of the two countries.
3a) *kerajaan bersetuju menggunakan angka Bank Negara yang memasukkan ***semua*** kategori jangka pendek, sederhana dan panjang.* The government has agreed to use Bank Negara's figure to insert ***all*** short, medium and long categories.	3b) *Tulislah surat atau hantarlah telegram kepada ***semua*** Raja-Raja Melayu* Write letters or send telegram to ***all*** the Malay Kings.

From the data above, there exists a pattern of specific and non-specific meaning. The sample sentences (1a), (2a) and (3a) involve non-specific numerical, which are recognized as grammatical. Meanwhile, the sentences (1b), (2b) and (3b) involve non-specific numerical, which are considered as violating the Malay grammatical formula. This view can be revised by stating that there exist two non-specifics numerical in the Malay language. One is the specified non-specific numerical. Here, the noun that follows the numerical is not reduplicated. The noun itself is a singular noun. The notion of specified here shows that the singular noun used has a clear reference and can be supplied from the sentence context discussed. Here, BCR is applicable. The evidence, *territory chiefs, village heads, Village Safety and Development Committee Members* (1a), *factories management* (2a), *short, medium and long categories* (3a) have shown specific information to be delivered. These given evidences with the additional premises that behave as a reference assignment will help the hearer to interpret the message conveyed. Thus, JKKK (1a), safety equipment (2a) and Bank Negara (3a) are the additional premises that are part of the context used to explicate the meaning and its function.

When compared to the unspecified non-specific numerical, which uses reduplicated nouns; all examples do not give a specific reference but give general attributes instead. We can still ask, 'where is the officer from?' (1b), 'which issue?' (2b) and 'which Malay king? (3b)' There exists a clear statement with general attributes.

Hence, the division can be made by categorizing any specific non-specified numerical using a singular noun, while unspecific non-specific numericals use reduplicated nouns.

***Kebanyakan* (Most)**

4a) ****kebanyakan**** *pembiayaan seperti Bai Murabahah, Bai Muajjal dan Bai Istisna adalah berdasarkan kaedah kadar faedah tetap.* ***most*** expenditure such as Bai Murabahah, Bai Muajjal and Bai Istisna are based on the fixed interest rate method .	4b) *Jarak berselangan di antara ****kebanyakan**** huruf-huruf adalah sama.* The space in between ***most*** letters are the same.
5a) ****kebanyakan**** *iklim dan keadaan cuaca dunia mula kembali normal.* ***most*** climate and the world's weather condition has gone back to normal.)	5b) ****Kebanyakan**** *corak-corak hiasan pada hasil-hasil kraf tersebut disusun berdasarkan kepada struktur formal.* ***Most*** design patterns of the craft products are arranged based on the formal structure.
6a) *tidak semua kes kemerosotan mutu kredit berkaitan kegawatan kewangan serantau dan ****kebanyakan**** penurunan taraf dilihat RAM sebagai sebahagian daripada perubahan profil risiko kredit korporat yang timbul* not all cases of the fall of credit value is related to the foreign financial crisis and ***most*** status decrease is looked upon by RAM as part of the changes in the corporate credit risk profile.	6b) ****Kebanyakan**** *pokok-pokok balak ditebang dengan begitu hebat.* ***Most*** timber trees are widely chopped.

The same case can be observed from the examples above. (4a), (5a) and (6a) involve predicates, which are extremely specific and they can be predicted based on the context of the sentence itself. (4a) is about the financing systems of a bank, (5a) is about climate and weather conditions and (6a) is on credit quality (financial system). Relevance Theory has stated that reference assignment is one of the important sub-tasks to determine the phrase to be understood by listeners. Since the reference clearly explains the object, the singular noun that follows the numerical can explain the specific matter being discussed. Please note that both non-specific numericals carry the meaning of plurality. The difference is that one is specified and the other is unspecified. On the

other hand, in reference to (4b), (5b) and (6b), the three examples show general objects in expressions such as '*huruf-huruf'* (letters)*, 'corak-corak'* (designs) and *'pokok-pokok'* (trees). The evidences and the reference assignments that linked the numerical and the nouns are too general. Listeners have to find additional information to determine the exact interpretation. According to Relevance Theory, any expression with general attributes carry meaning that does not have to be seriously paid attention to. Listeners will stop from processing certain information when it reaches its optimum. Here, the speaker might assume that there is nothing serious to be discussed, only to convey non-specific meanings.

***Sebilangan* (A few/Some)**

7a) *"Saya kesal dengan sikap ***sebilangan*** orang Pas kerana selepas kita memberi bantuan, mereka pula menimbulkan isu* "I regret the attitudes of ***some*** of the Pas members because after we gave them assistance, they brought up issues."	7b) *Namun demikian keuntungan daripada perdagangan komoditi ini mencakupi untuk menampung ***sebilangan*** besar pedagang-pedagang desa.* Nonetheless, the profit from this commodity trading is enough to accommodate ***most*** of the rural traders.
8a) *Hakikatnya inilah aktiviti ***sebilangan*** remaja kota terutama di bandar utama.* In reality, this is the activity for ***most*** urban youngsters, especially those in major cities.	8b) *Suatu amalan yang berlaku dalam masyarakat kita sekarang ialah keengganan ***sebilangan*** kecil anak-anak menjaga ibu bapa mereka yang sudah lanjut umur.* One habit occurring in our society today is the unwillingness of ***some*** children to take care of their elderly parents.
9a) ****Sebilangan*** wakil media juga turut bermain congkak bersama penduduk tempatan dan menyaksikan proses membuat kuih bahulu* ***A number*** of media representatives also joined in the '*congkak*' game with the local residents and watched the process of making '*kuih bahulu*'.	9b) *Kita mungkin ada pengalaman melihat bagaimana ada ***sebilangan*** anak-anak muda yang menghabiskan malam-malam Ramadan dengan kerja sia-sia.* We may have experience how ***a number*** of youngsters spent their Ramadhan nights involved in useless activities.

Next, the numerical '*sebilangan*' also presents the same patterns. Again, '*Pas members'* can be connected to their attitudes, *'urban youngters'* with their activities and '*media representatives'* compared to local residents. This information has clear references and has specific attributes. On the other hand, '*traders', 'children'* and '*youngters'* are more general. We still need more information to do the reference assignment that can relate between the numerical and the noun itself. Since the attributes are too general, then reduplication is used to identify the specified non-specific numerical.

***Sebahagian* (A part)**

10a) ****sebahagian*** ahli Jawatankuasa Masjid Nurul Hidayah mendengar ucapan Shamsul Baharin.* ***A part*** of the Nurul Hidayah Mosque Committee members listened to Shamsul Baharin's speech.	10b) *Sebenarnya perbezaannya hanya sedikit dan ***sebahagian*** besar bid ang-bidang yang dibincangkan adalah sama.* Infact, the differences is minor and a major ***part*** of the field discussed is the same.
11a) ****sebahagian*** projek Sri Puteh di Arena Kepayang Putra yang memberikan keselesaan kepada penghuninya.* ***A part*** of the Sri Puteh project at Kepayang Putra Arena has given comfort to its residents.	11b) ****Sebahagian*** daripada peniaga-peniaga Melayu ini memang tidak bersedia u ntuk menghadapi cabaran dan saingan sengit dalam pemasaran koko.* ***A part*** of the group that comprise Malay entrepreneurs is not ready to face the challenges and the tight competition in the cocoa market.
12a) ****sebahagian*** besar penggambaran dibuat di Padang, Sumatera Barat, paling mencabar kemampuannya.* ***A big portion*** of the shooting made at Padang Sumatera Barat, really challenges its competency.	12b) *menjadi ***sebahagian*** daripada unsur-unsur sastera pascamodenisme dalam beberapa hal yang bertentangan dengan nilai-nilai Islam.* It becomes ***a part*** of the postmodernist literary elements related to a few matters, which are against Islamic values.

(10a), (11a) and (12a) also illustrate the same matter. The specificity of the references can be retrieved directly from the context of the sentence itself when placing the singular noun as part of the numerical '*sebahagian*' (a part). Meanwhile, the unspecified non-specific parts present those phrases, which are not considered to have any specific

information that needs to be given attention to by hearers. The phrases '*Mosque Committee members, Sri Puteh project'* and '*shooting made at Padang Sumatera Barat* ' are very specific attributes and have very clear references. Hence, these phrases are considered clear and obvious and extra information is not needed. Meanwhile, general statements that do not need specific attention are '*fields, Malay entrepreneurs* and *postmodernist literary elements.'* The data only describe general matters, which do not need a lot of effort to reach its relevance.

***Segala* (All)**

13a) *"***Segala*** tanggungan dan aset syarikat kewangan itu akan ditanggung oleh syarikat induk perbankan dan ini termasuk kontrak sedia ada pelanggan dengan firma kewangan terbabit," katanya .* ***"All*** the financial company's liabilities and assets will be borned by the financial holding company and this includes the customers' – contracts with the related financial firm," he/she said.	13b) *Kerja-kerja pembinaan empangan dijalankan oleh sekumpulan petani yang akan melantik seorang ketua yang bertang gungjawab untuk mengawas ***segala*** kerja-kerja dan setelah siap, penjagaannya.* The dam construction works are done by a group of farmers, who will appoint a responsible leader to supervise ***all*** works and after completion, its maintenance.
14a) *Pertama, pengakuan keesaan Tuhan serta penolakan terhadap ***segala*** macam dan bentuk kemusyrikan baik berupa patung, bintang, bulan dan matahari* First, the confession of God's oneness and the rejection towards ***all*** kinds of polytheism, be it in the form of statues, stars, moon and the sun.	14b) *Ini dijalankan beberapa kali sehinggalah ***segala*** rumput rumpai itu reput.* This is conducted a few times until ***all*the ** grass rot.
15a) *Seha berkata ***segala*** persiapan untuk AIM'98 kini berjalan lancar dengan menampilkan Harith Iskandar sebagai pengacara.* According to Seha, ***all*** preparation for AIM'98 is in order with Harith Iskandar as the host.	15b) *Enakmen ini telah menghakis ***segala*** batasan-batasan yang menghalang proses pengeluaran untuk pasaran.* This enactment has eroded ***all*** boundaries and have blocked the production process for marketing.

Again, the same pattern can be seen from the examples of numerical '*semua*'. The model is specified versus unspecified. The information of '*financial company's liabilities and assets* is connected to '*financial firm'* and this can verify the BCR analysis. Additional information which becomes the reference assignment will strengthen the information that needs to be conveyed. The same goes with '*polytheism*', which is connected to information such as '*statues', 'stars', 'moon'*, which act as antecedent to strengthen the form, which is considered as polytheist. Meanwhile, '*AIM preparation*' is connected with '*Harith Iskandar*' as the host, which is the completion for the AIM preparation itself. Hearers will automatically receive the specific information from the available context. Therefore, the singular noun, which follows the numerical, need not be reduplicated. However, it is a different matter with examples (13b), (14b) and (15b). '*Construction works, grass* and *boundaries'* are general information, which do not need attention from listeners. With this, the reduplicated noun shows the unspecified numerical characteristics.

Conclusion

The overall discussion, which touches upon the linguistic and corpus relationship, has brought new insights into the discussion of Malay numericals. The role of corpus data in presenting authentic corpus can assist linguists to look at language patterns more academically. Evidence of available patterns can provide guidelines in forming grammatical rules. In fact, more interesting is the presence of cognitive mechanism behind the existing language patterns. This happens in all languages in the world. Hence, it is not accurate if language is given prescriptive description only, which is sometimes incompatible in the usage of the language. This research has shown that sometimes prescriptive description of grammar might sometime miss the wisdom of why such rules exist in a grammar such as in the case of the reduplicated form of numericals that carry a distinctive role on its own. Through corpus data and linguistics analysis we can now explain why the non accepted forms can now be accepted as part of Malay grammar. The Malay corpus has validated it.

What is more important is that linguistic studies allow language patterns to be described scientifically. There is no right or wrong with regard to the use of a language. The role of the linguist is to explain why the language behaves as it does. Indirectly, the description and changes found in grammar will be improved upon. Restrictions to language creativity will produce a grammar that will oppose the natural growth of the language itself. When language creativity is restricted, users will feel restrained and may not want to use the language, and they will find ways and means to form their own language patterns. We certainly do not want this to happen to the Malay language.

Notes

[1] *Memper-kan* is a verb forming affix in Malay, which can combine with verbal, nominal and adjectival roots to derive transitive verbs by causation.

[2] ‘ Paling’ and ‘cantik’ are words that denote the superlative. Prescriptive grammarians claim that it is ungrammatical to have two superlatives in one sentence. It is considered redundant.

[3] The affix ‘men-‘ indicates active transitive verbs.

REFERENCES

Abdullah Hassan and Liaw Yock Fang. (1993). *Nahu Melayu Moden*. Kuala Lumpur : Fajar Bakti.

Abdul Jalil Anuar. (2000). *Pangkalan Data Korpus dan Penyelidikan Bahasa: Kajian Kes Memper*. *Persidangan Linguistik ASEAN 1*. Bangi, Universiti Kebangsaan Malaysia.

Asmah Hj. Omar. (1982). *Nahu Mutakhir Bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Hawiyah Baba. (2003). Perluasan Makna Kata Sifat Tabii dan Kata Sifat Indera Sentuh*:* Analisis Semantik Kognitif. *Kertas Projek Sarjana*. Bangi, Universiti Kebangsaan Malaysia.

Kamus Dewan. (1994). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Kempson, R. (1986). Difinite Np's And Context – Dependence. Dlm. T. Myers et. al. (pnyt). *Meaning and Interpretation.* Oxford: Basil Blackwell.

Nik Safiah et.al. (1986). *Tatabahasa Dewan Jilid* 1. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Nik Safiah et.al. (1989). *Tatabahasa Dewan Jilid 2.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

Nor Hashimah Jalaluddin. (2002). Kata Ganda Bahasa Melayu: Analisis Semantik dan Pragmatik. *Majalah Bahasa.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

Nor Hashimah Jalaluddin. (2003). Masalah Polisemi Dalam Perkamusan: Analisis Semantik Kognitif, *International Workshop of Lexicography*. Jakarta, Universiti Indonesia.

Nor Hashimah Jalaluddin, Zaharani Ahmad, Idris Aman. (2004). Penguasaan Bahasa Melayu di Kalangan Pelajar Melayu dalam Arus Globalisasi. *Jurnal Bahasa,* Jil 4 (1), 63-95.

Nor Hashimah Jalaluddin. (2006). Nahu Praktis Gantinama 'IA' dan 'IANYA' dalam Data Korpus: Analisis Rangka Rujuk Silang, dalam Zaharani Ahmad, *Aspek Nahu Praktis Bahasa Melayu*. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Pangkalan Data Korpus Berkomputer, Jabatan Bahasa. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Sperber, D & D. Wilson. (1995). *Relevance Theory: Communication and Cognition*. Oxford: Basil Blackwell.

McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Zaharani Ahmad. (2002). Perubahan Morfologi dalam Kamus Dewan Tiga Edisi. *Jurnal Bahasa* 2(3): 1-18.

Zaharani Ahmad. (2006). Pengimbuhan Memper-kan Dengan Kata Adjektif: Analisis Data Korpus Berkomputer, *Aspek Nahu Praktis Bahasa Melayu*, Bangi, Penerbit Universiti Kebangsaan Malaysia.