Compounding in Malay: A descriptive analysis

Yasir Azam

A thesis submitted for the degree of

Doctor of Philosophy in Psycholinguistics and Neurolinguistics

Department of Language and Linguistics

University of Essex

January 2016

Acknowledgements

All prises to Allah for giving me the strength and ability to complete this thesis.

I owe my deepest gratitude to my supervisor Professor Andrew Spencer for his support and guidance throughout the course of my study. I am grateful for all of his insightful discussions and valuable suggestions for this study. I thank him for having the confidence in me even in my time of doubts. He is truly a great mentor and I can only hope that I could be as good as he is someday.

I would also like to acknowledge Dr. Doug Arnold and Prof. Marian Klamer for their critical suggestions for the betterment of this study. Special thanks to Doug for being extra patient with my shortcomings during the final stages of my writing.

I am also grateful to Universiti Sains Malaysia (USM) for their financial funding and to the School of Humanities (USM) for their support throughout the years.

It is a pleasure to thank a few of my dearest friends during the course of my study. Ebrahim Safaie, Abdullah Thalji, Afiq Syahmi and Aiman Rahimi. Thank you for the wonderful memories.

I am exceptionally grateful to my parents, Azam Hussain and Che Maheran Maina, parents-in-law and siblings, who are always there for me in my time of need. Thank you for all the prayers.

Finally, I would like to thank my loving family, my wife Hadura Abu Hasan, and our children Husna and Hamzah. They have endured me during the good and the bad times. I would not have made it without their never-ending support and unconditional love. This thesis is dedicated to them.

Abstract

This study concerns Malay compounding. The aim is to have an in-depth description and analysis of the topic which will create a more comprehensive and systematic understanding of the phenomenon in the language. Various features and issues in relation to compounding are identified and explored in order to achieve this aim. Given that Malay compounds and phrases are structurally similar, the question of whether compounding is a morphological or syntactical product is first entertained. Discussion on this issue favours the understanding that compounds can be distinct objects from those of structurally identical phrasal ones in Malay language. The focus is then given on the topics of definition, components, headedness, criteria and classification of compounds as the foundations of Malay compoundhood. It is agreed that Malay can have left, right and headless compounds, with the prototypical structures of [X Y] (X)(Y) for endocentric and [X Y] (Z) for exocentric compounds. It is also agreed that the measures for Malay compounds (with degrees of suitability) are the syntactical criteria (inseparability, modification, component switching, circumfixation and reduplication), phonological criteria (stress and assimilation), and semantic criteria (compositional/lexicalised status) of compoundhood. This study also supports the classification of Malay compounds base on the relationship between their components, i.e. under the subordinative, attributive or coordinative relationships. Based on these foundations, this study is able to analyse and organize the different types of compounds available from the corpus, among others the (NN, NV, NA) nominal compounds, the (VN, VV, VA) verbal compounds, the (AN, AA) adjectival compounds and the idiomatic compounds. In general, the attributive relationship has the most common occurrence throughout the analysis, followed by the coordinative ones, and finally the extremely limited subordinative relationship. The discussions and findings of this study have definitely enhanced the overall knowledge on Malay compounding.

Table of contents

Chapter 1: Introduction

1.1 The interest of compounding

Over the years, the phenomenon of compounding has attracted many linguists and scholars alike, making it one of the most prominent topics of interest in linguistics. This has resulted in the production of a number of serious compilations of work on the subject, such as Libben and Jarema's (2006) *The representation and processing of compound words*, Lieber and Stekauer's (2009) *The Oxford handbook of compounding* and Scalise and Vogel's (2010) *Cross-disciplinary issues in compounding*, to name a few. The question is then, why is there so much interest in compounding?

To begin with, compounding has long been recognised as a universally significant method of word formation in the world's languages. Its significance is extremely prominent − to the point that some scholars consider compounding as one of the earliest (if not the first) multimorphemic word formation processes of human language (Jackendoff, 2002; Libben, 2006; Dressler, 2006). The idea of compounding evolving earlier than other methods such as derivation, for instance, is neatly summed up by Libben (2006: 2), who states that "It is hard to imagine how the derivational enterprise could have started without positing an intermediate step of grammaticalisation, whereby roots become affixes. Thus, it seems reasonable to assume that the dawn of derivation must have been preceded by the dawn of compounding". Similar to such an assumption, Jackendoff (2002, 2009) argues that the principle of concatenating items together is the basis of multimorphemic word formation, and, as the essence of compounding embodies this very principle, the phenomenon itself thus acts as fossils or remnants reflecting the earliest multimorphemic structure that still remains in human language today.

Although such ideas are certainly debatable, the importance of compounding as a fundamental morphological method across languages is evident nonetheless. Some languages even depend exclusively on compounding as their only means of multimorphemic word formation. Dressler (2006: 23) points out that, in general, "(…) if a language has inflection, it also has derivation and compounding, and if a language has derivation, it also has compounding, but not vice-versa". This statement exemplifies how compounding can be a universally powerful standalone process, distinct from other morphological processes in the world's languages.

One of the main factors as to why compounding is a universal process very much has to do with its prototypical characteristics, i.e. the basic property of compounding is to build 'new' words by recycling 'old' words through concatenation. Thus, in theory, compounding allows for any existing lexical item to combine with any other lexical item, entailing virtually unlimited sets of possible combinations in a given language. This also makes compounding a more flexible method of word formation in comparison to the other available methods. Derivational morphology, for instance, can be quite restrictive as it often does not apply freely to the members of a given category, e.g. *unhappy* but not **unsad*, and *dislike* but not **dishate*. In contrast, one can take any two lexical items and combine them rather freely to create compounds. The concatenation of *coffee* and *table*, for instance, results in the compounded entity *coffee table*. Although the relationship between the constituents of a compound can get more complicated than this, the example of *coffee table* nonetheless illustrates how easy it is to concatenate two existing 'old' words into a 'new' prototypical compound. Concatenating words together to coin novel compounds is a simple enough concept even for children as young as two and half years of age (Clark et al., 1985). In this sense, compounds are arguably more flexible and less constraining (in comparison to affixation, for instance), making compounding a universally more effective and productive method of word formation.

Indeed, being a universal phenomenon is reason enough for compounding to be of interest to many. Nonetheless, other factors have also prompted the interest in compound studies, one of which concerns the structure of compounds. Scalise and Vogel (2010) noted that compounds are unusual constructions as they are words that have a complex internal structure similar to that of syntactical structures. For example, one needs to relate the constituents in a syntactic-like construction in *coffee table* to derive the reading of 'a kind of table' instead of 'a kind of coffee', for instance. Evidence of compounding having an internal syntactic-like structure is more prevalent in examples such as *[health and welfare] fund*, and *[foreign exchange] flow*, where the former exhibits conjunction relations and the latter adjective-noun relations (Jackendoff, 2009). Synthetic compounds have an even more complicated internal structure, e.g. in *truck driver*, *truck* is the internal argument of the verb *drive* (Lieber, 1983; Spencer, 1991). Compounds can thus be argued as compositional objects, i.e. the structure of compounds can be composed and computed to a certain degree, exhibiting similar properties to that of a syntactic phrasal structure.

Nevertheless, compounds are constructions of combined words with properties of morphological cohesiveness. In other words, compounds can be argued as lexicalised forms, stored as a whole in the lexicon like any other lexical items. What this means is that some compounds like *coffee table* can be argued not only to be a decomposable item, but at the same time can also be argued to be a lexicalised item as well. A clear support for its lexicalisation is that the combination of *coffee* and *table* immediately triggers a specific reference, i.e. a table usually placed in the living room, typically used for holding beverages and other common things. Compare this to an analogical concatenation – *orange table*, for instance: given the right context, the term can mean anything from 'a table for oranges' to 'an orange coloured table', or even 'a table shaped like an orange'. Clearly, not all concatenated forms are compounds. It takes more than simple concatenation of words for a structure to be considered as a

compounded form. It is in this sense that a true compounded form can be argued as a cohesive lexicalised item stored as a whole in the lexicon.

Compounds have therefore regularly being considered as objects with dual attributes, treading between the lines of being lexical words as well as syntactical phrases. Having such a unique quality has naturally led to compounds becoming ideal tools of investigation for different issues concerning complex word processing. For instance, it has often been questioned whether complex words are the products of the morphological or the syntactical domains of grammar. In general, scholars approach this issue from one of two positions, i.e. through the lexicalist or the non-lexicalist point of view. Proponents of the lexicalist approach adhere to the idea that the mental lexicon is the centre of idiosyncratic information, as well as a generator of complex words (Spencer, 1991, 2005; Lieber and Scalise, 2006). In other words, the morphological component not only functions as a storage system, but at the same time it also processes complex words. Conversely, proponents of the non-lexicalist approach assume that all complex word formation is subsumed under syntax (Chomsky, 1995; Embick and Noyer, 2007; Harley, 2009). This is to say that the morphology only functions as a storage system, while the processing of complex words is handled by the syntax. As compounds arguably tread between the two grammatical domains of morphology and syntax, they are more often than not the prominent candidates to account for such an issue.

Similarly, psycholinguists alike have often utilised compounds as tools of investigation to assist in their attempts to understand the workings of the mind during complex word processing. One underlying assumption is that different types of complex words undergo different processing routes in the mind. Compounds once again provide a good platform to account for such an assumption. For instance, it has been observed that, in English, a compounded form with an irregular plural non-head (e.g. *mice eater*) is more acceptable than a compounded form with a regular plural non-head (e.g. *rats eater*) (Haskell et al., 2003;

Cunnings and Clahsen, 2007). Experimental studies on these kinds of compounds have shown that it takes a significantly longer time to process compounds with regular plural non-heads in comparison to compounds with irregular plural non-heads[1] (Cunnings and Clahsen, 2007). One view in interpreting such findings is to argue that some word forms undergo rule-based decomposition, as evidenced from the longer processing time (e.g. regular non-head compounds), while other word forms undergo a direct memory retrieval process, as evidenced from the shorter processing time (e.g. irregular non-head compounds) (Clahsen, 1999; Pinker 1999; Pinker and Ullman, 2002).[2] In this sense, compounds have provided a window into the workings of the mind during complex word processing by being the object of investigation and analysis in prominent linguistic issues.

Yet another interesting aspect of compounding is the embodiment of one of the core characteristics of human language, i.e. the concept of creativity. This is largely mirrored through the meaning of compound words themselves. As has been observed, compounds are able to generate relatively unrestricted interpretations with multiple meanings as long as the meaning is within logical limits (Lieber, 2004; Jackendoff, 2009). This is another unique property exclusive to compounding in comparison to other morphological processes. Taking *coffee table* again as an example, besides having its typical meaning, other possible readings of *coffee table* are valid as well, such as 'a table used specifically for placing coffee drinks only', 'a table used only for brewing coffee', 'a coffee coloured table', 'a table shaped like a

---

[1] There are other reasons why the regular are less preferred to the irregular forms (e.g. due to several constraints). Here, we are merely pointing out one of the ways in which psycholinguists have tried to prove that the dispreference is not only an intuitive justification, but it can be shown empirically, i.e. by analysing the processing time of these types of compounds.

[2] An alternative view is to is to argue that all complex word forms either undergo a retrieval directly from the lexicon (as full forms), or all complex word forms undergo rule-based decomposition (Rumelhart and McClelland, 1986; Taft, 1988; Bybee, 1995). As the argument goes, if all complex words are processed in the same way in the mind, then there should not be any major difference in the processing time of the forms above. However, this has more often than not been proven otherwise (cf. Clahsen, 1999; Pinker 1999; Pinker and Ullman, 2002; Cunnings and Clahsen, 2007, amongst others).

coffee cup or a coffee bean', or even the more dubious but still plausible interpretation of 'a table made entirely out of coffee beans'. These possibilities show that, in order to interpret the intended meaning of a given compound, one cannot simply make a straightforward semantic encoding of the components involve. Deriving the intended meaning requires a supplement of pragmatic knowledge as well. In compounding, such knowledge is usually not overtly conveyed by the association between its constituents. It requires an understanding of the aspects of meaning that the combination of *coffee* and *table* refers to, 'a table for holding beverages', instead of the other plausible but unintended meanings.

To this end, we have highlighted several points to support the rhetorical question posed at the beginning as to why there is so much interest on compounding. As a method of word formation, compounding is indeed a versatile and effective process, which translates into a universally productive phenomenon. On the other hand, the nature of compounding itself presents multiple levels of linguistic inquiry: structural complexities at morphological and syntactical levels, and language creativity at semantic and pragmatic levels. Collectively, these attributes (and others not mentioned here) have made compounding a very attractive subject matter to linguists in their efforts to better understand complex word representation and processing.

Overall, the study of compounding clearly bears a great deal of importance, and such importance is doubly true for languages like Malay, where the phenomenon is understudied. In general, the small amount of available information on Malay compounding can be inconsistent and at times misinterpreted. It will be apparent that one of the core aims of this thesis is to have a better representation of Malay compounding, thus creating a better understanding of the phenomenon in the language. In the subsequent sections of this chapter, we will first look at some facts about Malay language in general and in relation to this study; this will be followed

by the aims and objectives of the study, and this first chapter will close with the overall outline of the thesis.

1.2 About the Malay language

It is fitting that we begin with some general knowledge about the Malay language before we get into the topic of Malay compounding. The aim of this subsection is to present some historical and contemporary facts about the language, which will in turn differentiate between the two largest varieties of Malay language which are used in Malaysia and Indonesia.

The Malay language is one of the many thousand Austronesian languages, more specifically grouped into the Western branch of the Malayo-Polynesian family (Tadmor, 2009). In terms of numbers, Malay is native only to a relatively small group of speakers compared to other Austronesian languages. However, the language has managed to attain a wider importance by playing a unifying role in the multilingual region of South East Asia. A quick look at its past is sufficient to understand how such status has been achieved.

Historically, native speakers of Malay inhabit the area of the Malacca straits, home to the ancient Malaccan Empire located in the southern region of peninsula Malaysia. The success of this empire was largely due to its control over the straits, a strategic and important maritime route connecting the east and the west regions of mainland Asia. The empire dominated a great part of South East Asia for many centuries and consequently Malay emerged as the lingua franca of the region (Hassan, 1987; Clark, 2009; Omar, 2010).

The many territories of the Malaccan Empire included her immediate southern province of Johor (of modern Malaysia) and the neighbouring Riau Archipelago (of modern Indonesia). This area was later the birthplace of the Johor-Riau Empire, successor to the Malaccan Empire.

The Malay language of the Johor-Riau Empire is commonly accredited as the origin of the modern Malay of Malaysia and Indonesia (Omar, 1992b).

As a lingua franca, Malay had been in contact with, and exposed to, a variety of different languages. Thus, it comes as no surprise that many languages have had their share of influence on Malay in one way or another. Orthographic records, for instance, have shown that the earliest old Malay was written in a form of the *Pallava* script of southern India, while the more recent classical Malay was written in the *Jawi* script derived from Arabic orthography (Tadmor, 2009; Karim, 2010). Only towards the end of the nineteenth century was the writing system standardised into the *Romanised* script of today (Tadmor, 2009). Such multilingual contacts have obviously enriched the Malay lexicon throughout its history. Many foreign words have been borrowed and assimilated into the language. It is estimated that about a third of all Malay content words are of foreign origin, mainly from Sanskrit, Arabic, Chinese, Tamil, Portuguese, Dutch and, of course, English (Hassan, 1987; Karim, 2010).

Today, the Malay language is recognised as the national language of Malaysia, Indonesia and Brunei, and also as one of the four national languages of Singapore (Omar, 1992b). Collectively, these four countries comprise an estimated 250 million speakers, making Malay the most widely used language in South East Asia (Clark, 2009; Tadmor, 2009). However, although Malay is the root of the national language of each of these countries, they nonetheless differ from one another with varying degrees (Omar, 1992b). In general, the main varieties of Malay can be divided into the Indonesian and Malaysian varieties.[3]

---

[3] Omar (1992b) argues for three main varieties of the Malay language, i.e. the Malaysian, Indonesian and Bruneian varieties. The Singaporean variety can be subsumed under the Malaysian variety due to historical and socio-geographical factors between the two nations. It is also safe to say that the Brunei variety is closer to the Malaysian than it is to the Indonesian variety. Hence, the Malaysian variety can be considered as the umbrella for *Bahasa Melayu* (Malay language) as the national language of Malaysia, Singapore and Brunei. Only the Indonesian variety is distinct enough to be considered a separate main variety known as *Bahasa Indonesia* (Indonesian language). Thus, we will briefly look at the two varieties of Malay, i.e. *Bahasa Melayu* of Malaysia and *Bahasa Indonesia* of Indonesia.

In Indonesia, the national language is known as *Bahasa Indonesia*, which literally means 'Indonesian language'. *Bahasa Indonesia* is a derived form of the Malay language originating from the native Malays of Indonesia's Riau Archipelago (Omar, 1992b). The Malays are a relatively small ethnic group, comprising only about 7% of the total Indonesian population (Prentice, 1987; Tadmor, 2009). In other words, Malay is not the first or native language to most Indonesians. There are many other larger ethnic groups practising other Austronesian languages as their first and native language, such as Javanese, Sundanese, etc. Thus, the larger majority of over 90% of the population actually need to learn the derived form of Malay as their national *Bahasa Indonesia* in a second language context, i.e. mainly in school settings (Prentice, 1987; Omar, 1992b).

The Malay root of *Bahasa Indonesia* has undergone many changes in its process of becoming the national language (Omar, 1992b). *Bahasa Indonesia* is not only affected by external influences, especially by the previous Dutch coloniser, but it also receives heavy internal influences, especially from Javanese, as Javanese people make up 50% of the total Indonesian population (Omar, 1992b). Therefore, in Indonesia, there is a clear distinction between the terms *Bahasa Melayu* (Malay language) and *Bahasa Indonesia* (Indonesian language), i.e. the former belongs specifically to the Malay ethnic group, and the latter belongs to everyone else as the national language.

The scenario is different in Malaysia. The national language in Malaysia is known as *Bahasa Melayu*, which literally means 'Malay language'. As the name implies, *Bahasa Melayu* is not exactly a derived form of Malay like that of *Bahasa Indonesia*, but rather it is the standardised version of the Malay language[4] itself (Omar, 1992b). In Malaysia, the native Malays are the largest ethnic group, comprising at least 50% of the total population, making

---

[4] The Malay language of Malaysia is commonly accepted as the standardised version of the Johor Malay language, originating from the Johor-Riau Empire (Omar, 1992b).

the Malay language the first and native language to a very large community (Hassan, 1987; Prentice, 1987). Thus, in terms of learning *Bahasa Melayu* as the national language, it is the minority population who have to learn it as a second language, while the Malay majority are learning the standardised version of their first and native language in a formal context. The distinction between the terms *Bahasa Melayu* (Malay language) and *Bahasa Malaysia*[5] (Malaysian language) can be unclear in Malaysia as both terms are used interchangeably when referring to the national language.

Obviously, *Bahasa Melayu* is not exempted from the influences of other languages. However, as *Bahasa Melayu* is dominant not only in terms of its user majority but also in terms of its constitutional status, the language is constantly being monitored and regulated accordingly to any influences from other languages. *Bahasa Melayu* can thus be said to be more rigid by maintaining a more purist approach in its development as the national language (Hassan, 1987). In this sense, the neighbouring *Bahasa Indonesia* is arguably more accommodating to changes in comparison to *Bahasa Melayu*. Such different attitudes adopted towards the development of *Bahasa Melayu* and *Bahasa Indonesia*, along with the different socio-political histories of the two countries, has distinguished the languages far enough for them to be considered as two distinct language varieties (Omar, 1992b).

Nevertheless, the fact remains that Malay is the root of both varieties, and thus it is often assumed that both *Bahasa Melayu* (BM) and *Bahasa Indonesia* (BI) are the same languages or at least mutually intelligible. This assumption is true, but only to a certain degree. The most distinctive difference between the two varieties concerns their vocabulary (Omar, 1992b). For instance, there are plenty of different words with the same meaning, such as *kedai*

---

[5] The term *Bahasa Malaysia* (Malaysian language) has been substituted by *Bahasa Melayu* several times since its establishment as the national language. This is largely due to socio-political issues, as it is felt that the term *Bahasa Malaysia* would better represent the multiracial ethnicity of the country (Omar, 1992b). Nevertheless, the two terms are used interchangeably when referring to the one national language of Malaysia.

/kədai/ (BM) vs. *toko* /toko/ (BI) for 'shop', and *lobak* /lobak/ (BM) vs. *wortel* /wortel/ (BI) for 'carrot'. The differences are not similar, for instance, to that of the American English 'couch' and British English 'sofa' analogy, where both speakers are aware of the words, only to differ in usage preference. The Indonesian word *wortel* (which is of Dutch origins), is simply foreign to speakers of *Bahasa Melayu*. There are also plenty of similar words with totally different meanings, such as *jeruk* /dʒəruk/ for 'pickle' in BM, but 'orange' in BI, *budak* /budak/ for 'boy' in BM, but 'slave' in BI, and *baja* /badʒa/ for 'fertiliser' in BM, but 'steel' in BI. There are many differences in terms of word formation as well. The word 'drugs', for example, is a single register in BM, i.e. *dadah* /dadah/, while in BI it is an acronym, *narkoba* /narkoba/, for *narkotika dan obat-obatan terlarang* (narcotics and illegal drugs). Similarly, the opposite can be true in BM as well. The word 'committee', for example, is a single translated register in BI *komite* /komiti/, but a combined form in BM, i.e. *jawatankuasa* /dʒawatan kuasa/ from *jawatan* 'post' + *kuasa* 'authority'.

The differences are also observed in other contexts as well. In the educational context, for example, grammar books concerning both languages differentiate the two varieties, e.g. Mintz (1994) *A student's grammar to Malay and Indonesian* clearly acknowledges the difference between the two varieties by illustrating two different glosses for every English example presented. Similarly in a real-world context, for example, Google translator also has two different translation choices, distinguishing Malay from Indonesian. Even in the entertainment context, Indonesian dramas or movies aired in Malaysia are provided with BM subtitles. All these examples illustrate that the two varieties are not exactly as mutually intelligible as they may appear to be, i.e. they do hinder the free flow of communication between their speakers (Omar, 1992b).

As the main focus of this thesis concerns compound words, and as the divergence between *Bahasa Melayu* and *Bahasa Indonesia* is most prominent at word level, it is essential

that we highlight that the two varieties are different. Thus, this thesis is interested in the compounding phenomenon in *Bahasa Melayu* use in Malaysia. We will not take into account compounds in *Bahasa Indonesia*. As such, it will be noted from this point onwards that all references to the use of 'Malay (language)' refer specifically to the Malays and *Bahasa Melayu* of Malaysia.

1.3 Standard Malay in Malaysia

Malay was declared as the national language of the federation of Malaya when the country gained its independence from the British Empire on the 31st of August 1957 (Hassan, 1987; Omar, 2010). The need for and importance of a single national language was driven by several political and nationalistic factors. Among them was to create a new identity away from the previous colonising image, and also to create unity between the multiracial ethnicities in the country (Hassan, 1987). Selecting Malay as the only national language was met with a degree of hostility since other ethnic groups wanted to uphold their native languages as well. Nevertheless, several factors prompted the decision. Among them was the fact that the Malays were – and still are – the largest in terms of ethnic group population, and they also had the majority control over the political scenario at the time (Hassan, 1987; Omar, 2010). However, the more significant factor for choosing Malay as the national language lies in the fact that the language had been established and utilised for many centuries in the country not only by the Malays, but also by all other races, and as such Malay was deemed to be deserving of the status (Hassan, 1987).

Since independence, there have been many efforts to regulate and mould the language to make it worthy of national status. One of the earliest concerns was to establish a standardised version from the different varieties of Malay used in the country. Omar (1993) divides the

language into two main varieties, i.e. the royal and the non-royal varieties. Each variety can be further subcategorised into the non-standard and standard versions respectively. The royal variety is basically practised within limited environments and events, i.e. within royal settings. The focus is then on the non-royal variety, which is the main variety used elsewhere.

As stated, the non-royal variety can also be subcategorised into the non-standard and standard versions. The non-standard non-royal variety is considered as colloquial Malay. The colloquial variations in peninsula Malaysia can be roughly divided into several main regional dialects, namely the north-west, north-east, eastern and southern dialects (Omar; 1991, 1993). Among them, the southern dialect[6] (more specifically the state of Johor dialect) has been commonly considered as the standard non-royal variety, and is thus recognised as the root of standard Malay for the national language (Omar; 1988, 1993, 2010).

As Omar (1988, 1992a) points out, the general difference between these regional dialects has more to do with phonological variation, rather than morphological or syntactical differences. Omar (1988, 1991, 1992a) categorises the phonological variation as the a-variety and the schwa-variety. The former reflects the northern dialect, while the latter reflects the southern regional style of pronunciation. The most significant difference is in the way the vowel /a/ and the consonant /r/ are realised in word final positions. In the a-variety, both /a/ and /r/ are realised as in *apa* /apa/ 'what', and *kotor* /kotor/ 'dirty', but in the schwa-variety, /a/ is realised as a schwa, and the /r/ is dropped in word final position, i.e. /apə/ and /koto/ (Omar; 1988, 1992a). The two varieties have co-existed for a very long time, and thus the a-variety has

---

[6] The southern dialect became the convenient choice based on several historical reasons. According to Omar (1992a), the media in the early days was transmitted into peninsula Malaysia via Singapore, at which point the Johor dialect was the regional dialect used in Singapore. Similarly, the first broadcasting station in Malaysia was based in Kuala Lumpur, and since Kuala Lumpur falls within the southern jurisdiction, parallel southern dialect was used in the broadcasts. Accordingly, the southern dialect became the preferable variety to be upheld as the basis of standard Malay during the establishment of Malaysia (Omar; 1992a, 1993).

been assimilated with the schwa variety as part of the effort to have a better representation of standardised Malay (Omar; 1988, 1992a).

Issues related to the morphology of the language have also been one of the main concerns in the standardisation efforts. Hassan (1987) points out that, in the early days, standardisation efforts were grounded by a strong purist approach, and this attitude is still prevalent, especially on matters concerning foreign influence on Malay. In affixation, for instance, Malay has a relatively small set of native affixes, and this may cause some difficulties in matching foreign affixes to the language. Thus, English prefixes such as *in-*, *im-*, *un-*, *an-*, *ab-*, *ir-*, *dis-* and *non-*, for example, can be problematic as these prefixes have only brought a single transfer in Malay (Hassan; 1987, 1997). These prefixes are mainly represented by the Malay negating word *tidak* 'no/not', e.g. *insane* is *tidak waras* 'not sane', *imperfect* is *tidak sempurna* 'not perfect' and *irregular* is *tidak sekata* 'not uniform'. This sort of limited matching can have an impact on the development of the language to a certain degree.

The purist approach is also prevalent in word translation and loanword adaptation. New words or terminology must first and foremost be translated into existing native Malay words before anything else. For example, an English word such as *technique* will be translated into the native word *cara* /tʃa.ra/ or *kaedah* /ka.e.dah/ (lit. way, method, manner, etc.). Similarly, single words might need to be translated into a compounded form, e.g. *latitude* becomes *garis lintang* (lit. line + horizontal), so that a native-like terminology can be maintained. Exceptions are allowed when there are no other suitable native words available to replace the donor word. Even so, the new loanword must conform to the phonotactics of Malay. This involves adjusting the donor word to fit Malay native sounds, permissible phonemic sequences and syllabic structures. Native Malay is not equipped with certain phonemes, consonant clusters are not permissible and words are preferably of disyllabic structure in general (Hassan, 1974, 1997; Maris, 1980). Such characteristics are mirrored in loanword adaptations. For example, the /v/

phoneme is not native to the language, thus in a word such as *governor* /gʌvənə/, the /v/ sound is substituted by the closest native phoneme, namely the /b/ sound, to become *gabenor* /gabənor/ upon adaptation. Another common method of loanword adjustment to fit Malay phonotactics is through epenthesis. Monosyllabic structures usually undergo epenthesis upon adaptation to create the more preferred structure in the language, i.e. the disyllabic structure. For example, *book* /bʊk/ undergoes vowel epenthesis to become *buku* /bu.ku/, while other words like *norm* /nɔːm/ undergo both consonant and vowel epenthesis to become *norma* /nor.ma/. Epenthesis is also common to break up consonant clusters, e.g. *glass* /glɑːs/ becomes *gelas* /gə.las/ and *film* /fɪlm/ becomes *filem* /fi.ləm/.

Nonetheless, in comparison to such loanword adaptation, priority is still given to native word substitution or translation, as mentioned earlier. Having said that, language planners have learnt to become more flexible and open to this issue over the years. This is partly due to the fact that substitution or translation can be problematic as it affects the donor word's accuracy of meaning. This problem is prevalent especially with scientific terminologies. In today's context, it can be preferable to adapt words by phonotactics modification; thus, previous examples such as *technique* /tɛkniːk/ can be preferably adapted into *teknik* /tek.nik/, and *latitude* /lætɪtjuːd/ is adapted to *latitud* /la.ti.tud/ in Malay. This sort of adaptation not only complies with the phonotactics of the language, but, more importantly, it retains the original meaning of the donor words rather than them being translated into less accurate meanings.

In line with this flexibility, the phonological structure of the language has also undergone several general alterations to accommodate further morphological demands (Omar, 1988; Hassan, 1997). As has been pointed out above, there are several phonological attributes that are not native to the language. However, in order to accommodate the influx of new words, more non-native qualities are now being accepted. Modern Malay has adopted several borrowed phonemes, accepted consonant clusters and become more lenient to multiple syllabic

structures. For instance, the adapted word *struktur* /struk.tur/ (from *structure*) has an initial consonant cluster of CCCVC.CVC, and the word *eksperimen* /eks.pe.ri.men/ (from *experiment*) has four syllables, VCC.CV.CV.CVC (Hassan, 1997). To some linguists, these adjustments to the language, i.e. adapting foreign phonological and morphological structures, are seen as a form of language corruption, as it is believed that Malay should strive to preserve its native qualities in one way or another. Others like Hassan (1987), however, see such an attitude as a healthy approach instead, arguing for a balance between preservation to hinder corruption, and at the same time being flexible enough to ensure development and sustainability of the language.

This subsection started with a brief historical account of how Malay came to be recognised as the national language of Malaysia. The focus then turned to the efforts of standardising the language, i.e. some phonological and morphological regularisations were presented. Obviously, there are other issues involved in regulating the language which have not been mentioned here. Nevertheless, the ones that have been presented give us a glimpse of how such efforts have contributed to the development of the standard language in one way or another. It is in this same spirit that we view the study of compounding in this thesis, as one of the many ongoing efforts that will contribute to the further betterment and development of standard Malay in Malaysia.

In this thesis, all references to compounding relate to the grammar of standard Malay. The term standard grammar here refers to both prescribed and described grammar as written in the literature by reputable Malay scholars over the years. The term standard Malay can thus be defined as the Malay used as the standard medium in formal and official contexts; as the language of governance and administration, political discourse, medium of education in schools and universities, modern literature, mass media and communication at the national

level (Hassan, 1987; Tadmor, 2009; Omar, 2010). Thus, the study of the compounding phenomenon in this thesis is of compound forms used in standard Malay as defined.

1.4 Why Malay compounding?

This section justifies some of the main reasons why the study of Malay compounding is worthwhile. We will consider this study as having a twofold beneficial outcome, i.e. first for the language itself, and second for the larger body of compounding knowledge.

With regard to Malay morphology, it has long been claimed that, besides affixation and reduplication, compounding is the language's other main method of word production. However, if one was to explore the literature on this topic, it would be quite difficult to find a comprehensive account of Malay compounding. Somehow, there seems to be less emphasis on detailing the phenomenon in comparison to the other major word processes. This, of course, does not do justice to the topic when it is commonly claimed to be an important part of the language. This current study is beneficial in the sense that it is an attempt to put the neglected topic of compounding back into the spotlight of Malay morphology.

In the literature, it is common for the topic of compounding to be presented as one of the subtopics within a larger work on Malay morphology. Presented in this manner, description of the topic often lacks proper explanation and elaboration, which can have a negative impact. Such an impact can be seen in ill-defined matters concerning compounding definition, feature properties and compounding categorisation, etc. The shortcomings of such have affected the literature, stopping it from having a unified conception of Malay compounding. In addition, it does not help that different scholars also have different takes on what compounding actually is, e.g. what constitutes a compound, what compounds should look like, etc. This has led to

fuzzy outcomes of what really is a compound in Malay: misinterpreting other constructions for compounds, and missing constructions that actually are compounds. To this end, this study will contribute by recognising the problematic issues that have caused these outcomes, and will try to resolve them by coming up with a better-organised conception of Malay compounding. Having an organised conception of the topic will benefit not only the understanding of Malay compounding as a whole, but will also benefit other branches of Malay linguistic studies as well. A sound concept of Malay compounds, for instance, can be utilised as a tool in experimental studies such as bilingualism, psycholinguistics, translational and computational linguistic studies, just to name a few.

The second beneficial outcome of this study is hopefully a contribution to the larger body of compounding knowledge. Although a growing number of studies are being conducted on different languages, the fact remains that most studies on compounding are based on only a few languages, predominantly European languages. In this sense, theories and models of compounding are based on rather limited language data. It is good to have other varieties of data on this topic from other languages, and this is where this current study can fill this gap. As a typologically distinct language, Malay can contribute to the insufficient compounding data from different languages. In this sense, this study hopes to contribute by enriching the larger knowledge of compounding phenomenon.

To this end, the objectives of this study are as follows:

(i) To identify the pertinent issues in relation to the concept of Malay compounding.

(ii) To establish an organised approach to account for Malay compounding.

(iii) To recognise all the possible constructs that can qualify as compounds in Malay.

Accordingly, the research questions for this study are as follows:

(i) What are the issues related to the concept of Malay compounding?

(ii) What is an organised approach to account for Malay compounding?

(iii) What are the different types of compounds in Malay?

To sum it up, the aim of this thesis is to have an in-depth analysis of the compounding phenomenon in the Malay language. By doing so, it is hoped that we will have a more comprehensive and systematic understanding of the topic, which will in turn provide us with some beneficial outcomes. The outcome of this study will not only contribute to the gap in Malay literature on compounding, but it is hoped that it will also contribute to the knowledge of compounding phenomenon at a larger scale.

1.5 Overview of the thesis

The outline of the thesis is as follows:

Chapter 2: Theoretical and typological overview of compounding

The first part of this chapter looks at the issue of complex word formation, i.e. where exactly processes such as compounding take place in the grammar. We opt for Ackema and Neeleman's (2004) theory which argues for both morphological and syntactical modules as the generative systems for complex word formation. The second part of this chapter attempts to provide an overview of the universals of compounding. The literature comes from studies of the representation and processing of compounds across various languages. The main aim here is to draw an outline of compoundhood in terms of its universal characteristics. Issues of definition, components, headedness, criteria and classification of compounding are discussed

in order to create the outline. This outline is then used as the guideline to analyse the compounding phenomenon in the Malay language in the subsequent chapters.

Chapter 3: Malay morphological processes

This chapter presents the main morphological processes in Malay, namely affixation, reduplication and compounding. The aim here is to provide a general overview of the major morphological processes in relation to Malay compounding.

Chapter 4: Studies of Malay compounding

Central to this chapter is a review of several previous studies on Malay compounding. The studies are Hassan (1974, 1986, 2006), Musa (1993), Karim (1995), Karim et al., (2008), Sew (2007), and Ismail and Jalaluddin (2008). These studies provide further information on how Malay compounds have been understood and represented in the literature over the years. The reviewed information from this chapter (along with the other chapters) is used to assist the analysis in the subsequent chapter.

Chapter 5: Analysis of Malay compounds

This chapter consists of three subsections. The first part revisits the issues of complex word formation, applying the principles of the competition model to account for the compounding phenomenon in the Malay language. The second part focuses on the universal characteristics of Malay compounding. Issues of definition, headedness, criteria and classification of Malay compounds are discussed. Based on the discussion on these issues (along with the other collective discussions from the previous chapters), a checklist of Malay compoundhood is drawn up. The final part of the chapter utilises the compoundhood checklist to analyse the compiled corpus of structures exemplified as compounds in the works of Hassan (1974), Musa (1993), Karim (1995), Sew (2007) and Karim et al. (2008). The analysis identifies and

classifies the compounds available according to their respective types beginning with nominal, verbal, adjectival and idiomatic compounding.

Chapter 6: Summary

An overall summary of this study is presented in this final chapter.

Chapter 2: Theoretical and typological analysis of compounding

In the first part of this chapter, we begin by questioning where in the grammar does complex word formation such as compounding take place, i.e. is it processed in an independent morphological module, or is it processed in a more dominant syntactical module within the grammar. We attempt to answer this question by reviewing some of the theoretical developments concerning the issue. By the end of this first part, we will establish how compounding in general can be accounted for in the grammar through Ackema and Neeleman's (2004) competition model.

The second part of this chapter provides an overview of the universals of compounding. The main aim here is to draw an outline of compoundhood in terms of its universal characteristics. Issues of definition, components, headedness, criteria and classification of compounding are discussed in order to create the outline. This outline will then be used as the guideline to analyse the compounding phenomenon in the Malay language in the subsequent chapters. It should be mentioned here that the literature in this chapter mostly comes from cross-language studies of compounding. The fact remains that the body of knowledge on this topic mainly comes from the many studies of European languages. The intention of reviewing such literature is to create a platform or framework that will be adapted to the analysis of Malay compounding.

2.1 Issues of complex word formation

One of the central issues in the study of complex word formation is the question of where in the grammar does word formation actually take place, i.e. where do complex words

fit in the grammar? The basic recurrent views in relation to this question are: they are handled by the morphological module, or they are handled by the syntactical module of the grammar. Proponents of the former defend word formation as a process that is governed by principles that are independent or distinct from syntactic principles; thus, morphology is a separate module to syntax. Conversely, proponents of the latter defend word formation as a process that can be explained by syntactic principles; thus, morphology is a submodule of syntax.

The view that morphology should be an independent module has long been contested. During structuralism's prime, the study of morphology was indeed a distinct branch of linguistics, where it could be separated into the study of *morphotactics* and *allomorphy*[7] (Anderson, 1988; Katamba and Stonham, 2006). However, when generative grammar became popular, morphology was pushed aside. Anderson (1988: 147) points out that early generative grammars do away with morphotactics by assigning "(…) the arrangement of all items into larger constructions to the syntax (…)" and simplified allomorphy to the "(…) listing of arbitrary suppletions" by reducing "(…) all variation in shape of unitary linguistic elements to a common base form (…)". "With nothing of substance left to do in morphology, generative linguists had to be either phonologists or syntacticians" (p. 147). Thus, the process of word formation (morphology) in early generative grammar was handled through phrase structure and transformational rules in the syntactical module, while word form variations (allomorphy) were handled in the phonological module (Lees, 1960; Chomsky, 1965; Chomsky and Halle, 1968).

Some of the classical examples of how word formation is handled in the syntax include Lees (1960) deriving the noun *appointment* from the verb *appoint* from the sentence (a) The committee *appoints* John, to (b) The committee's *appointment* of John (p. 67), or the noun *priesthood* from the noun *priest* via (a) John is a *priest*, to (b) John's *priesthood* (p. 110).

---

[7] From a structuralist view, *morphotactics* is the arrangement of morphological elements into larger structures, while *allomorphy* is the variations in the shape of the 'same' unit (Anderson, 1988: 147).

Chomsky (1965: 184) on nominalisation explains how *destroy* and *refuse* "will be entered in the lexicon with a feature specification that determines the phonetic form they will assume (by later phonological rules) when they appear in nominalised sentences", i.e. via transformation, [nom + *destroy*] produces *destruction* and [nom + *refuse*] produces *refusal*. Similarly, Chomsky and Halle (1968) proposed that regular and irregular past tenses are handled in the syntactic module with the feature 'past', (e.g. [*mend* + past] and [*sing* + past]), and through readjustment rules they are converted into *mended* and *sang* in the phonological module.

With regard to compounding, Lees (1960), for instance, discusses nominal compounds as constructions generated from underlying deep structures via transformational methods as well. In other words, compounds are considered as sentences reduced by transformation. The surface structure of the compound *windmill*, for example, is derived from the deep structure of *wind powers the mill*, while the compound *flour mill* is derived from *mill grinds flour*. However, there are problems with such an account of compounding. *Windmill* and *flour mill* are two examples of compounds with similar noun-noun surface structure, but the semantic representations between the constituents of the two compounds are not the same. The relationship between the constituents of the former is that of the wind powering the mill, while the latter is the mill grinding the flour. It will take too many transformational processes in order to properly account for the abstract and complex relationship between the constituents of a given compound (Chomsky, 1970). Therefore, explaining compounds (and other morphological processes) through a syntactic mechanism as such can be problematic and obviously has had its share of criticism. Nevertheless, the point here is to illustrate how complex word formations were deemed describable via syntactic principles (even if this is complicated or insufficient) without the need for a morphological account. This has led to the understanding that a separate morphological module was unnecessary in early generative grammar.

Yet another common argument for a single dominant syntactic module is reasoned by the observation of overlapping features that are present in both morphological and syntactical accounts of complex structures. As the argument goes, if both modules share parallel features and operations to generate complex constructions (e.g. the merger of constituents seen in compounds has very similar principles to that of phrases), it will then be redundant and uneconomical to have a separate morphological module when the syntactic module alone can be held accountable. However, showing that morphological processes have the same properties as syntactic constructs does not mean that morphology should be subsumed under syntax. It could very well just mean that morphological processes have the same principles and use the same operations as syntax in generating new structures. Chomsky (1970) in 'Remarks on Nominalization' suggested a separate lexical account of morphologically complex words, as Aronoff (1976: 6) reiterates "(…) all derivational morphology is isolated and removed from the syntax; it is instead dealt with in an expended lexicon, by a separate component of the grammar". Since then, Chomsky's 'remarks'[8] have often being credited as the catalyst for the so-called lexicalist approach.

The lexicalist approach has been generally categorised into two forms, namely the *Strong Lexicalist Hypothesis* and the *Weak Lexicalist Hypothesis*. The former approach cuts off syntax from all morphological phenomena, i.e. word formation (including inflection) is fully realised in the lexicon. The latter approach has some leniency to it as it considers

---

[8] Marantz (1997) argued that the assumption made by most linguists from Chomsky's 'remarks' on how there should be a separate lexical property is not entirely correct: as he puts it, "Chomsky proposes no special 'lexical rules' or special lexical structure meaning correspondences in his 'Remarks'" (pp. 215).

derivations (including compounds) as operations of the lexicon,[9] while inflections[10] are generated by the syntax (Spencer, 1991; Scalise and Guevara, 2005). The foundation of lexicalism is the Lexical Integrity Principle (LIP). This principle asserts two main generalisations, namely: (i) words are built by different structural elements and by different principles of composition than syntactic structures, and (ii) words are seen as atoms of syntactic structures which are invisible to syntax, and for that reason syntax is unable to access and operate on the internal structure of words (Di Sciullo and Williams, 1987; Bresnan and Mchombo, 1995; Spencer, 2005; Scalise and Guevara, 2005; Lieber and Scalise, 2006).

The LIP has been challenged through several arguments, among others from the observation of coordination in derivation and compounded forms. For instance, in English, some bound prefixes can be coordinated with each other while others cannot (e.g. *pro-* as opposed to *anti-*war, vs. **in-* or *ex-*port), and coordination within synthetic compounds is also possible (e.g. a [truck driver], vs. a [[car and truck] driver] (Spencer, 2005; Lieber and Scalise, 2006). Thus, the concept of coordination (which is more synonymous with syntactic structure) has been argued as evidence against the LIP (i.e. words should not be operating on the same principles as syntactic structures). Similarly, another piece of evidence for the violation of the

---

[9] From the perspective of lexicalism, the lexicon is not only a repository of entities, but it is also a generative component of the grammar where morphological processes can take place. A number of different models have been proposed to account for the workings of the lexicon, for example, the single mechanism and dual mechanism models of processing, among others. These models basically argue for the difference in how word forms are stored and retrieved from the lexicon, i.e. direct retrieval as full forms from the lexicon, or some sort of rule-based decomposition by rules. For more details on this, refer to Rumelhart and McClelland (1986), Taft (1988), Bybee (1995), Clahsen (1999), Pinker (1999), Pinker and Ullma (2002), Haskell et al. (2003), Cunnings and Clahsen (2007), and Berent and Pinker (2007).

[10] Inflections are commonly associated with the syntax module, while derivations are not. Although most derivational processes involve syntactic category change, and the syntactic category of a word is arguably associated with syntax, the syntax nonetheless does not differentiate whether a word is of its simple form (e.g. chauffeur) or of its complex derived one (e.g. driver); both forms are simply lexemes (nouns) to the syntax, and it is in this sense that derivation can be argued to be syntactically irrelevant (Plag, 2003). On the other hand, inflectional word-forms are normally determined or dependent on the syntactic environment in which they occur; they realise morphosyntactic features of words. Thus, inflections have often been associated as relevant to the syntax (Anderson, 1992). Indeed, there are other properties to distinguish between lexeme producing derivation and word-form producing inflection; the point here is to illustrate the regular understanding of derivation as morphology based, while inflection as syntax based.

LIP comes from the concept of recursion. Recursion has been considered as one of the distinguishing properties of syntactical structure (e.g. sentence embedding), and yet recursion can be argued as a morphological property as well. For instance, Pinker (1994) points out that the longest word in English (i.e. *floccinaucinihilipilification* – the action of estimating something as worthless)[11] could never really retain its record as the longest word due to the concept of recursion. One can always make a word longer by applying recursion (Pinker, 1994), e.g. *floccinaucinihilipilificational*, *floccinaucinihilipilificationalize*, *floccinaucinihilipilificationalization*, etc. Recursion can also be easily applied to compounds, e.g. *film society*, *student film society*, *student film society committee*, etc. (Spencer, 1991). Thus, in theory, morphological recursion has no limits and, in this sense, recursion does not seem to be exclusive only to syntax. Bauer (1983) points out that the recursiveness of syntactical structures is in fact limited due to several factors (e.g. computational and memory limitations). Consider the sentences below (Bauer, 1983: 67):

(1)   (a) This is the malt that the rat that the cat that the dog that the cow tossed worried caught ate.

        Compared to a morphological structure such as:

        (b) His great-great-great-great-great-great-great (…) -grandfather was killed in a Viking raid on Holy Island.

Although they are difficult to process, the recursion in both examples is nonetheless grammatical. Thus, both morphology and syntax can be argued to 'share' the same principles (e.g. coordination and recursivity) in their own limited manner. If this is the case, it then seems

---

[11] Definition from Oxford English Dictionary online (https://en.oxforddictionaries.com).

to pose a degree of challenge to LIP, since one of the generalisations is that words are built by different principles than syntactic structures.

The No Phrase Constraint (NPC) principle was introduced to complement the LIP. The main premise of the NPC principle is to disallow syntactic phrases from entering the root compound (Botha, 1983).[12] However, evidence against LIP and NPC (in particular) is the phenomenon of phrasal compounding. Phrasal structure occurring in the non-head position of compounds such as a '*pipe and slipper* husband' and a '*slept all day* look' violates one of the core ideas of the LIP, that syntax is not able to access and operate on the internal structure of words (Lieber, 1992). Nonetheless, this argument about phrases inside compounds has been countered by claims that the phrasal part is in some way lexicalised and not a freely formed phrase when it appears in the compound (Bresnan and Mchombo, 1995; Spencer, 2005). Others, such as Ackema and Neeleman (2004), propose that LIP will still be satisfactory if the phrases that appear inside compounds are treated as atoms instead. This is because they treat insertion as a process that works both ways between the morphological and syntactical module. In other words, morphological units inserted into syntactic structure are treated as syntactic atoms, while the syntactic units inserted into morphological structure are treated as morphological atoms. In this sense, they maintain that LIP is not violated in phrasal compounds.

Arguments against the LIP and NPC have indeed questioned the validity of the lexicalist approach, but, like any other theory, some arguments for lexicalism are simply not foolproof. However, such flaws certainly do not imply that the approach should just be dismissed. As pointed out, explaining and proving that certain word formations are accountable via syntactic principles does not mean that morphology holds a syntactic status and that it needs

---

[12] Nonetheless, this does not mean that phrases and compounds cannot share certain principles.

to be subsumed under syntax. There are many other reasons to believe in morphological autonomy. Some constructions are better off explained morphologically and some properties are arguably particular only to words. Among the particularity of words are: words are less regular and limited in productivity, words take on idiosyncrasies, and words can have paradigms (Katamba and Stonham, 2006; Scalise and Guevara, 2005). These sorts of properties are generally considered as exclusive to morphology, not to syntax. Similarly, some syntactic operations are not attested in morphological structures. Syntactic operations such as conjunction, movement, topicalisation and pronominal reference, among others, are considered exclusive to syntax and not to morphology (Ten Hacken, 1994; Spencer, 2005). Such operations are commonly taken as syntactic diagnostics for lexical status as they do not affect parts of words. For instance, it is not possible to apply the syntactic operation of coreferencing between a pronoun and a nominal expression inside a word, e.g. *John is a* [*taxi$_i$ driver*], **[It$_i$] is yellow in colour.* The pronoun 'it' cannot be used to refer to the noun 'taxi', which signifies that the 'taxi driver' is a unified word form, as evidenced by the fact that syntactic operations cannot access the internal structure of words. The existence of such particular operations between morphology and syntax renders the need for a division between the modules. Spencer (2003: 236-237) puts it adequately by saying that it is better to "(…) assume that morphology is at least partly autonomous and to investigate the principles that might be unique to it". What is meant here is that, if morphology is subsumed under syntax, any principles that are distinctive to morphology might be overlooked. Thus, it is safer to assume a separation between the two modules. If, however, it is discovered that both morphology and syntax are indeed the same module, the separated studies can be put together and nothing will be lost (Spencer, 2003).

In this spirit, most scholars today generally agree that, rather than having a superior syntactic module that can account for both morphological and syntactical processes,[13] it is better to understand morphology and syntax as two independent modules that have some sort of interaction between them, i.e. morphology sees syntax and syntax sees morphology (Lieber, 1992; Jackendoff, 1997, 2002; Ackema and Neeleman, 2004; Lieber and Scalise, 2006). Among others, Jackendoff's (1997, 2002) model of grammar illustrates this core view by asserting the idea of each grammatical module being exclusive to one another, but at the same time connected to each other. In Jackendoff's (1997, 2002) point of view, it is misleading to assume a single superior module governing the rest of the modules in the grammar (i.e. it is inaccurate to consider phonological, morphological and/or semantical structures as merely interpretive components of syntactic structure). This argument can be illustrated by presenting some mismatches between the modules. For instance, it can be observed that syntactic structures are not strictly equivalent to their phonological structure:

(2)   (a) Syntax: [a [[big] house]], [a [[[very] big] house]]

          (b) Phonology: [a big] [house], [a very] [big] [house][14]

The sentence above exemplifies that phonological structure does not merely follow syntactic structure, i.e. the phonological bracketing is not equivalent to the syntactic one. They each have their own principles. It can thus be argued that there is an absence of a one-to-one connection between the modules.

---

[13] We do not intend to review models or frameworks that assume superiority of the syntactic module governing both syntactical and morphological processing (such as proponents of the *Minimalist Program*, *Constructionism* and *Distributed Morphology*, to name a few). The interested reader can refer more to works on such an approach, e.g. Lees (1960), Halle and Marantz (1994), Chomsky (1995), Halle (1997), Harley and Noyer (1999), Embick and Noyer (2007), and Harley (2009) and others alike.

[14] Example in (2) is taken from Jackendoff (1997: 26).

Similarly, there can also be mismatches between syntax and semantics. The infamous sentence by Chomsky (1957), *colorless green ideas sleep furiously*, illustrates how structures can be syntactically but not semantically correct. Furthermore, syntax and semantics do not share the same units and principles. For instance, syntax runs on units such as noun and verb phrases with head-complement principles, whereas semantics runs on units such as predicates, events and quantifiers with predicate-argument principles (Jackendoff, 2002). Again, this illustrates that grammatical modules have their own principles.

Similarly, mismatches are also evident at word level. The occurrence of the so-called bracketing paradoxes (typically exemplified by the word *unhappier*), exemplifies the mismatching of phonology and semantics (Pesetsky, 1985; Spencer, 1988, 1991). In terms of phonology, the comparative suffix *-er* only attaches to one- or two-syllabic stems, e.g. *bigger*, *tinier*, but not **beautifuller*. Thus, in terms of phonology, the suffix *-er* cannot attach to the stem *unhappy* as the stem is three syllables long. Instead, it needs to merge with *happy* first before the final attachment of the prefix *un-* can take place, i.e. [un [happier]]. However, in terms of semantics, the adjective *happy* must attach to the prefix *un-* first before it can be attached to the suffix *-er* (i.e. [[unhappy] er]) to deliver the intended meaning of 'more unhappy' (and not the meaning of 'not more happy' as derived from the first structure of [un [happier]]). Mismatches then seem to indicate that grammatical modules are independent of each other at some level of degree, each with their own principles. Thus, although the thought of a single module like syntax to govern phonology, morphology and semantics is indeed an attractive idea, it is nonetheless an inaccurate assumption, as shown.

In line with Jackendoff's (1997, 2002) point of view, scholars such as Ackema and Neeleman (2004) argue for a model of grammar that has interface systems with mapping principles between the grammatical modules. In other words, grammatical modules are exclusive to one another but at the same time they are able to interact with each other. Ackema

and Neeleman's (2004) grammar assumes three main macro-modules, namely semantics, syntax and phonology. Each macro-module has two generative subsystems, one for generating phrasal structures and the other for generating word structures. A diagram of the grammatical system is shown below.

(3) Ackema and Neeleman's (2004: 4) diagram of grammatical system

SEMANTICS	SYNTAX	PHONOLOGY
Phrasal Semantics ↓ Phrasal semantic structure INSERTION Word semantic structure ↑ Word Semantics	Phrasal Syntax ↓ Phrasal syntactic structure INSERTION Word syntactic Structure ↑ Word Syntax	Phrasal Phonology ↓ Phrasal phonological structure INSERTION Word Phonological Structure ↑ Word Phonology
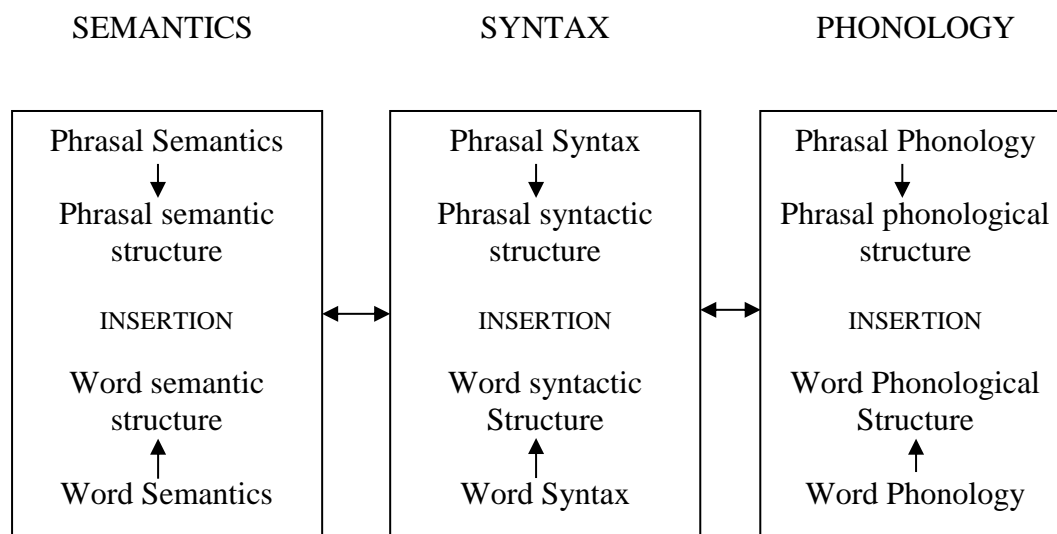
The diagram above shows how the word and the phrasal subsystems of semantics, syntax and phonology have their own principles of combination within each respective macro-module. The two subsystems of word structure and phrase structure within the (big) syntax macro-module, are commonly labelled as 'morphology' and 'syntax' in the general literature. Since both morphology and syntax are within the same macro-module, they are interrelated and can interact with each other.

Ackema and Neeleman's (2004) model supports the assumption that morphology and syntax are two independent generative systems. In this sense, complex structures can therefore be merged in either the morphological or syntactical subsystems. Hence, given a language such as English, an item α can either merge with an item β in the morphological system, to form a

word [β α], or in the syntactical system, to form a phrase [α βP].[15] An abstract illustration of the merger is shown below:

(4)   (a) Syntax merger   (b) Morphology merger

$$\alpha P$$
$$\alpha \qquad \beta P$$
$$\text{(head)} \qquad \vert$$
$$\beta$$

(dependent)

$$\alpha$$
$$\beta \qquad \alpha$$
$$\text{(dependent)} \quad \text{(head)}$$
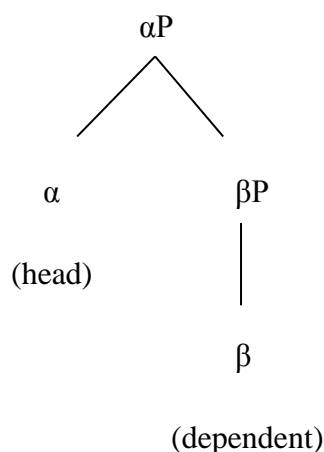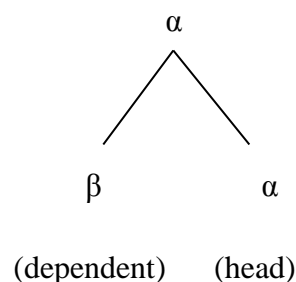
The abstract above illustrates the two possibilities of merging items in English to create complex structures, i.e. a syntax merger (hence producing a syntactic phrase) or a morphological merger (hence producing a morphological word).[16] The question then is, in which system will the merger actually take place?

According to Ackema and Neeleman (2004), since both morphology and syntax are capable of generating complex structures, they are therefore in competition with each other to do so (hence the name *competition model*). Ackema and Neeleman (2004: 51) provide the conditions for competition as below:

(5)   Let α_1 and α_2 be syntactic representations headed by α. α_1 blocks α_2 iff

       (i)   In α_1 (a projection of) α is merged with (a projection of) β in syntax, while in

---

[15] In the English headedness system, the morphological head is on the right, while the syntactic head is on the left. Assuming α is the head item to be merged, it will thus be placed on the right for word formation (i.e. [β α]), and it will be placed on the left for phrasal formation (i.e. [α βP]).

[16] We can exemplify this illustration using the words *shelf* [α] and *book* [β], for instance. A morphological merger produces [N *book* N *shelf*] N, while a syntactical merger produces [N *shelf* [PP *for books*]]NP.

α ₂ (a projection of) α is merged with (a projection of) β in morphology, and

(ii)   the semantic relation between α and β is identical in α ₁ and α ₂

The above conditions project that competition will occur if the merging of α and β involves the same category and has the same semantic relation between them in both syntactical and morphological structure. If all of the conditions in (5) are observed as being equal, competition takes place and the preference for the merger will be the unmarked structure, i.e. the unmarked blocks the marked structure from occurring. The unmarked structure (of either syntax or morphology merger) is language-specific. In a language such as English, the syntactic merger is more preferable than the morphological one.[17] A concrete exemplification of how the competition works is as follows. Say, for instance, an item [α] *drive* is to combine with an item [β] *truck* to form a complex structure. The two possible outcomes according to the model will hence be:

(6)   (a) Syntactic merger   or   (b) Morphological merger

---

[17] Syntactic structure can be considered as the unmarked form in the English language. We can once again refer to the comparative formation to support this argument. The most common way of forming comparatives is to morphologically attach the suffix -*er* or to syntactically attach modifying words (such as *more*) to adjectives and adverbs. As previously mentioned, the suffix -*er* attaches only to monosyllabic or disyllabic stems (e.g. *taller*, *prettier*, but not *beautifuler*). In contrast, the modifier *more* is, in principle, not limited in the type of syllabic stems to which it is allowed to be attached (e.g. *more fun, more tangled, more beautiful*). In this sense, we can argue that comparative formation via morphological construction is a constrained process (marked), while comparative formation via syntactic modifier is a more general one (unmarked). Thus, by extension, syntactic constructions in the English language can be assumed as the unmarked structure of complex construction.

Looking at the example above, we will notice firstly that both syntax and morphology structure involve the merger of items with the same categories (i.e. noun and verb), and, secondly, the semantic relation of the merged items is also the same (i.e. the noun in each structure is being interpreted as the object of the verb). In this sense, both structures in (6) epitomise the condition of competition in (5), and thus both structures can be considered as being in a state of competition with each other. The outcome or the winner of the competition in this case is therefore the syntactic structure (as evident from the grammaticality of the structure 'they *drive trucks'*, as apposed to the non-existent morphological structure 'they **truck drive'* in the English language).

Having said this, when will the non-preferable structure (i.e. the morphological structure in the case of English) be possible, then? Ackema and Neeleman (2004) argue that morphological structure can occur when there is no syntactic competitor. In other words, the conditions that will allow for a morphological merger are the opposite of the conditions in (5), i.e. the elements that merge in the syntax structure are different in terms of their category and semantic relations to the elements that merge in the morphology structure. Let us look further into this by exemplifying the syntactic structure of '*driver of trucks'* vs. its morphological counterpart '*truck driver'* below.

(7) Syntactic merger of the phrase '*driver of trucks'*.
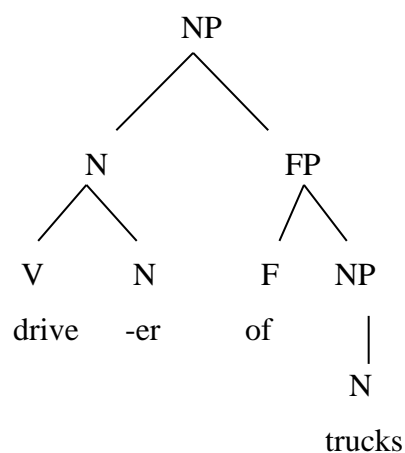
In the above, the combination of the V *drive* and nominalising suffix *-er* generates the head N *driver*, which then combines with its argument N *truck*, hence the phrase in (7).[18] This phrase, *driver of trucks*, has a morphological counterpart, *truck driver*, e.g.:

(8) Morphological merger of the root compound[19] '*truck driver*'.
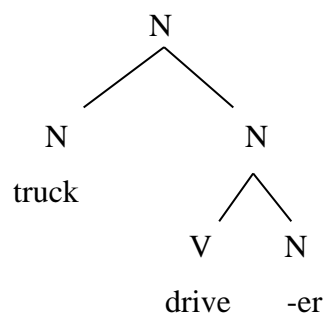
Here, the N *truck* combines with the deverbal N *driver* (generated from the V *drive* and suffix *-er* derivation) to form the compound in (8) *truck driver*. Although the combination of the elements is of the same category in both (7) and (8) (i.e. N+N combination), the elements here in (8) are not involved in argument linking like that of the elements in the syntactic counterpart in (7). This is to say that *truck* is the internal argument of the N *drive* in (7), but it is not in (8). Unlike in (7), the V *drive* in (8) cannot assign its internal argument to the N *truck*, because it is unable to transfer its argument structure to the dominating node which is now being a different category due to the category-changing suffixation of *-er*. There is no argument linking taking place in structure (8) and *truck driver* here can be analysed as a root/primary compound. In this sense, the semantics in the syntactic structure of (7) *driver of trucks* differs from the

---

[18] Functional projection such as 'of' is not counted in Ackema and Neeleman's (2004) analysis.

[19] Basically, a prototypical root/primary compounding is the combination of two (or more) words, consisting of a head (modified) and non-head (modifier) relation. In general, endocentric root compounds usually have a fairly transparent meaning derivable from their constituents and the semantic relation between them, e.g. *book* 'modifier' + *case* 'modified' = *bookcase*. However, there are many others with a less transparent semantic relationship between the head and its modifier, e.g. a *butterfly net* is used to catch butterflies, but a *mosquito net* is used to deter mosquitoes. More on root/primary compounds will appear later.

morphological structure of (8) *truck driver*, i.e. in (7) the noun *truck* is the argument of *drive*, but in (8) it is a modifier. This means that there is no competition going on and thus the morphological structure in (8) is not blocked by its syntactic structure counterpart in (7); thus, both structures can coexist. The compound *truck driver* here is interpreted as a root compound with unpredictable semantics[20] such as 'a driver owning a truck' or 'a driver wearing a shirt with a picture of a truck on it'.[21]

On the other hand, there is also another morphological structure for *truck driver*, i.e.:

(9) Morphological merger of the synthetic compound[22] '*truck driver*'.

```
              N
             / \
            V   N
           / \  -er
          N   V
        truck drive
```

---
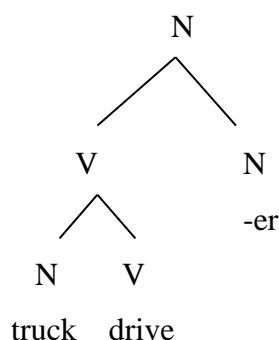
[20] As Ackema and Neeleman (2004: 52) put it, "Morphological merger of α and β may result in a semantics that cannot be expressed by the result of syntactic merger of the two".

[21] This line of argument mirrors Lieber's (1983) analysis of such compounded structure. She argued that, if *truck driver* is to be analysed as $[[truck]_N [[drive]_V -er]_N]_N$ like in (8), it can only have the interpretation of a root/primary compound of non-argument-taking elements. This is because the verb is contained within a noun (*-er* nominalises the verb into a deverbal noun), hence the verb's argument structure is blocked, and the V *drive* cannot assign its internal argument to the N *truck*. Therefore, the N *truck* is not incorporated as an argument of V *drive*, i.e. no argument linking takes place in this structure. Since the relation between the elements here is 'free' (i.e. N+N), Lieber's (1983) analysis thus acknowledges a multiple reading of *truck driver* in (8) with meanings such as 'driver owning a truck' or 'driver wearing a shirt with a truck'. In this sense, the characteristics of the compound here are similar to root compounding as they can have multiple meanings/readings given the right context.

[22] Synthetic compounds are different in several ways from root/primary compounds. The three most fundamental properties of synthetic compounds are: (i) a complex head adjective or noun, which is derived from a verb, (ii) a non-head constituent, which is interpreted as a syntactic argument of the deverbal noun or adjective head, and (iii) the meaning of the compound is transparent (Spencer, 1991; Katamba and Stonham, 2006). For more on synthetic compounding, refer to: Roeper and Siegel (1978), Selkirk (1982), Lieber (1983), Spencer (1991) and Katamba and Stonham (2006), among others.

Here, the N *truck* combines with the V *drive* to generate the form *truck drive*, which subsequently combines with the nominalising suffix *-er* to derive the compound in (9) *truck driver*. Although English does not have verbal compounds with incorporated objects such as '*to *truck drive*', this NV combination is nonetheless crucial in this construction here in order to derive a synthetic compound reading. As noted, one of the main properties of synthetic compounds is having a non-head constituent, which is interpreted as a syntactic argument of the deverbal head (cf. footnote 23). In order for the compound *truck driver* to have this criterion, it is necessary that the N *truck* is interpreted as the internal argument of the V *drive*. To do so, the NV merger must first take place before combining with the nominalising *-er*. Only by this can the V *drive* directly assign its argument to the N *truck* (as apposes to the structure in (8) where the nominalising *-er* attaches first, hence blocking the verb from assigning its internal argument to the noun). The compound *truck driver* in (9) can thus be analysed as a synthetic compound with the reading of 'someone who drives trucks'.[23]

Although both the morphological merger in (9) and its syntactic counterpart in (7) have the same meaning (i.e. the N *truck* is the argument of V *drive* to mean '*someone who drives trucks*'), the two structures are nonetheless not in competition. This is because the morphological merger in (9) has different merging categories from its syntactic counterpart in (7), i.e. in (9) the categories that merged in (9) are V (*truck drive*) + N (*-er*), while the categories that merged in (7) are N (*driver*) + N (*of trucks*). This means that the condition of competition is not fulfilled,[24] and thus the morphological structure in (9) is not blocked by its syntactic counterpart in (7). It can thus be said that the morphological structure of *truck driver* is able to

---

[23] Lieber (1983) also favours the structure in (9), i.e. [[[truck]N [drive]V]V *-er*]N, if a structure such as *truck driver* is to be analysed as a synthetic compound. Lieber (1983) similarly argues that the NV base (*truck drive*) is important as it allows the argument structure of the V *drive* to percolate to the highest V node, by which the N *truck* is then able to be assigned as the argument of the V *drive*. Only in this sense can the compound *truck driver* have the synthetic compound reading of 'someone who drives trucks'.

[24] Refer to the conditions in (5).

co-exist with its syntactic structure, one as a synthetic compound in (9), and the other as a root compound in (8).

To this end, the competition model of grammar as proposed by Ackema and Neeleman (2004) seems to offer an appealing approach to account for the workings of complex word formation, especially concerning the phenomenon of compounding. However, as we have only seen how this model can work for English compounding, we cannot therefore say the same for other languages such as Malay (for the time being). For that reason, this model will be taken up again with regard to Malay compounding later, in Chapter 5. We leave it here for now and move on to the subsequent topic of the universals of compounding.

2.2 The universals of compounding

The concern of this section is to give an overview on issues of defining and classifying the concept of 'compoundhood'. In order to do so, this section is divided into several parts addressing the topics of definition, headedness, criteria and classification of compounding. A discussion on these topics will assist in our attempts to draw an outline of compoundhood. The literature on these topics sources different studies on compounding from different languages. It should be mentioned here that the literature on Malay compounding will not be incorporated into this section. Similarly, Malay will not be used to illustrate the points and issues raised in this section. A dedicated discussion on Malay compoundhood is presented in Chapter 5. The discussion here is intended to create a universal outline of the compounding phenomenon which will be used as a tool or guideline in the analysis of Malay compounding. In this sense, it will be seen later whether the issues and topics discussed here have any bearing on the discussion of Malay compounding that will follow later, in Chapter 5. We now begin this section with the first topic concerning the definition of compounding.

2.2.1 Definition of compoundhood

As briefly mentioned in the introductory chapter, there has always been a substantial amount of discussion and debate with regard to the definition of compounding. Compounding may seem somewhat of a straightforward phenomenon, and thus the task of defining it should not present much of a problem. However, this is simply not the case, as it has been rather difficult to come up with a satisfactory and universally agreed upon definition of compoundhood.

Many scholars will point out the fact that compounds have been given either a 'too strict' or a 'too loose' definition. One of the main premises in defining compounding is to identify the foundational components that function as its building blocks. In fact, this issue has long been one of the main reasons why the attempts at a compounding definition have been complicated. Bauer (2006: 719) states that "(…) the forms in which the individual subwords appear may be differently defined in different languages: a citation form in one, a stem in another, a specific compounding form in yet a third, a word form in a fourth", pointing out problems that can occur when identifying the components of compounds. The point here is that linguists have been using different linguistic units as the building blocks of compounding, varying from *words* to *lexemes*, along with *roots*, *bases* and *stems,* influenced not only by personal preference or perhaps an adhered to framework, but also due to the particular language under consideration. Adopting the presentation style of Scalise and Vogel (2010), let us begin by illustrating some examples from the literature on how these different terms have been utilised to define compounds.

(10) The different types of linguistic units used to define compounds[25]

Words

"In simple cases, compounding consists of the combination of two **words**, in which one word modifies the meaning of the other, the head".

(Booij, 2012: 77)

Lexemes

"A compound is a complex **lexeme** that can be thought of as consisting of two or more base **lexemes**. In the simplest case, a compound consists of two **lexemes** that are joined together".

(Haspelmath and Sims, 2010: 137)

Roots

"Compounding is a word-formation process that combines two or more **roots** inside the same word".   (Fabregas and Scalise, 2012: 111)

Bases

"A compound word contains at least two **bases** which are both words, or at any rate, root morphemes".   (Katamba 1993: 54)

Stems

"When two (or more) elements which could potentially be used as **stems** are combined to form another stem, the form is said to be a compound".   (Bauer, 1983: 28)

---

[25] The emphasis in bold is our addition to the quoted authors' words.

It is apparent from the quotations above how the definition of compounds varies in the literature, as different linguists use different units, causing irregularities in definition. As mentioned, different linguists have different preferences for the units of compounding, and, although some may have justified their reasons for using particular units, equally, others may not have done so. Nonetheless, this sort of inconsistency has allowed some room for a well-founded disagreement, as the need to pin down the exact components of compounding is an important issue. One of the main reasons is that (a strict) adherence to a specific linguistic unit may (heavily) influence the outcome of a given form, i.e. whether a form is considered as a compound or otherwise.

Bauer (1983), for instance, points out that, if the components of compounds are defined as made up only of lexemes, forms such as *fishmonger* or *warmonger* cannot be interpreted as a compound. The reason for this is that only the first element is a lexeme; the second element, *monger*, is not an independent lexeme. Compare this to *war supporter*, where *supporter* is an independent lexeme. Thus, in his view, a construction as such can only be a compound if stems are defined as their compounding components. Bauer (1983) also points to neo-classical compounds, forms like *Anglophobe*, where both *anglo-* and *-phobe* are not lexemes nor are they stems (best classified as bound roots). Thus, defining compounds as being exclusively build of lexemes or stems will exclude *Anglophobe* as a compound form. Similarly, Plag (2003) also touches on the issue of considering different components as units of compounding. He points out that forms such as *astrophysics* have a bound root component (*astro-*), forms like *parks commissioner* have a grammatical word component (*parks*), and forms such as *pipe and slipper husband* have a whole syntactic phrase component, and yet they are all commonly accepted as compounds. The point is, it is essential to recognise the units involved in compounding, as exclusive adherence to one over the other could result in the acceptance or rejection of a given form as a compound or otherwise.

It is no surprise then for some linguists to utilise multiple linguistic units to define compounds; among others, this includes Lieber's (2010: 43) definition that "compounds are words that are composed of two (or more) bases, roots, or stems" and Plag's (2003: 135) definition of a compound as "(…) a word that consist of two elements, the first of which is either a root, a word or a phrase, the second of which is either a root or a word". Recognising the importance of identifying the units of compounding, what follows then is to ensure that these units, i.e. *words*, *lexemes*, *roots*, *bases*, *stems* and *phrases*, are properly defined themselves.

2.2.2 Components of compounding

Let us begin with the first two commonly used components of *word* and *lexeme*. The concept of 'word' is arguably one of the most basic concepts in linguistics. In general, virtually anyone, even without any proper linguistic background, will have some sort of intuition of what the term 'word' means (Aronoff and Fudeman, 2005). In this sense, the term has universal value, which makes it an excellent linguistic component for morphological approaches (Dressler, 2006; Montermini, 2010). Nonetheless, it is also well known for being highly problematic in that it can be difficult to precisely define what exactly is this concept known as 'word' (Bauer, 2000; Booij, 2012). The notion is relative in at least the senses of: (i) what is a word in morphology is not necessarily identical to a word in the sense of phonology or syntax, and (ii) what is defined as a word in one language may be insufficient or incompatible in the definition of another language (Bauer, 1983; Spencer, 2006; Lieber and Stekauer, 2009; Montermini, 2010).

Matthews (1974) divided 'word' into three main senses, namely the *lexeme,* the *word-form* and the *grammatical word*, and these sorts of division have been commonly accepted and

adopted by many (Lyons, 1968; Bauer, 1988; Coates, 1999; Bauer, 2000; Haspelmath and Sims, 2010, among others). The term *lexeme* is given to define the word in its abstract sense, the term *word-form* is used to define the word in its concrete sense, while the *grammatical word* refers to the representation of the word associated with morphosyntactic properties (Katamba and Stonham, 2006; Booij, 2012). For instance, the noun lexeme TABLE and the verb lexeme WALK respectively realise different word-forms (i.e. *table*, *tables* and *walk*, *walks*, *walking*, *walked*), and these word-forms represent the grammatical words of their lexemes (e.g. *tables* [noun, plural], *walks* [verb, third person singular, present] etc.). Grammatical words are not new lexemes; they are word-forms that represent the appropriate forms or particular grammatical context of a given lexeme, governed by the rules of inflection.[26]

On the other hand, take, for example, a word-form such as *walker*. This word-form is not to be confused as one of the (inflectional) word-forms of WALK. *Walker* does not express the grammatical function of the verb lexeme WALK; instead *walker* [noun, singular] is the word-form of the separate noun lexeme WALKER – an altogether new lexeme (and concept) which refers to the device used to support babies or disabled people while walking. It is obvious that the lexeme WALKER is related to the lexeme WALK, but nonetheless lexemes are abstract entities, thus "(…) strictly speaking, one lexeme cannot be derived from another […] form *a* (e.g. reader) corresponding to lexeme *A* (READER), is derived from form *b* (read), relating to lexeme *B* (READ)" (Haspelmath and Sims, 2010: 18). Such phrasing can be somewhat inelegant, and, as the basic function of the derivational process is to create new lexemes, the relation between them is thus commonly simplified as *derived from* or *derived lexeme* for ease

---

[26] Inflection is traditionally regarded as a change in the grammatical or morphosyntactic form of a word (or lexeme) as opposed to derivation, which is the formation of a new lexeme from another lexeme (Spencer, 1991: 193), or, as Aronoff and Fudeman (2005) put it, inflection expresses morphosyntactic information, syntactic information that is expressed morphologically.

of explanation. The simple core lexeme WALK and the complex new lexeme WALKER are two different but related lexemes (Haspelmath and Sims, 2010; Booij, 2012). The importance of the lexeme is apparent: lexemes can represent the core primary units of morphology, in turn allowing them to function as the base units of morphological processes such as compounding.

As mentioned earlier, the term 'word' is a cover term for *lexeme*, *word-form* and *grammatical word*, and the function of these subdivisions is basically to further specify the concept of 'word' (Matthews, 1974). In this sense too, lexemes are also defined as having the three basic properties of phonological representation, semantic representation and morphosyntactic information (Amiot, 2005; Montermini, 2010; Booij, 2012). It is important to have such specifications, since there are times when it is necessary to use them specifically. However, linguists more often than not use the terms 'word' and 'lexeme' synonymously. The general sense of the word concept is commonly associated with the concept of lexeme, and thus words are usually understood and to be understood as meaning lexemes (cf. Bauer, 2006; Montermini 2010; Booij, 2012, among others). This principal view of word and lexeme is assumed to be the intended compounding element in the compounding definitions above.

Moving along, we now look at the remaining units of *roots*, *bases* and *stems* as the components of compounding. Generally, these three components are associated together even though they can be argued as distinct units. This is evident through their traditional definitions: (i) a root is the smallest irreducible core of morphological form (related to the word/lexeme) with nothing else attached to it, (ii) a base is the unit or form to which affixes can be added, and (iii) a stem is the part of the base before inflectional affixes are added to it (Coates, 1999; Lieber, 2010; Radford et al., 2009; Aronoff, 2012). A quick example from the word *printer*: the root is *print*, which is also the base for *printer*, and *printer* is the base for *printers*; the base *printer* is also known as the stem for *printers*.

Montermini (2010) points out that one of the main reasons why roots are commonly used is partly to ensure that components smaller than words/lexemes are taken into account in compounding, echoing Bauer's and Plag's points mentioned at the beginning of the section. Compounds with either one or both components that are not independent lexemes include the so-called neo-classical compounds, e.g. *pathology*, *psychopath*, *dermatitis*, etc. This type of compound where the elements are of Latin or Greek origin can be found in English and other languages. The status of these forms as compounds may be influenced by certain circumstances, as experience, knowledge of languages and the mental lexicon are not the same for everyone (Katamba and Stonham, 2006; Booij, 2012). For instance, a person with some understanding of Greek can recognise *dermatitis* as a compound form: Greek *derma* 'skin' and *itis* 'disease'. Alternatively, the components involved may also be understood by comparison. Thus, if someone understands that *-logy* basically means 'study', by comparison other forms ending with *-ology*, like *morphology*, *biology*, etc., would be understood as 'the study of' a particular subject. Linguists have generally recognised the elements in neo-classical compounds as bound roots. Some divide them further into initial (*bio-*, *psycho-*, etc.) and final (*-logy*, *-graphy*, etc.) bound roots and label them as combining forms (Booij, 2012, among others). Although these elements do combine with lexemes, e.g. *tele*-camera, magnet-*ometry* and music-*ology*, which may make them appear to be like forms of affixes, they nonetheless are more commonly combined with other bound roots (Bauer, 1983; Carstairs-McCarthy, 2002; Lieber, 2010). As Bauer (1983) and Booij (2012) point out, if these elements are taken as affixes, it would mean that words could be made solely out of affixes without any base, e.g. neither *bio-* or *-logy* are the base of *biology* (at least in English). Furthermore, these elements differ from affixes as they are more restricted in usage, e.g. the prefix *un-* is easily attached to create negative meanings, but *psycho-* would require more specific attachments. They also can occur in multiple positions, e.g. *path-* can precede the *-ology* to create *pathology*, and also come

after *psycho-* to give *psychopath*, which is not a common characteristic of affixes (Katamba and Stonham, 2006: Lieber, 2010). Having all these properties distinguishes these elements from affixation, making them better treated as (bound) root compounding. Therefore, if neo-classical compounding[27] is to be considered as compounds (in a language like English), non-affixal bound roots have to be considered as one of the compounding units of the language (Booij, 2012).

The idea of stems as the building blocks of compounding is insignificant for languages with a poor or without an inflectional system. English is an example of a language with a relatively weak inflectional system where only a few grammatical distinctions are marked; thus, stems are arguably insignificant to the language (Lieber, 2010). Similarly, the concept of stems does not have an effect on compounding in a language like Malay, which is entirely without an inflectional system. However, the role of stems is very important in inflectional languages. Take Italian, for example: *gatt-o* 'cat'/*gatt-i* 'cats' and *macchin-a* 'machine'/*macchin-e* 'machines' illustrate how, besides the plural, the singular forms must also be inflected, which means that both *gatt-* and *macchin-* are stems (Haspelmath and Sims, 2010). The importance of stems is also apparent in compounding for inflectional languages. Certain types of compounds have their internal components as existing lexemes, but appear differently when in compounded forms. Greek compounds, for instance, have their first constituent as stems. These stems are existing lexemes with different inflectional endings when used independently, but appear with a suitable linking element when in compounded form, e.g. *kukl-a* 'doll' and *spito* 'house' become *kukl-o-spito* 'doll house', *psom-i* 'bread' and *tir-i* 'cheese' become *psom-o-tiri* 'bread (and) cheese', and *sime-a* 'flag' and *stolizm-os* 'decoration' become

---

[27] Alternatively, Montermini (2010: 82) argued that, although these forms are special since they differ from the native compounds, they should nonetheless be treated similarly to normal compounds as "(…) they can be considered as particular (forms of) lexeme bearing some sort of [+bound] feature in their lexical representation". Thus, in Montermini's point of view, neo-classical compounds are a combination of lexemes, not of roots.

sime-o-stolizmos 'flag decoration' (Ralli, 1992; Ralli, 2010). These examples illustrate how stems have a more significant role in some languages, which is why they have always been considered as one of the components of compounding.

The final term that has been used as a unit of compounding is *phrase*. This type of phrasal compound is common in Germanic languages. Recall how the insertion of phrases inside compounds has been supported or argued against based on the *Lexical Integrity Principle* along with *No Phrase Constraint*. As mentioned, some argue that the phrases allowed in compounds are lexicalised phrases such as idioms, clichés and quotations, as Carstairs-McCarthy (2002; 82) puts it: "(…) lexically listed phrases (i.e. idioms) or institutionalised ones (i.e. clichés) can appear in some contexts where unlisted phrases cannot". Examples include [*fresh air*] *fanatic*, [*open door*] *policy* and [*sexually transmitted disease*] *clinic* (Carstairs-McCarthy, 2002). Nonetheless, non-lexicalised phrases do occur in the non-head position of compounds (at least in English). Some examples include *a* [*pain-in-stomach*] *gesture*, *a* [*pipe and slipper*] *husband* and *a* [*slept all day*] *look*, etc. (Lieber, 1992; Bauer, 1983). However, Booij (2012: 81) points out that "the possibility for phrases to appear within compounds does not mean that all kinds of phrases are allowed in this position", e.g. *a* [*French history*] *teacher* is acceptable, but *a* [*the French history*] *teacher* is not. This can be taken as some support for both the lexical integrity and phrase constraint.

To this end, we have identified several morphological units (i.e. *words*, *lexemes*, *roots*, *bases*, *stems* and *phrases*) that are commonly used as the elements of compounding. We have also looked at how and why the units are being used within a particular language. Clearly, it is necessary for a language to utilise different units of compounding to accommodate the peculiarities of that particular language. It is thus easy to see why it can be difficult to come up with a single universal definition of compoundhood to represent all languages. Guevara and Scalise (2009: 107) point out the fact that there are "many fundamental notions in linguistics

that are ill-defined but, nevertheless, constantly used in the literature in an intuitive way (sentence, phrase, word, etc.), and compounding may very well be one of them". Therefore, in their opinion, one should not look too much into this 'chaos' but instead begin looking for more general aspects of the world's languages. With that in mind, Guevara and Scalise (2009) propose the basic prototypical compounding schema of [X *R* Y] Z structure.[28] This schema assumes that X and Y are constituents representing one of the major lexical categories, and the *R* links X and Y by a grammatical or semantical relationship. The overall category of the compound is represented by Z, which may be the same as the category of X or Y or a different category altogether. The outcome of this schema produces:

(11)   (i) Right-headed compounds [X *R* Y] Y

         (ii) Left-headed compounds   [X *R* Y] X

         (iii) Exocentric compounds   [X *R* Y] Z

Some prototypical features of compounding must be attached to this schema for completeness: that compounds observe syntactic atomicity and lexical integrity (Scalise and Vogel, 2010). This proposal allows for a more universal definition and representation of compound structure; as Guevara and Scalise (2009) claim, this schema delimits 'canonical' instances of compounding in the world's languages. With that in mind, we now leave this discussion on components of compounding, to which we will return later in the analysis of Malay

---

[28] Booij (2012: 210) also sketched a similar schema for Germanic language in general as [XY]Y, where Y = N, A, V. The leftmost component (X) is the modifier, the rightmost component (Y) is the head, and the syntactic category of the whole is similar to the head (which also holds other features such as gender and inflectional locus). (X) can be content word, function word and even phrases. The non-head and the head have a relationship referred to as *R* with the general meaning of 'has some relation to'. The nature of this relation is of world knowledge (encyclopaedic knowledge) and also on the context in use, especially of new compounds, while established compounds have a fixed meaning.

compounding in Chapter 5, and move on to the subsequent topic within this discussion of the universals of compounding, i.e. the headedness in compounds.

2.2.3 Headedness in compounding

The notion of 'head' has been one of the central concerns in the studies of complex word formation. The literature on headedness in the morphological domain regularly highlights the fact that the notion originated from its use in syntax (i.e. the main constituent of a phrase) which is then extended and applied to word formation (Selkirk, 1982; Zwicky, 1985; Bauer, 1990; Hoeksema, 1992; Katamba and Stonham, 2006; Booij, 2012). Linguists in general have applied some of the typical criteria used to determine heads of syntactic construction directly onto word formation. Utilising several definitions of syntactic heads,[29] Bauer (1990) argued that, in general, there is no clear convergence between the representations of syntactic heads against morphological heads. This leads Bauer to question the validity of morphological heads, concluding that "(…) heads have no place in morphology. Certainly, if they have a role to play, this role needs to be defined much more carefully (…)" (1990: 30). Imposing a set of criteria that determines headedness in syntax directly onto morphology can create a distorted image of a morphological head. Although the two heads do overlap and share common similarities, they are nonetheless different entities. Basically, what it means to acknowledge complex words as having heads presupposes that a word has an internal structure of units, and that the prominence of these units differs (Scalise and Fabregas, 2010). In other words, the head is more prominent then the other units in a given structure. In this sense, the head is the only component that

---

[29] The definitions include hyponymy, subcategorisation, government, distribution, morphosyntactic locus and characterisation (among others).

cannot be taken off if an intended meaning (or grammaticality) of a given complex word is to be maintained.

There have been many attempts to identify the head component in morphology, one of which is through its position within a given construction. One of the classical attempts is Williams' (1981) *Right-hand Head Rule*. As the name implies, the right-most component in a morphologically complex construction is considered as the head. Despite some counter examples, most English compounds conveniently meet the terms of the rule, e.g. *black*$_A$ + *board*$_N$ becomes *blackboard*$_N$. The right-most component shows headedness qualities, i.e. it is the component that determines the category of the whole structure.[30] Nevertheless, this right-hand rule is at best a language-specific rule. It is obvious that this rule is not applicable to languages like Malay where the most important component (i.e. the head) of constructions is predominantly on the left-hand side.

Thus, identifying the head in compounds is arguably straightforward. Firstly, one needs to identify the position of the unit that is the most prominent component, and secondly identify the grammatical category of this component (which will then determine the overall category of the compound). However, it is common for compounds to have multiple head positions, i.e. left, right, headless, and even double-headed compounds. Ceccagno and Basciano (2007) reported that Chinese compounds can have three head positions of right, left and double-headed, all of which are productive in the language. In Chinese, most endocentric noun-noun compounds are right-headed, like *fang-xing* (lit. house model) 'layout of a house', while other verb-verb compounds like *ruz-hu* (lit. enter stop) 'move into' are left-headed (Ceccagno and Basciano 2007: 220). In a sample of 36 languages, Bauer (2001) reported that, although the

---

[30] From this viewpoint, derivational suffixes are justified as heads as they are the right-most and category-changing components, while prefixes are not heads as they are left-most and category-neutral components. Similarly, inflectional affixations are not considered as heads due to the fact that they are not category changers.

overall preference is indeed for right-headedness, nevertheless almost half of the sample had a variable head-modifier order. In another more recent study of compound typology, Scalise and Fabregas' (2010) findings on the type of compound headedness in the languages of the world, the preferences were as follows: right-headed compounds (66.7%), headless compounds (16.3%), left-headed compounds (6.8%) and double-headed compounds (5.9%). These studies generally conclude that, since there is no exclusive preferred head position (even within a given language), it is not then possible to identify the head simply by its position inside compounds. Therefore, the positioning of constituents cannot be taken as an absolute parameter of identifying headedness in compounds.

An assumption that has been better accepted than positioning identification is the second point given above, i.e. the head is the component responsible for the overall category of the compound. Thus, by identifying the overall category of the compound, one can infer that the constituent with the same category will be the head as well. This is true for most endocentric compounds: the head of *blackboard* must be *board*, not *black,* since the whole compound behaves as a noun. However, what happens when the constituents of the compound are of the same lexical category? One cannot simply determine which constituent is the head of the compound by consulting the overall category (as both constituents are of the same category). Take French noun-noun compounds *salades-sante* 'health salads' and *maitres-mots* 'master words', for instance: the former is a left-headed compound while the latter is right-headed (Rosenberg, 2007). The head of these compounds cannot be determined only by consulting the grammatical category of their components. If this approach is taken, one could equally argue for either the rightmost or the leftmost constituent as the head. It has been suggested that a better way of identifying the head of such compounds is through the meaning of the compound. To identify the head, one must recognise *salades* 'salads' - *sante* 'health' becomes *salades-sante* 'health salads', as a type of salad, not a type of health. Similarly, the head of Italian noun-

noun compound *capo* 'master' - *stazione* 'station' becomes *capostazione* 'station master', which is better explained through its meaning, i.e. *capo* is the head because the compound as a whole is a kind of person, not a kind of station (Scalise and Fabregas, 2010). The meaning of the heads *salade* and *capo* respectively determines the kind of compounds they denote, or, to put it in another way, *salades-sante* is a hyponym of *salade* not of *sante*, and *capostazione* is a hyponym of *capo* not of *stazione*. Semantics is indeed one of the most important features and methods of head identification in compounding; i.e. in terms of semantic, a compound is the hyponym of its head (Hoeksema, 1992; Haspelmath and Sims, 2010; Scalise and Fabregas, 2010).

Another relevant property for identifying the head is through its morphological features, i.e. the head is the grammatical locus of the compound. Morphological features such as inflection and gender have been assumed to be transferred from the head to the whole compound (Lieber, 1981; Zwicky, 1985; Hoeksema, 1992; Scalise and Fabregas, 2010). For example, in *apple pie*, *pie* is arguably the head because it is the locus of inflection, i.e. *apple pies*, not **apples pie* (Dressler, 2006). Similarly, the earlier Italian *capostaione*s shows that the head component imposes its gender onto the whole compound, i.e. *capo* (masculine) - *stazione* (feminine) becomes *capostazione*, which is a masculine noun because the head, *capo*, is masculine. These features are not necessarily applicable to all languages, especially those without grammatical markers such as gender or inflection, such as Malay. Nonetheless, even when they are relevant, some problems may still occur, as we will see below.

So far, we have seen that heads can have at least three important properties to identify headedness, namely categorical, semantic and morphological features. The assumption is that

the head (when there is one) will transfer these features to the whole compounded form.[31] However, as shown above, it is not efficient to only consider individual features when identifying headedness. Then again, if we take them all together, they will not coincide as well. In other words, the ideal condition is to have categorical, semantic and morphological features all transferred from the same head component, but this is, however, not always possible.

Mismatches of the categorical feature transfer can be seen through compounds where both of the constituents have a similar lexical category. In a coordinative compound such as *singer-actor*, it can be unclear as to which noun constituent is exactly the head of the compound, i.e. which constituent projects its categorical features to the whole compound, as both noun constituents are arguably of equal status. This sort of compound is not the same as the *salades-sante* 'health salads' type, because the noun-noun compound here has a dominant constituent; as mentioned, *salades-sante* is a hyponym of *salade* not of *sante*. A *singer-actor* does not necessarily mean someone who is mainly a singer and only acts occasionally or vice-versa, but someone who is equally a singer and an actor at the same time. Thus, it can be unclear as to which component exactly percolates its lexical features to the entire compound. In Chinese, mismatches of the categorical features can be seen clearly in examples such as *cai-feng* 'tailor', where both *cai* 'cut' and *feng* 'sew' are verbal constituents, thus neither one imposes its lexical category on the nominal compounded form (Scalise and Fabregas, 2010). The English example *singer-actor* shows that more than one component can be accounted for in the overall category; while the Chinese example *cai-feng* shows neither can be accounted for.

---

[31] Feature transfers are usually explained through the concept of *percolation*, i.e. a general well-formedness condition that ensures the features of the head are transmitted and interpreted as the features of the entire compound (cf. Lieber, 1983).

Mismatches of the semantic features can be seen through Italian compounds such as *porta-lettere* 'postman', where its overall noun category is not exactly inherited from the noun *lettere* 'letters' component. This is because the meaning of the compound is more relevant to the verb component *porta* 'to carry' as the compound denotes 'someone who carries something' rather than 'a type of letter'. Therefore, the semantics of the compound suggest that the head should be the verb *porta*, but the grammatical category of the compound is not a verb.

Concerning morphological features, as mentioned, it is assumed that the heads are the locus of inflection. This feature is not relevant to languages without inflection such as Malay. Nonetheless, a few examples are shown here on how mismatching of the morphological feature can occur in other languages. For instance, exocentric headless compounds such as *pickpocket* are inflected as *pickpockets*, even when *pocket* is not the actual head. Scalise and Fabregas (2010) exemplify that Italian *pomodoro* 'tomato' (originally derived from the phrase *pomo doro* 'apple of gold') used to have the plural *-i* marked on the head *pomi doro*, but, due to lexicalisation, the plural marking is at the rightmost of the compound, i.e. *pomodor-i*, even though it is not where the head is. This is the same for *capostazione*: the inflection is marked as *capostazion-i*, instead of on the head, *capo* (Dressler, 2006). These examples show that morphological features are not transferred even when the head component transfers its categorical features onto the whole compound.

Clearly, the concept of headedness is definitely useful and important in our attempt to understand the nature of compoundhood. However, identifying the head in compounds is not a straightforward process. The mismatches of categorical, semantic and morphological features illustrate how the head of a compound does not necessarily transfer its entire features to the compound as a whole. Many have observed such a property of non-uniformity in feature transfer in compound heads. The notion of head will not hold if we want to consider the head as the one single constituent that transfers all kinds of features to the overall compound. The

notion of headedness must be seen as dependent on the particular feature in consideration in order for it to work, even if all the features do come from a single head (Di Sciullo and Williams, 1987; Dressler, 2006; Scalise and Fabregas, 2010). It is in this respect that Scalise and Fabregas (2010) proposed a tripartite definition of headedness (i.e. distinguishing them into categorical, semantic and morphological heads) and to refer to them separately when required. We leave the issue of compound headedness now, only to return to it in Chapter 5 for our analysis of headedness in Malay compounding, and move on to the next topic in this section, i.e. the criteria of compoundhood.

2.2.4 Criteria of compoundhood

One of the popular approaches taken in the attempts to define compoundhood is to create a set of criteria to which a construction can adhere. If indeed the criteria are met, the construction can thus be considered as a compound. The common aim of these criteria is to prove that compounds are single unified lexeme units (lexical integrity), and, in this sense, the proposed compoundhood criteria are very much similar to the criteria of wordhood. Donalies (2004: 76)[32] suggested 10 criteria of compoundhood, namely that compounds are: (i) complex, (ii) formed without affixes, (iii) spelled together, (iv) have specific stress, (v) have linking elements, (vi) right headed, (vii) inflected as a whole, (viii) syntactically inseparable, (ix) syntactico-semantic islands, and (x) conceptual units. Obviously, some of the proposed criteria are of relevance, but others can be either less important or, rather, language-specific. For example, points (ii), (iii) and (vi) are not quite true for Malay compounding. The language is not without affixed compounds (e.g. *ke-daya serap-an* (absorptivity),[33] compounds are

---

[32] Cited in Lieber and Stekauer (2009: 6-7).

[33] Example, *daya + serap* (strength + absorb) = *daya serap* 'absorbance' = *ber-daya serap* (absorptive), *daya serap-an* (absorption), *ke-dayaserap-an* (absorptivity), etc.

regularly spelled separately (e.g. *papan hitam* 'blackboard'), and compounds are predominantly left-headed in Malay. On the other hand, points (v) and (vii) are simply not applicable as Malay does not have linking elements and inflections. This small point here shows how a fix list of criteria (i.e. a universal criteria) of compoundhood such as the one proposed above will never truly work. Having said this, we can nevertheless separate the criteria for compoundhood under a larger canopy in terms of orthographic, phonological, morphological, syntactic and semantic criteria (Spencer, 2003; Dressler, 2006; Bauer, 2006; Lieber, 2010; Haspelmath and Sims, 2010).

We begin by addressing the two arguably less reliable criteria of compoundhood, namely the orthographic and semantic criteria. In terms of orthography, the main concern is the issue of spelling convention. The argument lies in the fact that compounds are argued to be unified units, and thus they should also project this unification through a single word-spelling form. This property can be true for some languages and not for others. Lieber and Stekauer (2009) point out that spelling convention can play a major role in, for instance, Czech and Slovak, since all compounds are written as single-word units (distinguishing them from phrases) in these languages. However, spelling of compounds in other languages is commonly unpredictable. Bauer (2006) exemplifies the inconsistency of *flower pot*, *flowerpot*, *flower-pot*, or in *rain forest*, *rainforest*, *rain-forest,* all of which are attested examples in English. There is no particular reason as to why they are spelled as such; it can very well be due to stylistic or any other reasoning. In general, scholars have largely agreed that an orthographic convention such as spelling is not a reliable measure of compoundhood (Lieber, 2010).

Yet another less reliable criteria usually mentioned by scholars is the semantic measure.[34] This measure of compoundhood mainly concerns lexicalisation, i.e. if a compound

---

[34] The semantic measure here is not to be confused with the semantics of compounds, i.e. the meaning of the compound derived from the relation of or between its constituents.

is truly a single unified unit, it too should then be lexicalised in the lexicon as single lexemes are (as mentioned in the previous chapter). Although this lexicalisation is true to a certain extent, compounds, especially new ones, are nevertheless compositional by nature (Lieber, 2005). Bauer (2006) briefly argued that, if a syntactic sequence such as '*how do you do?*' is still considered as a compositional sentence even though it is very much lexicalised, then there is no reason for lexicalised compounds to lose their compositional status after lexicalisation. The semantic measure alone is definitely not sufficient to be a foolproof measure of compoundhood. We will see how the orthographic and semantic criteria fair against Malay compounding later.

A much more effective criterion of compoundhood concerns phonological aspects. Different languages have different phonological criteria in characterising their compounds. For instance, Japanese has a process known as *rendaku*, where the initial consonant of the second component is voiced in a certain class of compounds, e.g. *iro* 'colour' + *kami* 'paper' = *irogami* 'coloured paper', and *ike* 'arrange' + *hana* 'flowers' = *ikebana* 'flower arranging' (Bauer, 2006). In English, it is the stress patterns that are considered as a significant phonological criterion of compoundhood. English compound stress patterns are a well-studied topic. In English, the general pattern of words is to have one main stress; again, since compounds are considered as single unified units, the stress pattern should also reflect likewise, i.e. a compound should have one main stress (Haspelmath and Sims, 2010). Typically, English compounds carry stress on the left non-head constituent, e.g. `*blackbird,* `*gold fish*. This pattern distinguishes them from phrases where stress falls on the right head constituent (or both), e.g. *black* `*bird,* `*gold* `*medal* (Bauer, 2006; Haspelmath and Sims, 2010).

Although this is the common pattern in English, it is nonetheless not always the case. Counterexamples such as adjectival compounds *blue green* or *icy cold* seem to have the same stress on both constituents, and nominal compounds like *apple* `*pie* have their stress on the

right constituent, but `*apple cake* is argued to have a left stress (Lieber, 2005). Even within a single compounded form, the stress pattern can be random, as Spencer (2003) points out; one can have different readings merely from the change of stress pattern: a `*toy factory* stress may be read as 'a factory where toys are made', while a *toy* `*factory* stress may be read as 'a factory toy'. Furthermore, Bauer (1983) also argues that stress patterns can be affected by the surroundings, i.e. the compound stress pattern within a sentence context may be different when in isolation. Therefore, compound stress in English does not necessarily have to be on the left-hand constituent.[35]

In this sense, asserting that compounds have or need to have a particular stress pattern can be rather difficult to maintain, even in a well-established language like English. In general, phonological criteria as evidence of compoundhood are arguably language-specific. They can be more relevant in some languages in comparison to others. For a language such as English, reliance on a criterion such as stress pattern alone can be misleading being that it is not always consistent. Although the impact of the phonological criteria is apparent, it is still not the only requirement needed as a complete representation of compoundhood. Again, we will see how this issue fairs against Malay compounding later, in Chapter 5.

Moving along, we now look at yet another group of criteria which are always associated with compounding, i.e. the morphological criteria. One of the main concerns is how compounds are affected by the inflectional system of a language. Similarly, reflecting the characteristics of a single entity, it is thus understood that inflection of compounds too should be on the whole compound instead of its individual components. For instance, Booij (2010) exemplifies Dutch where pronominal adjectives are inflected in phrasal form, *zure* 'sour' *kool* 'cabbage' is a phrase literally meaning 'sour cabbage' (*zure* is inflected agreeing with the noun); instead, *zuur*

---

[35] Spencer (2003) further suggests that English stress patterns in compounding are perhaps better associated with lexicalisation rather than compoundhood.

kool is the compounded form of '*sauerkraut*' (*zuur* is an uninflected form when compounded).

In English, the non-head component of compounds is generally not inflected. Inflectional markings are on the head (as previously mentioned, the head is the grammatical locus). Thus, even nouns that seldom appear in the singular form, such as *trousers*, will not be marked in the compound *trouser-press* (Bauer, 2006).[36] In addition to inflection, linking elements is yet another concern of morphological criteria. Many languages have some sort of meaningless (semantically empty) extension, linking the two components of the compound. Ralli (2009) shows that modern Greek has a clear linking element, *-o*, after the first component of a compound. The linking element in German, however, has been argued as an inflection instead. For example *liebe.s.lied* 'love.LE.song' has the element '*s*', which derives from the genitive and plural markers, when the language also has the suffix *-s* for genitive and plural markers, e.g. *die auto-s* 'the car-PL'. However, the suffix genitive or plural marker cannot occur with *liebe* since it is the wrong gender for *-s* genitive and the wrong declension for *-s* plural in the language (Bauer, 2009: 346). Having said this, morphological criteria can be highly language-specific. Not all languages have inflections nor do they have linking elements. We will see whether morphological criteria have any relation to Malay compounding later, in Chapter 5.

Moving along, we now look at the syntactic group of criteria. Among the main concerns are the uninterruptability or inseparability of the components involved, and the inability to modify them (Bauer, 1988). Again, these test functions to demonstrate that a given structure is a unified unit (lexical integrity) rather than a linear sequence of components. Take the compound *blackboard*, for instance, insertion of other elements, e.g. **black large board*, is

---

[36] Nonetheless, this is not always the case as there are compounds with internal plural marking as well, such as *pants-loving*, *suggestions box*, *weapons inspector*, *programmes list*, *children's hour*, *girls' club*, etc. (Lieber, 2005; Bauer, 2006; Lieber and Stekauer, 2009). It has been argued that the markings function to force a reading of plurality of the non-head, such as *programmes* in the compound *programmes coordinator* emphasises that the coordinator is handling more than one programme. However, this does not explain marked examples such as *programmes list* where an unmarked *programme list* will also have the same plural reading - it is not quite right to think that a list only has one programme (Lieber and Stekauer, 2009).

not allowed between constituents if the compound is to retain its intended meaning. However, in the phrasal sequence of *a black board*, it is permissible to modify the phrase, for example, *a black large board*. Bauer (1998:73) points out that, even in compounds with a complex second element, e.g. [*Chinese* [*jade figure*]], modifiers are not allowed to be inserted between elements *[*Chinese dirty* [*jade figure*]]. A related test is the inability to modify the components of compounds. Again with *black board*, adding an adjective to the phrasal *a very black board*, for instance, can be understood as a board which is very black in colour. The same is not applicable for the compound as a **a very blackboard* is ungrammatical.

Another common syntactic test is considering compounds as anaphoric islands, i.e. the coreferential of the compound head (e.g. by a pro-form) should not be possible in a compounded structure, unlike in a phrasal structure (Spencer, 1991; Bauer, 1998; Lieber, 2005). For example, a *black one* is possible for the phrase structure *black board*, but a **blackone* is not for the compounded *blackboard* structure. Similarly, Spencer (1991: 420) shows how the pronoun *it* cannot refer to the *tea* in the compound *teapot* in **He took the tea$_i$ pot and poured it$_i$ into the cup*. Nonetheless, such a 'pro-form test' can be countered as in '*He wanted a riding horse, as neither of the carriage ones would suffice*', with *riding horse* and *carriage horse* otherwise having the appearance of compounds (Lieber and Stekauer, 2009:12). In short, the task of a syntactic criterion is basically to prove that a given compound is a single unified structure, so that it can be categorised as a compounded structure. We will discuss how syntactic criteria play their part in Malay compoundhood later, in Chapter 5.

All of the criteria discussed above (orthography, phonological, morphological, syntactic and semantic) have definitely assisted linguists in their attempts to identify compoundhood. The general premise of the discussion above illustrates how dependence on just a single criterion is insufficient to account for the concept of compoundhood as a whole. Furthermore, certain criteria can be more relevant to some languages in comparison to others.

A balanced acceptance of the appropriate ones is needed in accordance to the particular language involved. As mentioned, we have purposely omitted any discussion on Malay compounding here in this section. This is partly because the aim of this section is to draw a cross-linguistic outline of the common criteria of compoundhood. We will look into these criteria again with regard to Malay compounding in Chapter 5. We now move on to the topic of compound classification.

2.2.5 Classification of compounds

This section looks at the general literature on classification of compounds. A more specific focus on the classification of Malay compounding will be taken up in Chapter 5. Compound classification has been one of the main concerns of compounding studies. To a certain extent, compounds have been classified rather differently over the years by different linguists (cf. Spencer, 1991; Fabb, 1998; Plag, 2003; Booij, 2005; Haspelmath and Sims, 2010; among others), and thus one can expect some sort of inconsistency to be apparent in the literature. Scalise and Bisetto (2009) point out three main issues that have caused the inconsistency in compound classification, namely: (i) exclusive focus on certain categories (thus neglecting others), (ii) the use of different terminology, and (iii) the inconsistent criteria used to classify compounds.

The third point on the use of inconsistent criteria to identify compounds has been discussed in the previous section (cf. 2.2.1 and 2.2.2). With regard to the first point, it is apparent that the literature in general has, in one way or another, neglected certain compound formations. This is true, as nominal compounds have been focused on more often than other types of compounds. It is obvious that there are many other types but this bias is supported on the basis of nominal compounding being the more productive type of compounding in general.

If we look at English, for example, the fact that nominal compounds are the most productive ones in the language makes it obvious why more focus is given to this particular type of compound (while other types are somewhat neglected). The same can be said of Malay compounding as well, as we will see later on.

With regard to the second point, the use of different terminology may also cause inconsistency in compound classification. Even familiar classifications such as 'root' and 'synthetic' compounds can cause problems as well. This is mainly because this sort of labelling may not fit other languages. Other languages may not have the exact classification required by these compounds. Unless the definitions of such labels are adapted cross-linguistically, they are better off as language-specific terms. Even a commonly used language-specific term can be problematic as well, for instance, the use of Sanskrit terminology such as *bahuvrihi* and *dvandva*. These two terms have been widely used in the literature by many linguists, but it has been argued that this usage is erroneous, differing from their original meanings (Scalise and Bisetto, 2009).[37] *Dvandva* is commonly used as a synonym for coordinate compounds, when, in current terms, coordinate compounds can also be compounds with two properties associated with an entity (e.g. *fighter-bomber*), which is not a true *dvandva* (Bauer, 2003). Among others, Plag (2003) does differentiate between appositional (*fighter-bomber*) and coordinative (*doctor-patient*) compounds subsumed under copulative/dvandva, but others may not do so as they consider the two types are essentially in a coordinating relation. Similarly, *bahuvrihi* has been used as a synonym for exocentric compounds, when it originally meant a specific subclass of exocentric compounds, i.e. possessive compounds (Scalise and Bisetto, 2009).

---

[37] Scalise and Bisetto (2009) detail that *dvandva* originally indicated either coordinated elements (*candradityau* 'moon and sun') and/or multiple separate and coordinated elements (*itapabahanavyasanani* 'disease, pain, grief, captivity, and misfortune'), while *bahuvrihi* originally indicated nominal compounds with possessive interpretation.

The chart below is a short summary of how differing terminologies have been used to differentiate between the different types of compounds:

(12) Compound classification summary chart

Spencer (1991) Types:	Fabb (1998) Types:	Plag (2003) Types:	Booij (2005) Types:	Haspelmath and Sims (2010) Types:
1) Endocentric e.g. *film society*	**1) Endocentric** e.g. *sneak thief*	**1) Endocentric** e.g. *blackboard*	**1) Endocentric** e.g. *travel office*	**1) Endocentric** e.g. *lipstick*
2) Exocentric (*Bahuvrihi*) e.g. *pickpocket*	**2) Exocentric** (*Bahuvrihi*) e.g. *red head*	**2) Exocentric** (*Bahuvrihi*) e.g. *pickpocket* (Possessive) e.g. *redhead*	**2) Exocentric** (*Bahuvrihi*) e.g. form Italian '*lavapiattiï*' 'dishwasher'	**2) Exocentric** e.g. *pickpocket*
3) Dvandva e.g. *mother-child* (Appositional) e.g. *learner-driver*	**3) Co-ordinate** (*Dvandva*) (Appositional) e.g. *student price*	**3) Copulative** (*Dvandva*) Subtype: a) Coordinative e.g. *doctor-patient* b) Appositional e.g. *fighter-bomber*	**3) Copulative** (*Dvandva*) e.g. *washer-dryer* **4) Appositive**	**3) Coordinative** e.g. from Korean '*o-nwui*' 'brother and sister' **4) Appositional** e.g. *student worker*

From the chart above, it is apparent that the two most common divisions are between endocentric and exocentric types, which are determined by the property of headedness. For the rest of the compound types, they can be different in terms of their grouping and labelling. Spencer's (1991) classification is in general through headedness and the relations between the elements, i.e. (i) endocentric compounds with head-modifier relations, (ii) exocentric compounds with predicate-argument relations and (iii) *dvandva* (appositional) compounds without any dependency between the elements. Fabb's (1998) classification is not as diverse as the rest, as it only has three basic types: single headed (endocentric), no-heads (exocentric) and both elements with equal head-like characteristics (co-ordinate/dvandva/appositional) compounds. Plag (2003), on the other hand, refines *bahuvrihi* by possessive compounds, while the terms copulative and dvandva are use synonymously as umbrella terms for two subtypes of coordinative and appositional. Haspelmath and Sims (2010) further classify coordinative and appositional compounds, where the former is considered as having both elements with equal footing, while the latter has both elements with the same referent. In general, the classifications of compounds have been similar in spirit throughout, but the individual linguists may not use the same terminology to define what essentially is the same type of compound. We will see later, in Chapter 5, how Malay compounds fare against this issue.

A much more systematic effort at classifying compounds is made by Bisetto and Scalise (2005) and Scalise and Bisetto (2009). They claim that their classification is based on consistent criteria that aim to be universally valid.[38] Bisetto and Scalise (2005) propose that compounds can clearly be classified under three major headings of subordinate, attributive and coordinate. These three headings are basically the grammatical relations between the elements of the

---

[38] They base this claim on the fact that they have tested their proposed criteria on more than 20 languages in the Morbo/Comp projects (Scalise and Bisetto, 2009).

compound. A second level is then added distinguishing between endocentric and exocentric compounds. A diagram can be sketched as below.

(13) Bisetto and Scalise's (2005) compound classification diagram

```
                              Compounds
            ┌────────────────────┼────────────────────┐
        Subordinate          Attributive           Coordinate
         ┌──────┐             ┌──────┐              ┌──────┐
      Endo.    Exo.        Endo.    Exo.         Endo.    Exo.
        │       │            │        │            │        │
    apple cake kill joy   key word white collar actor author mind brain
```
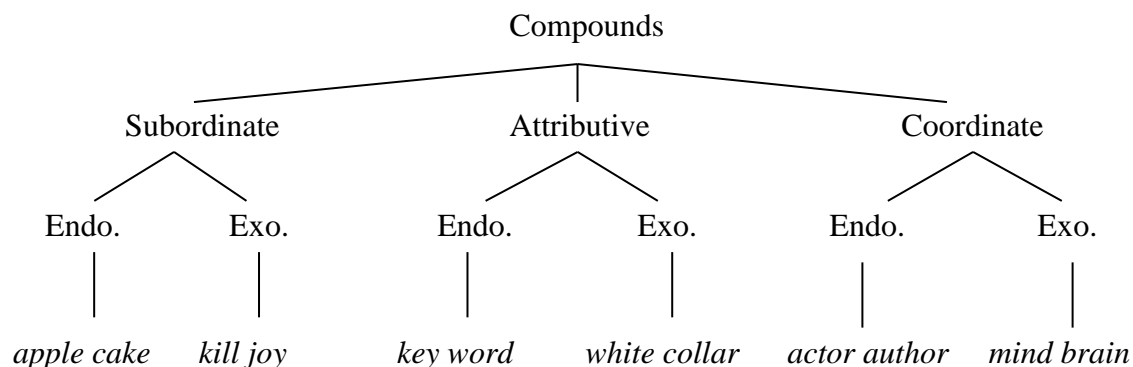
The main points of this classification are:

(i)   Each level of analysis and classification must be consistently based on a single, homogeneous criterion.

(ii)   The first level is based only on the implicit grammatical relation between the constituents.

The constituents of *subordinate* (SUB) compounds have the grammatical relation of 'complementation', i.e. head-complement relation; for example, *catfood* interpreted as 'food for cats' or *apron string* 'string of an apron'. Compounds like *taxi driver* also fall under the subordinate type where *taxi* is interpreted as the complement of the deverbal head *driver*. These kinds of compounds are in a head-complement relation, i.e. subordination. In addition to these endocentric compounds, Scalise and Bisetto (2009) also exemplify exocentric ones such as *cutthroat* and *pickpocket*, arguing that even these sorts of compounds have subordinate relations; the complement relation is determined by the verb (even though it is not the head) as it selects the other constituent. In contrast, the constituents of an *attributive* (ATR) compound

have the relations of attribution or modification, i.e. the non-head conveys a property of the head. These are usually nominal heads modified by nouns, adjectives, or verbs; for instance, *blue cheese* where *blue* modifies *cheese*, or *snail mail* where *snail* modifies *mail* but here in a metaphorical interpretation. Some exocentric ATR compounds include *greybeard* and *loudmouth*. Finally, the constituents of *coordinate* (CRD) compounds have the grammatical relation of coordination, e.g. *poet painter* where the relation 'and' coordinates between *poet* and *painter*. Semantically coordinative compounds can be considered as having two heads.

The classification outlined above is supported by the background of Lieber's (2004) framework of lexical semantics. In this framework, the lexemes of the compounds are characterised by two levels of representation, a skeleton and a body. The skeleton contains syntactically relevant information, while the body contains the encyclopaedic, holistic and idiosyncratic information. The illustration can be represented as below (cf. Bisetto and Scalise 2005: 329-330):

(14)   (i) Coordinate (CRD) compounds

	actor		*director*
skeleton	[+ material, dynamic]		[+ material, dynamic]
body	\<human, professional\>	⟷	\<human, professional\>
	\<show business\>	⟷	\<show business\>

(ii) Subordinate (SUB) compounds

	apple		*cake*
skeleton	[+material]		[+material]
body	\<physical\>		\<physical\>
	\<edible\>	⟷	\<edible\>
	\<can be an ingredient\>	⟷	\<made with ingredients\>

(iii) Attributive (ATR) compounds

	snail	*mail*
skeleton	[+material]	[+material]
body	<gastropod>	<institution>
	<secretes slime>	<means of communication>
	<very slow> ⟷	<takes time>

The general idea of this framework is that the meaning of a compound can be derived or interpreted from the selection and matching of informational pieces between its head and its non-head. Coordinate compounds such as *actor director* have both of their constituents matched on both levels of representation, i.e. the skeletal and the body features. In subordinate compounds such as *apple cake*, the skeletal features are irrelevant, while, for the body features, <edible> is the matching feature between the two constituents. This is sufficient as, at the least, only one matching body feature of the head constituent is needed to match the body feature of the non-head constituent. In attributive compounds such as *snail mail*, again the skeletal features are irrelevant and, similarly, only one matching body feature is needed: *snail* has the feature of <very slow> which matches the encyclopaedic feature of <takes time> in *mail*. This selection system of head selecting a different non-head constituent in each of the compound types is taken as support for the division of the compounds into the proposed SUB, ATR and CRD.

Nonetheless, the system of the head selecting the non-head as shown above can be unclear. Firstly, the entry matching between the features of the head and the non-head is confusing. For instance, on the difference between SUB and ATR compounds, the matching features between the components of these two types are minimal, i.e. for SUB compounds "At least one of the features of the head constituent must be matched by the encyclopaedic features

characterizing the non-head constituent", and for ATR compounds "What matters is that the non-head matches at least one of the encyclopaedic features of the head." (Bisetto and Scalise, 2005: 330). In this sense, there is not much difference between how the heads of these two types of compounds are matched with their respective non-heads, which arguably questions their distinctiveness.

The system also does not account for other types of compounds; for instance, unlike the root/primary compounds illustrated, synthetic compounds (e.g. taxi driver) do not select their non-heads through the matching of body features; instead, their non-heads satisfy the argument structure of the deverbal head. In other words, more specifications are needed in order to account for other types of compounding. Bisetto and Scalise (2005) mentioned that even though the first level of grammatical discrimination (i.e. SUB, ATR, CRD) does provide a preliminary grouping of compounds, it is nonetheless insufficient to properly account for other types of compounds, as they put it (331):

"We do believe that this first step is basic and that it should be kept separated from other possible criteria such as the internal structure, the semantic relation between the constituents, the origin of compound constituents or the categorical status of the constituents; all these criteria have to be ordered, so to speak, after the grammatical level of classification".

Scalise and Bisetto (2009) added an extra subclass level of analysis below the first level (i.e. specification of this first level) which further refined other types of compounds. The SUB is divided into two subclasses. The first subclass of SUB compounds is the type where the relations between the constituents are determined by the semantico-encyclopaedic information in Lieber's (2004) semantic framework shown above. The second subclass of SUB compounds is in accordance with the head-argument and head-complement/adjunct relation. These new subclass subcategories are named as *ground* and *verbal-nexus* respectively, which essentially

account for root/primary compounds and secondary/synthetic compounds in the traditional terms. The ATR type is separated into *attributive* and *appositive* compounds, and renamed ATAP (attributive/appositive). Attributive compounds are defined as those with modifying adjectival or verbal non-heads, expressing the quality/property of the heads. Appositives are compounds with modifying nominal non-heads as an attribute expressing the property of the heads. With regard to coordinate compounds, the components are mainly seen as having either a synonym or an antonym relation. There is no further division made to coordinate compounds in the renewed proposal. Thus, a new diagram can be illustrated as below:

(15) Scalise and Bisetto's (2009) renewed compound classification diagram
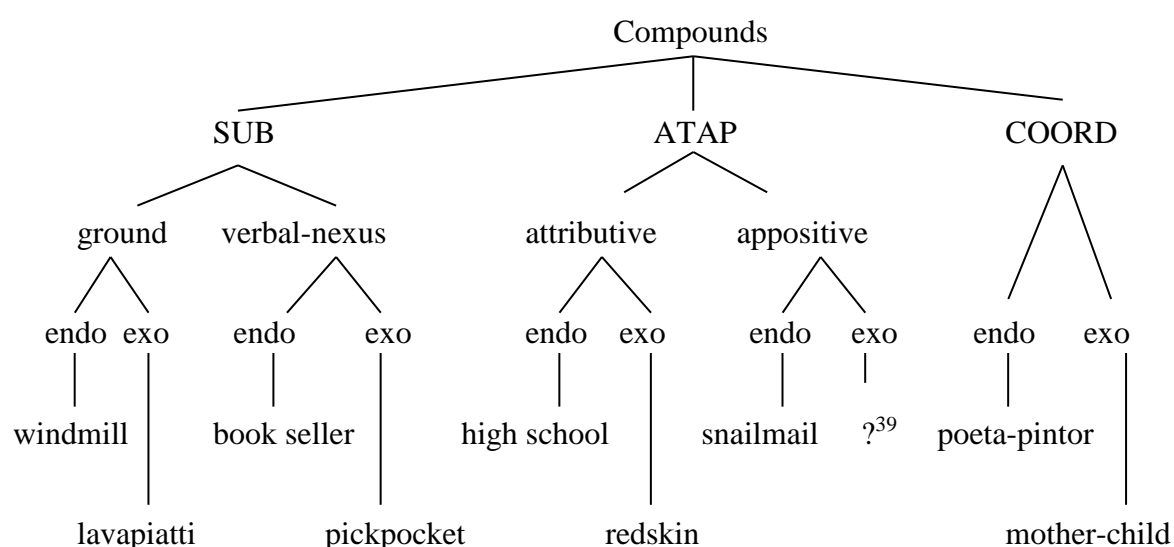
This manner of dividing compounds here in (15) has been claimed to be better than the previous classification in (13). This is partly because the schema arguably tries to account for a more detailed and specific classification of the many different types of compounds universally possible. The schema also stresses its universal applicability as it has been tested

---

[39] Scalise and Bisetto (2009: 50) point out that this type of compound is not apparent from their data, only some possible examples of Norwegian *kryssord* (crossword) 'crossword puzzle' and Chinese *ren she* (people snake) 'illegal immigrant'.

on more than 20 different languages, and all the tests have been claimed to produce satisfactory results. Having said that, it will be interesting to see how Malay compounding fits into this sort of classification. We leave this topic here only to return to it again later with regard to classification of Malay compounding in Chapter 5.

2.3 Chapter summary

This chapter is divided into two main parts. The first part concerns the issue of complex word formation, while the second part concerns issues of the universals of compounding. We begin this section by summarising the first part.

The question of where exactly it is in the grammar that complex word processing such as compounding takes place was raised at the beginning. The core argument to this question is that complex words are either handled by the morphological module, or they are handled by the syntactical module of the grammar. If word formation is considered as a process that is governed by distinct principles, morphology can be argued as a separate individual module in the grammar. Conversely, if word formation is considered as a process explainable through syntactic principles, thus morphology can be argued as a submodule of syntax.

To understand this issue, we first looked at some arguments on how syntactical approaches have been claimed to be able to represent the morphological phenomenon. We looked at how early generative grammar was deemed to be able to handle the process of word formation (e.g. derivation, nominalisation, compounding, past tense, etc.) through phrase structure and transformational rules or methods (Lees, 1960; Chomsky, 1965; Chomsky and Halle, 1968). Through such a point of view, a separate morphological module was deemed unnecessary. However, there are some obvious problems in maintaining such an account. For

instance, it can be a complicated (or even unnatural) process to explain the abstract and complex relationship between the constituents of a given compound through a transformational method. Yet another common supporting argument for the unnecessary existence of a morphological module is based on the observation of overlapping features observable in both morphological and syntactical modules. From this point of view, it is redundant and uneconomical to have two separate modules if both modules can account for the same features or principles. However, the argument that both modules share similar properties does not mean one should be subsumed under the other.

We discussed this viewpoint through the lexicalist approach (Spencer, 1991, 2005; Di Sciullo and Williams 1987; Bresnan and Mchombo, 1995; Scalise and Guevara, 2005; Lieber and Scalise, 2006). The lexicalist approach cuts off syntax from the morphological phenomenon, arguing for the basic principle of lexical integrity, which asserts that: (i) the components and principles of word composition are distinct from that of syntax, and (ii) words are atoms of syntactic structures (hence the internal structures of words are inaccessible to syntax). We looked at some evidence of these principles through some examples on affixation, coordination, recursion and compounding, which were also used as counter examples to deny the lexicalist approach. Nevertheless, we came to agree that the lexicalist approach adheres to the idea of morphology being an independent module different to syntax.

In addition to the lexicalist principles, there are also many reasons to believe in a separate morphological module. For instance, it has been observed that some properties are particular only to words, e.g. words are idiosyncratic, paradigmatic and limited in productivity (Katamba and Stonham, 2006; Scalise and Guevara, 2005). We came to understand that, if morphology is subsumed under syntax, one is in danger of missing out any principles or properties that are exclusive to morphology and not to syntax. By looking at several mismatches between the modules, it can be argued that these mismatches indicate the

exclusiveness of each module in one way or another. As such, it is better to assume a separation between morphology and syntax, and assume that they are two independent modules with some sort of interaction between them (Lieber, 1992; Jackendoff, 1997, 2002; Ackema and Neeleman, 2004; Lieber and Scalise, 2006).

To this end, we adopted what is known as the competition model, as proposed by Ackema and Neeleman (2004). This model supports the assumption of two separate generative modules (i.e. the morphological and the syntactical module) that are connected to each other (cf. grammatical system diagram in (3)). In this model, it is assumed that both morphology and syntax are equally capable of generating complex structures. In other words, complex structures such as compounds can be realised in either the morphological or syntactical subsystems. The two modules are in competition with each other to realise a given structure. Competition occurs when both morphological and syntactical structures merge the same categorical items and also have the same meaning relation between the items (cf. competition condition in (5)). The unmarked structure (which is language-specific) will be the winner of the competition, hence the realisation of either morphological or a syntactical structure is realised accordingly. We have come to agree that a model such as Ackema and Neeleman's (2004) competition model can sufficiently account for complex word formation such as compounding (in general).

The second part of this chapter provides an overview on the universals of compounding. We looked at topics concerning issues of definition, components, headedness, criteria and classification of compounding, all of which has been provided to help us draw an outline in terms of the universal characteristics of compoundhood.

First we explored the issue of definition of compounding and the issue of identifying components involved in compounding. We looked at how it can be rather challenging to come

up with a universal definition of what a compound is and is not. Similarly, it can also be difficult to say that certain linguistic items are suitable to be the components of compounding while others are not. This is mainly because such issues are very much language-specific (Bauer, 2006). It is thus understandable why different linguists define compounds differently and utilise different linguistic units: the components of a compound are in accordance to the peculiarities of the language involved. We further looked at the varying forms of units commonly used, among others the use of *words, lexemes*, *roots*, *bases*, *stems* and *phrases* as units of compounding. We recognised the importance of identifying these units appropriately, as the adherence to a specific linguistic unit may influence the outcome of whether a form is considered as a compound or otherwise. Taking everything into consideration, it is easy to understand why it can be difficult to have a single definition to universally represent compoundhood. We ended the section with Guevara and Scalise's (2009) proposal of the basic prototypical compounding schema of [X *R* Y] (X)(Y)(Z) structure. The X and Y represent a categorical unit, while the *R* represents the grammatical or semantical link between them, and (X)(Y)(Z) represents the overall category of the compound (cf. schema (11)). This proposed schema is deemed as a universal representation of compound structure. All of the issues discussed in this subsection (i.e. definition, compounding units and compounding schema) will be returned to with regard to Malay compounding in Chapter 5.

Next we examined the notion of headedness in compoundhood. It was acknowledged that this concept originated from a syntactical concept adapted to the morphological domain (Selkirk, 1982; Zwicky, 1985; Bauer, 1990; Hoeksema, 1992; Katamba and Stonham, 2006; Booij, 2012). Even so, the concept of headedness in morphology must be differentiated from that of the syntactical concept. Basically, a head presupposes that, within an internal structure of a word, there is one unit that is more prominent than the others (Scalise and Fabregas, 2010). In terms of compounding, a compound can have multiple head positions, i.e. left, right,

headless and even double-headed compounds. Scalise and Fabregas (2010) found that, universally, there are more right-headed compounds, followed by headless compounds, left-headed compounds and finally double-headed compounds in the world's languages.

Identifying the head of a compound is not simply a case of identifying the positioning of an item inside a compound. There are better ways to identify the head; among others: (i) the head is the component responsible for the overall category of the compound, (ii) the head defines the meaning of the whole compound (a compound is the hyponym of its head), and (iii) the head is the grammatical locus of the compound. In other words, there are at least three important properties in identifying headedness, namely categorical, semantic and morphological features. These properties are not necessarily present all together in the constituent that is to be considered as the head of the compound. Instead, we have come to understand that, in order for the concept of headedness in compounding to work, the head constituent must be seen as dependent on the particular property under consideration (Di Sciullo and Williams, 1987; Dressler, 2006; Scalise and Fabregas, 2010). It is in this respect that Scalise and Fabregas (2010) proposed a tripartite definition of heads, separating them into the categorical head, semantic head and morphological head, referred to accordingly when discussing headedness in compounding (Hoeksema, 1992; Haspelmath and Sims, 2010; Scalise and Fabregas, 2010). We will have more to say about the concept of headedness in Malay compounding later, in Chapter 5.

We then moved on to the issues of compoundhood criteria. We discussed, among other things, the orthographic, phonological, morphological, syntactic and semantic criteria of compounding (Dressler, 2006; Bauer, 2006). The general idea is that a structure can be considered as a compounded structure when certain conditions are met. As discussed, some criteria might be more relevant than others, all of which are dependent on the particular language involved. For instance, in terms of orthography, some languages have a single

spelling convention, while others may have a separate spelling convention for their compounded forms. Similarly, some languages utilise phonological aspects such as stress patterns to identify compounded forms, while other languages might not have a reliable stress pattern to do so. Nonetheless, some compounding criteria are rather universally acceptable. For instance, the syntactic criteria usually concern the issue of uninterruptability or inseparability of the compounded components. The inability to modify the components demonstrates that a given structure is a unified unit (lexical integrity), hence a compound. All in all, we have come to agree that it is insufficient to relay on one single criterion to account for the concept of compoundhood. It is also impractical to enforce the idea that all of them must be applicable to a given structure only for it to be considered as a compounded form. What is needed is appropriate criteria that are in accordance with the particular language involved. Once again, we will return to this topic later to discuss the appropriate criteria needed in Malay compounding.

Finally, we looked at the issue of compound classification. We have come to understand that compounds have been classified differently by different linguists. The inconsistency in compound classification is partly due to an imbalanced focus on certain compound categories in comparison to others, and the use of different terminologies to classify compounds (cf. compound classification summary chart in (12)). Again, all this can be attributed to the language-specific attributes of compoundhood. We put forward the suggestion made by Bisetto and Scalise (2005) and Scalise and Bisetto (2009) that compounds can classified under three major headings of subordinate, attributive and coordinate (cf. diagrams (13) and (15)). These headings represent the type of relation that can occur between the components involved in a given compound. Scalise and Bisetto (2009) insist that compounds in general can be universally accounted for through such subclassification, as it is claimed to be able to account

for the many different types of compounds universally possible. We will see later how Malay compounding fits (or can be accounted for) through such classification.

All things considered, we have managed to achieve the aim set out for this chapter, i.e. to provide a universal outline of the concept of compounding. We accomplished this by discussing the fundamental issues of *definition* (*compounding components*), *headedness*, *criteria* and *classification* of compounding. Taking into account these issues has clearly led us to the conceptualisation of compoundhood. Therefore, we can summarise a (universal) outline of compoundhood in the chart below:

(16) Outline of compoundhood

```
                    ┌─────────────────────────┐
                    │  Complex word formation │
                    └───────────┬─────────────┘
                                │
                    ┌───────────▼─────────────┐
                    │      Compounding        │
                    └─────────────────────────┘
```

Definition	**Headedness**	**Criteria**	**Classification**
- *components of compounding*	- *categorical* - *semantic* - *morphological*	- *orthographic* - *phonological* - *morphological* - *syntactic* - *semantic*	- *subordinate* - *attributive* - *coordinate*

```
                    ┌─────────────────────────┐
                    │      Compoundhood       │
                    └─────────────────────────┘
```
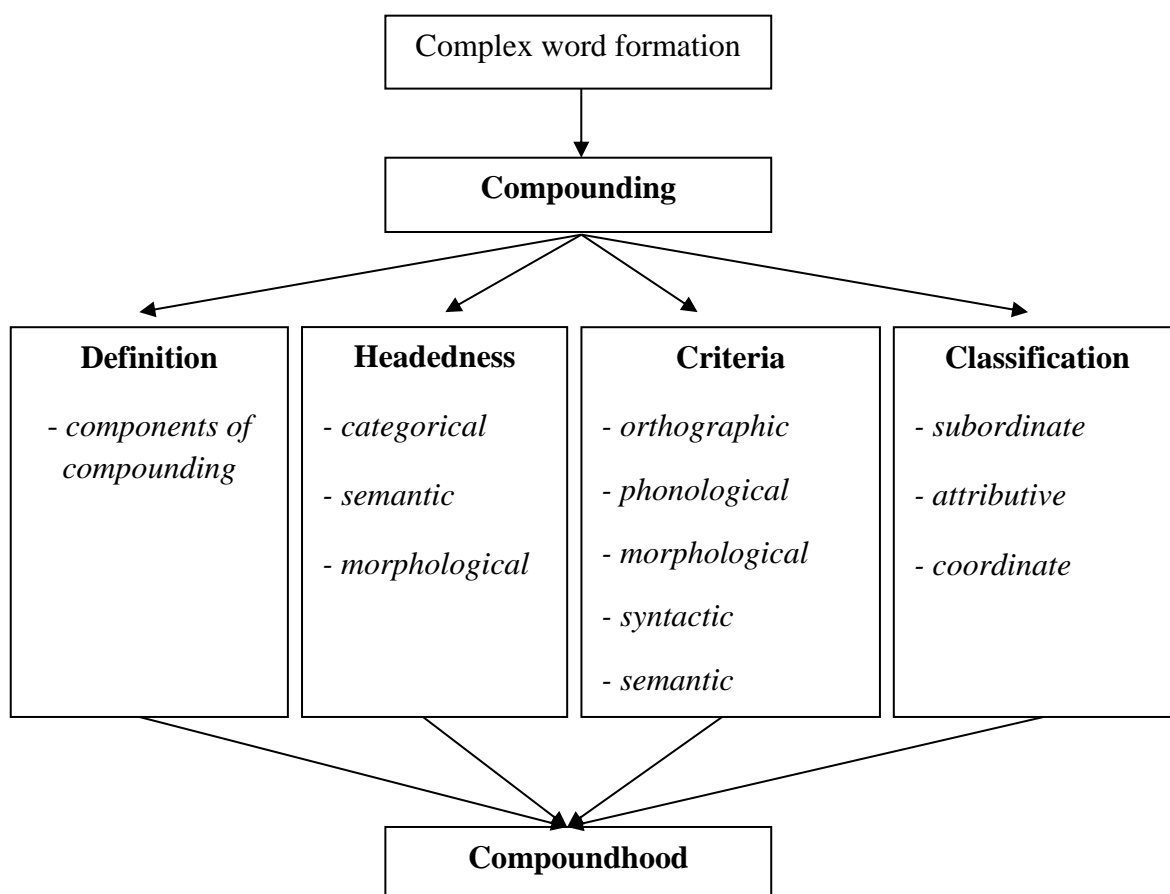
Given that we now have a workable outline of 'compoundhood', we can thus utilise it as a tool of analysis in our attempt to understand the compounding phenomenon in the Malay language.

In Chapter 3, we look at the two other main word formations in Malay, namely affixation and reduplication, and see how they interact with compounds in the language. We then seek to understand how the concept of Malay compounding has been perceived in the literature. This will be the task of Chapter 4, where we review how Malay compounds have been understood by Malay scholars throughout the years. The discussions in chapters 3 and 4, along with this outline that we have built here in Chapter 2, will be the foundation that will assist us in our main analysis of Malay compounding in Chapter 5.

Chapter 3: Malay Morphological processes

The aim of this chapter is to outline the morphology of the language. The main focus will be on the major lexical classes of nominal, verbal and adjectival compounds. Each category is then divided into the two main morphological processes of affixation and reduplication. Compounding, being the third main morphological process of the language, is not included here as it will be further discussed in subsequent chapters. It is to be noted here that this section has left out other classes such as adverbials and particles because they are not largely involved in compounding. Unless specified, the resources (and the examples) in this chapter are mainly from Hassan (1974), Omar (1980), Musa (1993) and Karim et al. (2008).[40]

3.1 General structure of Malay words

Before going into the details of the nominal, verbal and adjectival morphology, it is fitting that we begin this chapter with a brief introduction to the orthography and sound system of Malay words. In terms of orthography, standard Malay utilises the common 26 letters of the alphabet (similar to the English alphabet from A to Z) in its Romanised writing system. The letters represent a total of 32 underlying phonemes (26 consonant sounds /p, b, t, d, θ, ð, k, g, f, v, s, z, h, ʃ, x, ɣ, tʃ, dʒ, m, n, ŋ, ɲ, l, w, r, j/, [41] and 6 vowel sounds /i, e, a, u, o, ə/) and 3 diphthongs (/au, ai, oi/) (Hassan, 1986; Harahap, 1991; Omar, 1991; Onn, 2014). In general, the orthographic system can be considered as having a high grapheme to phoneme

---

[40] The reason for presenting certain information or examples without a specific reference is because it is the type of information that is generally agreed on among Malay scholars.

[41] The sounds of /f, v, θ, ð, ʃ, z, x, ɣ/ were originally considered as foreign phonemes borrowed from other languages such as Arabic and English to compensate for the influx of loan words entering the language (Hassan, 1986; Omar, 1991). Nonetheless, over the years, their status has been reinforced and integrated to be widely accepted as part of the sound system of the Malay language.

correspondence. This is to say that each letter in use has an almost perfect one-to-one matching to its phonemic counterpart in the International Phonetic Alphabet (IPA) representation (Harahap, 1991). For instance, the letter 'a' corresponds to its counterpart /a/ phoneme in the IPA, and the letter 'b' corresponds to the /b/ phoneme, and so on and so forth. In this sense, the mapping between the spelling and the pronunciation of Malay words is usually transparent, regular and predictable. Thus, the word *batu* 'stone', for instance, is phonetically pronounced as /ba.tu/, i.e. the orthographic spelling is virtually identical to the phonetic transcription.

Nonetheless, there are a few exceptions to this regularity, as Hassan (1986) and Onn (2014) point out, among others:

(i) Single letters:

'c' corresponds to the sound /tʃ/, 'j' to /dʒ/, 'x' to /ks/, and 'y' to /j/.

(ii) Sequence of letters:

'sy' corresponds to the sound /ʃ/, 'th' to /θ/, 'dz/dh' to /ð/, 'ng' to /ŋ/, 'ny' to /ɲ/, 'kh' to /x/ and 'gh' to /ɣ/.

(iii) The orthographic 'e' represents either the /e/ or the /ə/ phoneme across different contexts.[42]

In comparison to other languages with more complex grapheme to phoneme matching, these exceptions can be seen as minor occurrences. This is why Malay in general is considered as a language with a transparent and regular spelling and pronunciation system.

---

[42] The sounds of /e/ and /ə/ cannot be considered as allophones of the sound /e/ (represented by the letter 'e' in writing). This is because the two sounds have different meanings when used in different words. For instance, the word *sepak* can be pronounced as either /sepak/ 'kick' or /səpak/ 'slap', and similarly the word *bela* can be either /bela/ 'defend' or /bəla/ 'care'. Knowing when to use the appropriate sounds requires knowledge of the language. Even within a single word with more than one letter 'e', language user needs to know which appropriate phoneme is required for the correct pronunciation, e.g. *kereta* 'car' is pronounced as /kəreta/ (and not as /kərəta/ or /kereta/ or /kerəta/).

Malay words can either be in the simple form or the composite form. Simple words are root words, which are also the base for composite word formation derived via processes of affixation, reduplication and/or compounding. The syllabic structure of root words ranges from one to four syllables, but they generally do not exceed that. In general, Malay root words are more commonly disyllabic in comparison to other syllabic structure. Hassan (1974) in his descriptive morphology study argued that the majority of words in his corpus of approximately 4500 root words are disyllabic. Likewise, Onn (1980) observed that more than 740 root words (out of 926) in his corpus of spoken dialogues are disyllabic. To this end, both Hassan (1974) and Onn (1980) agree that the most common syllabic structure in Malay is the disyllabic structure with the template of (C)V(C).CV(C).

The preference for the disyllabic structure can also be seen through loanwords or borrowings where words from the donor language that have more than two syllables will usually be reduced to two syllables upon being borrowed by Malay (as briefly mentioned in section 1.3). Nevertheless, adapted foreign words or scientific terms with multiple syllables usually retain their original structure, for instance, *universiti* /u.ni.vər.si.ti/ from 'university' /yunəvɜrsɪti/ and *vegetarian* /ve.ge.ta.ri.an/ from 'vegetarian' /vɛdʒɪtɛərɪən/. Some examples of root words in the language with the various syllabic structures are given below.

(17) Examples of root word syllabic structures

	Syllabic structure	Orthographic symbols	Phonetic symbols	English gloss
Monosyllabic	CV	*yu*	/ju/	shark
	CVC	*bab*	/bab/	chapter
Disyllabic	V.CV	*ubi*	/u.bi/	potato
	V.CVC	*emak*	/ə.mak/	mother
	CV.CV	*bulu*	/bu.lu/	feather
	CV.CVC	*sukan*	/su.kan/	sports
	CVC.CV	*lembu*	/ləm.bu/	cow
Trisyllabic	V.CV.CV	*udara*	/u.da.ra/	air
	CV.CV.CV	*kerusi*	/kə.ru.si/	chair
	CV.CV.CVC	*belalang*	/bə.la.laŋ/	grasshopper
Four syllabic	CV.CV.VC.CV	*keluarga*	/kə.lu.ar.ga/	family
	CV.CV.CV.CVC	*hulubalang*	/hu.lu.ba.laŋ/	commander
	CV.CV.CVC.CVC	*semenanjung*	/sə.mə.nan.juŋ/	peninsula

Having now understood the spelling and pronunciation of Malay words and how they are generally regular and transparent, we will henceforth not be too dependent on phonetic transcriptions to transcribe the Malay words used throughout this thesis. We now proceed to the next section, which is the morphology of Malay affixation and reduplication with regard to the major word classes of the language.

3.2 Nominal morphology

It is apparent that nouns can be classified into two basic categories, namely root and composite/complex forms. The nominal roots can be divided into three main subclasses, i.e. nouns, pronouns and classifiers. The derivation of nominal forms employs all of the main

morphological processes of the language, i.e. affixation, reduplication and compounding. Nominal morphological processes are presented in the following subsections.

3.2.1 Nominal affixation

Affixation is a productive and well-established system in the language. In addition to being extensively used as a method of word formation, affixation also provides the means of expressing grammatical relationship in the language. Scholars normally categorise Malay affixes in relation to their combining roots, i.e. grouping them according to their derivations after combining them with root words from different lexical categories. The nominal affixes are listed below:

(18) Types of nominal affixes[43]

(i) Prefixes	*peN-*, *per-*, *ke-*
(ii) Suffixes	*-an*, *-wan*, *-man*, *-wati*
(iii) Infixes	*-el-*, *-em-*, *-er-*
(iv) Circumfixes	*ke-…-an, peN-…-an, peR-…-an*

(i) Nominal prefixes

(a) *peN-*

The prefix *peN-* can be attached to most lexical categories (but it is most often attached to verbs) to derive nominal words. The prefix *peN-* has several allomorphic variations, namely

---

[43] Capital letters in certain prefixes indicate that the particular prefix contains allomorphs.

/pə/, /pəm/, /pən/ and /pəŋ/, which change in relation to the phonological environments of the combining words.

Prefix *peN-* becomes *pe-* (/pə/) if the first letter (phonemes) of the root or stem it precedes is one of the following: m (/m/), n (/n/), ny (/ɲ/), ng (/ŋ/), r (/r/), l (/l/), or w (/w/). For example, *peN + masak = pemasak* 'cook'. Furthermore, *peN-* will also become *pe-* (/pə/) if the root it precedes begins with p (/p/), t (/t/), k (/k/) or s (/s/), but the first letter (phoneme) of the root/stem will be dropped and replaced with: m (/m/) for p (/p/), n (/n/) for t (/t/), ng (/ŋ/) for k (/k/), and ny (/ɲ/) for s (/s/), e.g. *peN + pukul = pemukul* 'beater'.

PeN- becomes *pem-* (/pəm/) if the initial letter (phoneme) of the word it precedes is either b (/b/), f (/f/), or v (/v/), e.g. *peN + baca = pembaca* 'reader'. On the other hand, *peN-* becomes *pen-* (/pən/) if the initial letter (phoneme) of the word it precedes (phonemes) is d (/d/), c (/tʃ/), j (/dʒ/), sy (/ʃ/), or z (/z/), e.g. *peN + curi = pencuri* 'thief'. Finally, *peN-* changes to *peng-* (/pəŋ/) if the initial letter (phoneme) of the word it precedes is one of the following: consonants g (/g/) or h (/h/) or vowels a (/a/), e (/e, ə/), i (/i/), o (/o/), or u (/u/), e.g. *peN + ukir = pengukir* 'carver'.

Prefix *peN-* derivation results in several grammatical functions such as agentive, qualitative, habitual, instrumental, abstract and units of measure (Karim et. al, 2008; Jalaluddin and Syah, 2009). In terms of grammatical functions, *peN-* derives:

Agentive nouns, e.g.:

adil 'to be just'	=	*pengadil* 'judge/referee'
bawa 'to carry'	=	*pembawa* 'one who carries something'

Qualitative and habitual nouns, e.g.:

kasih 'to be loving' = *pengasih* 'one who is loving'

rokok 'cigarette' = *perokok* 'smoker'

Instrumental nouns, e.g.:

sapu 'to sweep' = *penyapu* 'broom'

besar 'to be large' = *pembesar* 'something that makes large'

Abstract nouns, e.g.:

dapat 'to get' = *pendapat* 'idea/thought'

sakit 'to be ill' = *penyakit* 'illness/disease'

Units of measure, e.g.:

peluk 'to embrace' = *pemeluk* 'circle of arms length'

(b) *per-*

Per- is less productive in comparison to *peN-*. The prefix combines with a limited number of nouns and verbs to form agentive and instrumental nouns, e.g.:

asap 'smoke' = *perasap* 'container to burn incense'

(c) *ke-*

Similarly, the prefix *ke-* is unproductive, occurring with several nouns and verbs to derive agentive nouns, e.g.:

kasih 'to be loving' = *kekasih* 'someone who is loved'

> *tua* 'old'   =   *ketua* 'someone who is elder (leader)'

(ii) Nominal suffixes

(a) *-an*

The suffix *-an* is very productive, occurring with nouns, verbs, particles and adverbs to derive nouns that function as locative, variety (in plurality), nouns that are results of actions, and a few fruit names:

Locative, e.g.:

> *tepi* 'side'   = *tepian* 'place at the edge/river bank'

> *lapang* 'to be spacious'   = *lapangan* 'place that is spacious/airstrip'

Nouns that are results of actions, e.g.:

> *ajar* 'to teach'   =   *ajaran* 'something which is being taught'

> *kotor* 'to be filthy'   =   *kotoran* 'something that is dirty/filthy'

Fruit names, e.g.:

> *duri* 'thorn'   =   *durian* 'a kind of thorny fruit'

> *rambut* 'hair'   =   *rambutan* 'a kind of hairy fruit'

The suffix *-an* also occurs with reduplicated noun base forms where the reduplicated form conveys plurality while the suffixation adds the meaning of variety (discussed further in the next section) or, e.g. *buah-buah* 'fruits' = *buah-buahan* 'all kinds of fruits', *sayur-sayur* 'vegetables' = *sayur-sayuran* 'different kinds of vegetables'.

(b) *-wan*, *-man*, *-wati*

These suffixes are not productive; *-wan* occurs after vowel /a/, *-man* occurs elsewhere, and *-wati* replaces *-wan* and *-man* in some cases to indicate femininity, e.g.:

harta 'property'	=	*hartawan* 'one with lots of property'
juta 'million'	=	*jutawan* 'millionaire'
seni 'art'	=	*seniman/seniwati* 'actor/actress'

(iii) Nominal infixes

Infixes are not a productive form of affixation. Hassan (1974) estimated a closed set of approximately 50 words (of all lexical categories) that can undergo infixation in the language. Nonetheless, they still produce several noun forms with agentive, variety and instrumental functions, for example:

(a) *-el-*

e.g.:	*patuk* 'to peck'	=	*pelatuk* 'woodpecker'
	tunjuk 'to point'	=	*telunjuk* 'ponter'

(b) *-er-*

e.g.:	*suling* 'to whistle'	=	*seruling* 'something that whistles/flute'

(c) *-em-*

e.g.:	*kuncup* 'to fold'	=	*kemuncup* 'something that folds/a kind of grass'

(iv) Nominal circumfixes

(a) *ke-...-an*

The circumfix *ke-...-an* may be affixed to all classes and certain compound base forms to derive nouns with abstract and locative functions, e.g.:

tuhan 'god'	=	*ketuhanan* ' belief in god/faith'

raja 'king' = *kerajaan* 'kingdom'

menteri 'minister' = *kementerian* 'ministry'

simpul 'to tie a knot' = *kesimpulan* 'conclusion'

(b) *peN-...-an*

In terms of allomorphic variations on the circumfix *peN-...-an*, the same rules of the prefix *peN-* are applied to the prefix part of the *peN-...-an* circumfix. The circumfix can be attached to most lexical categories to derive verbal nouns that show the process of some action carried out by an agent (Hassan, 1974; Mintz, 1994), e.g.:

hamba 'slave' = *penghambaan* 'the process of slavery'

beritahu 'to tell' = *pemberitahuan* 'the process of informing'

lari 'to run' = *pelarian* 'refugee'.

(c) *per-...-an*

In contrast to *peN-...-an*, the circumfix *per-...-an* produces nouns that are the result of some action and locative (Hassan, 1974; Mintz, 1994), e.g.:

satu 'one' = *persatuan* 'something that is united/association'

kubur 'grave' = *perkuburan* 'grave yard'.

3.2.2 Nominal reduplication

Reduplication is also a productive process of the language. It can be described as being either a simple or a complex construction. A simple reduplication means the duplicated constituent is identical to the root without any phonemic alterations; while a complex reduplication means that the duplicated constituent undergoes phonemic changes or additions

(Rubino, 2005; Lieber, 2010). In general, reduplication is divided into three types, namely full reduplication, partial reduplication and rhyming/chiming reduplication (Hassan, 1974; Omar, 1975; Onn, 1980; Musa, 1993; Mintz, 1994; Karim et al., 2008).

(i) Full reduplication of nouns

In a full reduplication process, the duplicate is an exact copy of the whole root. The general structure of full reduplication can be illustrated in a schema, as below:

$$\text{Full reduplication} = [\,X\,]_{\text{ROOT}} - [\,X\,]_{\text{DUPLICATE}}\,{}^{44}$$

(X_{ROOT} = major lexical categories, and the dash '-' symbol marks the separation between the root and the duplicate).

This schema is also applicable to the other lexical classes (verbs and adjectives) in terms of full reduplication. Studies with corpus data have shown that full nominal reduplications are the most frequent form of reduplication in the Malay language (Omar, 1975; Sew, 2007). The full reduplication of nominal base forms is quite straightforward, whereby the reduplicated nouns derived may possess some of the subsequent functions, plurality, resemblance and a small number of animal names, as described below.

(a) Plurality

Plurality is the most common function of nominal reduplication and all types of nouns (simple or composite) may be reduplicated to convey plurality, e.g. *meja* 'book' becomes *meja-meja* 'books', *telur* 'egg' becomes *telur-telur* 'eggs', etc. However, full reduplication of

---

[44] Diagram adapted from Musa (1993:176).

composite words is limited to derived nouns, where the reduplicated outcome also serves to indicate plurality, e.g.:

Root	Affixed stems (nouns)	Full reduplication
kasih 'love'	*ke+kasih* 'lover'	*kekasih-kekasih* 'lovers'
pakai 'to wear'	*pakai+an* 'clothing'	*pakaian-pakaian* 'clothes'
satu 'one'	*per+satu+an* 'association'	*persatuan-persatuan* 'associations'

(b) Resemblance

Full reduplication of some simple nouns may produce nouns that resemble objects similar to the base noun forms, e.g.:

anak 'child' = *anak-anak* 'doll'

kuda 'horse' = *kuda-kuda* 'wooden horse/pillion'

(c) Animals and insects

A number of bound roots can only appear in fully reduplicated constructions. Hassan (1974) listed 11 bound noun roots that always appear in reduplicated forms denoting insects and small animals, i.e.:

anai ---	*anai-anai* 'termite'	*kupu* ---	*kupu-kupu* 'moth'
katu ---	*katu-katu* 'flying ant'	*rama* ---	*rama-rama* 'butterfly'
amang ---	*amang-amang* 'hermit crab'	*kura* ---	*kura-kura* 'tortoise'
korok ---	*korok-korok* 'fruit fly'	*labi* ---	*labi-labi* 'turtle'
kunang ---	*kunang-kunang* 'firefly'	*labah* ---	*labah-labah* 'spider'
layang ---	*layang-layang* 'swallow'		

The bound roots themselves do not carry any meaning, thus *labah* is meaningless, but *labah-labah* refers to 'spider'. The reduplicated forms may intrinsically connote plurality (as some insects are always found in groups, e.g. *anai-anai* 'termites'), but this is not always the case as some can be referred to as a single entity, e.g. *kura-kura* can refer to a single tortoise.

In general, linguists agree that plurality is one of the more important and regular functions among the many grammatical functions of full nominal reduplication (Heah, 1989; Karim et al., 2008). In her study of Malay reduplication, Omar (1975) established that 49.27% of the reduplicated items in her data conveyed the function of plurality.[45] Sew's (2007) study of reduplication also shows that plurality is the most common meaning of nominal reduplication. However, plurality through reduplication is not obligatory in Malay as plurality/quantity are not necessarily expressed via reduplication. They can also be realised through context or by using other indicators such as quantifiers and classifiers, e.g.:

(i)	*(satu) se buah rumah*	(iii)	*rumah - rumah*
	numeral + classifier + Noun		noun + noun
	one + *buah* + house		house - house
	'a house'		'houses'

(ii)	*dua buah rumah*	(iv)	*banyak rumah*
	numeral + classifier + Noun		quantifier + noun
	two + *buah* + house		many + house
	'two houses'		'many houses'

---

[45] The other functions include intensity at 13.80%, the meaning of continuity at 10.08% and repetition at 10.06%.

The above show how the noun is never marked for number since number is not a grammatical category in the language. Therefore, although reduplication is semantically parallel to the concept of plurality, it is nonetheless not syntactically justified as an inflectional process (Hassan, 1974). However, noun reduplication does have some inflectional properties in the sense that, in general, it is a transparent and not a category-changing process (unlike derivation). Even so, full reduplication of the noun may change the lexical categories of the reduplicated forms nonetheless, e.g. noun reduplication can cause changes from nominal to adjectival and adverbial reduplicated forms, e.g.:

Nouns → *Adjectives*

e.g.   *hati* 'liver'   =   *hati-hati* 'careful'[46]

Nouns → *Adverbs*

e.g.   *subuh* 'dawn'   =   *subuh-subuh* 'in the dawn'

(ii) Partial reduplication of nouns

In partial reduplication, the process of reduplication only takes place on a specific section of a given word. Partial reduplication comes in two forms, namely partial reduplication of roots (which is exclusive to nouns) and partial reduplication of affixed forms (Omar, 1975; Musa, 1993; Karim et al., 2008). In noun root partial reduplication, only the first consonant of the noun is duplicated. The vowel 'e', which is phonetically a schwa [ə] epenthesis, immediately follows the reduplicated consonant, e.g.:

---

[46] Omar (1975) considers *hati-hati* as the only isolated example since it does not fit into any pattern for the generation of other adjectives of such nature.

Root	Partial reduplication	Orthographic realisation
langit 'sky'	/lə -laŋit/	*lelangit* 'mouth palate'
rambut 'hair'	/rə -rambut/	*rerambut* 'capillary'
jari 'finger'	/jə -jari/	*jejari* 'radius'

Alternatively, Musa (1993) and Karim et al. (2008) view the process as involving an initial repetition of the first syllable, which is then followed by a reduction of the reduplicated vowel to the central vowel (schwa), e.g.:

Root	Partial reduplication	Vowel change	Orthographic realisation
langit 'sky'	/la -laŋit/	/lə -laŋit/	*lelangit* 'mouth palate'
rambut 'hair'	/ra -rambut/	/rə -rambut/	*rerambut* 'capillary'
jari 'finger'	/ja -jari/	/jə-jari/	*jejari* 'radius'

There is some ambiguity in partially reduplicating certain derived forms such as *sayuran* 'vegtables' to *sayur-sayuran* 'variety of vegetables'. Some agree that these formations involve the process of omitting the affixes in *sayuran* and copying only the root *sayur* to create the duplicate (Musa, 1993; Karim et al., 2008). Others view this as a process of attaching the suffix *-an* to the fully reduplicated construction *sayur-sayur* (Hassan, 1974; Omar, 1975). There is no clear distinction on this as the lack of clarity is apparent in other examples, e.g. *tumbuh-tumbuhan* 'plants' is arguably derived from *tumbuhan* (grow + *-an*) 'plant' since **tumbuh-tumbuh* is ungrammatical. On the other hand, *buah-buahan* 'fruits' may have *-an* attached to *buah-buah* 'fruits'. This is also a similar issue in verbal partial reduplication, as we will see later.

(iii) Rhythmic reduplication of nouns

The final type of nominal reduplication is the rhythmic reduplication (Musa, 1993; Karim et al., 2008), also known as rhyming/chiming reduplication (Onn, 1980; Musa, 1993). Both terms refer to the same reduplication but with some differences in description. The 'rhyming' forms refers to the duplication of the initial or final syllables of the root, e.g. initial rhyming *bukit* 'a hill' becomes *bukit-bukau* 'hills', final rhyming *sayur* 'vegetable' becomes *sayur-mayur* 'variety of vegetables', while, in 'chiming' forms, the duplicate maintains the consonants but changes the vowels of the root, e.g. *tanah* 'soil' becomes *tanah-tanih* 'variety of soils'.

Conversely, the rhythmic reduplication perspective divides the process into three types, namely:

(a) Vowel harmony: the vowels of the root are repeated in the duplicate constituent,

      e.g.:   *lauk* 'dish' becomes *lauk-pauk* 'variety of dishes'

(b) Consonant harmony: the consonants of the root are repeated,

      e.g.:   *gunung* 'mountain' becomes *gunung-ganang* 'range of mountains'

(c) Free rhythmic: there are no similarities between the root and the duplicate,

      e.g.:   *ipar* 'brother-/sister-in law' becomes *ipar-lamai* 'in-laws'

Omar (1975:191), dismisses this type of reduplication by asserting that it "(...) does not invite such an easy deduction (...)" and that "(...) any component, be it a syllable, a consonant or a vowel can be reduplicated". As the entity of the reduplication is what is more important, the more general term 'rhythmic reduplication' is used (instead of rhyming/chiming reduplication) to represent these types of reduplication in the language.

Rhythmic nominal reduplication is exclusive in both number and function (Omar, 1975; Karim et al., 2008). There are approximately 15 noun reduplications of this sort, all of which

function to indicate plurality/variety. It is exclusive in the sense that, for each rhythmic form, there is also a full reduplication form as well. For instance, the word *gunung* 'mountain' can be fully reduplicated into *gunung-gunung* 'mountains', and also undergo the rhythmic reduplication of *gunung-ganang* 'mountains'. Although both versions refer to mountains, the rhythmic version has an extra semantic feature of variety, bringing the meaning of something like a 'range of mountains'. A final feature to be noted is that the rhythmic duplicated components (e.g. *ganang*) never appear as independent words in the language.

3.3 Verbal morphology

The verb in Malay is commonly categorised into the two main types of intransitive and transitive. Intransitives are without complements, while transitive are either active or passive forms. Only a number of intransitive verbs may occur in their simple bare form without any affixation, e.g. *duduk* 'sit', *jatuh* 'fall', *terbang* 'fly', etc. (Karim, 1995). Simple transitive verbs occur in either the active form with *men-* or in the passive form with *di-* affixes. Similar to nouns, verbs undergo the usual morphological process of affixation, reduplication and compounding. Again, only affixation and reduplication will be presented here.

3.3.1 Verbal affixation

Listed below are the verbal affixes:

(19) Types of verbal affixes

Prefixes	*meN-, beR-, di-, ter-*
Suffixes	*-kan, -i*
Infixes	*-el-*
Circumfixes	meN-…-kan, *beR-…-an, meN-… -i*

(i) Verbal prefixes

(a) *meN-*

The prefix *meN-* has several allomorphic variations, namely /mə/, /məm/, /mən/ and /məŋ/, which change in relation to the phonological environments of the combining words, e.g. *meN + masak = memasak* 'to cook', *meN + basuh = membasuh* 'to wash', *meN + curi = mencuri* 'to steal', *meN + ukir = mengukir* 'to carve'. The prefix *meN-* attaches to transitive verbs to derive the active transitive voice, e.g. *meN- jual* 'sell', *meN- beri* 'give', etc. The prefix *meN-* attaches to certain intransitive verbs, nouns and adjectives to derive the active intransitive voice, e.g. *meN-nyanyi* 'sing', *meN-gunung* 'to rise (like a mountain)', *meN-kuning* 'to become yellow', etc.

(b) *beR-*

The prefix *beR-* occurs with transitive verbs, nouns and adjectives to indicate middle voices.

Transitive voice:

Indicates a state (rather than action), e.g. *ber-sawah* 'to work on a rice farm', *ber-kedai* 'to run a shop', etc.

Reflexive voice:

Indicates action by the doer, e.g. *ber-cukur* 'to shave (oneself), *ber-hias* 'to prepare oneself', etc.

Reciprocal/repetitive voice:

Occurs with verbs, e.g. *ber-tumbuk* 'to box one another', *ber-tikam* 'to stab one another', etc.

Productive voice:

Occurs with nouns, e.g. *ber-telur* 'to lay eggs', *ber-anak* 'to give birth', *ber-buah* 'to produce fruit'.

Other functions of *ber-* include to indicate: action (*ber-lari* 'to run'), possession (*ber-harta* 'possessing wealth'), and state (*ber-duka* 'state of sadness), among others.

 (c) *di-*

The prefix *di-* occurs with verbs to indicate the passive voice, e.g. *di-jual* 'sold', *di-makan* 'eaten', *di-tendang* 'kicked', etc.

(d) *ter-* [47]

The prefix *te*r- occurs with transitive and intransitive verbs to indicate unintentional actions, e.g. *ter-ambil* '(unintentionally) took (something)', *ter-makan* '(unintentionally) eaten (something)', *ter-jatuh* '(unintentionally) fell', etc.

(ii) Verbal suffixes

The suffix *-kan* occurs with any word class to derive transitive verbs with the function of causation and benefaction. Hassan (1974: 97) glossed the 'causative benefactive' nature of this suffix as 'to cause to… (for)', e.g. *hamba* 'slave' = *hambakan* 'to cause to be a slave (for)', *satu* 'one' = *satukan* 'to cause to be one (for)', etc. On the other hand, the suffix *-i* also occurs with any word class to derive causative transitive verbs, but it is more specific than the *-kan* suffix, e.g. *hambai* 'to cause someone to be a slave', *satui* 'to cause something to be one', etc. (Hassan, 1974). These suffixes also occur with other derived forms with a general meaning as the above, e.g. of 'to cause (something) to be (something)'.

(iii) Verbal infix

As with nouns, infixes are no longer productive in verbs. There is arguably only one verbal infix, i.e. *-el-*, as in *jajah* 'colonise' = *jelajah* 'to explore', *sodok* 'to shove' = *selodok* 'to bulldoze', etc.

---

[47] As an aside, some scholars have argued that not all affixations in the language are exclusively derivational. In a complex affixation structure, the prefixes *ter-*, *di-* and *meN-* will occur at the outermost layer of a structure. For example, the prefix *di-* in *main* 'to play' to *dipermainkan* 'to be fooled' [*di* + [*per* + [[*main*] + *kan*]]: the prefix *di-* is attached to the base word at its outermost layer, as opposed to *[*per* [*di* [[*main*] *kan*]]]. To a certain degree, this shows similar qualities to inflection and thus affixations have been suggested by some to be the inflections of the language (Md Salleh, 1993; Syed Jamaluddin, 2005). However, affixations in the language are not being used to mark inflectional categories (number, person, case, etc.), as they are marked by other indicators such as context and specific words. We will not be looking into the argument of whether Malay has inflections or not in this thesis. Therefore, we leave this here as a side note for the interested reader.

 (iii) Verbal circumfixes

There are actually several kinds of verb-forming circumfixes in the language. Karim et al. (2008) listed 11 of them, including *meN-...-kan*, *beR-...-kan*, *di-...-kan* and *ke-...-an*. However, only two will be briefly mentioned here for illustration. The *meN-...-kan* is used to derive transitive verbs carrying several functions, such as to indicate causative action and benefactive action. For instance, *besar* 'big' = *membesarkan* 'to make something bigger' and *baca* 'read' = *membacakan* 'to read (something) to (someone)'. Among the functions of *ber-...-an* is to indicate reciprocity, such as *salam* 'greet' = *bersalaman* 'to greet/shake hands with each other', and to indicate two things in close proximity, *dekat* 'near' = *berdekatan* 'to be near each other'.

3.3.2 Verbal reduplication

There is some disagreement about verb reduplication in terms of it not having a fully reduplicated form. Hassan (1974) and Karim et al. (2008) do attest to some fully reduplicated verbs, but arguably there are not many examples, as both of them only manage to give seven examples between them.

 (i) *Full reduplication of verbs*

From Hassan (1974) and Karim et al. (2008), fully reduplicated verbs have at least the function of repetition (*gelak-gelak* 'to laugh repeatedly') and continuity (*pusing-pusing* 'to spin continuously').

(ii) *Partial reduplication of verbs*

Karim et al. (2008) maintain that partial reduplication of verbs consists of duplication of derived verbal base, i.e. only the root is duplicated (e.g. *berlari* 'to run' = *berlari-lari* 'to run (somewhat casually)'), and not from *lari-lari* = *berlari-lari*. Other examples include: *mencari-cari* 'to keep on searching', *tersenyum-senyum* 'to smile repeatedly (rather shyly)', *berbisik-bisik* 'to whisper with one another', etc. There is also the type where the second constituent of the reduplication contains the affix with the usual function of reciprocation, e.g *peluk-memeluk* 'to hug each other', *gigit-menggigit* 'to bite each other', etc. There are also plenty of suffixed verbs with partial reduplication, e.g. *doa-doakan* 'to (say) a prayer (to someone)', and circumfixed verbs, e.g. *bersalam-salaman* 'to greet/shake hands with one another'. As mentioned, partial reduplication of verbs is a large subject on its own, and this section does not do justice to the topic.

(iii) *Rhythmic reduplication of verbs*

The final form of verbal reduplication is the rhythmic reduplication. In general, the grammatical functions of rhyming verbs are those of repetition and intensification. Some examples taken from Hassan (1974: 86) are given below:

(a) *initial rhyming*

  beli 'to buy'   =   *beli-belah* 'to buy lots of things'

  lalu 'to pass by'   =   *lalu-lalang* 'to pass to and fro'.

(b) *final rhyming*

  kacau 'to stir'   =   *kacau-bilau* 'commotion'

  hurai 'to untie'   =   *hurai-barai* 'to be loose and disentangled'.

(c) *chiming reduplication*

      mundar 'to toddle'   =   *mundar-mandir* 'to toddle to and fro'

      kumat 'to murmur'   =   *kumat-kamit* 'to murmur repeatedly'.

There are plenty of other examples of rhythmic reduplication. Similar to nouns, the duplicated components in rhyming verbs are specialised forms that do not appear independently in the language, i.e. they are meaningless forms based on the rhyming pattern of the base root.

3.4 Adjectival morphology

The adjective has been generally defined as a word that describes a noun or a noun phrase. Karim (1995) points out that an adjective in Malay can be distinguished from nouns and verbs since it can be preceded or followed by an intensifier, e.g. *sangat besar* (very big) 'very big', *cantik sungguh* (beautiful very) 'very beautiful', etc. Karim et al. (2008) classify several types of adjectives according to their meaning function; among others, adjectives describe colour, measurement, time, form/shape, distance, manner, situations, feeling, senses, etc. Derivations of complex adjective forms are of the usual morphological process of affixation, reduplication and compounding. Again, only affixation and reduplication will be presented here.

3.4.1 Adjectival affixation

In adjectival affixation, there is also no suffixation. Adjectival affixation also provides the means of expressing grammatical relationship in the language. Listed below are the adjectival affixes:

(20) Types of adjectival affixes[48]

Prefixes	*ter-, se-*
Infixes	*-el-, -em-, -er-*
Circumfixes	*ke-...-an*

(i) Adjectival prefixes

(a) *ter-*

The prefix *ter-* attaches to the adjective, forming the superlative measure, e.g.:

    muda 'young'   =   *termuda'* youngest'

    baru 'new'   =   *terbaru* 'latest'

(b) *se-*

The prefix *se-* also attaches to the adjective, forming the comparative degree, e.g.:

    muda 'young'   =   *semuda* 'as young as'

    cantik 'beautiful'   =   *secantik* 'as beautiful as'

(ii) Adjectival infixes

As mentioned, infix affixations are no longer productive. Even so, there are some adjectival infixations applicable to adjectives, nouns and verbs, with the added function of repetition and intensity, e.g. *-el-* as in *gembung* 'bloated' = *gelembung* 'bubbles', *-er-* as in *kelip* 'twinkle/blink' = *kerelip* 'to twinkle', and *-em-* as in *guruh* 'lightning' = *gemuruh* 'rolling of thunder'.

---

[48] Prefix *meN-* and suffix *-kan* (cf. verbal affixation (19)) are also applicable to adjectives.

(iii) Adjectival circumfixes

This is limited as there is only one circumfix for adjectives, i.e. *ke-...-an*, generally attached to nouns referring to a condition or nature of the noun, e.g.: *barat* 'west' becomes *kebaratan* 'western-like', *cina* 'Chinese' becomes *kecinaan* 'Chinese-like'.

3.4.2 Adjectival reduplication

Similarly, adjectival reduplication is manifest in all three forms of Malay reduplication, i.e. full, partial and rhyming/chiming forms.

(i) Full reduplication of adjectives

All adjectival roots are capable of full reduplication, which mainly derives from adjectives with intensified meaning or with the meaning of comprehensiveness, for example:

gila 'mad'	=	*gila-gila* 'very mad'
benar 'true'	=	*benar-benar* 'truly'
besar 'big'	=	*besar-besar* 'to be (generally) big'.

Omar (1975) proposed that a reduplicated form that has an adjective as its underlying root functions as an adverb, mostly of manner, duration and time. For example:

Adjective → Adverb of manner

e.g.:	*dalam* 'deep'	=	*dalam-dalam* 'deeply'
	tajam 'sharp'	=	*tajam-tajam* 'sharply'

Adjective → Adverb of distance

e.g.: *dekat* 'near' = *dekat-dekat* 'close by'

jauh 'far' = *jauh-jauh* 'in the distance'

Adjective → Adverb of measure

e.g.: *panjang* 'long'= *panjang-panjang* 'in great length'

(ii) Partial reduplication of adjectives

Similar to the arguments for partial reduplication of nouns, it is unclear whether partial reduplication is a process of reduplicating the un-affixed part of the given stem or whether affixes are being attached to fully reduplicated forms. Omar (1975) and Karim et al. (2008) seem to agree on the former process. In view of such an opinion, examples of derived partial reduplicated adjectives are given below:

Adjective → Adverb of manner

e.g.: *pandai* 'clever' = *sepandai-pandai* 'with all the cleverness'

besar 'big' = *besar-besaran* '(done) on a large scale'

terang 'clear' = *terang-terangan* '(done) clearly in the open'

Adjective → Adverb of measure

e.g.: *panjang* 'long' = *sepanjang-panjang* 'all the way'

Adjective → Verb of state

e.g.: *kuning-kuning* '(very) yellow' = *kekuning-kuningan* 'in the state of yellowness'

(iii) Rhythmic reduplication of adjectives

The final form of verbal reduplication is rhythmic reduplication, which is fairly straightforward as it brings about the function of intensification (Hassan, 1974; Karim et al., 2008). This sort of reduplication is also considered as unproductive, with quite a limited number of examples, including:

(a) initial rhyming

    bengkok 'crooked'   =   *bengkang-bengkok* 'to be very crooked'.

(b) final rhyming

    ramah 'friendly/talkative'   =   *ramah-tamah* 'to be very friendly/talkative'.

c) chiming reduplication

    huru 'to be tumultuous'   =   *huru-hara* 'to be very tumultuous'

    gopoh 'to be hasty'   =   *gopoh-gapah* 'to be very hasty'.

There are also a few examples of free rhythmic adjectival reduplication in the language, such as *kusut* 'to be tangled/haywire' = *kusut-masai* 'to be very tangled/haywire'. Similar to nouns and verbs, most of the duplicated components in rhythmic adjectives are specialised forms that do not appear independently in the language.

3.5 Chapter Summary

The first part of this chapter introduces some basics features of the orthography and sound system of the Malay language. This subsection is beneficial in the sense that it provides a basic knowledge of Malay word spelling and pronunciation. In general, Malay spelling and pronunciation is transparent, regular and predictable. Knowing the basics of the system reduces

the less fluent reader's dependency on a phonetic transcription for every Malay word in this thesis from here on.

The next section of this chapter provides an overview of the two other main morphological processes in the Malay language (apart from compounding), namely the affixation and reduplication processes. Affixation is undeniably a well-developed and productive system in Malay. The language possesses a considerable amount of affixes ranging from prefixes, suffixes, infixes and circumfixes. The common method of categorising Malay affixes is by grouping them according to the types of lexical categories to which they can attach to derivate new forms. It is extensively utilised as a means to express grammatical relationship in the language. Similarly, reduplication is a productive process in the language. In general, the language distinguishes at least three types of reduplication form, i.e. full reduplication, partial reduplication and rhythmic reduplication. Reduplication may cause nominal, verbal or adjectival roots to derive into other lexical categories subsequent to the process. Additionally, the derived reduplicated constructions may possess several grammatical functions, for example, to indicate plurality, variety, resemblance, repetition, continuity, intensity, extensiveness, emphasis, etc.

The presentation is ordered in a systematic manner. The three major lexical categories (i.e. nominal, verbal and adjectival) are divided into the way in which they are linked to the morphological processes of affixation and reduplication respectively. For instance, in the discussion of nominal affixation, we began by looking at the types of affixes that are available to derive nominal words. We then looked at how the affixes were attached in relation to the phonological environment and the type of roots to which they attach in order to derive the appropriate complex nominal form. Finally, the discussion focused on the grammatical function and meaning of the affixation. The same presentation sequence was also provided for the other two lexical categories of verbal and adjectival affixation.

Similarly, in the discussion of nominal reduplication, for instance, we first presented the types of reduplication available for nouns, followed by the way in which they reduplicate in terms of the structure and phonological environment involved. This was then followed by the final discussion on the grammatical function and meaning of the reduplicated form. Again, the same presentation sequence was also provided for the verbal and adjectival reduplication. We can illustrate our overview of this in the table below.

(21) Overview of Malay affixation and reduplication processes

Nominal/Verbal/Adjectival Morphology	
Affixation	Reduplication
Types of affixation: Nominal (cf. 3.2.1) Verbal (cf. 3.3.1) Adjectival (cf. 3.4.1)	Types of reduplication: Nominal (cf. 3.2.2) Verbal (cf. 3.3.2) Adjectival (cf. 3.4.2)
The affix relation to the structure and environment of the root	The reduplication relation to the structure and environment of the root
Grammatical function and meaning of the affixation	Grammatical function and meaning of the reduplication

The morphological processes presented in this chapter have focused on each process in relatively separate conditions. Therefore, it may appear as if the Malay morphological process only allows a single layer of affixation and reduplication. However, this is not always the case. Both morphological processes can correlate and interact with each other to create complex multi-layer word structures.

For instance, root words can have the common single layer of affixation, e.g. prefixation, suffixation or circumfixation as illustrated below.

(22) Affixation of root words (e.g. root *ajar* 'teach')

Root	Single layer affixation	Derivation
ajar 'teach'	[*pel* + (*ajar*)] prefix + root	*pelajar* 'student'
	[(*ajar*) + *an*] root + suffix	*ajaran* 'teachings'
	[*pel* + (*ajar*) + *an*] prefix + root + suffix	*pelajaran* 'subject'

The structure above can be further affixed to build a cluster of prefixes, e.g.:

Root	Multiple layer affixation	Derivation
ajar 'teach'	[*ber* + *pel* + (*ajar*) + *an*] prefix + prefix + base + suffix	*berpelajaran* 'educated'

Similarly, affixation can also interact with reduplication in the language, e.g.:

(23) Affixation of reduplicated words (e.g. root *main* 'play').

Root	Layers of affixation	Derivation
main 'play'	[*main-main*] reduplication	*main-main* 'playful'
	[*ber* + (*main-main*)] prefix + base	*bermain-main* 'playing continuously'
	[(*main-main*) + *kan*] base + suffix	*main-mainkan* 'to cause to play'
	[*mem* + *per* + (*main-main*) + *kan*] prefix + prefix + base + suffix	*mempermain-mainkan* 'to make fun of'

The exemplification above is the (general) maximum structure or layer of complex affixation possible in the language, i.e. (prefix) + (prefix) + root/base + (suffix). The core of the structure can be a simple root form or a complex structure (such as affixed or reduplicated forms shown above), and the overall category of the affixed form is set by the outermost prefix category.

Having understood this, we can take a quick look at some examples of the base being a compounded structure. The process of compound affixation adheres to several predetermined steps which allow affixes to be attached to both elements of a compound as prefix, suffix and circumfix, e.g.:

(24) Affixation of compounds

Compound base	Affixation	Derivation
lipat ganda 'multiply'	[ber + *lipat ganda*] prefix + base	*berlipat ganda* 'to multiply (more)'
daya serap 'absorbance'	[*daya serap + an*] base + suffix	*daya serapan* 'absorption'

As shown above, the single attachment of either a prefix or suffix will not influence the structure of the compound constituents. However, circumfixation will cause the compound constituents to merge, for example:

Compound base	Affixation	Derivation
surat khabar 'newspaper'	[*per + surat khabar + an*] prefix + base + suffix	*persuratkhabaran* 'press/journalism'

Similar to any other root/base affixation, an affixed compound form can derive a different lexical category and grammatical function from its base compound form. On the other hand, in terms of compound reduplication, only the first element (i.e. the head) is reduplicated, e.g.:

(25) Reduplication of compounds

Compound	Reduplicated compound
kertas kerja 'paperwork' (singular)	*kertas-kertas kerja* 'paperwork'(plural)

Reduplication of the whole compound form[49] is applicable if it is a circumfixed compound base, e.g.:

Circumfixed compound	Reduplicated compound
persuratkhabaran 'press/journalism (singular)'	*persuratkhabaran-persuratkhabaran* 'press/journalism (plural)'

There will be more to say about compounding and its relation to affixation and reduplication in the chapters to come. This summary here is meant to create a sense of closure to the relatively separate discussions on affixation and reduplication in this chapter. We believe that this summary has managed to encapsulate the foundations of how affixation, reduplication and compounding interact with each other to create even more complex structures in the language. It is important to understand this sort of interaction as it will be a recurrent theme (in one way or another) in our build up to the analysis of Malay compoundhood. We leave this chapter here to begin the next chapter on the studies of Malay compounding.

---

[49] Pre-conjoined compounds (such as the established forms) are reduplicated as a whole as well, for example, *warganegara* (people + country) 'citizen' is reduplicated into *warganegara-warganegara* 'citizens'.

Chapter 4: Studies of Malay compounding

The aim of this chapter is to draw an overview of how the subject of compounding has been addressed in the Malay literature. The works and studies presented here are in chronological order, i.e. they are presented according to the year of publication, from past studies to more recent ones. By doing so, we are not only presenting the studies in an orderly manner, but, more importantly, we are able to reflect the development (if any) of how the matter has been approached through the years. A different study is presented in each subsection, which is then immediately followed by some comments on the particular study.

4.1 Abdullah Hassan's (1974) *The morphology of Malay***.**

Hassan's (1974) study is anchored on a corpus of about 5000 different items of the major lexical categories in the language. The study is divided into individual chapters on nominals, verbals, adverbials and particles, all of which are analysed with regard to the three main processes of affixation, reduplication and compounding.

With regard to compounding, the basic definition is given as "(…) a construction that has two or three free forms as its constituent. Each of the constituent forms may either be a root or a derived form" (1974: 46). Hassan (1974) recognises compounds as being like syntactic structures (i.e. predominantly head + modifier constructions). To this end, he classifies compounds as constructions that are either combined syntactically (head + modifier) or asyntactically (modifier + head). The former are called syntactic compounds and the latter are called asyntactic compounds. Based on the findings of his corpus, Hassan states that syntactic compounds are more common to the language than asyntactic ones.

Hassan (1974) also distinguishes compounds as being either endocentric or exocentric constructions. Endocentric compounds are defined as compounds with the head belonging to the same lexical class (having a similar syntactic function) as the constructions as a whole, e.g. *kayu api* (wood (head) + fire (modifier)) 'firewood'; i.e. the head is a noun, thus the whole compound is also a noun. On the other hand, exocentric compounds are compound constructions that, as a whole, do not belong to the same word class (or have the same syntactic function) as any of their constituents, e.g. *hirup darah* (to suck + blood) 'to exploit/extort' (here it is a noun, 'person who exploits/extorts'). Similarly, from the findings of his corpus, he also claims that, in general, there are more endocentric compounds than exocentric compounds in the language.

Hassan (1974) exemplifies the types of compounds through the main lexical categories as below.

(i) *Nominal compounds*

As mentioned, Hassan (1974) claims that endocentric compounds are more dominant than exocentric ones and this pattern is apparent in nominal compounds obtained from his corpus. Below are some examples taken from his study:

Endocentric nominal compounds

In endocentric nominal compounds, the head precedes the modifier, and the head is always nominal, while the modifier may be nominal, verbal, adverbial or even a particle, e.g *abu rokok* (ash + cigarette) 'cigarette ash', *padang tembak* (field + shoot) 'shooting range', *kapal terbang* (ship + fly) 'aeroplane', *rumah sakit* (house + sick/ill) 'hospital', etc. These are considered as the *syntactically endocentric* ones as they follow the head-modifier structure.

On the other hand, Hassan exemplifies some *asyntactically endocentric* ones as well, where the modifier precedes the head, e.g. *bumi putera* (earth + son/prince) = *bumiputera* 'native', *maha guru* (great + teacher) = *mahaguru* 'great teacher/professor', *perdana menteri* (prime + minister) = *perdana menteri* 'prime minister'. Hassan (1974) points out that these asyntactically endocentric compounds are limited because they were originally loanwords taken as a whole without modification to suit the language's general structure. Therefore, they do not follow the general head + modifier order of the Malay structure.

Exocentric nominal compounds

Exocentric nominal compounds have heads preceding the modifiers, and the head is always a transitive verb and the modifier is always a noun, e.g. *cucuk sanggul* (to pierce + hair bun) 'hair pin', *kaji bumi* (study + earth) 'geology', *sapu tangan* (to wipe + hand) 'handkerchief'. Hassan does not give any asyntactically exocentric nominal compounds.

Semantically exocentric nominal compounds

These types of compound are basically idiomatic constructions. Hassan points out that they are grammatically endocentric but semantically exocentric. Idioms are semantically exocentric since the overall meaning is not produced by the meanings of the constituents. However, Hassan (1974) maintains that they are always endocentrically combined into a compound. Some examples given are: *panjang tangan* (long + hand) 'thief', *tangkai jering* (stalk + a kind of fruit) 'miser', *berat tulang* (heavy + bone) 'lazy'.

(ii) *Verbal compounds*

Hassan (1974) describes the process of verbal compounding as similar to that of the nominal. Again the claim is that there are more endocentric patterns than exocentric ones. The chapter on verbal compounding is concise and mainly consists of examples of the constructions.

Endocentric verbal compounds

In endocentric verbal compounds, the head precedes the modifier, e.g. *ikat mati* (to tie + (die) completely) 'to tie a dead knot', *pandang rendah* (to look + low) 'to underestimate', *tolak tepi* (push + side) 'to put aside', etc. Hassan (1974) gives one example of an asyntactically endocentric verbal compound, i.e. *temu duga* (to meet + to anticipate/guess) 'interview'.

Exocentric verbal compounds

In exocentric verbal compounds, the constituents are combined in a verb plus object construction, or a head plus an attribute. The head is a verb and the object may be a nominal. Some examples given are *buang negri* (to banish + state) 'to expatriate', *bawa diri* (to carry + self) = 'to sulk', etc. Hassan (1974) recognises some *semantically exocentric verbal compounds*, i.e. their meaning as a whole is not derivable from the meaning of their constituents. These are also considered as idioms, e.g. *tumbuk rusuk* (to pound + flank) 'to bribe', *cuci mata* (to clean + eyes) 'to enjoy a scene', etc.

(iii) *Adjectival compounds*

Hassan (1974) considers adjectives as *Stative Intransitive Verbs*, thus there is no separate chapter on adjectival compounding. The explanation and examples of such compounds are combined together within the verbal compounding section. Nonetheless, the basic process and characteristics of adjectival compounding are very much the same as for nominal and verbal compounds. Similar to the others, endocentric adjectival compounds have their heads preceding their modifiers. The head is an adjective, while the modifier can be adjectival or nominal, e.g. *biru laut* (blue + ocean) 'ocean blue', *biru tua* (blue + old) 'dark blue'. An example given of an *asyntactically endocentric adjectival compound* is *duka cita* (to be sad +

feeling) 'to be sorrowful'. One example is given of a *semantically exocentric adjectival compound*, i.e. *tinggi hidung* (high + nose) 'arrogant'.

(iv) *Adverbial compounds*

There are very few adverbial compounds. Nonetheless, Hassan (1974) categorises them similarly to the rest, i.e. either endocentric or exocentric, formed by adverbial heads with modifiers of other lexical classes. Endocentric adverbial compounds have their heads preceding their modifiers, e.g. *acap kali* (often + time) 'often times', *dahulu kala* (before + time) 'olden times'. An example given of an asyntactically endocentric adjectival compound is *demi kian* (by + way/manner) = *demikian* 'that way'. The discussion on adverbial compounding is short due to the few examples available.

In addition, Hassan (1974) also claims that there are a few prepositions that occur as compounds. Among others are *dari pada* (from + in the direction of) = *daripada* 'from somewhere' and *ke pada* (to + in the direction of) = *kepada* 'towards'.

4.1.1 Comments

Hassan's (1974) work is indeed an important contribution to the overall knowledge of modern Malay morphology and in particular of Malay compounding.

One of the key points that we would like to highlight from Hassan's (1974) work is his recognition of the headedness in compounding. He elaborates that regular Malay compounds have the head + modifier structure, while non-regular compounds (other scholars label them as established compounds) have a modifier + head structure instead. Hassan (1974) recognises the former as syntactically combined compounds, as he argues that they follow the general rules of the language in terms of structure order, while the non-regular compounds are

considered as having an asyntactically combined structure as they do not follow the regular structure order of the language.

In extension to this, Hassan (1974) also recognises compounds as being either endocentric or exocentric constructions. This is an important distinction to establish. Through this, he can classify compounds as either having a head identifiable by one of the constituents of the compound, or as having a head that is 'outside' the constituents of the compounded form. With such a classification in hand, he manages to identify from his corpus the different types of compounds and classify them according to the endocentricity and exocentricity attributes they possess. Hassan's (1974) study also manages to contribute to the understanding that the language (in general) has more endocentric compounds than exocentric ones (as he observed from his corpus).

On the other side, we find several claims made by Hassan (1974) to be somewhat inaccurate. Take, for instance, the categorisation of compounds with non-transparent components, i.e. the *panjang tangan* (long + hand) 'thief' types. In general, Hassan categorises these compounds as *semantically exocentric compounds*. From the labelling itself, it is quite clear that Hassan is saying that these sorts of structures have their heads 'outside' of the construction (hence the term *exocentric*), and that the meaning of the construction is dependent on one's 'knowledge' of the language to understand them (hence the term *semantically*). However, compare, for instance, *panjang tangan* 'thief' with a combination such as *rabun ayam* (short/long-sighted + chicken) 'short/long-sighted (only) during dusk'. The degree of semantic opaqueness in *panjang tangan* is obviously higher than that of *rabun ayam*. In other words, the meaning of *panjang* (long) and *tangan* (hand) is not literally related to the meaning of 'a thief', but *rabun ayam* (although partially opaque) is literally related to the state of being *rabun* 'short/long-sighted'). In this sense, it is important to distinguish between and/or separate compounds that are fully (non-transparent) idiomatic and those that are partially (transparent)

idiomatic. To this end, we will say that Hassan (1974) did not set a clear boundary between these types of compounds in the language.

On another point, Hassan (1974) also claims that compounds in the language can be of prepositional structures. For instance, *dari pada* (from + in the direction of) = *daripada* 'from somewhere', etc. Although the examples given are indeed single preposition units, i.e. *dari* (from) and *pada* (in the direction of) are prepositions in their own right, nonetheless, most scholars recognise *daripada* (and the others alike) as single preposition units as well (and not as a combination of *dari* and *pada*).This is because every one of these prepositions carries a different function (or at least slightly different) when preceding nominal structures (Karim et al., 2008). Furthermore, these sorts of propositional constructions are not only limited in numbers, but they are also constrained in the sense that prepositions never combine with any other lexical categories in the language to form new compounds. To this end, we will not support the idea of prepositional combinations as true compounds of the Malay language. We leave Hassan's (1974) work here only to continue with more of his other works on the subject of compounding.

4.2 Abdullah Hassan's (1986) *Penerbitan kata dalam bahasa Malaysia*

Hassan (1986) opens his discussion on Malay compounding with the statement below:

"Compounding has become a controversial issue in recent years. Some believe the Malay language does not have compounds at all, and that the forms considered as compounds are

merely combination of words. What separates compounds from word combinations is rather

unclear as well".[50]   Hassan (1986: 180)

This statement encapsulates how linguists (such as Hassan) are aware of the controversial issues surrounding the Malay compounding phenomenon. Hassan (1986) reiterates a seminar that he attended in 1982 (by the National Language and Literary Agency), discussing the issues and controversies of Malay compounding. The outcome of the seminar detailed a list of nine criteria of compound words shown below.

(26) Criteria of compoundhood [51]

(a) Consisting of at least two free morphemes.

(b) Combination of constituents does not involve any affixation.

(c) The relationship between constituents is not directly correlated to Head-Modifier: the two constituents are not Head and Modifier but the sequence is still Head-Modifier.

(d) Production of a compound word is not in relation to "and" in between the constituents.

(e) The meaning is dependent on the constituents, and from this the specific features of the meaning are added to.

(f) Cannot be separated by insertion between the constituents.

(g) Compound (should be) written as one word; reduplication of it must also then be as a whole.

---

[50] Translated from Hassan (1986: 180): "*Kata majmuk telah menjadi satu isu kontroversi dalam masa kebelakangan ini. Ada pihak yang mengatakan bahasa Malaysia tidak mempunyai kata majmuk dan apa yang dianggap kata majmuk itu merupakan gabungan kata saja. Apa yang membezakan kata majmuk dan gabungan kata pula agak kabur juga*".

[51] These criteria are our own translation from Hassan's reiteration of the Malay compounding criteria discussed in a seminar held in 1982 by *Dewan Bahasa dan Pustaka* (cf. Hassan, 1986, p. 81).

(h) The Head-Modifier relationship in a phrase is more direct than the Head-Modifier relationship in compounds.

(i) The characteristic of the meaning comes from the "outside" and not directly from the Head-Modifier relationship, and the order of words is "grammatical".

Accordingly, Hassan (1986) commented on every point presented above as below:

(27) Hassan's (1986) comments on the proposed criteria of compoundhood

(a) Neither constituent particularly has to be a free morpheme, e.g *keluar* (*ke* 'to' + *luar* 'out') 'to emerge/go out'.

(b) The existence of affixation does not define a compound, e.g *surat khabar* (letter + news) = *suratkahbar* 'newspaper', *per-suratkhabar-an* 'press/journalism' (affix+compound+affix).

(c) There are some compound words that follow the head-modifier order like *rumah sakit* (*rumah* 'house' (H) + *sakit* 'sick/ill' (M)) 'hospital' and there are some that do not, like (*maha* 'great' (M) + *guru* 'teacher' (H)) *mahaguru* 'great teacher/professor'.

(d) On point (e), the meaning of a compound is indeed dependent on the Head, but this is not always the case in exocentric compounds and idioms (which Hassan considers as compounds, as mentioned in his previous work).

(e) On point (g), the orthography must not be a criterion to define compounds as this is a stylistic convention which has nothing to do with the morphology of the phenomenon.

(f) Hassan is unclear on point (h). According to Hassan (1986), the head-modifier rule was introduced by Malay linguists to assist phrase sequencing in the language. Although compounds do adhere to such sequencing, it should not be imposed as it is a syntactical concept more than it is a morphological one. Counter examples are *menteri* 'minister' (H) +

besar 'large/big' (M) = *menteri besar* 'chief minister' vs. *perdana* 'prime' (M) + *menteri* 'minister' (H) = *perdana menteri* 'prime minister'.

(g) On point (i), Hassan is also unclear on what is meant, so does not elaborate on it.

From his comments above, Hassan (1986) only truly supports the proposed criteria of (d) and (f), which are essentially the same point, i.e. a compound word must be in unity and work as a single entity, and not able to be separated by other forms. Hassan (1986) leaves his comments on the matter as they stand, and carries on discussing his take on Malay compounding.[52]

Let us begin by mentioning that, in this part of Hassan's (1986) work, he makes quite a lot of reference to his previous study, *The Morphology of Malay* (1974). Therefore, explanations and examples that are similar to his previous study will not be thoroughly repeated here. The intention here is to present any sort of development and/or changes in the subject in comparison to the previous study.

We can divide the discussion on compounding in Hassan's (1986) work into three main topics, namely: (i) characteristics of compounding, (ii) orthography/spelling convention, and (iii) types of compounds.

(i) *The characteristics of compounding*

With regard to the characteristics of compounding, Hassan (1986) elaborates further on four subtopics, namely:

---

[52] Our comments on these criteria will be given later, in section 4.2.1.

(a) *Syntactic or asyntactic construction*

A compound word is said to be a syntactic compound when it adheres to the head modifier construction sequence, and if otherwise it is said to be an asyntactic compound.

(b) *Endocentric or exocentric construction*

An endocentric compound is said to be a compound belonging to the same word class as its head constituent, while an exocentric compound does not belong to the same word class (or have the same syntactic function) as any of its constituents.

(c) *Semantically exocentric constructions*

Some compounds may be regarded as grammatically endocentric but semantically exocentric. These types of compounds are essentially idioms.

(d) *The unity of compound words*

In Hassan's point of view, compound words are clearly distinct from phrases, i.e. compounding is a morphological construction while a phrase is a syntactical construction. Hassan believes that the difference between them can be seen through the function within sentences, and open/close-endedness.

Function

The syntactic functions of compound forms are elements of/for phrases, i.e. compounds are words and words feeds syntax, while phrases are elements of/for sentences.

Open/close-endedness

The composition of the constituents cannot be extended. For example, *rumah sakit* (house + sick/ill) 'hospital' cannot be extended without changing its meaning, such as *rumah*

yang sakit 'house which is sick/ill' (*yang* - relative clause marker). In this sense, compound forms are close-ended. On the other hand, phrases can be extended, e.g. *budak gemuk* (kid + fat) becomes *budak yang gemuk* 'a fat kid', making phrases open-ended constructions.

The composition of the constituents is also limited. For example, the compound *rumah sakit* 'hospital' can be used as an analogical structure to derive similar patterns, such as *rumah rehat* (house + rest) 'rest house', *rumah makan* (house + eat) 'restaurant', *rumah urut* (house + massage) 'massage parlours', *rumah ibadat* (house + worship) 'house of worship', and a few others, but is nonetheless limited. On the other hand, as for the pattern of a noun + adjective phrase such as *budak gemuk* 'fat kid', any noun can be combined with an adjective to produce hundreds of similar phrases, *budak kurus* (kid + skinny) 'skinny kid', *budak hitam* (kid + black) 'black kid', *budak nakal* (kid + naughty) 'mischievous kid', *budak kecil* (kid + small) 'small kid', and so on and so forth, making the construction an open structure.

(ii) *Orthography/spelling convention*

According to Hassan (1986), the spelling of Malay compounds is largely dependent on stylistic conventions, as some are written as separate constituents (spelled as separate words) while others are combined (spelled as single words). However, what is more important than the convention is the fact that the constituents of compounded forms are interconnected synthetically (and semantically). Hassan (1986) emphasised the importance of understanding this concept (i.e. the connection between the constituents), rather than the stylistics of compounding orthography in determining whether a given form is a compound in the language.

(iii) *Types of Malay compounds*

Hassan (1986) carries on describing nominal, verbal, adjectival and adverbial compounds, very similar to the description and examples in his previous (1974) study. There

is a slight difference, however, as he no longer consider adjectives as stative intransitive verbs, i.e. acknowledging the lexical entry of 'adjectives' in the language, giving adjectival compounds their own subsection in this current work. Hassan still recognises function words like prepositions *daripada* (from somewhere) and *kepada* (towards) as compounds when other scholars do not. The following subsections summarise the types of compounds described by Hassan in this (1986) work.

(a) *Nominal compounding*

The definition given of a compound is a combination of two or more words forming a new syntactic nominal entity. Similar to his previous study, Hassan (1986) recognises both endocentric and exocentric nominal compounds. Endocentric nominal compounds are divided into two types. The first type adheres to the syntactic rules of the language, i.e. the head precedes the modifier, the *kapal terbang* (ship + fly) 'aeroplane' types. The second type goes against the rules, compounding as a modifier-head sequence, the *perdana menteri* (prime + minister) 'prime minister' types.

Exocentric nominal compounds are categorised as syntactically exocentric and semantically exocentric. The syntactically exocentric nominal compounds are formed by a transitive verb (head) + object/complement (modifier), e.g. the *sapu tangan* (to wipe + hand) 'handkerchief' types. On the other hand, semantically exocentric nominal compounds are exocentric in terms of meaning, the *panjang tangan* (long + hand) 'thief' types.

(b) *Verbal compounding*

Hassan divides verbal compounds into endocentric and syntactically/semantically exocentric forms as well. In endocentric forms, the head precedes the modifier, the *ikat mati* (to tie + (die) completely) 'dead knot' types. Syntactically exocentric verbal compounds are

formed by combining a verb (head) + object/complement (modifier), the *buang negri* (to banish + state) 'expatriate' types. Finally, semantically exocentric verbal compounds (idioms) are constructions that are exocentric in meaning, the *tumbuk rusuk* (to pound + flank) 'to bribe' types.

(c) *Adjectival compounding*

This section is very concise. Hassan provides hardly any examples or detailed elaboration. Nonetheless, the basics are similar to the nominal and verbal compounds. Syntactically endocentric adjectival compounds typically consist of an adjective (head) + adjectival or nominal (modifier), the *biru tua* (blue + old) 'dark blue' types. As Hassan did not discuss this type of compound in detail, reference for the other types of adjectival compounds is deduced from his previous (1974) study (cf. asyntactically endocentric, syntactically exocentric and semantically exocentric adjectival compounds respectively).

(d) *Adverbial compounding*

Similar to adjectival compounds, adverbial compounds are not particularly discussed, perhaps because they do not occur very often. Nonetheless, the basics are similar, i.e. formed by adverbial (head) + other lexical classes (modifier). Again, references are from his previous work (cf. syntactically endocentric and asyntactically exocentric adverbial compounds respectively).

4.2.1 Comments

Hassan's (1986) opening statement is important as it asserts the fact that the topic of compounding must not be overlooked in Malay morphological studies. The statement recognises two main issues, namely: (i) the compounding phenomenon is (to a certain degree)

a controversial subject in the Malay language, and (ii) Malay linguists are aware (or at least should be aware) of the controversy. Keeping this in mind, Hassan (1986) continued his discussion by scrutinising the proposed nine criteria of Malay compound words mentioned above.

Out of the nine points given, Hassan (1986) basically only agrees with one of them, that is the compounds being a unified structure, hence they cannot be separated by other forms. As for the rest of the proposed points, he basically argued otherwise by providing counter-examples. On the first point, Hassan (1986) argues that compounds do not necessarily have to be free morphemes. He exemplifies prepositional compounding such as *ke* 'to' + *luar* 'out' = *keluar* 'to emerge/go out', as evidence of compounds not having a 'free morpheme' constituent. To this end, we disagree with Hassan (1986) that prepositions are units of compounding based on the same reasoning that we argued for in section 4.1.1. The second point argues that affixation should not be involved in compounding. We agree with Hassan on this point: that there is no reason why compounded structures cannot undergo affixation, e.g. s*uratkahbar* 'newspaper' affixed to *per-suratkhabar-an* 'journalism' is not a problem. The third point concerns the relationship between constituents, i.e. whether the constituents involved have a modified-modifier relation. On this issue, we will agree that, in general, the constituents of Malay compounds have the modified-modifier correlation. Those that have the opposite correlation (e.g. *maha* 'great' (modifier) + *guru* 'teacher' (modified) *mahaguru* 'professor') are in fact limited in number due to their origins of being a borrowed form, and thus they are in a way an exception to the phenomenon. The fourth point is that the constituents of a compounded form should not have the relation 'and' between them. This is basically saying that Malay does not have coordinated compounds. We do not agree with this point. As will be seen later on in this chapter and in the analysis of Malay compounding, it is quite apparent that Malay does have the potential to produce coordinative compounds.

The fifth point is that the meaning of a compound is dependent on its head. We agree with Hassan on this as a compound usually is the hyponym of its head. In the case of exocentric compounds, the meaning or the head is not necessarily dependent on the constituents involved. The sixth point is that the constituents of compounds cannot be separated by insertions. We agree with this point, as it is one of the ways to show the unity that exists between the components of a compounded structure. The seventh point concerns the orthography of Malay compounding. On this point, we agree with Hassan (1986) and other scholars (cf. 2.2.4) on how using orthographic criteria to define compounding is not a reliable method in a language such as Malay. The eighth point argues that the head-modifier relationship of a compounded structure is less direct in comparison to the head-modifier relationship of a phrasal structure. As with Hassan (1986), we agree that the relationship between the head-modifier of compounds should be viewed as a distinct relationship from that of phrases, even if they look the same. This is because the headedness of compounds and that of phrases are essentially two different entities (cf. section 2.2.3 on syntactic vs. morphological heads). The final point is unclear to Hassan (1986). We believe that it means to include the idea of endocentric vs. exocentric compounding (hence the use of meaning coming from the 'outside' (cf. (26) (i)). If this is indeed what it is intended to mean, than we are in agreement that compounds should be distinguishable (at the very least) in terms of endocentric and exocentric types.

It is clear that attempting to set a list of criteria for Malay compoundhood is not as easy as it may seem. Having a rigid list of criteria of 'what a compound should be/look like' can actually make matters more complicated (as evident from the discussion above). Some measures can be unclear, while others can be outright rejected. Clearly, creating a workable list of criteria is not merely the process of listing as many measures as possible. There needs to be a balance not only in the quantity but also in the quality (i.e. the significance) of the measures involved in defining compoundhood.

We now move on to comment on the next subsection in Hassan's (1986) work. As mentioned, the subsection echoes his previous (1974) study, and therefore we only looked at the more 'current' point of view on the subject matter. Hassan (1986) divides the subsection into three main topics discussing issues of compound characteristics, orthography/spelling, and the different types of compounds available.

With regard to the characteristics of compounding, Hassan (1986) addresses the issues of endocentric and exocentric compounds, which also includes the issue of headedness. These are the same issues that he raised and touched upon in his previous study, and on which we have also commented (cf. 4.1.1). Similarly, Hassan (1986) still recognises idioms as what he terms 'semantically exocentric compounds'. Again, believe that there should be a separation between fully idiomatic compounds and partially idiomatic compounds in the language (cf. 4.1.1).

Still within the section of compounding characteristics, Hassan (1986) gives some emphasis to the issue of compounds vs. phrases this time around. In his view, Malay compounds are clearly distinct from phrases. He views compounds as morphological constructions, while phrases are syntactical constructions. What this means is that Hassan (1986) understands compounds as elements of phrases, i.e. like words, compounds feed syntax. This understanding is very similar to Ackema and Neeleman's (2004) point of view where compounds are seen as atoms that can be inserted into syntactic structures (cf. 2.1).

Hassan (1986) also argues that compounds can be differentiated from phrases based on their ability to build their construction, i.e. compounds are 'close-ended' in comparison to phrases, which are more 'open-ended'. This basically means that the constituents of a given compound are unified in a way that constituents of a phrasal form are not. For instance, the meaning of a compound cannot be expanded by inserting additional elements in between its

constituents. On the other hand, adding an extra element within a phrasal structure is not a problem. We agree with this understanding of compounds as unified entities.

Furthermore, compounds are argued to be 'close-ended' in the sense that the expansion of compounds can be limited. Hassan (1986) exemplifies that the compound *rumah sakit* 'hospital' can analogically produce a limited set of similar patterns, such as *rumah rehat* (house + rest) 'rest house', *rumah makan* (house + eat) 'restaurant', and a few others, while phrases are not constrained by this limitation, e.g. *budak gemuk* 'fat kid', *budak kurus* (kid + skinny) 'skinny kid', *budak hitam* (kid + black) 'black kid', so on and so forth. We disagree with Hassan (1986) on this point. There is no reason to believe that compounds should be limited in generating other compounds based on an analogy from an existing compound form. In fact, we believe that it should be one of the natural courses of development for the phenomenon. Furthermore, compounds also have the ability to be recursive (cf. 2.1). This means that a given compound is more than capable of undergoing expansion, and should not be limited by such constraint as above.

In the final section of his work, Hassan (1986) discusses the topic of orthography/spelling and the different types of Malay compounds. With regard to the issues of orthography and spelling convention, we agree with Hassan's (1986) argument that it is more important to recognise that it is the bond between the constituents which defines whether a given combination is a compound, not the manner in which it is written or spelled. Orthography and/or spelling are not a reliable indicator of compoundhood in Malay (cf. 2.2.4). On the topic of types of compounds, the discussion in Hassan (1986) is very similar to the description and examples in Hassan's (1974) study (compare 4.2 (iii) against 4.1 above). There is not much that we can add here that we have not said in 4.1.1. Therefore, we now move on to the next study by Hassan (2006).

4.3 Abdullah Hassan's (2006) *Morfologi*

Hassan's (2006) work is essentially a reiteration with some alterations made (as an update) to his previous (1986) work discussed above. Therefore, our aim here once again is to highlight the changes.

In general, Hassan (2006) made a number of changes in this current work in relation to several issues. There is a notable difference in his discussion on the characteristics of compounding With regard to 'the unity of compound words' (cf. 4.2 (i) (d) above), Hassan (2006) no longer acknowledges certain constructions as compounds. In his previous work, Hassan (1986) exemplified how the noun *rumah* (house), for instance, can be used in a similar pattern to the compound *rumah sakit* (house + sick/ill) 'hospital', to analogously derive other similar structures such as *rumah rehat* (house + rest) 'rest house', *rumah makan* (house + eat) 'restaurant', *rumah urut* (house + massage) 'massage parlours', etc. Hassan (1986) accepted all of these constructions (deriving from the head *rumah*) as compounds.

However, in his 2006 work, he no longer acknowledges such analogical constructions as compounds. Instead, he argues that they merely have the appearance of a compounded structure, but in fact they are actually phrases. Hassan (2006) loosely explains that the insertion of *untuk* (for) and *tempat* (place) between the constituents of such constructions will not disrupt the meaning of the construction (proving that the constituents involved are not unified like those of compounded constituents). Hence, they are not to be considered as compounds, but as phrases, i.e. *rumah untuk rehat* (house for resting)*, rumah tempat makan* ((house) place for eating), *rumah tempat urut* ((house) place for massages), etc.

Hassan (2006) gives a lot more attention to the issue of spelling convention in this work in comparison to the previous one (cf. 4.2 (ii) above). He lists 27 compounds found in the *Tatabahasa Dewan* (Karim et al., 2004) and 34 compounds found in the *Tatabahasa Dewan*

(Karim et al., 2005),[53] comparing their spelling conventions. Hassan (2006) points out that, when a compounded form is spelt separately in isolation, it will be spelt together in reduplicated form. For example, the compound *uji kaji* (to test + to study) 'experiment' will be spelt as *ujikaji-ujikaji* 'experiments' when reduplicated. This shows that the constituents are bonded as a single entity (which is why they will reduplicate as a whole even when their constituents are initially separated by spelling convention). On the other hand, Hassan (2006) argues that a phrasal structure will not be reduplicated in this way, e.g. *pasar malam* (market + night) 'night market', becomes *pasar-pasar malam* (market-market night) 'night markets'. Hassan considers this as yet another method to differentiate between compound and phrases.

Based on this view, Hassan (2006) also rejects his previous acceptance of idioms as compounds in the language (cf. 4.2 (iii) above). As Hassan (2006) now holds the view that compounded forms must be reduplicated as a whole, idioms can no longer be considered as compounds because they do not comply as such (e.g. *buah hati* (fruit + heart) 'sweetheart' becomes *buah-buah hati* 'sweethearts'). Therefore, there are no longer examples of idiomatic structures in any of the major lexical categories (i.e. nominal, verbal, adjectival and adverbal compounding) in Hassan's (2006) study.

4.3.1 Comments

It is apparent that some of the readjustments made here are attempts to refine the general concept of Malay compounding. We begin by addressing the issue of how Hassan (2006) tries to 'rectify' his previous understanding on generating (new) compounds from the analogy of existing ones. We find this argument slightly unnecessary. It is obvious that compounds can

---

[53] *Tatabahasa Dewan* is the main dictionary of standard Malay in Malaysia (similar to the OED by analogy).

occur via analogy that has arisen form an existing compound. This is indeed one of the reasons why compounds are very productive as a word formation process. Curbing (or at worst rejecting) this idea may cause stunt the productivity of the Malay compounding scene. We do not see any particular reason to do so.

However, we can assume that the real reason behind Hassan's (2006) argument on this issue is to illustrate how true compounds can be differentiated from look-alike structures. This issue is indeed important as the structures of compounds and phrases in Malay are very similar. In this sense, Hassan's (2006) argument on this issue is not without significance, as it is important for us to be able to differentiate between structures that are compounds and those that are not. In order to do so, Hassan (2006) is basically proposing a way to identify Malay compounds (and phrases) via the insertion test as previously mentioned.

However, with regard to the element of insertion to prove the test, the nature of the elements involved is unclear to us. From our observation, there is really no specification of the type of 'insertion' that is allowed between the constituents. From our observation, Hassan (2006), Karim et al. (2008) and Ismail and Jalaluddin (2008) have collectively proposed the insertion of *yang* 'which', *dan* 'and', *untuk* 'for', *dari* 'from', *kepada* 'to', *milik* 'own', *antara* 'between' and *tempat* 'place' (among others), as elements of testing. None of the linguists mentioned specify the type or limits the number of elements allowed for the testing. It seems very much a 'free' insertion as long as it makes sense to show that a given structure can be interpreted as a phrase or otherwise. Following this logic, we can easily argue that even a true compound such as *rumah sakit* (house + ill) 'hospital' can be filled with certain insertions to 'test' it into a phrase, e.g. *rumah untuk orang sakit* (house/home - for - people - ill) 'home for the sick'. This is similar to the exemplification Hassan (2006) made for *rumah makan* (house + eat), i.e. by inserting *rumah tempat makan* (house - place - eat) to derive the phrase 'a (house) place for eating' (denying the structure *rumah makan* (house + eat) of being a compound

'restaurant' instead). In our opinion, identifying compounds (and/or phrases) via the insertion test is useful provided that we understand the limitations of the test.

With regard to the discussion on compound spelling, we can once again conclude that Hassan (2006) is trying to put forth the idea of testing Malay compoundhood via reduplication, which also translates into a method of identifying phrases. The main aim of the test is to prove that the components of a compounded form are united/bonded in such a way that upon reduplication the components involved will all be conjoined and reduplicated together as a whole. This idea is supported by the observation about how compounds with initially separated components (by spelling convention) will join together when fully reduplicated, i.e. the example given of *uji kaji* (to test + to study) 'experiment' reduplicated as *ujikaji-ujikaji* 'experiments', and not as **uji-uji kaji*. This signifies unification between the two elements. Instead, a structure such as *pasar malam* (market + night) 'night market' can be argued as a phrase (and not a compound) because only the (phrasal) head component undergoes reduplication, i.e. *pasar-pasar malam* 'night markets'. This arguably signifies that there is weaker bond between the two elements involved.

Indeed, there is weight to this argument. If the elements in a structure are truly fused together, then it makes sense that the structure is also reduplicated as a single entity. However, this criterion can be problematic because there are many true compounds that only require the head component to be reduplicated. Again let us take the typical *rumah sakit* (house + sick/ill) 'hospital' as an example. Following the proposed test, the compound *rumah sakit* would have to be reduplicated as **rumahsakit-rumahsakit* to derive the plural form 'hospitals'. Any native speaker will say that this is an odd thing to do. It is odd enough even to pluralise the compound as *rumah sakit - rumah sakit* for 'hospitals', let alone to reduplicate it as required by the proposed test, i.e. **rumahsakit-rumahsakit* for 'hospitals'. As such, reduplicating the head noun (i.e. *rumah-rumah sakit* 'hospitals') is still the preferred pattern. Therefore, if we were to

adhere to the proposed test, *rumah sakit* will lose its status as a compound. To this end, we will agree that using reduplication as a test can be useful, and at the same time unreliable, to a certain degree. Further refinement is needed if the reduplication test is to be taken seriously as a solid test of Malay compoundhood.

4.4 Hashim Musa's (1993) *Binaan dan fungsi perkataan dalam bahasa Melayu*

Musa (1993) adopts four characteristics of compounding in his attempt to define compoundhood in Malay.

(28) Characteristics of compounds[54]

    (i) Disallows insertion of any element between its constituents.

    (ii) Produces a single unified form which has a different meaning to the literal meaning of all its constituents.

    (iii) Modifier is not of predicative or genitive status.

    (iv) Meaning of one of its constituents is not a literal or specific meaning.

Adhering to the characteristics in (28) above, Musa (1993) recognises several different types of Malay compounds. The compounds are of the main lexical categories, namely nominal compounds (e.g. *papan* (board) + *hitam* (black) = *papan hitam* 'blackboard'), verbal compounds (e.g. *kenal* (to recognise) + *pasti* (sure) = *kenal pasti* 'to identify') and adjectival compounds (e.g. *kelabu* (grey) + *asap* (smoke) = *kelabu asap* 'smoky grey').[55] However, Musa

---

[54] Oddly, the principles of (28) are an adaptation of English compounding principles.
[55] Cf. appendix (2) for Musa's (1993) list of compounds.

also recognises that there are many combinations in the language that seem to be like compounds but do not comply with all of the four principles in (28) above, for example:

(29) Examples of compound-like combinations

Category	Compounds
Nominal	*aneka warna* (variety + colour) 'multicolour'
Verbal	*kahwin lari* (to marry + to run) 'elope'
Adjectival	*suka rela* (like + willing) 'voluntary'

Musa (1993) argues that the examples in (29) are more compound-like than they are phrasal-like structures in the language. Thus, as he points out, if one adheres to the characteristics in (28), the examples in (29) cannot be considered as compounds. This is because they violate, among others, the principles of (28 (ii)) where they do not have a different meaning to the literal meaning of all the constituents involved, and they also violate the principle of (28 (iii)) where their modifier is of a predicative or genitive status.

As such, Musa (1993) deems that the concept of 'compound words' as characterised by the principles in (28) cannot work in the Malay language. He argues that the characteristics listed in (28) are too restrictive for Malay compounding (point in case, it will not be able to account for the compound-like structures in (29)). Therefore, he opts for the term *kata gabungan* (lit. 'word + combination') instead of the term *kata majmuk* (lit. 'word + compound') which is commonly used by other Malay linguists. In his point of view, the term *gabungan* (combination) is more flexible to cover a much larger syntactic relation and meaning between a given formation, rather than the strict meaning of *majmuk* (compound).

Musa's (1993) overall discussion on Malay compounding touches on several issues, which include the topics of spelling, headedness and compound meaning.

(30) Issues of Malay compounding

(i) *Spelling*

Malay compounds should be conjointly spelled due to:

(a) common prevalence, e.g. *matahari* (eye + day) 'sun'

(b) borrowing influence, e.g. *fotokopi* (photo + copy) 'photocopy'

(c) scientific terminologies, e.g. *isipadu* (content + solid) 'volume'

Malay compounds should be spelled separately due to:

(a) direct borrowing, e.g. *universal konkrit* (universal + concrete) 'universal concrete'

(b) translated borrowing, e.g. *roda tenaga* (wheel + energy) 'flywheel'.

(ii) *Head-modifier convention*

Musa recognises both endocentric and exocentric compounds, touching briefly on the modified-modifier concept, i.e. headedness and modifiers, e.g. endocentric *purba* (H) + *kala* (M) (ancient + time) = *purbakala* 'ancient times', and exocentric, *urus* + *setia* (to manage + loyal) = *urusetia* 'secretariat'.

 (iii) *Compound meaning*

In Musa's (1993) point of view, word combinations have different degrees of transparency, and the degree directly affects the meaning of the combination. In other words, the meaning of a combined form can either be transparent (i.e. meaning can be literally derived from the constituents involved, e.g. phrases), or opaque (i.e. meaning cannot be derived from

the constituents involved, e.g. idioms). A compounded structure falls in between this scale of transparency. For instance, the combination of *mata lembu* (eye + cow) 'cow's eye' is considered as a phrase because both of its constituents have a (literal) transparent meaning, i.e. the structure produces the meaning 'eyes belonging to (a) cow'. On the other hand, a combination such as *mata air* (eye/point + water) 'well/spring' is considered as a compounded form because one of its constituents has a (literal) transparent meaning (i.e. *air* 'water'), while the other has a (figurative) opaque meaning (i.e. *mata* 'eye/point'), i.e. the structure produces the meaning of 'a hole/point in the ground where water emerges'. Compare this to the combination of *cahaya mata* (light + eye) 'child'. This combination is considered as an idiomatic construction since both of its constituents have a (figurative) opaque meaning, i.e. the structure *cahaya mata* (light + eye) produces the meaning 'child' (and not 'light from eyes').

4.4.1 Comments

Musa (1993) begins by addressing the commonly use term of *kata majmuk* (lit. 'word' + 'compound') as an inaccurate representation of the Malay 'compounding' phenomenon. He argues that *kata majmuk* is a concept based directly on the principles of English compounding, and that is why compounding has been inaccurately represented in the language (cf. (28)). According to Musa (1993), the principle is too restrictive for Malay compounding, as it fails to account for the many other compound-like constructions in the language (c.f. compound-like examples in (29) not complying with the principles of (28) (ii) and (iii)). To this end, he suggests a change to *kata gabungan* (lit. 'word + combination') as a better representation of the Malay compounding phenomenon.

There are several problems with Musa's (1993) discussion on this issue. Firstly, despite pointing out the drawback of adhering to the principles in (28), he did not actually elaborate on any alternative criteria that are specifically needed for Malay compounding. Instead, we are forced to assume certain things from his discussion. For instance, in order to allow structures in (29) to be interpreted as compounds, Musa (1993) needed to reject some principles in (28) that are preventing the structures in (29) from becoming 'compounds'. Two principles in (28) were rejected, namely: (ii) a compound has a different meaning to the literal meaning of all its constituents, and (iii) compound modifiers are not of predicative or genitive status. This leaves us with the remaining principles of (28): (i) disallows insertion of any element between its constituents to maintain unity, and (iv) the meaning of one of the constituents is not a literal/specific meaning. We are left to assume that Musa (1993) only considers principles (i) and (iv) of (28) as the remaining suitable criteria for Malay compounding.

Having said this, it is confusing that Musa (1993) initially uses all of the principles in (28) to classify the nominal, verbal and adjectival Malay compounds, before he begins to address the issue of compound-like examples in (29). Therefore, it is not clear to us whether he utilises a different set of principles to classify nominal, verbal and adjectival compounds, and yet another set of principles to classify the other types of Malay compounds. Musa (1993) should have clearly discussed the exact criteria that he used, and that he believes are needed for Malay compounding. Therefore, his argument on the importance of changing the term and concept of *kata majmuk* (compound words) to *kata gabungan* (combined words) to address Malay compounding phenomenon is actually insignificant. Clearly, it is more important to properly define and characterise what a compound actually is in Malay, which Musa (1993) arguably did not manage to do.

Musa (1993) also raised several issues of Malay compounding, namely of spelling, headedness and compound meaning (cf. (30)). With regard to the issue of spelling (cf. (30) (i)),

it is not clear to us what he meant by compounds having combined spelling due to borrowing influence (e.g. *fotokopi* (photo + copy) 'photocopy'), and at the same time compounds are also spelled separately due to direct borrowing (e.g. *universal konkrit* (universal + concrete) 'universal concrete'). There is no clear reason why one should have a different spelling convention from the other given that both are borrowed compounds. Similarly, not all scientific terminologies in the language must have a combined spelling convention as illustrated by Musa (1993) (cf. (30) (i)). We will once again agree with Hassan (1986, 2006), along with other similar arguments from other scholars, that spelling is not a reliable property of compoundhood (cf. 2.2.4)

On the issue of headedness, Musa (1993) touches on the concept of head-modifier relation, hence recognising the endocentric and exocentric distinction in Malay compounding. Similar to Hassan (1986, 2006), Musa (1993) associates the concept of headedness (i.e. head-modifier relation) in Malay compounds as a concept that originated from syntax and was adapted into morphology. Furthermore, Musa's (1993) point of view on morphology is based on a lexicalist foundation. Therefore, he understands that morphological heads (such as heads in compounds) are not the same entity as syntactical heads. This thesis takes a similar view on the issue of compound headedness. As discussed in Chapter 2 on the issue of headedness, we agree that the properties of morphological complex heads in general should not be equal to, or subsumed under, syntactical heads. They both may exhibit similar properties at times, but nonetheless they are two distinctive entities (cf. 2.2.3).

On the issue of compound meaning, Musa (1993) analyses compounds as structures with meanings which fall somewhere in between the two ends of structures with totally literal meanings and structures with totally opaque meanings (cf. (30) (iii)). Through this point of view, he allows for a rather neat identification of compounded structure in contrast to phrasal and idiomatic structures. This sort of analysis is somewhat different from other Malay scholars

(e.g. Hassan (1986) and Karim et al. (2008)), who basically consider idiomatic constructions as compounded structures as well. Furthermore, through such analysis, Musa (1993) is also able to accept that a given combined structure can have two different meanings or interpretations (i.e. as either a compound or a phrase structure).

(31) Examples of differences between compounds and phrases in Musa's (1993) discussion

Combination:	*ibu* (mother) + *bapa* (father)
Compound reading:	[*ibu bapa*]$_N$ = 'parents'
Phrasal reading:	[*ibu bapa*]$_{NP}$ = 'mother and father'
Combination:	*papan* (board) + *hitam* (black)
Compound reading:	[*papan hitam*]$_N$ = 'blackboard'
Phrasal reading:	[*papan hitam*]$_{NP}$ = 'a board which is black (in colour)'
Combination:	*tanda* (mark) + *tangan* (hand)
Compound reading:	[*tandatangan*]$_N$ = 'signature'
Phrasal reading:	[*tanda tangan*]$_{NP}$ = 'mark of (a) hand'

It is interesting to see Musa's (1993) analysis of this when other scholars rigidly opt for one over the other, and are also at times unclear about the division between compounds and phrases (e.g. Karim et al. (2008), cf. section 4.6.1 in this chapter). We agree with Musa on the idea that combined structures in Malay are flexible enough to have multiple readings. In other words, one should not treat Malay combined structures as rigid formations of either compound or phrase structures. This echoes the argument of Ackema and Neeleman (2004) on how the combination of structures can equally take place as either a morphological construction or a syntactical construction. The outcome of the competition between them is what allows them to be born as either a compounded or a phrasal construction (cf. 2.1). For now, it is sufficient to say that we are in favour of this idea. The flexibility of combined structures in Malay to be

interpreted either as compounded or phrasal structures can be taken as an indication that combined structures in the language may perhaps be in competition with each other (along the lines of Ackema and Neeleman's (2004) competition proposal, cf. 2.1). We will have more to say about this matter in Chapter 5.

We end this summary of Musa's (1993) work by reiterating two significant points from his study, namely: (i) a clearer idea (in comparison to other Malay scholars) on separating compounds and idioms in Malay, and (ii) the recognition that word combinations can be a compounded construction or a phrasal construction based on specific readings. We now move on to the next study on Malay compounding.

4.5 Nik Safiah Karim's (1995) *Malay grammar for academics and professionals*

Karim (1995) defines Malay compounds as the typical combination of two or more words creating a new form with a meaning of its own (and behaving as a single entity). Her section on compounding is essentially an exemplification of the three main types of compounds available in the language, namely the nominal, verbal and adjectival compounds.[56]

(i) *Nominal compounds*

Karim (1995) exemplified four types of nominal compounds:

Free forms

> *guru besar* (teacher + big/large) 'head teacher', *tengah hari* (middle + day) 'midday', etc.

---

[56] The full list of Karim's (1995) compound examples is given in appendix (3).

Academic or scientific terminologies

analisis teks (analyse + text) 'text analysis'*, keluarga asas* (family + basic) 'nuclear family', etc.

Idiomatic expressions

anak emas (child +gold) 'favourite person', *cakar ayam* (to scratch + chicken) 'scribbling', etc.

Established forms

Karim (1995) stated that there are 13 established compounds which by convention are spelt as single entities, among others: *bumiputera* (earth + son/prince) 'native', *kakitangan* (foot + hand) 'staff', *kerjasama* (work + together) 'cooperate', *warganegara* (people + country) 'citizen', etc.

(ii) *Verbal compounds*

According to Karim (1995), compound verbs form a very small proportion of Malay compounds. Some examples from Karim (1995) include *kenal pasti* (to recognise + sure) 'to identify', *ambil alih* (take + to remove) 'to replace', *tolak ansur* (to push + gradual) 'to tolerate', etc.

(iii) *Adjectival compounds*

Some examples given of adjectival compounds include *hijau daun* (green + leaf) 'leafy green', *merah tua* (red + old) 'crimson', *muda remaja* (young + teenage) 'youthful', etc. According to Karim (1995) many idiomatic expressions in Malay are also formed through adjectival compounding, e.g. *panjan tangan* (long + hand) 'thief', *ringan tulang* (light + bone) 'hard working', *terang hati* (bright + heart) 'clever', etc.

(iv) *Compound spelling*

According to Karim (1995), with the exception of the established compounds in ((i) c) above, Malay compounds in general are spelt (written) separately. However, she points out that affixation can have an effect on this general spelling convention. Firstly, when compounds undergo prefixation or suffixation, they maintain their separate spelling convention, e.g. *temu ramah* (to meet + friendly) 'a talk' = *menemu ramah* 'to have a talk/interview', *daya serap* (strength+ absorb) 'absorbance' = *daya serapan* 'absorption', etc. However, when compounds undergo circumfixation, they are spelt as single entities, e.g. *ibu bapa* (mother + father) 'parents' = *keibubapaan* 'parenting', *surat khabar* (letter + news) 'newspaper' = *persuratkhabaran* 'press/journalism', etc.

4.5.1 Comments

In her work, Karim (1995) briefly touched on three topics of Malay compounding. The first topic is on definition, i.e. compounds are seen as (new) word forms derived from the combination of two or more words in the language. The second topic is on some of the main types of compounds available in the language, i.e. nominal, verbal and adjectival compounds. The final topic is on the spelling convention, i.e. compounded constituents are separated except when undergoing circumfixiation where they are conjoined.

In addition to these three topics, there is basically no further detailed elaboration on Malay compoundhood. Karim's (1995) take on the topic is mainly concerned with providing some basic description along with some examples of compounded forms in Malay. Nevertheless, we can acknowledge two main outcomes from her work for this current study. Firstly, the examples of Malay compound she provides are important, serving as a useful source

of data for this current study. Secondly, we are made aware of the effects of circumfixiation on the spelling convention of Malay compounding, which is perhaps useful as one of the means of testing Malay compoundhood. To this end, we leave Karim's (1995) work on compounding and move on to the next study on Malay compounding.

4.6 Nik Safih Karim et al. (2008) *Tatabahasa Dewan*

Karim et al.'s (2008) *Tatabahasa Dewan* is a Malay grammar handbook regulated by the *Dewan Bahasa dan Pustaka* (Language Planning Agency of Malaysia). It is intended to be (and largely accepted as) the main handbook of standardised Malay grammar. This descriptive (and to a certain degree prescriptive) grammar book is divided into several chapters focusing on the phonology, morphology and syntax of the language.

Karim et al. (2008) describe Malay compounding through several subtopics, namely by the: (i) definition, (ii) spelling convention, (iii) types of compounds, (iv) affixation process, (v) reduplication process and (vi) the difference between compounded and phrasal constructions.

(i) *Definition*

Karim et al. (2008) define compounds as word forms created by combining two or more root words, creating a single unified entity that carries a specific meaning.

(ii) *Spelling convention*

With regard to spelling, the general rule is that the constituents involved are spelled separately. Exceptions apply to some established forms (some examples are provided below) and when compounds undergo a circumfixation process. In these cases, the constituents involved will be spelled jointly.

(iii) *Types of compounds*[57]

The handbook categorises three main categories of Malay compounds, i.e. according to the lexical categories of nominal, verbal and adjectival compounds. From these main categories, four subtypes of compounds are recognised, i.e. *free forms*, *specific forms*, *idiomatic forms* and *established forms*. Below are some examples of the categorisation:

(a) Nominal compounds

Free forms

Examples: guru besar (teacher + big/large) 'head teacher', *air mata* (water + eye) 'tears', *jam tangan* (watch + hand) 'wristwatch', etc.

Specific forms

Examples: analisis teks (analysis + text) 'text analysis', *titik buta* (spot + blind) 'blind spot', etc.

Idiomatic forms

Examples: anak emas (child + gold) 'favourite person', *buah hati* (fruit + heart) 'sweetheart' etc.

Established forms

Examples: bumiputera (earth + son/prince) 'native', *warganegara* (people + country) 'citizen', etc.

---

[57] Refer to appendix (5) for the full list of examples.

(b) Verbal compounds

Free forms

ambil alih (to take + to move) 'to replace', *tolak ansur* (to push + gradual) 'to tolerate', etc.

(c) Adjectival compounds

Free forms

Examples: merah jambu (red + red guava) 'pink', *hijau daun* (green + leaf) 'leafy green', etc.

Idiomatic forms

Examples: besar kepala (large + head) 'stubborn', *ringan tulang* (light + bones) 'hard working', etc.

(iv) *Affixation process*

Affixation of compound words involves prefixes, suffixes and circumfixes. According to Karim et al. (2008), the constituents involved in compounding maintain a separate spelling convention with prefixation and suffixation, but combine with circumfixation, e.g.:

Compound	Affixation	Derived compound
daya serap	*ber-*	*berdaya serap* 'absorptive'
(strength + absorb)	*-an*	*daya serapan* 'absorption'
'absorbance'	*ke-...-an*	*kedayaserapan* 'absorptivity'

(v) *Compound words reduplication*

According to Karim et al. (2008), the reduplication of compound words only involves the first constituent of the compound. Only the pre-conjoined compounds (such as the established forms) are reduplicated as a whole, for example:

Compounds		Reduplicated compounds
alat tulis (tool + write) 'stationery'	=	*alat-alat tulis* 'stationery'
kapal terbang (vessel + fly) 'aeroplane'	=	*kapal-kapal terbang* 'aeroplanes'
warganegara (people + country) 'citizen'	=	*warganegara-warganegara* 'citizens'

(vi) *Compound vs. phrase constructions*

According to Karim et al. (2008), compounds can be distinguished from phrases as compounds do not allow any elements to be inserted between their constituents. They briefly argue that, if an element is inserted and the intended meaning of the structure is distorted, this can be taken as proof that the structure is a compounded form (i.e. the elements involved are unified). For instant, the combination of *kerusi* (chair) and *malas* (lazy) to form *kerusi malas* (chair + lazy) 'lounge chair' is a compound because insertion of any element between the constituents will disrupt the intended meaning, e.g. insertion of *yang* (that) produces **kerusi yang malas* (a chair that is lazy). On the other hand, if insertion of elements does not distort the intended meaning of the structure, then it is a sign that the structure is a phrase. This is because phrases are more open in the sense that their constituents are not bounded unified structures, e.g. insertion of *yang* (that) to a combination such as *orang malas* (person + lazy) 'lazy person',

produces *orang yang malas* (a person that is lazy), signifying that the combination of *orang malas* is a phrase. This is the only distinction made by Karim et al. (2008) in discussing the issue of compound vs. phrasal structures in the language.

4.6.1 Comments

As mentioned, Karim et al. (2008) describe Malay compounding through the properties of definition, spelling convention, compound types, affixation and reduplication process, and also briefly through the difference between compounded and phrasal constructions.

With regard to definition, Karim et al. (2008) put forth a typical description of compounds, i.e. as a combination of words creating a new word form with a specific meaning. They also mention that Malay compounds have an *inti*, which is the 'core', and a *penerang*, which is the 'modifier'. In this sense, we can say that Karim et al. (2008) are referring to the 'head + modifier(s)' relationship of Malay compounds.

Compounding spelling convention seems to be emphasised in the handbook. This is mainly due to the fact that some changes are required in the way compounds are spelled with certain affixation and reduplication processes. As mentioned, the general spelling convention is to maintain the constituents involved separately. However, the constituents of a circumfixed compound will be jointly spelled. In theory, a given structure can thus be argued to be a compound if its constituents (initially separated) need to be jointly spelled after circumfixation. In other words, this observation can be used as a tool for testing Malay compoundhood.

With regard to reduplication, Karim et al. (2008) point out that reduplication only applies to the first constituent of a compound form. Although the authors do not specifically discuss headedness in Malay compounding, this observation is indeed addressing the issue of the 'head' component. In theory, a compounded structure should have a head constituent, i.e.

the most important component of the compound. Upon reduplication, only the most important component will be reduplicated, hence the reduplication of the head. Again, this observation on the spelling convention of compound reduplication can perhaps be a useful tool for testing Malay compoundhood. In other words, the way in which a given structure undergoes the reduplication process can be an indicator of compoundhood. A given structure can be argued as a compounded form if the process of reduplication has an effect only on the head and not on the other components. We will have more to say on these sorts of 'tests' in later sections.

Moving on with the types of compounds, Karim et al. (2008) largely focus on the compounding of the three main lexical categories of nominal, verbal and adjectival compounds. They further categorise four types of compounds i.e. *free forms*, *specific forms*, *idiomatic forms* and *established forms* within the three lexical categories. Among these types of compounds, the *free forms* can be considered as the most productive type, while the *established forms* are the least productive. This is evident from the examples provided in the handbook, i.e. *free forms* are apparent in all three lexical categories, while *established forms* are only found in the nominal compound category.

The *established forms* are very limited in number and hence arguably unique to the language. Like any other compounds, they are clearly combinations of two different root words. However, they are different from other compounded forms as their components are permanently linked together in terms of spelling. In our opinion, language users (to a larger extend) no longer see them as combinations of two separate words (although it is clear to any language user that they are). Instead, they are commonly perceived as single word forms. This is perhaps why Karim et al. (2008) and other scholars have labelled them 'established' compounds. To this end, we will consider them as compounds with a lexicalised meaning and spelling convention.

With regard to *specific forms*, Karim et al. (2008) state that these types of compounds are terminologies used in the scientific and/or any other specialised domains. To a larger degree, they are mostly words and/or terminologies that are directly translated from foreign languages.[58] For instance, the term *garis pusat* (line + centre) is a direct translation from the meaning of the word 'diameter'. Similarly, the term 'geology' is translated into Malay as *kaji bumi* (study + earth). These sorts of compounds are sometimes superfluous to the language. In principle, there is nothing wrong with adapting the words 'diameter' and 'geology' (with some adjustments to comply with the phonotactics of Malay)[59] to be used directly in the language. Many foreign words have entered the Malay language in such a manner. As a matter of fact, the more contemporary adapted terms of *diameter* and *geologi* are currently being used in standard Malay as opposed to the *garis pusat* and *kaji bumi* compounds. This is partly due to an awareness that the meaning of scientific and other specialised terminologies can be inaccurately or distortedly represented when translated. Therefore, although *specific forms* do comply with the definition of compounding (i.e. a combination of words producing a new unified meaning), some of them are actually terms that have been unnecessarily translated into the so-called *specific form* compounds. Perhaps the question here is, where do we draw the line between a structure which is a product of direct translation and a structure which is a true compound in the language?

Moving along, Karim et al. (2008) consider *idiomatic forms* to be compounds in the language as well. However, there is no further explanation or justification of why they are being considered as such. In essence, idioms (especially those with combinations of two words) mirror the structure and definition of compoundhood. However, the way in which language

---

[58] It is a common practice for foreign words and/or terminologies to be translated (as close as possible to their original meaning) into native Malay words. Cf. previous discussion (section 1.3) on the purist approach to word translation and loanword adaptation in the language.

[59] Cf. previous discussion in section 1.3 on phonotactic adjustments of foreign word adaptations.

users understand the meaning of idioms is very different from that of compounds. As discussed previously, idioms have a higher degree of opaqueness in comparison to regular compounds, which are more transparent in meaning.[60] This is also one of the main reasons why compounds can be more productive than idioms (i.e. idiomatic forms are mostly lexicalised forms). To this end, we will consider idiomatic forms as a special type of compounding with different attributes to that of the regular compounding in the language.

We now move on to the topic of compounds vs. phrasal structures. Karim et al. (2008) only emphasised distinguishing between compounds and phrases with the 'insertion test'. The principle of this test asserts that insertion of any element between the constituents of a given compound will disrupt the meaning of the compound.[61] The fact that insertion distorts the meaning of the compound shows that the constituents of compounds are unified and bonded to each other. If the insertion of another element does not disrupt the meaning of a given combination, then the construction can be considered as a phrase (as the constituents of phrases are not bonded in the way that compounds are).

However, holding such a view to discriminate between compounds and phrases can be problematic. Firstly, unlike other languages, the grammatical structures for compounds and phrases in Malay are exactly the same, i.e. compounds have a 'head + modifier(s)' relationship, while the constituent order of a phrase also follows the 'head - modifier(s)' rule. This means that the compounded structure and the phrasal structure have virtually the same appearance in the language.

---

[60] Cf. Musa's (1993) argument on the issue of transparency between compound and idiomatic structures (in section 4.4 (30)(iii) above).

[61] Karim et al. (2008) did not specify the type of element that can or cannot be inserted between the constituents. From our observation, it seems that there is no limitation to the kind of elements allowed. Provided that it makes sense to 'prove' or 'disprove' whether a given combined structure is a compound or otherwise, it is permissible to insert virtually any kind of element between tested structures. It is important that we are aware of the limitations of the insertion test as a measure of compoundhood.

Secondly, to argue that if an insertion disrupts the meaning of a combination it signifies that the combination is a compound (or otherwise a phrase) can be misleading. For instance, the combination of *tahi lalat* (faeces + fly) is arguably both a compound and a phrase, depending on how it is interpreted. On one hand, the combination is a compound as *tahi lalat* can mean 'mole', and on the other hand the combination is also a phrase as *tahi lalat* can have the literal meaning of 'fly's faeces'. Applying the test will show that insertion of an element, e.g. the preposition *dari* (from) to produce *tahi dari lalat* 'faeces (from) fly' clearly disrupts the compounded meaning of 'mole'. Therefore, to this end, the structure *tahi lalat* can be considered as a compound. However, at the same time, insertion of the preposition *dari* (from) to produce *tahi dari lalat* 'faeces (from) fly' also maintains the phrasal meaning of 'fly's faeces'. Thus, to this end, *tahi lalat* can also be considered as a phrase.

In this sense, the insertion test will not work on structures such as *tahi lalat* (and others alike) because the test forces us to (strictly) justify whether a given combination is either a compound or a phrase. Therefore, it appears that relying on a measure such as the 'insertion test' to validate and determine compounded structures against phrasal structures is not as clear-cut as Karim et al. (2008) have proposed. This uncertainty of whether a structure is a compound or a phrase is quite evident even throughout the handbook itself, as some combinations can be found listed in the handbook as examples for both compounds and phrases.

(32) Examples of mismatches between compounds and phrases in Karim et al. (2008)[62]

Structures listed as compounds	Structures listed as noun phrases
jam tangan (watch + hand) 'wristwatch'	*jam tangan* (watch-hand) 'watch for wrist/hand'
ibu bapa (mother + father) 'parents'	*ibu bapa* (mother-father) 'mother and father'
meja tulis (table + write) 'desk'	*meja tulis* (table-write) 'table for writing'
meja makan (table + eat) 'dining table'	*meja makan* (table-eat) 'table for eating'

There are also plenty of structures exemplified as phrases in the handbook, and yet which are

arguably compounds, as listed in the next table.

---

[62] There are also plenty of adjectival compounds in the handbook that are listed as examples of phrases as well. These include, among others, compounds of colour: *biru laut* (blue + ocean) 'ocean blue', *merah jambu* (red + red guava) 'pink', *hijau daun* (green + leaf) 'leafy green', and also idiomatic compounds: *panjang tangan* (long + hand) 'thief', *besar kepala* (big/large + head) 'arrogant', *ringan tulang* (light + bones) 'hardworking', etc. These and some other examples are listed twice as both compounds and phrases in the handbook.

(33) Examples of arguably mislabelled structures

Structures listed as phrases	Structures listed as compounds
Examples: *kapal layar* (ship - to sail) *kapal api* (ship - fire) Comments: Both *kapal layar* and *kapal api* are exemplified as phrases in the handbook. Does this mean that *kapal layar* is to be read as 'ship for sailing', instead of the compound 'sailboat'? Similarly, should *kapal api* be read as 'ship powered by fire (stim)', instead of the compound 'steamboat'?	Example: *kapal terbang* (ship/vessel + fly) Comments: *Kapal terbang* is exemplified as a compound in the handbook. This implies that *kapal terbang* is to be read as the compound 'aeroplane'.
Examples: *tali pingang* (rope - waist) *sarung kaki* (cover - feet) Comments: Both *tali pingang* and *sarung kaki* are exemplified as phrases in the handbook. Does this mean that *tali pingang* is to be read as '(rope) fastener for waist', instead of the compound 'belt'? Similarly, should *sarung kaki* be read as 'cover for feet', instead of the compound 'socks'?	Example: *bom tangan* (bomb + hand) Comments: *Bom tangan* is exemplified as a compound in the handbook. This implies that *bom tangan* is to be read as the compound 'grenade'.

Example: *air limau* (water - lime) Comments: *Air limau* is exemplified as a phrase in the handbook. Does this mean that *air limau* is to be read as 'water (made) from lime' instead of the compound 'lemonade'?	Example: *nasi minyak* (rice + oil) Comments: *Nasi minyak* is exemplified as a compound in the handbook. This implies that *nasi minyak* is to be read as the compound 'oily rice' (cooked rice usually for special occasions).
Examples: *minyak wangi* (oil - aromatic) *sekolah rendah* (school - low) Comments: Again, does this mean *minyak wangi* is to be read as 'oil which is sweet in smell', instead of the compound 'perfume'? Similarly, should *sekolah rendah* be read as 'school for lower education', instead of the compound 'primary school'?	

Table (32) shows how the same combined forms (e.g. *jam tangan*, *ibu bapa*, etc.) are confusingly used in the handbook as examples of compounded forms and also as examples of phrasal forms. Table (33) shows how combined forms containing comparably similar elements are exemplified as different categories from one another. For instance, how is the combination of *kapal* (ship/vessel) and *terbang* (to fly) any different from the combination of *kapal* (ship)

and *layar* (to sail) for the former to be categorised as a compound, 'aeroplane', while the latter is categorised as a phrase, 'ship for sailing'?

There are plenty of other mismatches and arguable examples of compounds vs. phrases throughout the handbook. This shows that the structure of compounds and phrases in the language overlaps, hence the difficulty to precisely differentiate one from the other.[63] Furthermore, it appears that relying on a measure such as the 'insertion test' to validate and determine compounded structures against phrasal structures can be misleading. This is not to say that the test is wrong or useless, but instead it should be applied with caution if it is to be used as a test for Malay compoundhood.

4.7 Jyh Wee Sew's (2007) *Reduplicating nouns and verbs in Malay*

The main concern of this work revolves around the issues of noun and verb reduplication, along which a section on compounding is included. However, the analysis and discussion only focus on nominal compounding. The inclusion of noun compounds in the reduplication study is on the basis of Malay compound reduplicates either partially or completely (to designate plurality). The discussion begins with a general introduction on compounding. Compounding is defined as the process of combining at least two words with a semantic attribute of a single unit. It is also stated that compound constructions do not accommodate any linguistic intervention, such as insertion of other elements, or modification of the whole compound. Furthermore, the typical concepts of head-modifier and

---

[63] This, of course, is not a problem that is exclusive to the Malay language. English compounds are also well known for being difficult to differentiate from phrases.

endocentric/exocentric compounds are also mentioned. These topics are discussed briefly as an opening to the main analysis of nominal compounding.

Sew (2007) looked at Malay compounding in relation to two main issues, namely: (i) the types of nominal compounds and (ii) plurality in nominal compounding, and we now turn to them below.

(i) Types of nominal compounds

In Sew's (2007) analysis, nominal compounds can be further categorised into three subtypes, as detailed below.

(a) Nominal Compound type A

These types of compounds are characterised by the typical head-modifier construction. They are divided into the Noun + Noun (NN), Noun + Verb (NV) and Noun + Adjective (NA) compound types. Some examples include:

NN types, e.g.:

> *air mata* (water + eye) 'tears', *ibu kota* (mother + city) 'capital', *kereta api* (car/cart + fire) 'train'.

NV types, e.g.:

> *nasi goreng* (rice + fry) 'fried rice', *tukang masak* (artisan + cook) 'cook/chef', *pisau cukur* (knife/blade + shave) 'razor'.

NA types, e.g.:

> *sekolah rendah* (school + low) 'primary school', *papan hitam* (board + black) 'blackboard', *kerusi malas* (chair + lazy) 'lounge chair'.

(b) Nominal compound type B

These types of compounds are characterised as having both constituents synonymous to each other, and are also known as appositional. Some examples given include:

hutan rimba (woods + forest) 'large/dense jungle', *jiran tetangga* (neighbour + neighbour) 'neighbours', *bala tentera* (troops + soldier) 'army'.

(c) Nominal compound type C

These types of compounds are characterised as having constituents that are either synonymous or antonymous to each other, and are also known also coordinate compounds. Some examples given include:

ibu bapa (mother + father) 'parents', *rumah tangga* (house + stairs) 'household', *periuk belanga* (pan + pot) 'pots and pans'.

(ii) Plurality in nominal compounding

Part of the study focused on the plurality of nominal compounds, which relates to how compounds are reduplicated. The results of an acceptability test concluded that the general preference was for a full reduplication of:

(a) Appositional compounds, e.g.:

jiran tetangga (neighbour + neighbour) 'neighbour(s)', reduplicates into

jiran tetangga - jiran tetangga 'neighbours'.

(b) Coordinate compounds, e.g.:

ibu bapa (mother + father) 'parent', reduplicates into *ibu bapa - ibu bapa* 'parents'.

(c) Endocentric NN compounds, e.g.:

kereta api (car/cart + fire) 'train', reduplicates into *keratapi-keratapi* 'trains'.

(d) Endocentric NV compounds, e.g:

tukang masak (artisan + cook) 'cook/chef', reduplicates into *tukang masak – tukang masak* 'cooks'.

On the other hand, the preference for partial reduplication is of:

(e) Endocentric NA compounds, e.g:

papan hitam (board + black) 'blackboard', reduplicates into *papan-papan hitam* 'blackboards'.

To this end, we have reviewed the essence of Sew's (2007) study on compounding. We now move on to the comments section below.

4.7.1 Comments

Sew's (2007) study is very interesting in the sense that it presents a rather different view on Malay compounding in comparison to the other linguists we have looked at so far. Firstly, Sew (2007) recognises at least two types of compounds that are virtually 'non-existent' in the vocabulary of other linguists on Malay compounding. The two compounds are appositional and coordinate compounding. Appositionals are characterised as compounds with synonymous

constituents (e.g. *hutan rimba* (woods + forest) 'large/dense jungle'), while coordinates are characterised as compounds with either synonymous or antonymous constituents (e.g. *ibu bapa* (mother + father) 'parents').[64]

Although some linguists have indeed exemplified these types of compounds in their discussions, none had categorised them explicitly as either appositional or coordinate compounds. One can argue that this is perhaps merely a 'labelling' issue. However, one needs to remember that 'labelling' can assist us in applying the appropriate attributes that come with the label.[65] As such, the recognition (i.e. the labelling) is important in the sense that it can (to a certain degree) clarify the status of a given structure. For instance, Karim et al. (2008) are unclear in their categorisation of structures such as *ibu bapa* (mother + father), exemplifying them both as the compound 'parents' as well as the phrase 'mother and father' (cf. 4.6.1 (32)). Having extra subcategories such as the ones proposed by Sew (2007) can provide more channels by which to categorise the many different types of Malay compounds. Indeed, we are in favour of such an approach.

With regard to pluralisation of compounds, the pattern of reduplication presented in Sew's (2007) study is not the 'typical' manner of compound reduplication. Recall that most linguists adhere to the general principle of having the head component reduplicated for compounds with separated constituents, and having the whole compound reduplicated for compounds with conjoined constituents (Hassan, 1986; Karim et al., 2008). Sew (2007), however, argues that the general preference for full reduplication of compounds is not

---

[64] It is unclear to us what Sew (2007) meant by the constituents of coordinate compounds being either synonymous or antonymous to each other. The constituents of appositional compounds are synonymous with each other; thus, it will be redundant for coordinate compounds to have similar characteristics. In other languages, compounds labelled as coordinative are usually described as compounds with their constituents having the relation 'and' between them.

[65] In other words, by labelling a structure as, for instance, an 'appositional compound', this will entail that the structure consists of bonded synonymous components. In a way, 'labelling' assists in affording a given structure more precise attributes.

dependent on whether the compounds have separated or conjoined components. The preference is argued to be influenced by the type of compound instead.

Sew (2007) identifies four types of compounds (i.e. appositional, coordinate, endocentric NN, and endocentric NV compounds) as preferring to be fully reduplicated regardless of whether their constituents are separated or not. For instance, the appositional compound *jiran tetangga* (neighbour + neighbour) 'neighbours', will be reduplicated as *jiran tetangga - jiran tetangga*, and not *jiran-jiran tetangga* (cf. 4.7 (ii) above for the other examples). This finding is different from the general principle whereby the reduplication of compounds with separated constituents will only involve the head constituent (e.g. *jiran-jiran tetangga*, and not *jiran tetangga - jiran tetangga*, cf. Karim et al. (2008) in sections (4.6) and (4.6.1), and Hassan (2006) in sections (4.3) and (4.3.1)).

Sew (2007) argues that the preference for full reduplication of the four types of compounds mentioned has to do with them being 'prototypical compounds'. What is meant by a prototypical compound here is not specifically explained, but we can confidently assume that he is referring to the compounds as being established (lexicalised) compounds.[66] Hassan (2006) has a similar outlook on this when he states that *uji kaji* 'experiment' becomes *ujikaji-ujikaji* 'experiments'.

In our opinion, Sew's (2007) analysis on the issue of plurality and reduplication is partly inaccurate. On pluralising the appositional and coordinate compounds, the result is questionable, being that compounds such as *jiran tetangga* (neighbour + neighbour) 'neighbours' and *ibu bapa* (mother + father) 'parents' are inherently plural. *Jiran tetangga* itself means neighbours, as in '*Kita perlu menghormati jiran tetangga*', which translates as

---

[66] The reason we can assume this is because Sew (2007: 84) argued that endocentric NA compounds are "(…) not yet fully entrenched (…)" in the language as opposed to the other four types of compounds. This statement gives us a clue that Sew (2007) considers the four types of compounds as established (lexicalised) compounds.

'We have to respect our neighbours', and not *'*Kita perlu menghormati jiran tetangga - jiran tetangga*'. Similarly, '*Mesyuarat ibu bapa dan guru-guru*' translates as 'Teachers and parents meeting'. It is not *'*Mesyuarat ibu bapa - ibu bapa dan guru-guru*'. In this sense, pluralising an inherently pluralised form is redundant. It is not necessary the case that these types of compounds need to be reduplicated to show plurality.

With regard to the argument that these compounds are lexicalised forms (hence fully reduplicated), we are not in favour of taking this point seriously. This is partly because Sew's (2007) study is based on a very limit number of compounds being judged by a small number of respondents (four native speakers). More importantly, the study did not properly explain what is meant by 'lexicalised' forms. A form that is commonly or regularly used may be considered as 'lexicalised' by some but not by others. By the same notion, a form that is less common and less used may also be considered as 'lexicalised' by some but not by others. How can we then be sure that a given form is lexicalised or otherwise in order for it to be fully or partially reduplicated?

Interestingly, Sew (2007) also noticed that endocentric NA compounds do not reflect the same outcome as the other compounds mentioned above. Instead, endocentric NA compounds are more likely to be partially reduplicated upon pluralisation (e.g. *papan hitam* (board + black) reduplicated into *papan-papan hitam*, and not *papan hitam - papan hitam*). Sew (2007: 84) argues that the reason this happens is because NA structures are "(…) intermediaries between Malay compounds and Malay phrases. Although it is a head-modifier construction, endocentric noun-adjective compounds are not yet fully entrenched and remain as phrasal structures". In other words, endocentric NA compounds are assumed to be structures that are yet to be 'fully lexicalised' in the language. Hence, they do not have to be fully reduplicated as a whole like other lexicalised compounds need to be.

Once again, we can question the concept of 'lexicalisation' here. How does one prove that NA structures (such as *papan hitam*) are 'less' or 'not yet' lexicalised in the language? How are NA structures different from that of NN and NV structures to the point that the latter combinations are considered as lexicalised structures while the former is not? What is it meant by saying that NA structures 'remain' as phrasal structures? Does it mean that there are no compounded NA structures in the language? These sorts of questions are not properly addressed in the study. For all we know, the form *papan hitam* 'blackboard' could be the perfect embodiment of a lexicalised form (given the fact that it is arguably one of the most common and recognisable items known to virtually anyone who has experienced classroom education in one way or another). To this end, we are once again reluctant to accept some of the arguments on pluralisation, reduplication and lexicalisation of Malay compounds discussed in this study. Therefore, we are forced to conclude that Sew's (2007) analysis of these concepts is inaccurate (to a certain degree).

4.8 Ismail and Jalaluddin's (2008) *Kata majmuk dalam perkamusan*

Ismail and Jalaluddin (2008) begin their discussion on compounding by mentioning that compounds have often been confused with other combination-type constructions such as reduplication, idioms and phrases in the language. This, they argue, is due to the different approaches taken by scholars in defining, characterising and classifying compounds. They also argue that Malay linguists have somewhat neglected the semantic aspect of compounding, and thus this will be the focus of their study (especially of the explicit/implicit meaning of compounds).

Firstly, Ismail and Jalaluddin (2008) maintain some of the main measures compoundhood, namely of: (i) inseparability, (ii) inability to switch constituents, and (iii) the

ability of circumfixation. In addition, they impose two extra measures: (iv) the meaning of a compound (as a whole) must convey an entirely different meaning from the original individual meaning of its components, and (v) a given compound must not be able to generate more compounds alike. They claim that the two extra measures are beneficial as they force a true compound construction for the language.

The article continues to the main part of the study, i.e. analysis of 10 combining structures containing the word *ibu* 'mother' as one of their constituents (e.g. *ibu saudara* (mother + relation) 'aunty', *ibu angkat* (mother + step) 'step mother', *rumah ibu* (house + mother) 'main part of a house', etc.). The 10 combining structures are 'tested' against the five proposed measures of compoundhood mentioned above.

First, the study applies the combining structures within a context sentence (with specific cues) to facilitate pragmatic understanding (i.e. to derive the possible meanings of the combining structures). For instance, the structure *ibu ayam* (mother + chicken) is placed within a context sentence such as:

> ***Ibu ayam*** *itu mengajar anaknya mencakar tanah mencari cacing.*
>
> (**mother chicken** DET meN-teach chick-POS meN-scratch earth meN-search worms)
>
> The **hen** teaches her chick to claw the ground for worms.

The context sentence above assists in the literal reading of *ibu ayam* as 'hen'. Yet another context sentence is used to assists in the implicit reading of the structure, e.g.:

> "*Siapa perempuan itu?*" "***Ibu ayam***" *Ali berbisik.*
>
> ("PRO women DET" "**Pimp**" Ali whispered).
>
> "Who is that women?" "**Pimp**" whispered Ali.

The context sentence above assists in the implicit reading of *ibu ayam* as 'pimp'. Therefore, the structure *ibu ayam* (mother + chicken) can be said to have two different meanings, a literal meaning of 'mother of chicken' (i.e. hen) and an implicit meaning of 'women (mother) in charge of young girls' (i.e. pimp).

To this end, *ibu ayam* is argued to potentially be either a compound or a phrase. In the next step, Ismail and Jalaluddin (2008) scrutinise the structure through the remaining proposed measures of compoundhood, namely of inseparability, constituents switching and circumfixation. Given that *ibu ayam* cannot be separated, **ibu yang ayam* (*mother which is chicken), cannot be switched, **ayam ibu*[67] 'chicken mother', and can be circumfixed, *beribuayamkan* 'to make someone her pimp', the authors consequently maintain that the structure of *ibu ayam* is a compounded form rather than a phrasal one. This process is repeated for the rest of the nine structures under study. They argue that, out of the 10 structures, only five can be considered as compounds. They state that this outcome is an effect of narrowing the compounding criteria, which helps to better clarify the status of controversial combined structures (as compounds or otherwise). Ismail and Jalaluddin (2008) conclude that their study has managed to offer an effective method (and measure) of identifying compounds in the language.

4.8.1 Comments

One of the main aims of Ismail and Jalaluddin's (2008) study is to clearly identify compounded structures against other combined structures in the language. To do so, they

---

[67] *Ayam ibu* has the meaning of a hen nurtured and used specifically for producing eggs.

basically proposed a five-point test of Malay compoundhood. They argued that, in order for a structure to be categorised as a compound, it must have the qualities below, i.e.:

(34)   (i) the components involved cannot be separated,

(ii) the components must not be able to be switched (and retain the intended meaning),

(iii) the structure must be able to be circumfixed,

(iv) the meaning of the structure (as a whole) must convey an entirely different meaning from the original individual meaning of its components,

(v) the structure must not be able to generate more analogical compounds.

To a certain degree, these criteria are arguably problematic to maintain for compoundhood. For instance, if we agree with point (34) (iv), that a derived compound should have a totally different meaning to any of its constituents, this will very much eliminate endocentric compounds as we know them. As discussed, the head is a hyponym of the compound, e.g. *guru besar* (teacher + big/large) 'head teacher' is a kind of *guru* 'teacher', *jam tangan* (watch + hand) 'wristwatch' is a kind of *jam* 'watch', and *kapal terbang* (ship/vessel + fly) 'airplane' is a kind of *kapal* 'ship/vessel'. These compounds will lose their status if based on the measure of (34) (iv) since the components involved (i.e. *guru*, *jam* and *kapal*) are all projected into the meaning of the whole compound. This is a common feature of endocentric compounds and to have such a restriction is definitely problematic. Compounding will not be a productive word formation process with this principle.

Similarly, the second measure, i.e. the inability to generate other compounds via analogy, will also hinder the productivity of compounding. By this principle, compounds such as *rumah sakit* (house + sick/ill) 'hospital' and *kerata api* (car/cart + fire) 'train' will also lose their status, as these heads are highly generative: *rumah api* (house + fire) 'light house', *rumah*

pangsa (house + segment) 'flats', *rumah haram* (house + illegal) 'brothel', *rumah panjang* (house + long) 'long house', *kerata kebal* (car/vehicle + strong) 'tank', *kerata tolak* (car/cart + push) 'pushcart', *kereta sorong* (car/cart + slide) 'wheelbarrow', etc. The restriction on generating compounds has also been mentioned by Hassan (1986, 2006) (cf. 4.2.1 and 4.3.1). Hassan (2006) also disagrees with the concept of free generation from compound heads. He argues that compounds should be more restricted in order to differentiate between the open-endedness of phrases. He also argues that most of the other generated compounds, such as *rumah rehat* 'rest house', *rumah makan* 'restaurant', etc., can be seen as phrases because they maintain their meaning after insertion, i.e. *rumah untuk rehat* 'house for resting', *rumah tempat makan* 'place for eating', etc. However, Hassan (2006) still considers *rumah sakit* (house + sick/ill) 'hospital' as a compound because the components are inseparable. This is not the stand taken by Ismail and Jalaluddin (2008), as they argue that *rumah sakit* must lose its compound status, because of its generative similarity to other *rumah* kind formations.

We now turn to the analysis of combining structures in the study. To a very large extent, the study mainly analysed the semantics aspect of the constructions. As shown above, the structure *ibu ayam* is scrutinised first in relation to its meaning (i.e. explicit/implicit), and subsequently to the other tests of inseparability, constituents switching and circumfixation. Additionally, as mentioned, Ismail and Jalaluddin (2008) concluded that *ibu ayam* is a compounded form. However, what is really confusing is the fact that the authors did not apply the two extra test they themselves had suggested earlier (cf. (34) (iv) and (v)). According to these criteria, *ibu ayam* will not be a compound. The head *ibu* (mother) can be argued as having the meaning mother-like authority, of being a 'lady pimp'. Furthermore, the head *ibu* (mother) definitely generates a lot of *ibu*-type structures (as evident from the 10 structures used as the data set in the study itself). This point alone is a contradiction of the study's own principles (i.e. not complying with their own proposed measures of compoundhood).

There are also some other uncertainties and contradictions in their arguments throughout. For example, in their analysis of *ibu kandung* (mother + carry) 'birth mother', the combination does not violate the principle of inseparability. In other words, insertion of any kind will disrupt the bond between the constituents to produce the meaning of 'birth mother'. For instance, the insertion of *yang* 'which/that' will not generate a proper phrase **ibu yang kandung* *'mother (who) carry' is ungrammatical. Clearly, this structure is a compound (by the principles of insertion) as the insertion of *yang* disrupts the intended 'compound' meaning, and at the same time does not create any phrasal meaning. However, Ismail and Jalaluddin (2008) analysed the combination by inserting the *yang* 'which' and further adding the affixation *men-kandung-kan*. This is then analysed as *ibu yang mengandungkan* 'mother who carried (pregnant with) the child'. By doing so, the combination can now be justified as being able to take on the insertion and still maintain the same intended meaning. Therefore, *ibu kandung* cannot be considered as a compound, but as a phrase instead. This is definitely not the regular method that they use on the other combined structures (i.e. they never add affixation, always bare insertions).

Further evidence of irregularity of analysis can be seen in their depiction of *rumah ibu* (house + mother). They argued that the combination can be both, i.e. a compound and a phrase at the same time. The combination is a compound because it conveys the meaning of 'central structure of a building'. So if, say, *kepunyaan* (belongs to) is inserted, the structure becomes *rumah kepunyaan ibu* 'mother's house' (reinforcing the compound status). In other words, the insertion *kepunyaan* disrupts the intended compound meaning of 'central structure of a building'. However, they also maintained that *rumah ibu* can also be argued as a phrase as well. This is because the insertion of *kepunyaan* (belongs to/ownership) produces the grammatical meaning of 'mother's house'. The researchers thus accept this as both a compound and a phrase. This flexibility is not shown in their analysis of the other structures (i.e. such findings

on other structures in their analysis were categorised as either a compound or a phrase, but not as both).

Furthermore, if we apply the test of constituent switching to *ibu* rumah, the converted structure will also mean 'the central structure or a house'. This clearly violates their second measure of compoundhood (cf. (34) (ii)). In this sense, the analysed structure (i.e. *rumah ibu*) is less of a compound than it is of a phrase. However, Ismail and Jalaluddin (2008) defend this violation by saying that the converted form *ibu rumah* is seldom used, and thus is somewhat irrelevant for their analysis. Again, this sort of inconsistency is questionable. It seems to us that the study is manipulating its own principles too conveniently. To this end, it is apparent once again that distinguishing between compounds and phrases is not as easy as it may seem. Although Ismail and Jalaluddin's (2008) study tries to refine the criteria of compoundhood as narrowly as possible, it is still evident that identifying compounds in Malay is not a clear-cut matter.

4.9 Chapter summary

As mentioned at the beginning, the aim of this chapter is to provide an overview on how the topic of Malay compounding has been addressed in the literature by Malay linguists. We have reviewed several studies and commented on each one of them individually. Indeed, the topic of compounding has been discussed and approached somewhat differently from one linguist to another. Even so, it is clear that there are certain issues that are common and recurrent throughout when discussing Malay compounding.

In general, a discussion on Malay compounding begins with a definition of what a compound is in the language. More often than not, this concerns a mention of the elements involved in compounding, which include roots (e.g. Hassan, 1974), morphemes (e.g. Musa,

1993; Ismail and Jalaluddin, 2008) and words (e.g. Karim, 1995; Sew, 2007; Karim et al., 2008). Some scholars explain the reason for using such elements, while others do not necessary do the same. Nonetheless, these elements are suitable as elements of Malay compounding in one way or another. Additionally, within the scope of the definition, scholars often touch on the issue of how compounds should be recognised as single unified entities (e.g. Hassan 2006; Sew, 2007; Karim et al., 2008). This shows that Malay linguists in general recognise that the compound should be treated as a morphological phenomenon, rather than a syntactic one.

Another common topic when discussing Malay compounding is the issue of headedness. Malay linguists in general recognise Malay compounds as having a modified (head) - modifier relation (e.g. Hassan, 1974; Sew, 2007; Karim et al., 2008), which is either endocentric or exocentric (e.g. Hassan, 1974; Musa 1993). This is an important characteristic to acknowledge as it is one of the main criteria by which to distinguish between compound and other 'similar' structures. In addition to the modified-modifier relation, Sew (2007) is the only person in the review who explicitly recognises Malay compounds as potentially having an appositional and/or coordinate relation between its constituents.

The third recurring topic on Malay compounding concerns the criteria of compoundhood. In general, Malay linguists usually focus on syntactic criteria such as the inseparability of constituents, inability to switch constituents and the inability to modify the constituents of compounds (e.g. Karim et al., 2008; Sew, 2007; Ismail and Jalaluddin, 2008). There are also others which are less discussed in general, but are mentioned by a few, e.g. restricting the ability to generate more compounds via analogy (e.g. Hassan, 2006; Ismail and Jalaluddin, 2008). On the other hand, some scholars gave more focus to the semantic criteria of compoundhood such as the issue of transparency and opaqueness of compound meaning (e.g. Musa, 1993; Ismail and Jalaluddin, 2008). Likewise, there are also discussions on the orthographic criteria, which usually relate to the spelling convention of compounds upon

circumfixation and/or reduplication (e.g. Hassan, 1986; Musa, 1993; Karim, 1995; Karim et al., 2008; Ismail and Jalaluddin, 2008). On the whole, these criteria are usually intended as 'tests' of compoundhood in the language.

The final recurring topic on Malay compounding concerns the types of compounds in the Malay language. In essence, Malay linguists usually recognise nominal, verbal, adjectival and idiomatic compounds as the main types of compounds in the language (e.g. Hassan, 1974, 1986; Karim, 1995; Karim et al., 2008). There are also others who describe the types of compounds in terms of how they are used in the language, such as for academic or scientific purposes (Karim, 1995; Karim et al., 2008). In addition to these types of compounds, Hassan (1974) is the only scholar who recognises a minor set of adverbial and prepositional compounds in the language. In general, recognising the types of compounds is an important matter as it helps to identify the possible compounding patterns available to the language.

To this end, we can summarise this chapter with a table on the topics and issues relating to Malay compounding that are commonly discussed by Malay linguists.

	Linguists / Topics	Hassan (1974) (1986) (2006)	Musa (1993)	Karim (1995) Karim et al. (2008)	Sew (2007)	Ismail and Jalaluddin (2008)
1.	Definition	Roots (1974). Words (2006).	Morphemes.	Roots.	Words.	Morphemes.
	Function as a single lexical unit/entity	Yes.	Yes.	Yes.	Yes.	Not mentioned explicitly.
2.	Headedness (modified – modifier relation)	Yes.	Yes.	Not mentioned explicitly.	Yes.	No discussion.
	Endocentric	Yes.	Yes.	No discussion.	Yes.	No discussion.
	Exocentric	Yes.	Yes.	No discussion.	Yes.	No discussion.
3.	Criteria	Syntactic: Inseparability. No insertion. Limited generation. Orthographic: Conjoined spelling upon reduplication.	Not discussed explicitly, but put some focus on the degree of transparency and opaqueness of compound meaning.	Syntactic: Inseparability. No insertion. Orthographic: Conjoined spelling upon circumfixation.	Syntactic: Inseparability. No insertion. No modification.	Syntactic: Inseparability. No switching. No projection. No generation.
	Interaction with other processes	Affixation. Reduplication.	No explicit discussion.	Affixation. Reduplication.	Reduplication.	Circumfixation.
4.	Types of compounds	Nominal. Verbal. Adjectival. Idiomatic. Adverbial. Prepositional.	Nominal. Verbal. Adjectival.	Nominal. Verbal. Adjectival. Idiomatic.	Nominal. Coordinate. Appositional.	Nominal.

Chapter 5: Analysis of Malay compounding

This chapter is divided into a two-section analysis. In the first section, we will reiterate several parts from the previous discussion in order for us to be able to tie up the loose ends on some issues and topics in relation to Malay compounding. The outcome of this section will be the overall outline of Malay compoundhood. This outline will subsequently be utilised in the second section of this chapter, as a tool for the analysis of the main compounding data in this study.

5.1 Competition analysis in Malay compounding

The main query in the first part of Chapter 2 is the question of where exactly in the grammar does complex word formation such as compounding takes place. In relation to this question, the recurring argument seems to focus on whether complex words are handled by the morphological module or by the syntactic module. We have looked at both sides of the argument on this issue. On one hand, advocates for the superiority of the syntax module have tried to undermine the significance of a separate morphological module (Lees, 1960; Chomsky, 1965; Chomsky and Halle, 1968). In other words, the syntactic module is considered to be able to account for both morphological and syntactical processes. On the other hand, we have also looked at arguments against the idea of morphology being subsumed under a syntax module (Spencer, 1991, 2005; Di Sciullo and Williams 1987; Bresnan and Mchombo, 1995; Scalise and Guevara, 2005; Lieber and Scalise, 2006). In other words, the syntactic model is not suitable for handling morphological processes. In sum, we have come to agree that it is better to assume separate modules for morphology and syntax, with some sort of an interaction

between them (Lieber, 1992; Jackendoff, 1997, 2002; Ackema and Neeleman, 2004; Lieber and Scalise, 2006).[68] To this end, we adopt the principles of Ackema and Neeleman (2004) on what is known as the competition model.[69]

As mentioned before, the competition model proposed by Ackema and Neeleman (2004) seems to be able to neatly account for the workings of complex word formation, especially in relation to compounding. We believe that the model can assist in our attempt to understand Malay compounding as well. Before we analyse the model against Malay compounding, let us begin by reiterating some of its key points. The model is in support of the idea that morphology and syntax are two independent generative systems, which at the same time are able to interact with each other. Since they are considered as two generative separate modules, Ackema and Neeleman (2004) argued that complex structures can therefore be merged in either of them.

For instance, if an item α is to combine with an item β, the model assumes that they can either be combined in the morphological module (hence producing a morphological word) or they can be combined in the syntactic module (hence producing a syntactic phrase). Thus, in a language like Malay where both morphological and syntactical heads are on the left, an abstract illustration of how two elements (e.g. α and β) can merge is shown as below:

---

[68] Refer to Chapter 2 for the discussion.
[69] Cf. 2.1 (3) for the elaboration of the model.

(35) (a) Syntax merger   (b) Morphology merger

```
        αP                                        α
       / \                                       / \
      /   \                                     /   \
     α     βP                                  α     β
           |
   (head)  β                           (head)     (dependent)

          (dependent)
```
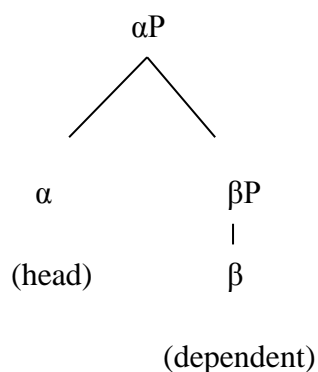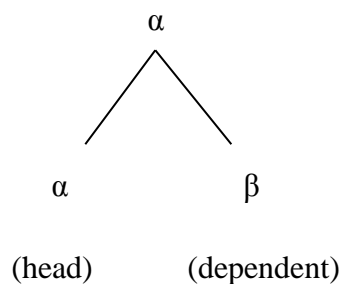
The illustration above abstractly exemplifies how two given items (or words) can merge in the Malay language when using the competition model. If there is an item α (let us assume it is the most prominent item, hence it will appear as the head in the merger) and an item β (let us assume it is the less prominent item, hence it will appear as the dependent in the merger) to be merged in the Malay language, they can either merge in the morphological system to form a morphological structure [α β], or merge in the syntactical system to form a phrase [α βP].[70]

The question is, in which module will the merger of the two given items (or words) actually take place? According to Ackema and Neeleman (2004), since both morphology and syntax are two independent subsystems, they are therefore in competition with each other to generate the complex structure. Competition[71] occurs when the merger of α and β involves the same category or has the same semantic relation in both the syntactic and morphological structures. When all conditions are observed as being equal, the preference of the merger will be for the unmarked option. The unmarked option of either syntax or morphology merger is language-specific. In a language such as Malay, we can assume that the syntax is the unmarked

---

[70] The Malay headedness system dictates that the head is always on the left in both morphological and syntactical structures.

[71] Ackema and Neeleman (2004: 51) provide the conditions for competition as below:
Let α_1 and α_2 be syntactic representations headed by α. α_1 blocks α_2 iff
(i) In α_1 (a projection of) α is merged with (a projection of) β in syntax, while in α_2 (a projection of) α is merged with (a projection of) β in morphology, and (ii) the semantic relation between α and β is identical in α_1 and α_2.

structure. We can make this assumption based on the fact that Malay constructions depend on rigid word order to convey the correct meaning (Othman, 1984; Hassan, 2006; Karim, et. al. 2008). This is partly because the language has no special or overt markers to differentiate morphological structures against syntactical ones. In this sense, syntactic structures are arguably more transparent in meaning, making them the unmarked choice of the language.

Having presented the outline above, we can now demonstrate on concrete examples in Malay. Take, for instance, the combination of *rumah* (house) and *papan* (plank), which derives *rumah papan* (wooden house):

(36)   (a) Syntax merger   (b) Morphology merger

```
        NP                                    N
       /  \                                  / \
      N    N                                N   N
    rumah  papan                          rumah papan
```
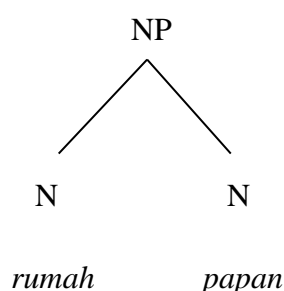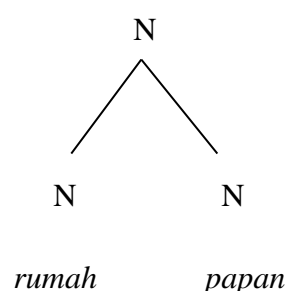
The model shows that the head *rumah* (house) and the non-head *papan* (plank) can be presumed to merge in either (36 a) syntax[72] or (36 b) morphology. In both mergers, the elements involved are of the same categories (Noun + Noun) and have the same semantic relation between them, i.e. both deriving the meaning of 'wooden house'. Since both are equal, the model dictates that they are in competition, and, given that syntax is the unmarked option, syntax wins the competition, i.e. *rumah papan* is a phrase and not a compound. A quick test of modifier

---

[72] One of the basic structures of the Malay noun phrase is NP -> N + (modifier) where the modifier may consist of either N, V, A or PP (cf. Karim et. al. 2008). The combination of N (*rumah*) and Modifier (*papan*) above does not need a functional projection (such as a prepositional phrase) to connect the meaning between them. In this case, the noun *papan* is sufficient to project a modification on the head to mean something along the lines of 'house made of wood'. Furthermore, a functional projection such as 'of' is not counted in Ackema and Neeleman's (2004) analysis.

insertion can illustrate how, for example, *merah* (red) modifies *papan* (plank) in [*rumah*] [*papan merah*] instead of [*rumah papan*] [*merah*], which shows that the *rumah papan* is not a unified combination, characteristics of a phrasal merger.

To this end, it seems like compounds will never win over a phrasal merger in Malay because syntactic structures are always identical to morphological ones. Before we look at other concrete examples, let us briefly illustrate the nature of Malay compound and phrase structures. Consider the combination of Noun + Noun, Noun + Adjective, and Noun + Verbs, all of which can potentially be the structure of root/primary noun compounds or equally the structure of simple noun phrases. These direct parings of head-modifier structures are very common, and they can be almost indistinguishable between being a phrase or being a compound. Malay linguists have often categorised them into several groupings based on the types of relationship held between the modifying constituent against the head. Karim et al. (2008: 364) listed at least 13 types of noun + noun structures with various thematic relations, among others:[73]

(37) Examples of noun + noun structures with different thematic relations

(a) Head + type modifier, i.e. the modifier specifies the type of the head:

Head ***burung*** 'bird'

e.g.: ***burung*** *gagak* (bird + gagak) 'crow', ***burung*** *helang* (bird + helang) 'eagle', ***burung*** *pipit* (bird + pipit) 'sparrow'

---

[73] The examples are partly from Karim (2008: 364 - 369) and partly derived from general knowledge of the language.

Head **pokok** 'tree'

e.g.: **pokok** *getah* (tree + rubber) 'rubber tree', **pokok** *kelapa* (tree + coconut) 'coconut

tree', **pokok** *pisang* (tree + banana) 'banana tree', etc.

(b) Head + energy-type modifier, i.e. specifies the type of energy used for the head:

Head **dapur** 'stove'

e.g.: **dapur** *gas* (stove + gas) 'gas stove', **dapur** *elektrik* (stove + electric) 'electric

stove', **dapur** *arang* (stove + charcoal) 'charcoal stove', etc.

(c) Head + hereditary modifier, i.e. specifies the heredity of the head, e.g.:

Head **orang** 'person'

e.g.: **orang** *Melayu* (person + Malay) 'Malays', **orang** *Cina* (person + Chinese)

'Chinese', **orang** *Inggeris* (person + English) 'English', etc.

(d) Head + sex modifier, i.e. specifies the sex of the head:

Head 'human' e.g. **budak** 'child'

e.g.: **budak** *lelaki* (child + male) 'boy', **budak** *perempuan* (child + female) 'girl'

Head 'animal' e.g. **ayam** 'chicken'

e.g.: **ayam** *jantan* (chicken + male) 'cock', **ayam** *betina* (chicken + female) 'hen'

e) Head + body parts modifier, i.e. specifies the body parts associated with the head:

Head **sarung** 'cover/cloth'

e.g.: **sarung** *tangan* (cover + hands) 'gloves', **sarung** *kaki* (cover + feet) 'socks',

sarung *lengan* (cover + arm) 'arm sleeves'

The point to note is that the heads and modifiers hold various types of thematic relation between them, and they are generally considered as noun phrases.

Let us consider the first examples of the *burung* (bird) + modifiers types. Here, the second modifying elements (i.e. *gagak* (crow), *helang* (eagle)*,* and *pipit* (sparrow)), are themselves proper names of the birds. They are plausible stand-alone words even without the head *burung*. In fact, *helang* (eagle), for instance, can be modified into several species of eagle, e.g. *helang merah* (red eagle), *helang laut* (sea eagle), etc. In this sense, the head *burung* 'bird' can be argued as being used loosely to refer to avian animals in general. Thus, the head here has less significance in such constructions, and is perhaps even redundant.

Now consider these two examples of *burung unta* (bird + camel) 'ostrich' and *burung hantu* (bird + ghost) 'owl'. Obviously, the modifiers are referring to the similarities in the attributes of the non-head constituents. Ostriches are flightless birds from a semi-desert environment which can be associated with camel-like attributes. Similarly, owls are nocturnal birds with the ability to make ghost-like sounds, attributes of ghostly activities. Nonetheless, the head *burung* here is syntactically very important as the modifying *unta* and *hantu* without the head will simply refer to a camel and a ghost. Does this make *burung unta* and *burung hantu* more of a compound in comparison to the other *burung* + modifier constructions?

Similarly, consider other combinations with the head *harimau* 'tiger'*,* e.g*. harimau bintang* (tiger + stars) 'leopard'*,* and *harimau kumbang* (tiger + beetle) 'panther'. The head *harimau* here refers to the generic 'big cats' rather than tigers, as neither leopards nor panthers really look like the common tiger. The modifying elements associate *bintang* (stars) with the spotted patterns of the leopard while *kumbang* (beetle) probably signifies the dark colour of the panther. Nonetheless, the head is more significant here as well since *bintang* (stars) and *kumbang* (beetle) are totally different entities without the *harimau* (tiger) head. Again, does

this make them more compound-like than, for instance, *harimau belang* (tiger + stripes) the '(common) tiger'? The point here is that, even within a regular pattern of groupings, the structure of Noun + Noun can be unclear, as illustrated by some of these exceptions. One must agree that they have more compound-like qualities in comparison to the other parallels that have been claimed as phrases.

Now let us look at an attested compound *orang hutan*[74] (person/people + jungle), a compound referring to the primate 'orangutan'. The structure of this compound is similar to an N+N phrase, e.g. *orang kampung* (person/people + village) 'village person/people', i.e. person/people who lives/live in villages. *Orang kampung* will never have the meaning of 'a kind of village primate'. Obviously, the *orang* in *orang hutan* does not carry the exact meaning of a 'person' or 'people', but is more to do with the fact that the primate embodies the characteristics of a human person. The point is, even the structure of an attested compound *orang hutan* has no difference to a regular phrase (i.e. *orang kampung*), i.e. there is little structural evidence to differentiate them. In this sense, given the right context and understanding, there is technically no reason to dismiss *orang hutan* from having the meaning of 'person/people who lives/live in jungles'.[75]

Distinguishing compounds from phrases can be tricky in Malay. This is partly because there are no special attributes to distinguish from one another, and partly due to issues of

---

[74] *Orang hutan* is also spelled as *orang-utan* (or *orangutan*) in Malay. In other words, the latter has undergone phonological assimilation (which contradicts the point of compounds not being overtly marked), but this is one of the few instances and will be considered as an exception. Additionally, a phrasal form, *orang hutan*, is also an unnatural structure to express 'people from/living in jungles; thus, the concatenation of *orang + hutan* is more likely to be a compound. Furthermore, *orang hutan* can be said to be headless or exocentric, being that *orang* here refers to the primate not the human, and is somewhat exclusively concatenated with *hutan* to specifically mean the primate.

[75] The phrase *orang yang tinggal di hutan* (people REL live in jungle) 'people who live in jungle' shows the further specification needed to distinguish between the animal and the human, which can very well refer to people who (for whatever reason) chose to live in the jungle. This is not to be generalised with the indigenous people as they do not necessarily live in jungles, the generic term (compound), i.e. *orang asli* (people + original), which refers to the indigenous/aboriginal people.

transparency between the conjoined constituents. In other languages, compounds and phrases are clearly distinguishable. For instance, Kornfeld (2009) exemplifies the concept of cobweb in Spanish, which can be distinguished as either *telaraña* (net-spider) or *tela de araña* (net of spider). The difference can be justified based on the fact that the constituents of the compounded form undergo a phonological merger (which is not the case with the phrase). Such overt markings are not as apparent in Malay. Furthermore, the heads of both morphological and syntactical constructions in Malay occur on the left. This makes compounds and phrases alike look identical to each other. In languages such as English, where the morphological head is not in the same place as the syntactical head, the two varieties can be adequately differentiated, e.g. *case for books* (syntactic structure) vs. *bookcase* (morphological structure). In other examples, a morphological counterpart of a given syntactic structure can clearly show ungrammaticality, e.g. *they drive trucks* (syntactic structure) vs. *they *truck drive* (morphological structure).

As phrases will always be identical to compounds in Malay, what are the chances then for a morphological merger through the competition model? Ackema and Neeleman (2004) argue that compounds can occur, but only when there is no syntactic competitor; in other words, when the elements that merge differ in their categories or semantic relations in the respective structures. Take, for example, the combination of *rumah* (house) and *api* (fire), deriving *rumah api* as shown below:
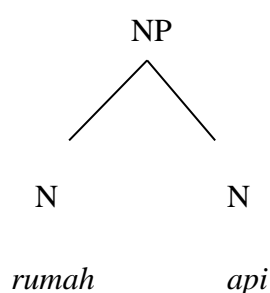
(38) (a) Syntax merger   (b) Morphology merger
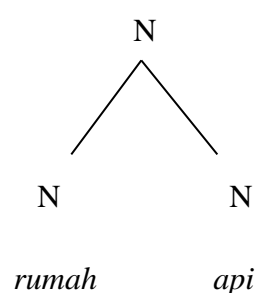
Again, the head *rumah* (house) and the non-head *api* (fire) can be presume to merged in either

(38a) syntax or (38b) morphology. Both mergers involve the same categories (Noun + Noun).

But the difference here is that they have different semantics.[76] The morphological merger of

rumah and *api* expresses semantics that cannot be expressed by the syntactic merger. *Rumah*

api (house + fire) does not mean a house on fire, a house containing fire, or even a house made

for fiery purposes, but rather a tower specifically built for producing light signals. In other

words, the semantic of the construction is not transparent in the syntax; instead, it has an

unpredictable semantics (to mean lighthouse). It is essential for *rumah api* to have an

unpredictable semantics, because otherwise it will be blocked by the syntactic counterpart since

they have the exact same merger of categories. In this sense, one can argue that the

morphological merger must be a triggered construct, and it can be triggered by having

unpredictable semantics. The assumption is also that *rumah api* is stored in the lexicon,[77] again

because it has an idiosyncratic meaning. Thus, it can be said that the merger results in the

root/primary compound *rumah api* (lighthouse). A quick test of modification by colour *merah*

(red) yields [*rumah api*] [*merah*] instead of [*rumah*] [*api merah*], where *merah* modifies the

whole *rumah api* construction, signifying a compounded form.

If we look back at some of the examples given earlier, we can see similar patterns as

well. For instance, it is clear that *burung hantu* (bird + ghost) 'owl' and *burung unta* (bird +

camel) 'ostrich' are not the same as the other head *burung* (bird) type constructions. They

clearly have some idiosyncrasy to them as they have unpredictable semantics. Similarly, *orang*

hutan (person/people + jungle) 'orangutan' forces a morphological merger as it triggers

unpredictable semantics, i.e. the primate rather than a person living in the jungle. Likewise,

---

[76] As Ackema and Neeleman (2004) put it, "Morphological merger of α and β may result in a semantics that cannot be expressed by the result of syntactic merger of the two".

[77] In general, most linguists agree that the entities of the lexicon must at least have some sort of information about the pronunciation, meaning, morphological and syntactic properties, and also the idiosyncratic information of the entities (Di Sciullo and Williams, 1987; Spencer, 1991).

orang kampung (person/people + village) 'village person/people', which has a clear literal

meaning, will undergo syntactic merger. Therefore, is it the case then that the only difference

between a compound and a phrase in Malay is on grounds of semantic unpredictability? We

can try to answer this question by looking at another type of compound that clearly has a

transparent meaning, i.e. synthetic compounds.[78] However, before we can do that, we need to

ask the question of whether Malay has synthetic compounds to begin with.

If we draw parallelisms with some structures in Malay, it will seem like the language

does have something similar to synthetic compounds (at the very least superficially similar).

Consider, for instance, *pemandu teksi* (driver + taxi) 'taxi driver'. This sort of construction is

very productive in Malay and commonly considered as a phrasal structure (Karim et al., 2008).

If we break down the structure of *pemandu teksi*, we can see that it possesses all the ingredients

needed for synthetic compounding, namely:

(39)   (a) Has an agentive prefix *peN-*.[79]

        The prefix is capable of deriving deverbal nouns (e.g. *peN + pandu* 'drive' $\rightarrow$ *pemandu*

        'driver').

        (b) Has noun + noun compounding.

        The structure here will be $[[peN\text{-}\ V]_N\ [N]]_N$ as in $[[peN\text{-}\ [pandu]_V]_N\ [teksi]_N]_N$

        (c) Has verb + noun compounding.[80]

        The second structure here will be $[peN\text{-}\ [[V][N]]_V]_N$ as in $[peN\text{-}[[pandu]_V[teksi]_N]_V]_N$

---

[78] Synthetic compounds are different from root/primary compounds. Their structure consists of a complex head adjective or noun (derived from a verb) and their non-head constituent must be interpreted as an argument of the deverbal head (Spencer, 1991; Katamba and Stonham, 2006). For more on synthetic compounding, refer to Roeper and Siegel, 1978; Selkirk, 1982; Lieber, 1983, among others). Cf. 2.1 (6) on English synthetic compounds with regard to the competition model.

[79] The prefix *peN-* is used because it is the closest analogy to the suffix *-er*, which will assist the analysis.

[80] There are some compounds of this structure but they are very limited, e.g. *cucuk sanggul* (to pierce + hair bun) 'hair pin' and *sapu tangan* (to wipe + hand) 'handkerchief', *kaji bumi* (study + earth) 'geology'.

Recall the two main properties of synthetic compounding: (i) a deverbal complex head and (ii) a non-head constituent interpreted as the argument of the deverbal head. If we look at the first possible morphological structure in 39 (b), we will see that it has the exact counterpart in the possible syntactic structure as well.

(40) Possible morphological mergers of [[*peN* [*pandu*]$_V$]$_N$ [*teksi*]$_N$]$_N$

a) syntax merger   b) morphology merger
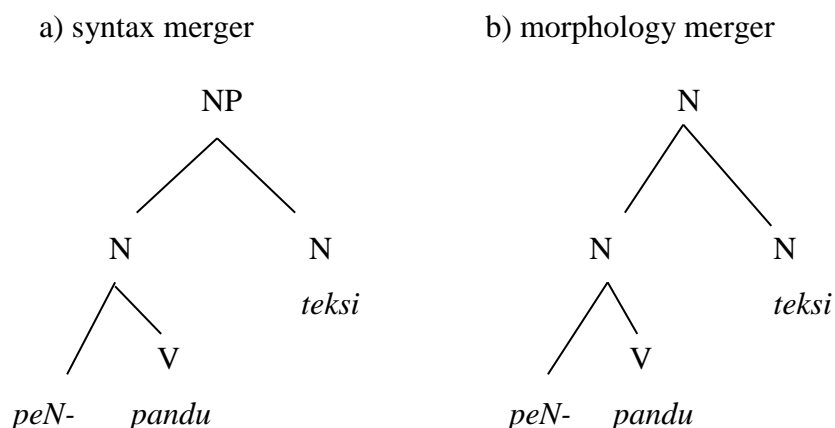
In (40a) the merger of V *pandu* and prefix *peN-* generates the head N *pemandu*, which in turn merges with the N *teksi*. This is exactly the same as in structure (40b). The semantics of both structures are also the same. As all conditions are equal, the compound merger is blocked by the syntactic merger, thus *pemandu teksi* is a phrase here. In this analysis, even if we consider *pemandu* 'driver' inherits its argument structure, and thus takes *teksi* 'taxi' as its argument, there will be no difference between syntax analysis and the morphological analysis, because all conditions between them are the same. They will merge the same lexical categories, with the same assumption of argument binding, along with the same semantic reading, and thus the result is still the same, syntax wins the competition.

However, if we take the second structure in 39 (c), i.e. [*peN-* [[V][N]]ᵥ]ₙ as in [*peN-*[[*pandu*]ᵥ[*teksi*]ₙ]ᵥ] ₙ, below:

(41) Possible morphological mergers of [*peN-*[[*pandu*]ᵥ[*teksi*]ₙ]ᵥ] ₙ

a) syntax merger

```
           NP
          /  \
         /    \
        N      N
       / \    teksi
      /   \
    peN-  pandu  V
```

b) morphology merger

```
           N
          / \
         /   V
        /   / \
      peN- pandu teksi
       V    V   N
```
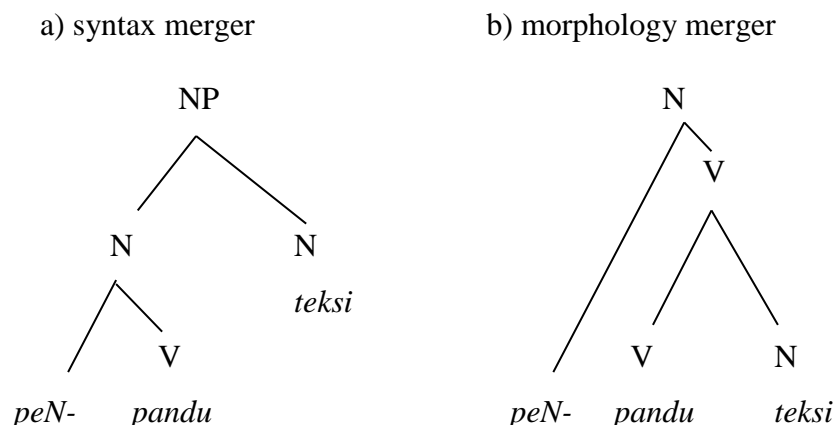
In this case, we will get two different analyses. First, if we accept the ability to inherit argument structure, both syntactic and morphological mergers will result in the same argument structure, i.e. the verb *pandu* in both (41a) and (41b) will take its respective *teksi* as the argument. The only difference between them is the merger of different lexical categories: (41a) is between (N) *pemandu* + (N) *teksi*, while (41b) is between the category-changing *peN-* with the VN compound *pandu teksi*. In this sense, both syntactic structure and morphological structure will be allowed to co-exist. The question of whether VN *pandu teksi* is allowed in the language can be illustrated by several attested compounds in the language such as *cucuk sanggul* (to pierce + hair bun) 'hair pin', *sapu tangan* (to wipe + hand) 'handkerchief' and *tarik tali* (to pull + rope) 'tug-of-war'. A possible counter argument is to say that these compounds are not verbal compounds in the sense that their overall category is a noun. Furthermore, are the nouns in these compounds treated as the argument of the verbs? Interestingly, if we add the prefix *peN-*

[81] to these compounds, they will produce the meaning of the same object that they were meant to derive without the prefix, i.e. *penyucuk sanggul* (*peN*-pierce + hair bun) 'hair pin', and *penyapu tangan*[82] (*peN*-wipe + hand) 'handkerchief', etc. Nevertheless, the point here is that the VN compound is valid for the Malay language (although limited), and the analysis in (41b) is also deemed as a valid analysis. If this is true, we might get a reading of something similar to the synthetic compound 'taxi driver'. We believe the reason why a reading such as (41b) is not acknowledged in the language is again very much to do with the structure of the language itself. As mentioned, both syntax and morphology project the same structure, unlike in English, where the difference between 'driver of a taxi' is overtly projected from the structure of 'taxi driver'.

Having said that, Malay does have constructions such as *peN-makan* (*peN*-eat) + *sayur* (vegetable) = *pemakan sayur* 'vegan/vegetarian' and *peN-tarik* (peN-pull) + *beca* (rickshaw) = *penarik beca* 'rickshaw runner (driver)', among others. These examples seem unified as single unit characteristics more than as compounds. Of course, one can argue that the first example is somewhat irrelevant as one can use the adopted *vegetarian*[83] form instead, and the other example may just be a lexicalised form as rickshaws are no longer pulled by their operators (drivers) as they used to be, but nowadays are driven by cycling. Nonetheless, both constructions are attested forms and they appear to be unified units rather than phrasal combinations, i.e. a *pemakan sayur* does not refer to someone who simply eats vegetables but to someone who strictly eats vegetables as his/her only diet. Similarly, a *penarik beca* is not someone who is doing the act of literally pulling or moving a rickshaw for whatever reason,

---

[81] The prefix *peN-* also derives instrumental nouns.
[82] *Penyapu* here is odd because it usually means 'broom', but, nonetheless, the meaning is still the same: something (in this sense cloth) to wipe the hands.
[83] Coincidently, the spelling of vegetarian complies with the Malay sound and syllable system, thus having a direct spelling and syllabification translation of /ve.ge.ta.ri.an/.

but refers to someone who operates or drives the rickshaw by pulling or cycling as an occupation. Their structure is similar to that of synthetic compounds, i.e. a deverbal head with the potential to have the following noun element interpreted as its object, mimicking the analysis of synthetic compounding. The deverbal heads also do not occur in isolation, i.e. *Ali seorang penarik* 'Ali is a puller', *Ali seorang pemakan* 'Ali is an eater', etc.

To sum up, the competition model effectively allows us to make a few observations on Malay compounding. The combination of lexical items with regular and transparent semantics will result in a phrasal reading (cf. (36) *rumah papan*). In contrast, the combination of lexical items with non-transparent semantics will result in compound reading (cf. (38) *rumah api*). This outcome is mainly due to the fact that both phrases and compounds bear the same head-modifier structure. To a certain degree, we can also extend our assumption that perhaps compounding in Malay is not so productive after all (at least for root/primary compounds), as it seems that the ability to be productive will incline towards a phrasal merger (as productive constructions are prone to have more transparent semantics). In addition, through the competition model, there is a potential reading of synthetic compounding forms (cf. (41b) *pemandu teksi*) in the Malay language. In this sense, we can build a four-way distinction:

(42) Four-way distinction of Malay compounding

	Transparent semantics	Non-transparent semantics
Phrase	*rumah papan* (36a) *pemandu teksi* (40a)	(idiomatic phrases)
Compound	*? pemandu teksi* (41b)	*rumah api* (38b) (idiomatic compounds)

To this end, we can say that Malay compounds are structures whose dependency on their meaning is higher than compounds with transparent structures. However, relying purely on semantics to distinguish Malay compounds from phrases can be misleading. Thus, we must have a proper definition of compounding on formal grounds, which will be the undertaking of the subsequent section.

5.2 Malay compoundhood

This section reiterates the topics of definition, headedness, criteria and classification, in order to define Malay compoundhood.

5.2.1 Definition

We begin this subtopic with a discussion on the Malay definition of compoundhood. As we have seen in section 2.2.1, there is considerable discussion on the attempts to define compoundhood. Defining the term is not an easy or straightforward task. One of the main issues is to identify the components or the units of compounding. Furthermore, different linguists use different units, causing irregularities in definition. Malay linguists can also be accused of succumbing to the same shortcomings. However, the units used to define Malay compounds do not vary too much. Some examples of how Malay compounds have been defined include: compounds are combinations of *roots* (Hassan, 1974), *bases* (Karim et al., 2008), *morphemes* (Musa, 1993; Ismail and Jalaluddin, 2008), and *words* (Karim, 1995; Hassan, 2006). Different linguists might have different reasons behind their definitions.

Let us begin with 'words' as the units of Malay compounding. Recall how Matthews (1974) defined the term word as having the abstract sense of *lexeme* and concrete sense of

word-form, e.g. the lexeme ₜₐBₗₑ realises the word-forms *table* and *tables*. In Malay, the distinction between lexeme and word form is not as obvious as it is in English. This is because the Malay language does not inflect for grammatical functions. In this sense, the grammatical word in Malay is not realised overtly, e.g. the lexeme ₘₑⱼₐ ₍ₜₐBₗₑ₎ realises the word-form *meja* 'table'. Thus, differentiating between lexemes and word forms in Malay is less significant as they both have virtually the same forms. This is also one of the reasons why all affixation and even reduplication for plural marking is considered as derivational in Malay (Hassan, 1974). To this end, we can argue that the Malay language can do without using the term *lexeme*. Instead, the term *word* will be more appropriate as the basic unit of linguistics. Nonetheless, many Malay linguists still use morphemes in the standard definition of being the smallest meaningful unit of language. These two terms are used synonymously in the language.

Now let us look at the remaining terms used as units of Malay compounding, i.e. *roots* and *bases*. Roots can be defined as the smallest core of a form (related to the word/lexeme) with no other attachments. Bases are the form to which affixes are attached. We have seen that one of the reasons roots are defined as compounding units is to account for units that are smaller than lexemes, such as those in neoclassical compounds. However, neoclassical compounding does not have a direct effect on Malay compounding Malay. The language regularly adopts these forms directly, e.g. 'biology' in Malay is *biologi*, 'morphology' in Malay is *morfologi*, etc. They are seen as single un-analysable words in the language. Nevertheless, it is obvious that roots are important building blocks in Malay compounding. Root-root compounds are simply abundant. Almost all compounds in the language consist of root-root compounding, as the previously used examples demonstrate: *papan hitam* (board + black) 'blackboard', *rumah api* (house + fire) 'lighthouse', etc.

The unit of bases might also have some prominence as the building blocks of compounding in the language. Malay has compound forms with derived components, e.g. *ke-*

tua (prefix *ke* + old) = *ketua* (chief), compounded into *ketua* + *menteri* (chief + minister) = *ketua menteri* 'chief minister. It is not the case that this compound is affixed after the merger, i.e. *tua + *menteri* = *tua menteri which is then affixed to *ke-tua menteri. This is to be differentiated from compound affixation, e.g. *daya serap* (absorbance) = *ber-daya serap* (absorptive), *daya serap-an* (absorption), *ke-dayaserap-an* (absorptivity). Thus, to a certain degree, Malay would definitely benefit from having bases as one of its units of compounding.

It is clear that there are several different units in Malay that can be used as its building blocks of compounding. As mentioned, one of the reasons it is difficult to come up with a unified definition on compounding (even within the same language) is because different linguists use different units of compounding. In our opinion, this issue might not be a large problem to Malay. In principle, the types of units used in Malay compounding are not as diverse, for instance, as in languages such as English. Perhaps it is not so much of an issue if we were to follow Plag (2003) and Lieber (2010), i.e. listing all the possible compounding units for the particular language, and specifically using them when necessary. In this case, Malay will be defined as having roots, bases, morphemes and words as its building blocks of compounding.

With that in mind, we can adopt Guevara and Scalise's (2009) schema of a prototypical compounding structure for Malay compoundhood as below:

(43)   (a) Endocentric left-headed compounds: [X *R* Y] (X)

(b) Endocentric right-headed compounds: [X *R* Y] (Y)

(c) Exocentric compounds: [X *R* Y] (Z)

From the above, X and Y are constituents of compounds, and the R has either a grammatical or a semantical relationship linking together the X and Y. The outer X (43a) and Y (43b)

represents the overall category of the compound, which is similar to the head of the compound, while the outer Z (43c) represents the overall category of exocentric compounds. X and Y can be of various categories of the language. As mentioned, Malay can have roots, bases, morphemes and words as its primary units of compounding.

5.2.2 Headedness

In section 2.2.3, we began by asserting that the notion of 'headedness' in the morphological domain has its origins in the concept of syntactic heads (Selkirk, 1982; Zwicky, 1985; Bauer, 1990; Hoeksema, 1992; Katamba and Stonham, 2006; Booij, 2012). We have also come to understand that, although both morphological and syntactical heads can overlap and share common similarities, it is important to recognise them as different entities nevertheless. Malay scholars such as Hassan (1974) also recognise the idea of headedness in Malay compounding as originating from the concept of syntactic heads. Although he did not go into detail on the topic, Hassan (1974, 1986) regularly exemplifies Malay compounds as being either syntactically constructed (i.e. adhering to the default head-modifier sequence of syntactic structure) or asyntactically constructed (i.e. having the modifier-head sequence, hence opposing the default syntactic sequence) (cf. sections 4.1 and 4.2). Identifying the head component of Malay compounds is essential, especially as morphological and syntactical structures in the language overlap significantly. In section 2.2.3, we discussed the several ways of identifying the head in compounds. We looked at how compound heads can be distinguished by identifying the features and properties of a morphological head, categorical head and semantic head (Di Sciullo and Williams, 1987; Dressler, 2006; Scalise and Fabregas, 2010).

It has been suggested that the head of a compound can be identified through morphological features (i.e. morphological heads). What this means is that the head of a given

compound is (usually) the component that bears the grammatical markers applied to the compound (i.e. heads are the locus of morphological features such as inflection, gender, number, etc.). However, as we have mentioned, such features are irrelevant to Malay as the language is without grammatical markers.[84] There are no particular examples of such occurrence in the language, and thus we will not rely on morphological features as a method of head identification in Malay compounding.

On the other hand, the concept of identifying a categorical head has more relevance in Malay compounding. The idea is to identify the position of the most significant component in a given compounded structure, which can then be assumed as the head component. In Malay, the position of the most significant component will (typically) be on the left (e.g. *papan* $_N$ + *hitam* $_A$ (board + black) = *papan hitam* $_N$ 'blackboard'). To this end, we can say that the default head of Malay compounds is the left-hand component.

However, this method of identifying headedness will not produce a regular outcome for the language. Similar to other languages, Malay consists of compounds with multiple head positions as well. There are at least three kinds of compounds with different head positioning in the language, namely: the default left-headed compounds, headless compounds and some right-headed compounds (Hassan, 2006; Karim et al., 2008).[85] Therefore, it is not necessarily the case that the 'left-hand component' will always be the head of Malay compounds. In this sense, we cannot fully rely on the parameter of the 'positioning' of components inside compounds as an indicator of headedness in Malay compounds.

---

[84] Reduplication for plural marking is not considered as an inflectional process in the language (Hassan, 1974).

[85] To a certain degree, this observation of a variable head-modifier order in Malay compounding supports the general conclusion that there is no exclusive preferred head position in compounds, even within a particular language (cf. Bauer, 2001; Ceccagno and Basciano, 2007; Scalise and Fabregas, 2010, among others).

As such, yet another proposed method is to identify the component responsible for the overall grammatical category of the compound. In other words, the component with the same category as the category of the whole compound is (therefore) the head of the compound. This is true for most Malay endocentric compounds. As exemplified above, the head of *papan* $_N$ + *hitam* $_A$ (board + black) = *papan hitam* $_N$ 'blackboard' must be the noun *papan* 'board', and not the adjective *hitam* 'black', since the whole compound behaves as a noun.

Then again, this method of identification can be unclear when the components of a compound are of the same lexical category, e.g. *kerusi* $_N$ + *roda* $_N$ (chair + wheel) = *kerusi roda* $_N$ 'wheelchair'. In this example, we cannot simply say that the category of the whole compound (i.e. noun) is determined by a 'noun constituent' within the structure. This is because there are two 'noun' constituents involved, and thus we must be clear about which of the two nouns is influencing the outcome of the compound's overall category.

To counter this problem, it has been proposed that we can identify the head by looking at the meaning of the compound instead (i.e. the semantic head). In this case, the compound *kerusi* + *roda* (chair + wheel) = *kerusi roda* 'wheelchair' is a kind of chair, not a kind of wheel. In other words, *kerusi roda* 'wheelchair' is the hyponym of *kerusi* 'chair' (and not of *roda* 'wheel'), and we can thus argue that the noun *kerusi* 'chair' is the head of the nominal compound *kerusi roda* 'wheelchair'. Indeed, the head in a compound can be identified by recognising the 'hyponym of' the relationship between the compound and its components (Hoeksema, 1992; Haspelmath and Sims, 2010; Scalise and Fabregas, 2010).

However, there are many compounds with components of a similar lexical category, but do not have such a clear 'hyponym of' relationship. Take, for instance, the compound *tanah* $_N$ + *air* $_N$ (land/soil + water) = *tanah air* $_N$ 'country/homeland'. It is unclear as to which noun is the head of the compound; in other words, which of the two nouns (exactly) is projecting its

categorical features to the whole compound (making it a nominal compound as a whole). This sort of compound is not the same as the *kerusi* ₙ + *roda* ₙ (chair + wheel) = *kerusi roda* ₙ 'wheelchair' type. The nouns involved in *kerusi roda* have a dominant component, i.e. *kerusi* 'chair' is arguably the core component which produces the meaning of 'a special kind of chair (with wheels)'. Instead, in *tanah air* (land/soil + water), the noun *tanah* 'land/soil' has no (necessary) higher importance than the noun *air* (water) as they both have equal significance in deriving the meaning of 'country'.[86] Thus, it can be unclear as to which noun component (i.e. *tanah* 'land/soil' or *air* 'water') transfers its lexical features to the overall nominal compound.

Even if the components involved are not of the same category, identifying the head component through semantic features can still be misleading. Take, for instance, the compound *sapu* ᵥ + *tangan* ₙ (to wipe + hand) = *sapu tangan* 'handkerchief'. The meaning of the compound is more relevant to the verb component *sapu* 'to wipe' as the compound denotes 'a piece of cloth used for wiping (hands)' rather than *'a hand used for wiping'. The semantics of the compound implies that the verb *sapu* 'to wipe' should be the head, but the grammatical category of the whole compound is not a verb, it is a noun instead.

Clearly identifying the head component is not a clear-cut process. Nevertheless, we can say that headedness in Malay compounds can be identified through at least two properties of the categorical and semantical features. Ideally, the head component in a given Malay compound should be able to transfer both of these features to the whole compounded form. This, however, is not always possible. As illustrated above in the mismatches of categorical

---

[86] Arguably the compound *tanah air* (land/soil + water) 'homeland' has a metaphoric reading, i.e. land/soil and water obviously represent some of the most important elements that define a country/homeland. The point here is that both components are as important as the other, to the extent that we cannot truly say (for instance) that the noun *tanah* 'land/soil' is the more prominent component in deriving the meaning of the compound *tanah air* 'country'.

and semantical features, the head of a Malay compound does not necessarily transfer both features to the compound as a whole. Therefore, it will be difficult if we are to say that the head of a compound must be the component that is able to transfer all of its available features to the overall compound. The notion of headedness must be perceived as the particular feature (i.e. categorical and/or semantical features) under consideration (Di Sciullo and Williams, 1987; Dressler, 2006; Scalise and Fabregas, 2010). Adhering to this idea, along with the evidence of Malay compounds (and their mismatches) presented above, we believe that a dual definition of headedness distinguishing between categorical and semantic heads is fitting to account for the concept of headedness in Malay compounds. This is the approach that we will take to identify the head component of Malay compounds in this study.

5.2.3 Criteria

In section 2.2.4, we saw how linguists have tried to define compoundhood through several properties, namely the orthographical, phonological, morphological, syntactic and semantic criteria. The main idea is to use these criteria as a tool to measure whether a given structure can be considered as a compounded form or otherwise. Indeed, the criteria have played an important role in defining the concept of compoundhood. However, their degree of importance and relevance varies from one language to another. In this section, we will look at how they fare against the concept of Malay compoundhood.

We begin with arguably the most irrelevant criteria to define Malay compoundhood; that is, the morphological measures. Concerning this topic, we have looked at how certain morphological properties such as inflectional markings and linking elements can be useful in identifying and differentiating between compounds and phrases in languages like English, German, Greek and Dutch (Bauer, 2006, 2009; Ralli, 2009; Booij, 2010). Nevertheless, we

have also come to understand that morphological properties as such can be highly language-specific. Not all languages have these sorts of properties with which to work. Case in point: Malay is a language without inflection and linking markers (Hassan, 1974, Karim et al., 2008). To this end, we cannot take the morphological criteria into account for Malay compoundhood as they are not applicable to the language.

With regard to the orthographic criteria, the main concern of this topic concerns the issue of spelling convention. The basic tenet is that a given compounded structure should be able to project the unification of its component through its spelling convention. In other words, the components involved should be spelled as a single conjoined unit. However, we have come to understand that compound spelling can be a consistent and predictable feature in some languages, but irregular and unpredictable in others (Bauer, 2006; Lieber and Stekauer, 2009).

To a larger degree, we can safely categorise Malay as a language with regular spelling patterns for its compound forms. This is because Malay does not suffer as much variation and/or irregularities in terms of its compound spelling principle in comparison to other languages (e.g. English *rain forest*, *rainforest*, and *rain-forest*). Except for a small set of compounds,[87] all compounds in Malay are always spelt separately (Hassan, 1974, 1986, 2006; Musa, 1993; Sew, 2007; Karim et al., 2008; Ismail and Jalaluddin, 2008). Hassan (1986) points out that there have been some attempts to regulate the issue, i.e. to force a combined spelling convention between the two constituents of compounded forms. This was to reinforce the idea of compounds being a single unit and therefore that they should be spelled as a single entity as

---

[87] Most linguists acknowledge 13 compounds in the language that are spelled as single words, i.e. the so-called 'established' compounds (cf. appendix (3)). What is implied by 'established' has never really been explained in the literature. However, from what we can understand, these compounds are either: (i) archaic in nature (e.g. borrowed or adapted into Malay from other languages a long time ago), (ii) some do not follow the regular head-modifier convention (e.g. *maha* 'great' (modifier) + *guru* 'teacher' (head) = *mahaguru* 'great teacher'), or (iii) they are simply spelled that way out of convention (e.g. *mata* + *hari* (eye + day) = *matahari* 'sun'). Regardless, the underlying point is that there is no particular reason why they are spelled as such.

well (cf. sections 4.2 and 4.3). Hassan (1986, 2006), however, argues that orthographic convention has no weight to determine Malay compoundhood, mainly because it has nothing to do with the grammar of the language. We agree on this matter and conclude that an orthographic convention such as spelling is not a reliable measure for Malay compoundhood (Hassan, 1986, 2006; Lieber, 2010).

Moving along, we now look at another less reliable criteria of Malay compoundhood, i.e. the semantic criteria. As mentioned in section 2.2.4, the topic of the semantic criteria of compoundhood is not to be confused with the semantics of compounding (i.e. the latter concerns the meaning relation between the constituents of a given compound). Here, the semantic criteria mainly concern the issue of lexicalisation. In other words, it has been argued that, if compounds are indeed unified entities, they should therefore be lexicalised (as single entities) in the lexicon, similar to any other singular word forms. Obviously, there are compounds in the language that are more lexicalised (hence less compositional) than others. These types of compounds are most evident through the established forms (e.g. *kakitangan* (foot + hand) 'staff') and the idiomatic forms (e.g. *panjang tangan* (long + hand) 'thief'). Such compounds do embody the characteristics of lexicalisation in the sense that their meanings are not the sum of their parts, they are unproductive, and they are limited in number.

However, the idea that all compounds must be lexicalised is a difficult position to hold. Compounds are by nature compositional, i.e. the overall meaning relates to the part of its components,[88] they are productive and not limited in number. Even lexicalised compounds arguably do not (totally) lose their compositionality (Lieber, 2005; Bauer, 2006). Take, for example, established compounds such as *matahari* (eye + day) 'sun' and *bumiputera* (earth +

---

[88] Recall section 4.8, where Ismail and Jalaluddin (2008) also argued that the meaning of a Malay compound should not be the sum of its parts, i.e. cannot be the hyponym of the head. We disagree with this understanding because adhering to such an idea will hinder the natural productivity of compounding.

son/prince) 'native'. Although it is unlikely that speakers are constructing these compounds word-for-word under regular circumstances (e.g. 'sun' = *mata* + *hari*, instead of 'sun' = *matahari*), the compounds still retain a certain degree of compositionality as speakers are nonetheless aware of the concatenation involved (i.e. the fact that these compounds are the combination of two words). To this end, we can say that the semantic measure of lexicalisation is indeed a feature of Malay compounding, but it cannot be taken as a compulsory measure of Malay compoundhood.

Moving on to phonological criteria, we have seen how phonological properties such as assimilation and stress pattern can assist in defining compoundhood in certain languages (Spencer, 2005; Lieber, 2005; Bauer, 2006; Haspelmath and Sims, 2010). Whether Malay has any particular phonological properties as such is questionable. Let us consider assimilation as a possible feature of Malay compoundhood. In Malay, a given compound can undergo assimilation if the final phoneme of its first component and the initial phoneme of its second component are the same sounds. For example, the compound *kereta* /kəreta/ (car/cart) + *api* /api/ (fire) = *keretapi* /kəretapi/ 'train'; in this case, the final /a/ sound of the first component matches the initial /a/ sound of the second component, hence the assimilation between the two components involved.

However, this sort of assimilation in Malay compounding can be considered as an exclusive occurrence as it is very limited and unpredictable. Firstly, if the converging sounds of the components (i.e. the final sound of the first component and the initial sound of the second component) are not the same, it is quite obvious that such assimilation will not happen (e.g. *papan* /papan/ (board) + *hitam* /hitam/ (black) = *papan hitam* /papan hitam/ 'blackboard', and not */papanhitam/ or */papanitam/). This observation is also apparent in compounds with components that are compulsorily spelled together (i.e. the 13 established compounds). In the established compounds, the final and the initial sounds of the converging components will not

assimilate, but will retain their individual sounds instead, e.g. *kaki* /kaki/ (foot) + *tangan* /taŋan/ (hand) = *kakitangan* /kakitaŋan/ 'staff' (and not */kakiaŋan/, etc.).

Having said this, even if the converging phonemes are of the same sound, it is not necessary that assimilation will take place, as evident through many examples in the language, e.g.:

(44) Examples of non-assimilation of similar phonemes convergence

Compounds	Orthographic transcription	Phonemic transcription
gambar + rajah /gambar/ + /radʒah/ (picture) + (diagram)	*gambar rajah* 'diagrammatic picture'	/gambar radʒah/ not */gambaradʒah/
tengah + hari /təŋah/ + /hari/ (middle) + (day)	*tengah hari* 'midday/noon'	/təŋah hari/ not */təŋahari/
pasar + raya /pasar/ + /raja/ (market) + (large)	*pasar raya* 'supermarket'	/pasar raja/ Not * /pasaraja/
alat + tulis /alat/ + /tulis/ (tool) + (write)	*alat tulis* 'stationery'	/alat tulis/ not */alatulis/
merah + hati /merah/ + /hati/ (red) + (heart)	*merah hati* 'maroon'	/merah hati/ not */merahati/

The above illustrates that, even when the assimilation condition are met, assimilation does not necessarily occurs. It is an unpredictable (and very much exclusive) occurrence, which is why assimilation cannot be taken as a reliable phonological measure of Malay compounding.

Yet another phonological property to be considered is the stress pattern in Malay compounding. Studies on this topic have mostly concentrated on syllabic stress patterns at word level in the language. Some Malay linguists have argued that the language in general has very weak or even no word stress at all (Maris, 1980; Mohd Don et al., 2008). Let us take a regular compound such as *rumah hijau* (house + green) 'greenhouse' as an example. There is just not enough evidence to say (for instance) that placing the stress on the head *rumah* (house) as in `*rumah hijau* will give the reading of 'green-coloured house', while stressing the non-head constituent *hijau* 'green' as in *rumah `hijau* (house + green) will give the reading of 'greenhouse' (or vice versa).

However, a compound such as *ibu bapa* (mother + father) 'parents', might just have some weight with regard to stress placing. For instance, an unstressed realisation of this compound produces the intended regular meaning of 'parents'. Arguably, if the first component *ibu* 'mother' is stressed as in `*ibu bapa*, it can actually convey the meaning of 'mother of (the) father' or 'father's mother' (i.e. grandmother). This is, however, a forced reading.[89] The regular way of addressing the relationship between a father and his mother would obviously be to use the actual word for 'grandmother' (i.e. *nenek*), or, if it really needs to be, the use of the preposition *kepada* 'to' as in *ibu kepada bapa* 'mother to (the) father' is more appropriate and clearer.

---

[89] It is a forced reading because stress placement to convey different readings is not a regular property of the language.

If we look at the neighbouring Indonesian language,[90] it has also been argued that, in general, there is no clear word stress in that language either (Halim, 1974; Van Zanten and Van Heuven, 1998). In an experiment on auditory recognition of Indonesian word stress, Van Zanten and Van Heuven (1998) reported that their respondents were not using stress information to differentiate between the words in the experiment. They concluded that "Word stress information was not used by our Indonesian listeners to differentiate between words. Our results indicate that stress is communicatively irrelevant and essentially free in Indonesian" (Van Zanten and Van Heuven, 1998: 142). Taking this assumption into account, along with the insufficient information on Malay compounding stress patterns, we are forced to conclude that stress is not a reliable property for Malay compoundhood. More studies will be needed on this topic if we want to seriously consider stress as an accountable measure of Malay compoundhood.

Malay linguists and others have commonly referred to syntactic criteria as the more prominent measure of compoundhood. As mentioned, some of the topics include the issue of modification and separation of the compounded components (Hassan, 1986, 2006; Bauer, 1988; Spencer, 1991; Musa, 1993; Karim, 1995; Karim et al., 2008; Ismail and Jalaluddin, 2008). Basically, if the components of a given structure are not affected by such issues, the structure can then be argued as a unified form (i.e. a compound) rather than a linear sequence of components (i.e. a phrase). Some syntactic 'tests' for Malay compoundhood include:

(i) *Inseparability of the constituents*

The test for inseparability works quite well on Malay compounding. It is argued that insertion of other elements is not allowed as it will not only disrupt the unity of the constituents

---

[90] We acknowledge our initial claim that this study is on the (Malaysian) Malay language, and not of the Indonesian variety. The reference here should then be taken as an analogy of a stress pattern of a closely related language (and not as an Indonesian stress pattern defining the Malay language).

but also the intended meaning of a given compound (Musa, 1993; Hassan, 1986, 2006; Karim et al., 2008). For example, the components of a compound like *kerusi* + *malas* (chair + lazy) = *kerusi malas* 'lounge chair', cannot be separated by modifiers such as *baru* 'new', as in **kerusi baru malas* *'a new chair (being) lazy'. The insertion clearly disrupts the cohesion between the two constituents and the intended compound meaning of 'lounge chair' is thus not realised. If, however, insertion does not disrupt the meaning of a given combined structure, the structure can then be argued as a phrase instead of a compound.[91]

However, we have also argued that the insertion test can be a misleading measure of compoundhood (cf. sections 4.3.1 and 4.6.1). For instance, the insertion test cannot properly work in Malay because most of the combined structures can be equally read as both a compound and a phrase.[92] In other words, insertion of an element within a given structure can disrupt the intended compound meaning while at the same time maintaining a phrasal meaning as well (cf. 4.6.1 on example *tahi lalat* (faeces + fly), which produces the compound 'mole' and the literal phrase 'fly faeces').

Yet another concern is the fact that the type of element use in the insertion test is not specified. As mentioned, Malay linguists such as Hassan (2006), Karim et al. (2008) and Ismail and Jalaluddin (2008) have all freely used *yang* 'which/that', *dan* 'and', *untuk* 'for', *dari* 'from', *kepada* 'to', *milik* 'own', *antara* 'between' and *tempat* 'place' as elements for the test. There seems to be no specification regarding the type of insertion allowed, as a given structure can be justified as either a compound or a phrase over any kind of insertion (cf. 4.3.1).[93]

---

[91] This argument is based on the understanding that components of phrases are not fused in the same way as components of compounds are. Therefore, insertion will not disrupt the given structure as phrases (unlike compounds) are 'open' constructions.

[92] Evidence of this argument can be seen in the mismatches between compounds and phrases in Karim et al. (2008), cf. section 4.1.6.

[93] From our observation, it seems that Malay linguists use the test of insertion to force out the inherent (grammatical and/or semantic) relation between the components (which is not overtly expressed) in a compound. For instance, Hassan (2006) argued that a structure such as *rumah makan* (house + eat) 'restaurant'

Furthermore, the concept of insertion has never been properly addressed either. For instance, Malay scholars regularly use *yang* 'which/that' as an element of the insertion test (e.g. Hassan, 2006; Karim et al., 2008; Ismail and Jalaluddin, 2008). For instance, Malay scholars will exemplify a structure such as *rumah besar* (house + big) 'big house' as a phrase because *yang* can be inserted into the structure without disrupting the intended meaning, as in *rumah yang besar* '(a) house that is big'. On the other hand, given a structure such as *rumah sakit* (house + sick/ill) 'hospital', insertion of *yang* derives *rumah yang sakit* '(a) house that is sick/ill', illustrating that the insertion disrupts the intended meaning of 'hospital', and thus the construction is considered as a compound. The problem, however, is that *yang* is a relative marker (hence forming a relative clause), and relativisation is not the same as insertion. In this sense, it is inaccurate for Malay scholars to characterise *yang* as an element of insertion to prove whether a given structure is a compound or a phrase.

(ii) *Inability to switch components*

This test argues that it should not be possible to switch the elements of a compound (Ismail and Jalaluddin, 2008). Again, this is to prove that the components are bonded, hence the inability to switch them. Thus, *kerusi malas* (chair + lazy) 'lounge chair' upon switching

---

is not a compound because one can insert *tempat* 'place' as in *rumah tempat makan* (house + place (to) + eat) and maintain the meaning of 'restaurant'. Since there is no disruption, the structure *rumah makan* is considered as a phrase instead of a compound. This sort of analysis can be problematic given that there is no specification in the type of insertion allowed as a test element. Since 'any' kind of insertion is allowed, there will therefore always be elements that can be inserted to reveal the inherent relationship between the components. In this sense, Malay will virtually have no compounds because all compounds naturally have this 'inherent' relationship between their components, and the relationship can obviously be explained through forced insertion (in one way or another). As Scalise and Bisetto (2009: 44) put it: "The peculiarity of compounds lies in their being word forms whose constituents are connected by a grammatical relation that is not overtly expressed (…)". Thus, the act of inserting (any kind of) element to force out the inherent relation between the components of a given compound (and then ironically justifying it as a phrase because it can undergo insertion and retain its meaning) is not necessarily a correct analysis.

becomes *malas kerusi* *'lazy (like a) chair'. The switching causes the intended compound to lose its meaning.[94]

(iii) *Ability to undergo circumfixation*

The ability of compounds to be circumfixed is a rather language-specific test of Malay compoundhood. The test assumes that, if a given two-component structure is a compound, it should therefore be able to merge its spelling and retain its intended meaning when circumfixed (Karim, 1995; Ismail and Jalaluddin, 2008). Such an outcome is seen as evidence that the components involved are unified entities, hence a compounded form. For instance, the structure *ibu saudara* (mother + relative) 'aunty' when circumfixed with *ber-…-kan* produces *beribusaudarakan* 'to have (someone) as an aunty'. The structure *ibu saudara* can therefore be argued as a compound because it retains its core meaning (i.e. aunty) and the components involved merge after the circumfixation (Ismail and Jalaluddin, 2008).

However, this test is not without problems. For instance, it does not specify which of the many circumfixes available in Malay can be (or should be) used for the test. As mentioned, Karim (1995) and Ismail and Jalaluddin (2008) randomly used the *peN-…-an*, *per-…-an*, *ke-…-an, ber-…-kan*, *di-…-kan*, and *meng-…-kan* circumfixes throughout their studies to accommodate the test. Does this mean that any kind of circumfixation can be used as long as it can justify in one way or another whether a structure retains or loses its meaning after the circumfixation?[95] Furthermore, the effect of circumfixation on two-component structures is not exclusive only to compounded structures. Take, for instance, the phrase *mudah sampai* (easy

---

[94] Nevertheless, there are some counter examples of structures that retain the same (or very similar) meaning even if the components are swapped, e.g. *ibu jari* (mother + finger) 'thumb', and similarly *jari ibu* (finger + mother) 'thumb' (although *jari ibu* can also mean 'finger belonging to (a) mother' i.e. 'mother's finger').

[95] Ismail and Jalaluddin (2008) dismiss the structure *rumah sakit* (house + ill) 'hospital' as a compound, arguing that the structure cannot be circumfixed with *ber-…-kan* (a verbal circumfix) as in **berumahsakitkan* '? to have (something) as a hospital'.

+ reach) 'easy to reach', when circumfixed with *ke-...-an* it produces *kemudahsampaian* 'accessibility'. This example illustrates how the phrasal structure not only maintains its intended meaning, but also requires the two components involved to be spelled jointly after the circumfixation. Therefore, the outcome of circumfixation on two-component structures will not necessarily clarify whether the structure is a compounded form or otherwise.

(iv) *Ability to be fully reduplicated*

This is yet another language-specific measure by which to test Malay compoundhood. Linguists like Hassan (2006) have argued that, if a structure is a compound, reduplication should therefore occur on both constituents as a whole (which proves the bond between the constituents involved). For instance, *uji kaji* 'experiment' reduplicates as *ujikaji-ujikaji* 'experiments', not **uji-uji kaji* (Hassan, 2006). A given structure can then be argued as a compounded form if the outcome is as such. If otherwise, a given structure can then be argued as a phrase instead. However, as we have argued before, this test is also not foolproof because there are many true compounds that only require the head component to be reduplicated, e.g. *rumah sakit* (house + sick/ill) 'hospital' reduplicated as *rumah-rumah sakit* 'hospitals', not **rumahsakit-rumahsakit* (cf. 4.3.1).

As we have pointed out, the syntactic measures presented here can be seen as 'tests' for Malay compoundhood. In one way or another, these tests have assisted linguists in the effort to define Malay compoundhood. As we have seen, however, they are not without their shortcomings. However, we do not necessarily see these shortcomings as a form of weakness, but instead more as a form of limitations of the tests. Therefore, when we utilise the tests, we are aware of the issues and limitations pertaining to each one, which will help us to provide an appropriate analysis outcome. We end this topic by asserting that syntactic measures such as

the ones discussed above are indeed very useful and arguably the most reliable measure of compoundhood available to the language.

We have seen and discussed how orthographical, phonological, morphological, syntactic and semantic criteria fare against Malay compounding. It is clear that each group of criteria has different degrees of relevance as a measure of compoundhood in the language. We conclude this section by summing up the degree of importance and relevance of the criteria that we have discussed above in defining Malay compoundhood.

We begin by looking at the orthographic criteria in terms of compound spelling convention. Although it is a regular and consistent feature in Malay compounding, it is nonetheless not distinctive enough to exclusively distinguish between compounds and other structures (such as phrases) in the language. With regard to the morphological criteria, the main issue concerns the understanding of compound heads as locus of inflections. We have dismissed this measures all together as Malay is arguably without such properties. We can conclude that orthographic and morphological criteria (in general) are not suitable criteria to account for Malay compoundhood.

With regard to the semantic measure (i.e. lexicalisation), we have come to agree that some compounds in Malay are 'more' lexicalised in the sense that they are less compositional than others. However, to fully support the idea that all compounds must be lexicalised forms is not accurate or even practical. Compounds are by nature compositional, productive and not limited in number (all of which are the opposite characteristics of lexicalisation). Therefore, lexicalisation as a measure of compoundhood in Malay is a valid one, but is only applicable to a certain (limited) degree. With regard to the phonological criteria, we have seen how properties such stress patterning and assimilation can be possible features in defining Malay compoundhood. Recall how structures such as *ibu bapa* (mother + father) can arguably have

different meanings of either 'parents' or 'grandmother' due to stress placement. Similarly, recall how structures such as *kereta* + *api* (car/cart + fire) can be assimilated to form *keretapi* 'train'. However, as we have argued, these features are very limited and unpredictable – to the point that they can be considered as exceptional occurrences. We conclude that semantic and phonological criteria can be useful measures of Malay compoundhood (to a certain degree), but under limited and specific circumstances.

We are left with the syntactic criteria of Malay compoundhood. We have seen how the issues of inseparability, inability to switch components, circumfixation and reduplication abilities, have helped to shape the concept of compoundhood in the language. We have also highlighted the shortcomings of each test. However, the shortcomings should not be taken as a deterrent to utilise the tests, but instead to ensure that one is aware of the limitations when applying such tests of compoundhood.

Clearly, it is not necessarily the case that a structure must comply with all of the proposed criteria in order for it to be categorised as a compound. A given structure can be justified as a compound by one measure but not by another. This raises the question of whether a structure is less of a compound than one that adheres to more (if not all) of the proposed criteria. Indeed, there are some linguists who only use a single criterion as a decisive measure of compoundhood (e.g. Karim et al., 2008). On the other hand, there are others who take a stricter route by only accepting structures as compounds when they comply with all of their selected criteria (e.g. Ismail and Jalaluddin, 2008). Clearly, a balanced and appropriate group of criteria by which to measure and account for Malay compoundhood is needed. Hence, through the discussion in this section, we can conclude that the suitable criteria for Malay compounding are the syntactical, semantic and phonological criteria.

5.2.4 Classification

We have looked at how compounds are classified differently and rather inconsistently by different linguists in the literature (cf. Spencer, 1991; Fabb, 1998; Plag, 2003; Booij, 2005, Haspelmath and Sims, 2010; among others). We also discussed the points raised by Scalise and Bisetto (2009) on the reasons for such inconsistency, i.e. (i) over focus on particular categories, (ii) use of different terminologies, and (iii) inconsistent criteria for compound classification.

In the previous section, we looked at the third point, i.e. the issue of inconsistent criteria for Malay compoudhood (and thus we can omit this point here). Concerning the first point, it is apparent that the literature on Malay compounding has focused more on certain categories in comparison to others. Malay linguists generally recognise three main categories of compounding, namely the nominal, verbal and adjectival compounds (e.g. Hassan 1986, 2006; Musa, 1993; Karim et at., 2008, among others). In general, nominal compounding has been given more attention than the other types. This is perhaps due to the general fact that nominal compounds are usually more productive than others. Arguably, there are a few other categories involved in Malay compounding. For instance, Hassan (1974) recognises some adverbal and prepositional compounds, while Sew (2007) recognises appositional and coordinative compounds in the language. However, these types of compounds are somewhat neglected in the sense that they are usually minor occurrences in the language (e.g. adverbal and prepositional compounds), while others are neglected because they are not acknowledged by the mainstream literature (e.g. appositional and coordinative compounds). Concerning the use of different terminologies to classify compounds, Malay does not seem to have much of a problem with this topic. In the literature, Malay linguists neither use Sanskrit terms (like *bahuvrihi* or *dvandva*) nor do they use regular ones (like synthetic, appositional and coordinative) to classify compounds. From the previous discussion, we can say that Malay linguists generally recognise root/primary compounds with the distinction of being either

endocentric or exocentric types.[96] In this sense, the topic on Malay compounding receives less complicated classification in comparison to other languages. However, there can be a downside to this simplicity, as it can actually cause the language to constrain or not acknowledge certain possible compound types in the language.

As we have hinted earlier, in section 2.2.5, this study will be adopting the classification of compounding as proposed by Bissetto and Scalise (2005). The classification claims that compounds can be clearly classified under three categories, namely as *subordinate*, *attributive* and *coordinate* compounding. These categories basically represent the grammatical relation that exists between the components of a given compound. The constituents of subordinate (SUB) compounds have the grammatical relation of complementation, i.e. the non-head complements the head. The constituents of attributive (ATR) compounds have the grammatical relation of attribution or modification, i.e. the non-head conveys a property of the head. Finally, the constituents of coordinate (CRD) compounds have the grammatical relation of coordination, i.e. both components coordinate with each other. A second level to each category distinguishes between endocentric and exocentric type of compounds. An abstract illustration of Bisetto and Scalise's (2005) classification is shown below:

---

[96] There are some linguists who separate Malay compounds into further subclassifications, for instance, into the established forms, idiomatic forms, scientific forms and so forth (e.g. Hassan, 1974; Karim, 1995; Karim et al., 2008, among others). Nevertheless, these classifications adhere to the endocentric or exocentric root/primary compounding distinction. Cf. chart (12) *Compound classification summary chart,* to compare how other linguists classify compounds in other languages.

(45) Bisetto and Scalise's (2005) compound classification diagram[97]

```
                          Compounds
            ┌─────────────────┼─────────────────┐
    Subordinate (SUB)   Attributive (ATR)   Coordinate (CRD)
        ╱    ╲              ╱    ╲              ╱    ╲
     Endo.   Exo.       Endo.   Exo.       Endo.   Exo.
```
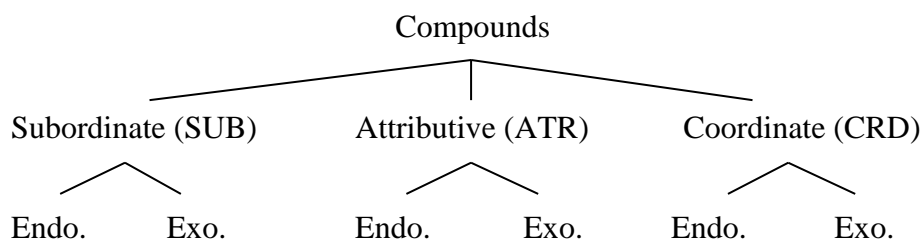
However, this classification was revised by Scalise and Bisetto (2009) (cf. 2.2.5 diagram (15)). From our understanding, one of the main reasons for the revision was to accommodate two types of compounding. First, the subordinate (SUB) is added with a second subclass known as *verbal-nexus*, i.e. to account for secondary/synthetic compounding. Another added subclass is within the attributive (ATR) category known as *appositive* compounding.[98] In the revised version, compounds with adjectival or verbal non-heads are known as attributive compounds, while compounds with nominal non-heads are known as appositive compounds. Nevertheless, the non-heads of both attributive and appositive compounds are still said to be expressing the quality/property of their respective heads.

We believe that it is unnecessary to impose the revised classification on to Malay compounding. As we have come to understand through our previous discussions, Malay technically has no synthetic compounds. Furthermore, we do not see a clear need to distinguish between attributive and appositive compounds in Malay. The two subcategories (in the revised version) essentially discriminate the non-head component as being either a verb or an adjective for attributive compounds, while the non-head is a noun for appositive compounds. However, they are described as having the same grammatical relation, i.e. the non-head modifies the head

---

[97] This is a reiteration of diagram (13) in section 2.2.5. Cf. this section for the discussion on Bisetto and Scalise's (2005) compound classification.

[98] Note that it is not necessarily the case that these two types of compounds (i.e. synthetic and appositive compounds) cannot be accounted for by the original classification. The revision is more of an attempt to overtly separate them from the categories under which they were previously subsumed.

in terms of expressing the quality/property of the head. Therefore, we will not complicate this matter for Malay compounding. We will follow the initial classification, i.e. compounds with non-heads (be it noun, verb and/or adjective) that modify and/or express the quality of the heads will be categorised simply as attributive (ATR) compounds. Overall, we believe that it will be more elegant for Malay compounds to be categorised according to the original compounding scheme. For this reason, we will adopt the classification as proposed by Bisetto and Scalise (2005) for the analysis of Malay compounding in this study.

5.2.5 Section Summary

Before we move on to the next part, let us have a quick summary of what we have discussed so far in this subsection. The aim here is to tie up the loose ends on the issues of Malay compoundhood. This concerns the issues of definition, components, headedness, criteria and classification of Malay compounding.

We started with the definition of and the units of compounding (cf. 5.2.1). We have seen how different linguists use different units for different reasons. We concluded that it is best to follow the likes of linguists such as Plag (2003) and Lieber (2010), i.e. to acknowledge all of the possible units of compounding for a particular language. We agreed that Malay can have roots, bases, morphemes and words as its primary units of compounding. We also agreed on the prototypical structure of [X R Y] (X) (Y) (endocentric) and [X R Y] (Z) (exocentric) for Malay compounds. X and Y are the components of the compound with R as the representation of either a grammatical or a semantical relationship linking between them. X and Y can be of various categories of the language, and the outer X, Y and Z represent the overall category as either endocentric or exocentric compounds.

On the topic of headedness, we acknowledge that Malay can have at least three kinds of compounds with different head positioning, i.e. the default left-headed compounds, headless

compounds and some right-headed compounds. We looked at how the head component of Malay compounds does not necessarily transfer all of its features to the compound as a whole. We argued that it can be difficult to defend the idea that the head of a compound must be the component that is able to transfer all of its features to the overall compound. Thus, we have come to agree that the concept of headedness in Malay compounds is best identified through a dual definition of headedness, i.e. distinguishing between categorical head (i.e. the head is the part that determines the whole category of the compound) and semantic head (i.e. the head is the hyponym of the compound) (cf. 5.2.2).

On the topic of compoundhood criteria, we have looked at how the orthographical, phonological, morphological, syntactical and semantical criteria fare against Malay compounding. We have argued that the suitable criteria for Malay compoundhood are the phonological, semantical and syntactical criteria (cf. 5.2.3). However, phonological and semantical criteria such as lexicalisation, stress patterning and assimilation are only relevant as measures of Malay compoundhood to a limited and specific degree. On the other hand, syntactical criteria such as inseparability of components, inability to switch components, circumfixation and reduplication abilities are more reliable measures of Malay compoundhood.

Finally, we looked at the issue of classification where we consider Bisetto and Scalise's (2005) proposal on classification of compounding (cf. 5.2.4). It is argued that compounds can be clearly classified by three main categories, i.e. subordinate, attributive and coordinate compounding. These categories represent the grammatical relation between the components of a compound. In other words, components of subordinate compounds have the grammatical relation of complementation, components of attributive compounds have the grammatical relation of attribution/modification, and components of coordinate compounds have the grammatical relation of coordination. Each category can be further classified as either

endocentric or exocentric compounds. We believe that such a manner of classification is sufficient to assist us in analysing the different types of Malay compounds in this study.

If we look back at the universal outline of compoundhood sketched in Chapter 2 (cf. section 2.3), we can thus aptly match Malay compoundhood to the outline as below.

(46) Outline of Malay compoundhood

```
                    ┌─────────────────────┐
                    │   Malay Complex     │
                    │   Word Formation    │
                    │ (Four-way distinction) │
                    │      (cf. 5.1)      │
                    └─────────┬───────────┘
                              │
                    ┌─────────▼───────────┐
                    │  Malay Compounding  │
                    └─────────────────────┘
```

Definition	**Headedness**	**Criteria**	**Classification**
- *components of compounding*	- *categorical*	- *phonological*	- *subordinate*
roots, bases, morphemes and words, with the prototypical structure of [X R Y] (X)(Y)(Z)	- *semantic*	- *semantical*	- *attributive*
		- *syntactical*	- *coordinate*
(cf. 5.2.1)	(cf. 5.2.2)	(cf. 5.2.3)	(cf. 5.2.4)

```
                    ┌─────────────────────┐
                    │  Malay Compoundhood │
                    └─────────────────────┘
```

This outline can be translated into a checklist which we will use as a guide for the subsequent analysis of Malay compounding. We end this section with the checklist of Malay compoundhood as below.

(47) Malay compoundhood checklist

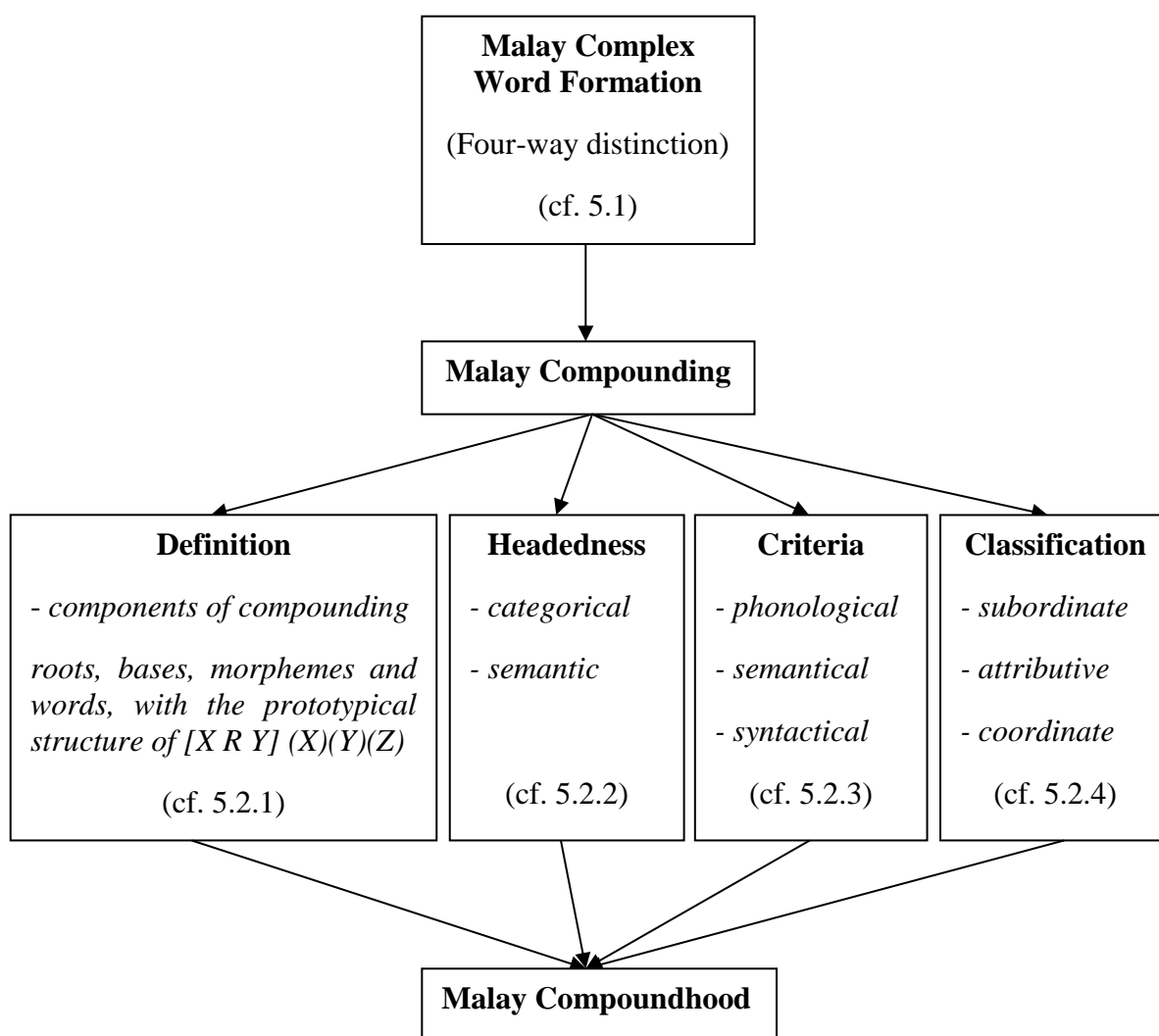|	Malay compoundhood	Comments
1.	Definition - components of compounding: - roots, bases, morphemes and words	Prototypical compounding schema for Malay: [X *R* Y] (X)(Y)(Z). Where, X and Y are the components, *R* represents the relationship, and (X), (Y) and (Z) represents the overall head/category.

2.	Headedness
	
	- Categorical head
	
	    (The head is the part that determines
	    the whole category of the compound)
	
	- Semantic head
	
	    (The head is the hyponym of the compound)
		Identify the head of a given compound, distinguish between endocentric and exocentric compounds.

3.	Criteria
	
	- Syntactic criteria
	
	    - insertion/modification
	
	    - inability to switch constituents
	
	    - circumfixation/reduplication
	
	- Semantic criteria (when applicable)
	
	    - lexicalisation
	
	- Phonological criteria (when applicable)
	
	    - stress and/or assimilation
		Use these tests to identify whether a given concatenated structure can be consider as a compounded structure.

4.	Classification - Subordinative (SUB) - Attributive (ATR) - Coordinative (CRD)	Classify compounds according to their types in accordance with Bisetto and Scalise's (2005) compound classification.

5.3 Analysis of Malay compounds

The aim of this section is to analyse the combined structures exemplified as compounds in Hassan (1974), Musa (1993), Karim (1995), Karim et al. (2008) and Sew (2009) (cf. appendices (1) to (5)).[99] The corpus consists of approximately 250 combined structures (along with appropriate supplements from our own native knowledge and general online resources). It is not the aim of this study to provide an exhaustive list of Malay compounds. The main concern is to clarify and comprehend the compounding phenomenon of the language (through the framework of this study). Thus, even though the number of combined structures being analysed is arguably small, we nevertheless consider this amount of data as sufficient in assisting us to achieve the intention of this study.

Each combined structure is run through the compoundhood checklist (cf. section 5.2.5, (47)) to analyse its compound status. Following the checklist, the first step establishes whether a given structure adheres to the prototypical schema of Malay compounding (i.e. the [X *R* Y]

---

[99] Note that the use of some combined structures (as examples of compounds) does overlap amongst the linguists in the literature. In other words, several structures are listed more than once in the appendix simply because they are coincidently the exact structures used as examples by different linguists. For instance, *surat khabar* 'newspaper' is exemplified as a compound by both Hassan (1974) and Karim et al. (2008). This study, however, is not particularly concerned with the number of combinations being declared as compounds by the many linguists. The concern is whether a given combined structure (which is declared as a compound) can actually be considered as a compound based on the arguments of this study.

(Z) schema). The next step clarifies the headedness of the structure by determining its categorical and semantic head. This will also help to establish whether the structure is an endocentric or exocentric compound. The next step tests the structure against the criteria of Malay compoundhood, i.e. the syntactic, phonological and semantic criteria. Finally, the structure can be classified in terms of the relationship held between its components, i.e. a subordinative, attributive or coordinative relation. Structures that are accepted as compounds are then grouped together according to their respective categories under the umbrella of nominal compounds, followed by verbal compounds and finally adjectival compounds.

5.3.1 Nominal compounds

To a large extent, nominal compounds are the most frequent type of compounds observable from the corpus. Different types of nominal compounds with a noun as their heads, namely the Noun-Noun (NN), Noun-Verb (NV) and Noun-Adjective (NA) compounds, are evident from the analysis. Furthermore, there are also examples of nominal compounds that either do not have nouns as their heads or where the noun component is not in the regular position within the compound. Accordingly, the analysis below is presented in the order of NN nominal compounds, NV nominal compounds, NA nominal compounds and other nominal compounds.

5.3.1.1 The NN nominal compounds

The Noun-Noun combination is the most common structure of all within the nominal compound category. Following the subordinate, attributive and coordinate classification, we can classify the Noun-Noun nominal compounds as follows.

 (i) Attributive (ATR)

Below is the list of combinations that have been analysed from the corpus as Noun-Noun (NN) attributive (ATR) endocentric (endo) nominal compounds.

(48) NN (ATR) (endo) nominal compounds

	Compounds	Gloss	Meaning
1	*jam tangan*	(watch + hand)	wristwatch
2	*bom tangan*	(bomb + hand)	grenade
3	*kanta tangan*	(lens + hand)	magnifier
4	*kereta api*	(car + fire)	train
5	*air mata*	(water + eye)	tears
6	*air hujan*	(water + rain)	rain (water)
7	*air tanih*	(water + soil)	soil water
8	*surat khabar*	(letter + news)	newspaper
9	*kelab malam*	(club + night)	night club
10	*nasi minyak*	(rice + oil)	oily rice
11	*batu kapur*	(stone + lime)	lime stone
12	*baja butir*	(fertiliser + granule)	granulated fertiliser
13	*abu rokok*	(ash + cigarette)	cigarette ash
14	*kertas kerja*	(paper + work)	paperwork
15	*gambar rajah*	(picture + diagram)	diagrammatic picture
16	*pita suara*	(tape + voice)	vocal cords
17	*deria rasa*	(sense + taste)	sense of taste
18	*had laju*	(limit + speed)	speed limit

19	*roda tenaga*	(wheel + energy)	flywheel
20	*garis pusat*	(line + centre)	diameter
21	*garis arus*	(line + current)	streamline
22	*kaji bumi*	(study + earth)	geology
23	*aneka warna*	(variety + colour)	colour assortments
24	*rukun tetangga*	(principle + neighbour)	neighbourhood committee
25	*lembu daging*	(cow + meat)	cow for producing meat
26	*ayam telur*	(chicken + egg)	chicken for producing eggs

Elaboration on how a given structure in the list above is analysed in order to be justified as an NN (ATR) (endo) nominal compound is exemplified below with the structure of *jam tangan* (watch + hand) 'wristwatch'.

(49) Example analysis of *jam tangan*

Structure: *jam tangan*	
1.	Components and schema [*jam* $_{N1}$ *tangan* $_{N2}$] (N_1)
2.	Headedness i) Categorical head: *jam tangan* is a noun similar to its head *jam* (N_1) ii) Semantic head: *jam tangan* is a type/kind of its head *jam* (N_1) iii) The structure has an endocentric noun head *jam* (N_1)
3.	Criteria i) Syntactic criteria:

	- Component modification: e.g. adding *lelaki* (man) = [*jam tangan*] [*lelaki*] 'men's wristwatch'. - Insertion between components: e.g. inserting *baru* (new) produces *[*jam baru*] [*tangan*] *'new watch (for) hand', or *[*jam*] [*baru tangan*] *'watch, then hand'. - Component switching: e.g. *jam tangan* 'wristwatch' to *tangan jam* 'hand (of) watch'. - Reduplication: e.g. *jam-jam tangan* 'wristwatches'. - Circumfixation: e.g. **berjamtangankan* *'to make something as a wristwatch'. ii) Phonological criteria: stress and/or assimilation are not applicable for this structure. iii) Semantic criteria: lexicalisation is not truly applicable for this structure.
4.	Classification - Attributive relation, i.e. the non-head component *tangan* (hand) modifies the head component *jam* (watch).

The table above demonstrates how the structure of *jam tangan* (watch + hand) is run through the proposed Malay compoundhood checklist in order to be justified as a compounded form.

To begin with, the structure of *jam tangan* clearly adheres to the prototypical compounding schema of [X *R* Y] (X), i.e. [*jam* $_{N1}$ *tangan* $_{N2}$] (N_1). The structure has an endocentric head, i.e. the left-hand component of the structure *jam* (N_1) is the head noun. This is justified by clarifying the semantic and categorical heads. The noun *jam* (watch) is the semantic head as it is the hyponym of the structure *jam tangan* (watch hand) 'wristwatch'; i.e. *jam tangan* is a kind of *jam* (watch), not a kind of *tangan* (hand). Accordingly, the noun *jam* (watch) is also the categorical head as it determines the category of the whole structure, i.e. *jam tangan* 'wristwatch' is a noun similar to the category of its noun head *jam* 'watch'. Once the headedness is established, the structure can be further analysed against the criteria of

compoundhood, beginning with syntactic criteria which are then followed by phonological and semantic criteria.

With regard to syntactic criteria, one of the concerns is the issue of modification. The structure of *jam tangan* noticeably disallows modification of its components. For instance, adding *lelaki* (man) to the structure, as in *jam tangan lelaki* (watch + hand + man), produces [*jam tangan*] [*lelaki*] 'men's wristwatch', and not *[*jam*] [*tangan lelaki*] *'watch (for) men's hands'. In other words, modification affects the structure as a whole, not just one of its components. This observation can be taken as an argument that the combination of *jam* + *tangan* is fused as a single compounded unit.

Another concern of the syntactic criteria is the issue of component separation. The components of *jam tangan* cannot be separated, i.e. any kind of insertion between the combination will interrupt the intended meaning of the structure. For instance, insertion of *baru* (new) in **jam baru tangan* ungrammatically produces either *[*jam*] [*baru tangan*] *'watch, new hand', or *[*jam baru*] [*tangan*] *'new watch, hand'. Again, this can be taken as an argument that the components are in a union (hence cannot be separated via insertion of other elements in between them).[100]

Yet another concern of the syntactic criteria is to test whether a given structure retains or changes its meaning upon component switching. In this case, switching the components of *jam tangan* to *tangan jam* will bring about a different meaning; i.e. *tangan jam* (hand watch)

---

[100] One can argue that it would be better to insert *untuk* (for) between the components (instead of *baru* (new)), results in *jam untuk tangan* 'watch for hands'. This, in principle, will show that the intended meaning of the original combination can be maintained even after insertion has taken place (hence proving that the components are not in total unification). However, such a claim can be misleading. If we look at it carefully, the argument is redundantly pointing out the existence of the unexpressed *R*-relation between the components of a given compound. This opaque relation between the components is exactly the peculiarity of compounding which distinguishes it from other morphological (and phrasal) constructions. If the R-relation of a given compounded form is purposely made clear, and then put forward as proof of non-compoundhood, all compositional compounds will fail to exist (since compounds inherently have this unexpressed *R*-relation).

metaphorically refers to 'the hands (the minute and hour pointers) of a watch'. Switching of components clearly interrupts the intended meaning. Once again the structure *jam tangan* can be perceived as functioning as a single compounded unit based on this point.

The final two concerns of syntactic criteria are the issues of circumfixation and reduplication. With regards to circumfixation, recall how it has been argued that the components of a compound should merge when circumfixed (e.g. *uji kaji* 'experiment' circumfixes into *mengujikajikan* 'to experiment', *pengujikajian* 'the process of experimenting', etc.). This, however, is not the case for *jam tangan* as it cannot be grammatically circumfixed, e.g. **menjamtangankan* *'to make (something) into a wristwatch', **berjamtangankan* *'to wear (something) as a wristwatch', etc. With regard to reduplication, again recall how it has been argued that a compound should undergo full reduplication with both components merging (e.g. *uji kaji* 'experiment' reduplicates into *ujikaji-ujikaji* 'experiments'). However, in the case of *jam tangan*, only the head of the structure is reduplicated, i.e. *jam-jam tangan* 'wristwatches' (and not **jam tangan - jam tangan*).

Having analysed *jam tangan* through the syntactic criteria, the analysis is then extended to the phonological and semantic criteria respectively. With regard to phonological criteria, the components of *jam tangan* do not show any signs of assimilation, i.e. *jam* (/ʤam/) + *tangan* (/taŋan/) combines as *jam tangan* (/ʤam taŋan/).[101] The structure is also not involved in any particular stress patterning. For instance, placing stress on the head component (/ˆʤam taŋan/) or on the non-head component (/ʤam ˋtaŋan/) will not differentiate it from being a compound (as opposed to a phrase).

---

[101] Obviously, the final bilabial nasal /m/ sound (in /ʤam/) is phonetically distinct from the initial alveolar plosive /t/ sound (in /taŋan/) for the two sounds to naturally assimilate.

On one hand, *jam tangan* can be considered as a lexicalised structure in the sense that its components have shown (to a certain degree) the possibility of being in a unified state (as evidenced from the syntactic test above). This can be taken as an argument that the compound might be a fixed structure lexicalised as a whole in the lexicon. However, *jam tangan* can also be considered as a compositional structure in the sense that the overall meaning is derivable from its components, i.e. *jam tangan* (watch + hand) is 'a watch for hands'. By analogy, there are many other combinations involving the head *jam* (watch/clock) with similar compositional property, e.g. *jam saku* (watch + pocket) 'pocket watch', *jam dinding* (clock + wall) 'wall clock', etc. This sort of analogous observation illustrates that the combination of *jam* and *tangan* is not necessarily a distinctive pairing (to the point that it needs to be lexicalised as a whole in the lexicon). Thus, in terms of semantic criteria, the idea of *jam tangan* being a truly lexicalised compound is questionable.[102]

The final step of the analysis it to establish the relation between the components of a given structure. In this case, the components of *jam + tangan* can be argued as holding a head + modifier relationship. In other words, the components are in an attributive relation, where the non-head component *tangan* (hand) modifies the head component *jam* (watch) to further express the meaning of 'a watch for hands' (i.e. wristwatch).[103] Taking all of the main points into account (i.e. the analysis of headedness, criteria and classification), we can classify the

---

[102] To this end, one can argue that the structure of *jam tangan* does not truly fulfil certain criteria as proof of compoundhood; namely, the syntactic criteria (reduplication and circumfixation), phonological criteria (assimilation and stress) and, to a certain degree, the semantic criteria as well. It will also be apparent as we go on that there are many more 'compounds' that do not fully comply to the proposed criteria and test of compoundhood. The question is then, does this mean that *jam tangan* and the other 'compounds' yet to be seen are perhaps not compounds after all? We will leave this issue for now and return to it in the concluding remarks.

[103] This hidden meaning between the two elements (i.e. the *R*-relation) needs to be interpreted through knowledge of the language (lack of which can distort the intended meaning of the compound). For instance, *jam tangan* (watch + hand): here is a kind of watch to be used as an accessory for hands. In contrast, a compound such as *bom tangan* (bomb + hand) 'hand grenade' is not a type of bomb for the hands, but a type of bomb to be launched using the hands. This subtle yet crucial understanding is important in order to derive the correct meaning of a given compound.

structure of *jam tangan* as an example of an NN (ATR) (endo) nominal compound in the language.

As illustrated with the structure of *jam tangan* above, the same manner of analysis is run through for the other combinations listed in (48). In general, they all adhere to the Malay compoundhood checklist as follows. They have the compound structure of [N_1 N_2] (N_1). The (N_1), which represents the endocentric nominal head of these combinations, is justified by the fact that: (i) they are the components that determine the overall category of the compound (categorical head), and (ii) they are the superordinate of the whole compound (semantic head). In this sense, the compounds in (48) are truly endocentric as they all have both endocentric categorical and semantic heads.

In general, the combinations disallow modification of their components. For instance, *kereta api merah* (car + fire + red) results in [*kereta api*] [*merah*] 'red train', and not *[*kereta*] [*api merah*] *'red-fire car'; or *surat khabar lama* (letter + news + old) results in [*surat khabar*] [*lama*] 'old newspaper', and not *[*surat*] [*khabar lama*] *'old-news paper', etc. The combinations also disallow insertion between the components. For instance, insertion of *baru* (new) in *gambar baru rajah* ungrammatically produces either *[*gambar*] [*baru rajah*] *'diagram-new, picture', or *[*gambar baru*] [*rajah*] *'diagram, new-picture', etc. Component switching will also change the intended meaning of the combinations. For instance, switching from *nasi minyak* 'oily rice' to *minyak nasi* produces the meaning of *'oil from rice'; or switching from *roda tenaga* 'flywheel' to *tenaga roda* produces *'power from wheel', etc. With regard to circumfixation, most of the combinations cannot be grammatically circumfixed.[104] Concerning reduplication, some combinations do not necessarily need to be reduplicated (as they are uncountable nouns and/or intrinsically plural). For the ones that can

---

[104] Except for *air mata*, *surat khabar* and *garis pusat*, which can be circumfixed.

be reduplicated, it is only the head component that undergoes reduplication, e.g. *surat-surat khabar* 'newspapers', *kelab-kelab malam* 'night clubs', etc.

With regard to phonological criteria, none of the combinations show any signs of assimilation.[105] Even the potential compound *gambar rajah* (/ɡambar raʤah/) does not assimilate to */ɡambaraʤah/. Similarly, none of the combinations show any evidence of stress patterning. Placing stress on the first and/or the second component of the compounds will not help to establish compoundhood. With regard to semantic criteria, there is no concrete argument to defend the compounds as true lexicalised structures. This is mainly due to the fact that the components and their combinations are transparent. In other words, the meaning of a compound can be worked out from the meaning of its components. Therefore, it is not necessary for the compounds to be memorised as a whole in the lexicon.

Finally, the relationship between the non-head and the head components of these combinations can be interpreted as holding a head + modifier relationship, i.e. attributive compounds. Overall, the structures in (48) above can be classified as examples of NN (ATR) (endo) nominal compounds in the language.

(ii) Coordinative (CRD)

Below is the list of combinations that have been analysed from the corpus as Noun-Noun (NN) coordinative (CRD) endocentric (endo) nominal compounds.

---

[105] Except for *kereta api*, which has been described before as showing evidence of assimilation from /kəreta api/ to /kəretapi/.

(50) NN (CRD) (endo) nominal compounds

	Compounds	Gloss	Meaning
1	*ibu bapa*	(mother + father)	parents
2	*periuk belanga*	(pan + pot)	pots and pans
3	*bala tentera*	(troop + soldier)	Army
4	*hutan rimba*	(wood + forest)	large/dense jungle
5	*jiran tetangga*	(neighbour + neighbour)	neighbours
6	*tanah air*	(land + water)	homeland
7	*rumah tangga*	(house + stairs)	household

Elaboration on how a given structure above is analysed in order to be justified as an NN (CDR) (endo) nominal compound is exemplified below with the structure of *ibu bapa* (mother + father) 'parents'.

(51) Example analysis of *ibu bapa*

Structure: *ibu bapa*	
1.	Components and schema [*ibu* $_{N1}$ *bapa* $_{N2}$] (N).
2.	Headedness i) Categorical head: the category of *ibu bapa* is a noun (N), but it is not necessarily projected from only one of its components (N$_1$) or (N$_2$). ii) Semantic head: the meaning of *ibu bapa* (N) derives equally from the meaning of its components (N$_1$) and (N$_2$). iii) The combination is an endocentric compound.

3.	Criteria
	i) Syntactic criteria:
	- Component modification: e.g. adding *saya* (I/me) = [*ibu bapa*] [*saya*] 'my parents'.
	- Insertion between components: e.g. inserting *saya* (I/me) produces *[*ibu saya*] [*bapa*] *'my mother (is a) father', or *[*ibu*] [*saya bapa*] *'mother, I am father'.
	- Component switching: e.g. *ibu bapa* 'parents' to *bapa ibu* 'mother's father'.
	- Reduplication: e.g. *ibu-ibu bapa, or **ibu bapa - ibu bapa* *'parents'
	- Circumfixation: e.g. *keibubapaan* 'parenthood'
	ii) Phonological criteria: stress is applicable, but assimilation is not applicable for this structure.
	iii) Semantic criteria: the structure is arguably lexicalised.
4.	Classification
	- Coordinative relation, i.e. both components are equally important to produce the intended meaning.

The table above demonstrates how the structure of *ibu bapa* (mother + father) is run through the proposed Malay compoundhood checklist in order to be justified as a compound. To begin with, the structure of *bapa ibu* adheres to the compounding schema of [*ibu* $_{N1}$ *bapa* $_{N2}$] (N). The nominal lexical category of the whole compound (N) is not necessarily derived specifically from either (N$_1$) or (N$_2$). This is because both (N$_1$) and (N$_2$) are considered as having equal headedness prominence within the combination. Similarly, it is unnecessary to determine whether (N$_1$) or (N$_2$) is the semantic head of the compound because both components are equally important in determining the meaning of the whole compound. In this sense, *ibu*

bapa can be considered as having two endocentric categorical heads and two endocentric semantic heads.

The structure disallows modification of its components. For instance, adding *saya* (I/me) to the combination produces [*ibu bapa*] [*saya*] 'my parents', and not *[*ibu*] [*bapa saya*] *'mother, my father'. Similarly, insertion of *saya* (I/me) disrupts the meaning of the compound, produces *[*ibu saya*] [*bapa*] *'my mother (is a) father', or *[*ibu*] [*saya bapa*] *'mother, I am father'. Switching components will also change the meaning of the compound, i.e. from *ibu bapa* 'parents' to *bapa ibu* 'mother's father'. The combination allows for circumfixation (e.g. *keibubapaan* 'parenthood'), but reduplication is unnecessary as the compound can be inherently plural within the context (i.e. *ibu bapa* 'parents', not *ibu-ibu bapa*, or *ibu bapa - ibu bapa*). As a whole, the observations can be taken as arguments for the unity between the components of *ibu* and *bapa* combination.

In terms of phonological criteria, there is no evidence of assimilation between the components of the compound. However, stress patterning can have an effect on the compound. For instance, placing the stress on the first component, as in /ˈibu bapa/, can actually convey the meaning of 'father's mother', in contrast to having a neutral stress to convey the meaning of 'parents'.

With regard to the issue of lexicalisation, the combination of *ibu bapa* arguably has more weight as a lexicalised form than it does as a compositional structure. On one hand, it is true that the meaning of the compound is derivable from its components (hence it can be considered as a compositional structure). On the other hand, the combination can be considered as a lexicalised form on account of its components showing unification as a single unit (as evident from the syntactic tests above). Take for instance the observation on the test of component switching. Given that the components are deemed as having equal status (i.e.

coordination), switching from *ibu bapa* (mother + father) to *bapa ibu* (father + mother) should therefore maintain the same meaning of 'parents'. This, however, is not the case. Instead, *bapa ibu* (father + mother) immediately means 'mother's father', without even having to place stress on the first component, /ˈbapa ibu/ (like that of placing stress on the first component in /ˈibu bapa/ to force out the meaning of 'father's mother'). This shows that there must be a special bond to the sequence of *ibu + bapa*, to specifically produce the meaning of 'parents'. We consider the compound *ibu bapa* as having a stronger argument for lexicalisation than for compositionality. As mentioned earlier, the components of *ibu* and *bapa* can be said to be holding a coordinative relationship, i.e. both components are equally responsible in expressing the intended meaning of the compound. All in all, we can classify the structure of *ibu bapa* as an example of an NN (CRD) (endo) nominal compound in the language.

In general, the rest of the combinations listed in (50) adhere to the checklist of compoundhood in more or less the same way as illustrated by *ibu bapa* above.[106] Although the existence of such coordinated combinations is well documented, Malay scholars do not recognise them as compounds in the language (Musa, 1993; Hassan, 2006; Karim et al., 2008; among others). The main argument for rejecting them as compounds is based on the idea that it is possible to insert the conjunction 'and' within the components (e.g. *ibu dan bapa*) and still maintain roughly the same meaning of the original combination (e.g. mother and father, i.e. parents). Once again, we reject this misleading idea of pointing out the inherent existence of the unexpressed grammatical relation between the components of a given compound (in this case 'and'), then claiming it as an evidence of non-compoundhood.

---

[106] Exception is noted for *hutan rimba, bala tentera, periuk belanga, rumah tangga and jiran tetangga*, none of which can undergo circumfixation. Also, stress placement is very much an exclusive event only for the structure of *ibu bapa*.

Karim et al. (2008) specifies these types of combinations as synonymous and antonymous (head + head) phrasal constructions.[107] According to Karim et al. (2008), in synonymous head + head combinations, the elements involved are a sort of repetition which brings out the same meaning (sense) as the whole of the given construction. On the other hand, antonymous head + head constructions bring out a single meaning from the contradicting meanings of the elements. Some examples of synonymous combinations from Karim et al. (2008) include *ribut taufan* (storm + hurricane) 'super storm/hurricane', *hutan rimba* (forest + jungle) 'large/dense jungle', and *pinggan magkuk* (plate + bowl) 'dishes/tableware'. Antonymous combinations are provided in examples such as *ibu bapa* (mother + father) 'parents', *kerusi meja* (chair + table) 'furniture', and *abang adik* (brother + younger brother/sister) 'siblings'.[108]

Karim et al. (2008) also mentioned two points about the structures, namely: (a) the heads are not modifiers to one another, and (b) the meaning as a whole is supported by both heads. In other words, neither component is superior to the other. Both components are equally important in conveying the intended meaning of the combination. Such characteristics strongly echo the conditions of coordinative compounding. Furthermore, we also observe that, not only are the combinations coordinated and produces new extended meanings, they are also exclusive in the sense that one cannot simply match two nouns to derive similar analogies. For instance, if *pinggan mangkuk* (plate + bowl) is accepted to mean 'dishes/tableware', coming up with an analogy such as **cawan gelas* (cup + glass) does not create the meaning of **'glassware'.

---

[107] Cf. Karim et al. (2008: 375) on the subtopic of *Frasa Nama: Inti + Inti*.

[108] Karim et al. (2008) only provided the examples of such combinations (e.g. *kerusi meja*); they did not specifically provide the meanings for them. The overall meaning of 'furniture' for the combination of *kerusi meja*, for instance, is provided by us based on our knowledge as a native speaker of the language. The overall meanings of the combinations are not something that we simply created out of intuition. They are indeed the actual implied meanings when native speakers use such constructions in the language. In general, some constructions are used more colloquially (e.g. *kerusi meja*), while others are regularly used in the standard language (e.g. *ibu bapa*).

Similarly, creating the combination of *cadar selimut* (bed sheet + blanket) will not necessarily bring about the meaning of 'bedding accessories' (as an analogy to the regular combination of *kerusi meja* (chair + table) 'furniture'). This shows that the combinations are not merely compositional constructions (like phrases), but more likely to be lexicalised constructions (in this case lexicalised compounds).

Taking into account the arguments and discussion so far, we strongly postulate that Malay actually does have coordinative compounding. To this end, the list of combinations in (50), along with the examples given by Karim et al. (2008), can be classified as examples of NN (CRD) (endo) nominal compounds in the language.

(iii) Idiomatic

There are several examples of idiomatic Noun-Noun (NN) attributive (ATR) endocentric (endo) nominal compounds.

(52) NN (ATR) (endo) partially idiomatic nominal compounds

	Compounds	Gloss	Meaning
1	*duit kopi*	(money + coffee)	Tips
2	*rabun ayam*	(short/long sighted + chicken)	short/long-sighted during dusk
3	*anak emas*	(child + gold)	favourite person

The compounds have the structure of $[N_1\ N_2]\ (N_1)$. Their semantic head is the (N1), e.g. *duit kopi* is a type of *duit* 'money', *rabun ayam* is a kind of *rabun* 'short/long sighted', and *anak emas* is a type of *anak* 'child/person'. Accordingly, (N_1) also projects its nominal category

onto the whole compound (i.e. as the categorical head). In this sense, the compounds are endocentric compounds.

Overall, the compounds are sensitive to modification, insertion and component switching. For instance, adding *saya* 'I/me' to *duit kopi* produces [*duit kopi*] [*saya*] 'my tips', not *[*duit*] [*kopi saya*] *'my coffee, money'. Similarly, insertion shows disruption as in *[*rabun saya*] [*ayam*] *'chicken, my-short/long sighted' or *[*rabun*] [*saya ayam*] *'I am a chicken, short/long sighted'. Switching the components from *anak emas* to *emas anak* *'child's gold' also disrupts the meaning of the combination. The compounds are not subjected to reduplication, and only *anak emas* can be circumfixed (e.g. *meganakemaskan* 'to treat someone with favouritism'). None of the compounds show evidence of phonological criteria (i.e. stress and/or assimilation) to assist in their compoundhood identification.

In terms of semantic criteria, the compounds have a fairly strong argument for lexicalisation. This is mainly because only the head component (N$_1$) is transparent, while the modifying component (N$_2$) is opaque. The relation between the head (N$_1$) and the non-head (N$_2$) needs to be interpreted idiomatically/metaphorically. For instance, the meaning of *kopi* (coffee) in *duit kopi* needs to be interpreted as 'extra' money given as 'tips', and not as 'money to purchase coffee'. Similarly, the meaning of *ayam* (chicken) in *rabun ayam* will not bring out the figurative meaning of 'during dusk' if it was to be read produces literally. Thus, the compounds can be said to be partially idiomatic/metaphoric, which suggests that they still need to be memorised and stored as a whole in the lexicon, rather than being fully compositional constructions.

Nevertheless, the components of the compounds have an attributive head + modifier relationship, i.e. the non-head noun component (N$_2$) metaphorically modifies the nominal head component (N$_1$). Overall, the combinations in (52) above can be classified as examples of NN (ATR) (endo) partially idiomatic nominal compounds in the language.

There is another type of idiomatic NN compounding which is slightly different from the ones in (52) above. These compounds are different in the sense that they are fully idiomatic/metaphoric compounds. Listed below are some of the examples from the corpus.

(53) NN (exo) idiomatic nominal compounds

	Compounds	Gloss	Meaning
1	*tangkai jering*	(stalk + a type of bitter/smelly fruit)	miser
2	*buah hati*	(fruit + heart)	lover/sweetheart
3	*kaki botol*[109]	(leg/foot + bottle)	drunkard
4	*buku lima*[110]	(fist + five)	brass knuckles

The compounds have the exocentric structure of $[N_1 \ N_2]$ (N). Their semantic head is neither the projection of the (N_1) nor the (N_2) component, e.g. *tangkai jering* is neither a type of *tangkai* 'stalk' nor is it a type of *jering* 'a bitter/smelly fruit', *kaki botol* is neither a *kaki* 'leg/foot' nor is it a *botol* 'bottle', etc. Similarly, the overall nominal category (N) of the compounds cannot be truly attributed to either the (N_1) or the (N_2) components. For this reason, the compounds above are truly exocentric, in both a categorical and a semantic sense.

Similar to the compounds in (52), the compounds here also disallow component modification, separation and switching, all of which demonstrates the intactness of the components within the compounds. However, the compounds here make a stronger case for lexicalisation in comparison to the previous type of idiomatic compounds in (52). Recall that the compounds in (52) are only partially opaque, i.e. the non-head component (N_2) is opaque but the head component (N_1) is transparent (e.g. *duit kopi* is still a type of *duit* 'money'

---

[109] The word *kaki* can also mean 'someone with a particular habit', e.g. *kaki* + *judi* (gambling) 'excessive gambler', *kaki* + *perempuan* (women) 'womaniser', etc.
[110] The compound *buku lima* can also simply mean 'hand fist'.

nonetheless). Conversely, the meaning of a compound here in (53) cannot be derived from any of the meanings of its components (e.g. the combination of *buah* 'fruit' + *hati* 'heart' is totally uncorrelated with the meaning of 'lover/sweetheart'). In this sense, the compounds are 'truly' idiomatic, hence necessarily need to be stored as a whole in the lexicon (i.e. lexicalised).

With regard to the relationship between the components, it is difficult to tell the actual relation (either as subordinate, attributive or coordinative) that occurs between them.[111] This is because the compounds lack a concrete head to enable us to make a clear connection between the components. For this reason, the combinations in (53) above are classified only as examples of NN (exo) idiomatic nominal compounds in the language.

5.3.1.2 The NV nominal compounds

In comparison to the Noun-Noun combinations, there are only a few examples of Noun-Verb combination from the corpus that can be analysed as nominal compounds.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Noun-Verb (NV) attributive (ATR) endocentric (endo) nominal compounds.

---

[111] Except perhaps for *kaki botol* 'drunkard' where *botol* can be argued as modifying the component *kaki*, since *kaki* also means 'someone with a particular habit', and the word *botol* here is metaphorically referring to 'alcohol bottle'. Thus, the components of *kaki botol* can be said to be in an attributive relation with the compound meaning of 'alcoholic person'.

(54) NV (ATR) (endo) nominal compounds

	Compounds	Gloss	Meaning
	Compounds	Gloss	Meaning
1	*padang tembak*	(field + shoot)	shooting range
2	*kapal terbang*	(ship + fly)	aeroplane
3	*meja makan*	(table + eat)	dining table
4	*meja tulis*	(table + write)	Desk
5	*alat tulis*	(tool + write)	stationery
6	*pisau cukur*	(knife/blade + shave)	Razor
7	*nasi goreng*	(rice + fry)	fried rice
8	*tukang masak*	(artisan + cook)	chef/cook
9	*daya cipta*	(ability + create)	Creativity
10	*hak cipta*	(rights + create)	copyright
11	*takat beku*	(point + freeze)	freezing point
12	*daya serap*	(strength + absorb)	absorbance

Elaboration on how a given structure in the list above is analysed in order to be justified as an NV (ATR) (endo) nominal compound is exemplified below with the structure of *padang tembak* (field/range + shoot) 'shooting range'.

(55) Example analysis of *padang tembak*

	Structure: *padang tembak*
1.	Components and schema [*padang* $_{N1}$ *tembak* $_{V1}$] (N_1)
2.	Headedness i) Categorical head: *padang tembak* is a noun similar to its head *padang* (N_1) ii) Semantic head: *padang tembak* is a type/kind of its head *padang* (N_1) iii) The structure has an endocentric noun head *padang* (N1)
3.	Criteria i) Syntactic criteria: - Component modification: e.g. adding *baru* (new) = [*padang tembak*] [*baru*] 'new shooting range'. - Insertion between components: e.g. inserting *baru* (new) produces *[*padang baru*] [*tembak*] *'new field, shoot', or *[*padang*] [*baru tembak*] *'field, new shoot'. - Component switching: e.g. *padang tembak* 'shooting range' to *tembak padang* '(to) shoot a field/range'. - Reduplication: e.g. *padang-padang tembak* 'shooting ranges'. - Circumfixation: e.g. **memadangtembakan* *'to make (a space) into a shooting range'. ii) Phonological criteria: stress and/or assimilation are not applicable for this structure. iii) Semantic criteria: lexicalisation is not truly applicable for this structure.
4.	Classification - Attributive relation, i.e. the non-head component *tembak* (shoot) modifies the head component *padang* (field/range).

The table above demonstrates how the structure of *padang tembak* (field/range + shoot) is run through the proposed Malay compoundhood checklist in order to be justified as a compounded form.

The structure of *padang tembak* adheres to the prototypical compounding schema of [X *R* Y] (X), i.e. [*padang* $_{N1}$ *tembak* $_{V1}$] (N_1). The noun *padang* (field/range) is the semantic head as it is the hyponym of the structure *padang tembak* (field/range shoot) 'shooting range'. In other words, *padang tembak* is a type of field/range (i.e. *padang*), and not a type of shooting activity (i.e. *tembak*). Accordingly, the head noun *padang* (field/range) also determines the category of the whole structure, i.e. *padang tembak* 'shooting range' is also a noun. Hence the noun *padang* (field/range) is also the categorical head of this structure. In this sense, the compound has both an endocentric categorical head and an endocentric semantic head.

The structure is examined through the criteria of compoundhood. With regard to the issue of modification, the structure of *padang tembak* disallows modification of its components. For instance, adding *baru* (new) to *padang tembak baru* (field + shoot + new) produces [*padang tembak*] [*baru*] 'new shooting range', and not *[*padang*] [*tembak baru*] *'field, new shoot'. The structure of *padang tembak* is also sensitive to the issue of component separation, i.e. insertion of other elements between the components will interrupt the intended meaning of the combination. For instance, insertion of *baru* (new) as in *padang baru tembak* ungrammatically produces either *[*padang*] [*baru tembak*] *'field/range, new shoot', or *[*padang baru*] [*tembak*] *'new field/range, shoot'. With regard to component switching, swapping from *padang tembak* to *tembak padang* (shoot + field) will bring about a different meaning, i.e. to literally 'shoot a field/range'. This clearly interrupts the intended meaning of the original combination. Overall, the test of modification, separation and switching argues for the structure of *padang* + *tembak* as a single compounded unit.

With regard to circumfixation, the structure *padang tembak* cannot be grammatically circumfixed, e.g. *memadangtembakan *'to make (an area) into a shooting range'. Concerning reduplication, only the head of the structure is reduplicated, i.e. *padang-padang tembak* 'shooting ranges' (and not **padang tembak - padang tembak*).

In terms of phonological criteria, the combination of *padang* (/padaŋ/) and *tembak* (/tembak/) combines as *padang tembak* (/padaŋ tembak/) with no occurrence of assimilation. Similarly, stress is not a strong argument to establish the structure as a compound. Placing stress on the head component (/ˋpadaŋ tembak/) or on the non-head component (/padaŋ ˋtembak/) does not assist in the identification of compoundhood.

In terms of semantic criteria, *padang tembak* on the one hand can be argued to be a lexicalised structure (as evidenced from the syntactic tests on the fusion of its components). However, the combination of *padang* (field/range) and *tembak* (shoot) is also compositional in the sense that the meaning of the compound is transparently derivable from its components (i.e. *padang tembak* is a 'range' for 'shooting' activities). Thus, it is arguable that the combination of *padang* and *tembak* is not necessarily an exclusive combination which needs to be lexicalised as a whole in the lexicon.[112]

The final part of the analysis concerns the existing relationship between the components of the compound. The components of *padang tembak* can be argued as having a head + modifier relationship. The non-head component *tembak* (shoot) modifies the head component *padang* (field/range), expressing the meaning of a field/range for shooting activities, i.e. a 'shooting range'. In other words, we can classify the components as having an attributive relation. As a

---

[112] There are many other combinations in the language involving the head *padang* (field/range) with comparable compositionality and transparency to *padang tembak*, e.g. *padang bola* (field + ball) 'football pitch', *padang sekolah* (field + school) 'school field', etc.; illustrating that *padang tembak* is not necessarily lexicalised.

whole, the structure of *padang tembak* is justifiable as a NV (ATR) (endo) nominal compound in the language.

In general, the combinations listed in (54) all adhere to the Malay compoundhood checklist as follows. They all have the compound structure of $[N_1 V_1] (N_1)$. The (N_1) represents the endocentric nominal head, which is the superordinate of the compound (semantic head), and also the overall category of the compound (categorical head).

With regard to the criteria of compoundhood, the combinations are consistent with the criteria of modification, insertion and component switching (i.e. the compounds adhere to these criteria as proof of unification between their components). However, the compounds cannot be grammatically circumfixed, and only the head components can be reduplicated.

In general, neither phonological nor semantic criteria are applicable to the compounds. There is no evidence of assimilation, even when there is a potential candidate for one (e.g. *alat tulis* 'stationery' is /alat/ + /tulis/ = /alat tulis/, and not */alatulis/). Similarly, there is no evidence of a significant stress pattern on any of the compounds (i.e. stressing either the first or the second components will not alter the meaning of the compounds). Furthermore, there is also no strong argument for the compounds to be considered as lexicalised structures. This is because they are transparent compositional combinations, i.e. their meaning can be derived from the meaning of their components. Thus, the necessity of the compounds being stored (i.e. lexicalised) as a whole in the lexicon is debatable.

Finally, the relationship between the head nouns and the non-head verbs of the compounds can be classified as having an attributive relationship, where the verb functions as a modifier of the head noun. Overall, the combinations in (54) above can be categorised as NV (ATR) (endo) nominal compounds in the language.

5.3.1.3 The NA nominal compounds

There are quite a number of Noun-Adjective combinations in the corpus that can be analysed as nominal compounds.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Noun-Adjective (NA) attributive (ATR) endocentric (endo) nominal compounds.

(56) NA (ATR) (endo) nominal compounds

	Compounds	Gloss	Meaning
1	*kerusi malas*	(chair + lazy)	lounge chair
2	*papan hitam*	(board + black)	blackboard
3	*rumah sakit*	(house + sick)	Hospital
4	*kuih kering*	(cake + dry)	Cookies
5	*guru besar*	(teacher + big/large)	head teacher
6	*menteri besar*	(minister + big/large)	chief minister
7	*kerani kanan*	(clerk + right)	senior clerk
8	*sekolah rendah*	(school + low)	primary school
9	*orang luar*	(person + outside)	outsider/alien
10	*jalan raya*	(road + great)	main road
11	*balai raya*	(hall + great)	community hall
12	*bandar raya*	(town + great)	City
13	*mogok umum*	(strike + general)	general strike

14	*hak milik*	(rights + own)	ownership
15	*segi tiga*	(angle + three)	Triangle
16	*angka ganda*	(figure + multiple)	multiplication figure
17	*titik buta*	(spot + blind)	blind spot
18	*isi padu*	(content + solid)	volume
19	*had kenyal*	(limit + elastic)	elastic limit
20	*kata dua*	(word + two)	ultimatum
21	*kata kunci*	(word + key)	Keyword
22	*darah panas*	(blood + hot)	Mammal

Elaboration on how a given structure in the list above is analysed in order to be justified as an NA (ATR) (endo) nominal compound is exemplified below with the structure of *kerusi malas* (chair + lazy) 'lounge chair'.

(57) Example analysis of *kerusi malas*

Structure: *kerusi malas*	
1.	Components and schema [*kerusi* $_{N1}$ *malas* $_{A1}$] (N_1).
2.	Headedness i) Categorical head: *kerusi malas* is a noun similar to its head *kerusi* (N_1) ii) Semantic head: *kerusi malas* is type/kind of its head *kerusi* (N_1) iii) The structure has an endocentric noun head *kerusi* (N_1)
3.	Criteria i) Syntactic criteria:

	- Component modification: e.g. adding *kuning* (yellow) = [*kerusi malas*] [*kuning*] 'yellow lounge chair'. - Insertion between components: e.g. inserting *baru* (new) produces *[*kerusi baru*] [*malas*] *'new chair, lazy', or *[*kerusi*] [*baru malas*] *'chair, then lazy'. - Component switching: e.g. *kerusi malas* 'lounge chair' to *malas kerusi* *'lazy (like) chair'. - Reduplication: e.g. *kerusi-kerusi malas* 'lounge chairs'. - Circumfixation: e.g. *mengerusimalaskan* *'to make (something) into a lounge chair'. ii) Phonological criteria: stress and/or assimilation are not applicable for this structure. iii) Semantic criteria: *kerusi malas* is a lexicalised structure.
4.	Classification - Attributive relation, i.e. the non-head component *malas* (lazy) modifies the head component *kerusi* (chair).

The table above demonstrates how the structure of *kerusi malas* (chair + lazy) is analysed in order to be justified as a compounded form.

The structure of *kerusi malas* adheres to the compounding schema of [*kerusi* $_{N1}$ *malas* $_{A1}$] (N_1). The noun *kerusi* (chair) is the semantic head, i.e. *kerusi malas* (chair + lazy) is a type of *kerusi* (chair). Accordingly, the noun *kerusi* (chair) is also the categorical head as evidenced by the lexical category of the whole structure (i.e. *kerusi malas* is a noun). The compound thus has both endocentric categorical and semantic heads.

The structure of *kerusi malas* disallows modification of its components, e.g. adding *kuning* (yellow) to *kerusi malas kuning* (chair + lazy + yellow) produces [*kerusi malas*] [*kuning*] 'yellow lounge chair', and not *[*kerusi*] [*malas kuning*] *'chair, yellow-laziness'. The

structure also disallows separation of its components, e.g. insertion of other elements between the components will interrupt the intended meaning of the combination. For instance, insertion of *baru* (new) as in **kerusi baru malas* ungrammatically produces either *[*kerusi*] [*baru malas*] *'chair - new lazy' or *[*kerusi baru*] [*malas*] *'new chair - lazy'. Furthermore, swapping the components from *kerusi malas* to *malas kerusi* will bring about an odd meaning, something along the line of '*lazy like a chair' (which is clearly not the intended meaning of the original combination). On the whole, the structure of *kerusi malas* adheres to the criteria of modification, separation and switching, arguing for the unification of its components as proof of compoundhood.

The structure of *kerusi malas* cannot be grammatically circumfixed (e.g. **mengerusimalaskan* *'to turn (something else) into a lounge chair'), nor can it be reduplicated as a whole (i.e. only the head can be reduplicated, as in *kerusi-kerusi malas* 'lounge chairs', not **kerusi malas - kerusi malas*). The proposed phonological criteria are not applicable to the structure either. The combination of *kerusi* (/kerusi/) + *malas* (/malas/) is without any signs of assimilation (i.e. /kerusi malas/, not */kerusimalas/), nor are there any signs of stress patterning (i.e. stressing the head component (/ˆkerusi malas/) has no significance over stressing the non-head component (/kerusi ˋmalas/)).

In terms of semantic criteria, *kerusi malas* can be argued as having a stronger claim for lexicalisation, and some supporting arguments for this have been hinted at earlier. For instance, the fact that the components of *kerusi* + *malas* show evidence of bound unification can be taken as a sign that the combination might be stored as a whole in the lexicon. Furthermore, the combination of *kerusi* + *malas* is arguably not as compositionally transparent as other compounded structures.[113] Although the combination has a literal noun head (i.e. *kerusi*

---

[113] Compare with the previous exemplification of *jam tangan* 'wristwatch' and *padang tembak* 'shooting range', where the components of the compounds are more transparent.

'chair'), the modifying non-head adjective (i.e. *malas* 'lazy') cannot be read literally. *Kerusi malas* (chair + lazy) does not mean 'a chair that is lazy' or 'a chair for a lazy person', but rather 'a reclinable chair used (mainly) for relaxing'. The actual meaning of the structure is unpredictable (to a certain degree) since it is not truly transparent through the literal interpretation of the components. We can thus assume that the compound needs to be stored as a whole in the lexicon in order to preserve its idiosyncratic meaning.

The final part of the analysis concerns the relationship between the components. The components of *kerusi* and *malas* have an attributive (head + modifier) relationship. In other words, the non-head component *malas* (lazy) further modifies the head component *kerusi* (chair) bringing about the meaning of 'a chair for relaxing/lazing', i.e. a 'lounge chair'. As a whole, the structure of *kerusi malas* can be categorised as a NA (ATR) (endo) nominal compound in the language.

In general, all of the combinations listed in (56) adhere to the Malay compoundhood checklist as follows. They all have the compound structure of $[N_1 A_1] (N_1)$. The (N_1) represents the endocentric nominal head of these compounds. (N_1) is the categorical head in the sense that it determines the overall category of the compound. (N_1) is also the endocentric semantic head in the sense that it is the hyponym of the compound.

All of the compounds above conform to the criteria of modification, insertion and component switching; illustrating some evidence for the unification of their components. However, most of them do not measure up to the criteria of circumfixation, reduplication, stress and assimilation. Exceptions are noted for the structures of: *guru besar* 'head teacher' (can be circumfixed to *bergurubesarkan* 'to have (someone) as a head teacher'), *menteri besar* 'chief minister' (can be circumfixed to *bermenteribesarkan* 'to have (someone) as a chief minister'), *kata kunci* (word + key) 'keyword' (can be circumfixed to *berkatakuncikan* 'to make

(something) as a keyword' and *bandar raya* (/bandar/ + /raja/) (can be assimilated to *bandaraya* (/bandaraja/), e.g. *Bandaraya Kuala Lumpur* - 'Kuala Lumpur City').

In terms of semantic criteria, some of the compounds in the list exhibit a stronger argument for lexicalisation than others. These compounds are: *kerusi malas* (chair + lazy) 'lounge chair', *papan hitam* (board + black) 'blackboard', *rumah sakit* (house + sick) 'hospital', *guru besar* (teacher + big) 'head teacher', *menteri besar* (minister + big) 'chief minister', *kerani kanan* (clerk + right) 'senior clerk', *orang luar* (person + outside) 'outsider/alien', *titik buta* (spot + blind) 'blind spot', and *darah panas* (blood + hot) 'mammal'. This is partly because their non-head adjectives are less transparent as modifiers. In other words, the meaning of the modification cannot be taken literally if the correct meaning of the compound is to be achieved. For instance, the meaning of *papan hitam* (board + black) is not 'a board which is black in colour', but 'a kind of board for writing or drawing on with chalk'. Similarly, *guru besar* (teacher + big) is not 'a teacher who is big in size', but instead 'a teacher who is the head of the other teachers'. These sorts of compounds can be assumed as being more lexicalised (i.e. stored as a whole in the lexicon) in order to preserve their idiosyncratic meaning.

In contrast, the other remaining compounds have a weaker argument for lexicalisation. These compounds are: *kuih kering* (cake + dry) 'cookies', *jalan raya* (road + great) 'main road', *balai raya* (hall + great) 'community hall', *bandar raya* (town + great) 'city', *mogok umum* (strike + general) 'general strike', *hak milik* (rights + own) 'ownership', *segi tiga* (angle + three) 'triangle', *angka ganda* (figure + multiple) 'multiplication figure', *titik buta* (spot + blind) 'blind spot', *isi padu* (content + solid) 'volume', *had kenyal* (limit + elastic) 'elastic limit', *kata dua* (word + two) 'ultimatum'. The non-head adjectives of these compounds are transparent enough to convey their intended modifying meaning. In other words, the meaning

of the whole compound can be derived from the components involved. These sorts of compounds can be assumed as being more compositional than lexicalised.

Finally, the relationship between the head nouns and the non-head adjectives of the compounds can be interpreted as having an attributive relationship, i.e. the non-head adjectives further modifies the head nouns. To this end, the combinations listed in (56) can be classified as examples of NA (ATR) (endo) nominal compounds in the language.

5.3.1.4 Other nominal compounds

From the corpus, there is also evidence of nominal compounds where the components basically do not follow the regular (Noun) head - (Noun, Verb, Adjective) non-head structure we have seen thus far. Below is the analysis of nominal compounds with the structures of Verb-Noun and Adjective-Noun, and the so-called established compounds respectively.

(1) *The VN nominal compounds*[114]

There are several Verb-Noun combinations in the corpus that can be considered as nominal compounds.

---

[114] There are two examples of nominal compounds consisting of the root verb *urus* (to manage) as the head, namely *urus niaga* (to manage + business) and *urus setia* (to manage + loyal). However, the compound *urus niaga* does not mean 'to manage a business', instead it means 'business matters/affairs'. Similarly, the compound *urus setia* does not mean 'to manage loyalty', instead it means 'loyal matters/affairs', i.e. 'secretariat' (the adjective *setia* 'loyal' here is opaque, hence *urus setia* is a partially idiomatic compound). In this sense, the root verb *urus* (to manage) is actually functioning as the derived noun *urusan* (matters/affairs) in both compounds. Therefore, we are not entirely sure whether to analyse them as a [V N/A] (N) construction or an [N N/A] (N) construction.

(i) Subordinate (SUB)

Below are the Verb-Noun (VN) subordinate (SUB) exocentric (exo) nominal compounds available from the corpus.

(58) VN (SUB) (exo) nominal compounds

	Compounds	Gloss	Meaning
1	*sapu tangan*	(to wipe + hand)	handkerchief
2	*cucuk sanggul*	(to pierce + hair bun)	hair pin
3	*jana kuasa*	(to generate + power)	generator
4	*reka bentuk*	(to create + shape)	design

The compounds have the basic structure of $[V_1 \ N_1]$ (N). They have mismatched semantic features where the overall nominal category (N) is not exactly inherited from the noun component (N_1). For instance, the meaning of the compound *sapu tangan* (to wipe + hand) 'handkerchief' is more relevant to the verb element *sapu* 'to wipe' as the compound denotes 'an instrument to wipe (something)' rather than 'a type of hand'. The semantics suggest that the head should be the verb (V_1), but the grammatical category of the compound is not verbal. Thus, the compounds have endocentric semantic heads, but exocentric categorical heads.

In general, the compounds disallow modification (e.g. *sapu tangan baru* produces [*sapu tangan*][*baru*] 'new handkerchief', not *[*sapu*][*tangan baru*] *'wipe new hands'), separation (e.g. **sapu baru tangan* ungrammatically *[*sapu baru*][*tangan*] *'new-wipe, hands', or *[*sapu*][*baru tangan*] *'wipe, new-hands') and switching (e.g. **tangan sapu* *'hand for wiping'). The compounds are reduplicated as a whole (e.g. *sapu tangan - sapu tangan*), while circumfixation is grammatically inapplicable. The phonological criteria of stress and

assimilation are not applicable to the compounds. Component stress placement is also not significant and thus cannot be used to differentiate these compounds from being phrasal structures. Similarly, there are no signs of assimilation between the components to further support compoundhood.

To a certain degree, the meanings of the components and their combination are transparent. Therefore, the idea of the compounds being lexicalised is debatable. In other words, the compounds can very well be compositional as it is possible to derive their meaning literally from the components. The components of (V_1) and (N_1) can be interpreted as having a subordinative relationship, i.e. the noun (N_1) as the complement of the verb (V_1). These compounds are perhaps the only type of true subordinate compounds in Malay. Thus, the combinations listed in (58) can be classified as examples of VN (SUB) (exo) nominal compounds in the language.

(ii) Idiomatic

There are also several combinations that can be analysed as Verb-Noun (VN) exocentric (exo) idiomatic nominal compounds.

(59) VN (exo) idiomatic nominal compounds

	Compounds	Gloss	Meaning
1	*hirup darah*	(to suck + blood)	extortionist
2	*makan angin*	(to eat + wind)	sightseeing
3	*gulung tikar*	(to roll + mat)	bankrupt

The compounds have the exocentric structure of [V_1 N_1] (N). Their semantic head is neither the projection of the (V_1) nor of the (N_2) components (e.g. *makan angin* 'sightseeing'

is neither the act of *makan* 'to eat' nor is it a type of *angin* 'wind'). Similarly, the overall nominal category of the compound (N) is not a projection of the inner (N$_1$) component. Thus, the compounds above have exocentric heads in both a semantic and a categorical sense.

In general, the compounds here disallow component modification, separation and switching. Reduplication and circumfixation are not applicable to these compounds. Also, there is no evidence of stress patterning and/or assimilation to further support compoundhood.

The meaning of the compounds cannot be derived from the meaning of their components as both components are opaque. For instance, *hirup darah* (to suck + blood) does not mean to literally suck blood, but instead a leechlike person who is metaphorically being compared to the act of sucking blood (i.e. an extortionist). Given that these idiomatic compounds have metaphorical meanings and that they cannot be derived from the meanings of the components, it is necessary for them to be stored in the lexicon as lexicalised forms.

To a certain degree, the components of the compounds can be argued as having a subordinate (SUB) relationship, i.e. the noun (N$_1$) being the complement of the verb (V$_1$). If this is true, the compounds here can be categorised similarly to the previous VN (SUB) (exo) nominal compounds in (58) (i.e. the *sapu tangan* 'handkerchief' types). However, the verbs (V$_1$) in (58) are (arguably) the endocentric semantic heads of the compounds.[115] Therefore, the verbs (V$_1$) and the nouns (N$_1$) there have a more apparent subordinate relationship. Conversely, given that the verbs (V$_1$) here in (59) are not the semantic heads of the compounds,[116] it can thus be misleading to claim that the components of (V$_1$) and (N$_1$) are in a subordinate (SUB)

---

[115] Recall that the compounds in (58) are arguably exocentric in terms of the categorical heads only. This is because the nominal category of the whole compound does not match with the category of its semantic head as these compounds have verbal endocentric semantic heads (V$_1$).

[116] As stated, the compounds in (59) have exocentric heads in both the semantic and the categorical sense. Lack of a concrete head makes it difficult to establish the relationship between the components of the compounds.

relationship. For this reason, the combinations in (59) above are classified only as examples of VN (exo) idiomatic nominal compounds in the language.

(2) *The AN nominal compounds*

There are several Adjective-Noun combinations in the corpus that can be considered as nominal compounds.

(i) Attributive (ATR)

Below are the Adjective-Noun (AN) attributive (ATR) endocentric (endo) nominal compounds available from the corpus.

(60) AN (ATR) (endo) nominal compounds

	Compounds	Gloss	Meaning
1	*perdana menteri*	(prime + minister)	prime minister
2	*timbalan menteri*	(deputy + minister)	deputy minister
3	*naib canselor*	(vice + chancellor)	vice-chancellor
4	*mahaguru*	*maha* (great) + *guru* (teacher)	professor
5	*tengah hari*	(middle + day)	midday
6	*ibu kota*	(main (mother) + city)	capital city
7	*setia kawan*	(loyal + friend)	solidarity
8	*purbakala*	(*purba* 'ancient' + *kala* period')	ancient times

The compounds have the structure of $[A_1 N_1] (N_1)$. Although (N_1) is not in the common left-hand position, it is nonetheless the head of these compounds. This is evidenced from (N_1)

being the semantic head of the compounds, e.g. *perdana menteri* is a kind of *menteri* 'minister', *setia kawan* is a state of *kawan* 'friendship', and *purbakala* is a frame of *kala* 'time'. Accordingly, (N_1) is also the categorical head of the compounds (i.e. the compounds are all nouns). In other words, the compounds all have endocentric right-hand nominal heads.

The compounds are sensitive to modification, insertion and component switching. For instance, adding *baru* 'new' to *perdana menteri* 'prime minister' produces [*perdana menteri*][*baru*] 'new prime minister', not *[*perdana*][*menteri baru*] *'prime, new minister'. Similarly, insertion shows disruption as in *[*perdana baru*] [*menteri*] *'new-prime, minister' or *[*perdana*] [*baru menteri*] *'prime, minister-new'. Switching the components from *perdana menteri* to *menteri perdana* will also disrupt the meaning of the combination.[117]

Concerning reduplication, some of the compounds above cannot be reduplicated while some can undergo the process. For instance, *setia kawan* 'solidarity' and *purbakala* 'ancient times' cannot be reduplicated as they are abstract nouns. In contrast, a compound such as *timbalan menteri* 'deputy minister' reduplicates as *timbalan-timbalan menteri* 'deputy ministers'.[118] In general, the compounds above can be grammatically circumfixed (e.g. *berperdanamenterikan* 'to have (someone) as the prime minister', *kesetiakawanan* 'to have the feeling of solidarity', etc.).[119]

There is no evidence of phonological criteria (i.e. stress and/or assimilation) to assist in compoundhood identification of the compounds above. In terms of semantic criteria, the

---

[117] Following the grammar of Malay, the correct structure for the compound prime minister should be *menteri perdana* (i.e. head + modifier), and not *perdana menteri* (i.e. modifier + head) as it is used in the language today. This is evident from other analogical combinations with the noun *menteri* 'minister' as the head, e.g. *menteri besar* 'chief minister', *menteri pendidikan* 'education minister', etc. Hassan (1974) mentioned that the reason why compounds such as *perdana menteri* are right-headed is because they were directly taken from other languages without modification to suit the general structure of the Malay language.

[118] The component that undergoes reduplication is the non-head left component. Such irregular exception only occurs to these kinds of right-headed compounds in the language.

[119] Except for *tengah hari* 'midday' and *purbakala* 'ancient times', which cannot undergo circumfixation.

compounds have a fair argument for lexicalisation. Although the meaning and the combination of the components are transparent, the fact that their heads are not in the regular left-hand position might suggest the need for them to be fully memorised rather than being compositional. The components can be argued as holding an attributive modifier + head relationship, i.e. the non-head adjectival component (A_1) modifies the nominal head component (N_1) in all of the compounds. Overall, the combinations in (60) above can be classified as examples of AN (ATR) (endo) nominal compounds in the language.

(3) *Established compounds*

In the Malay language, there are several combinations that are known as the 'established' compounds, i.e. combinations that are considered as stable and solid compounded forms. Below is the list of these so-called established compounds that can be found from the corpus.

(61) Established compounds

	Compounds	Gloss	Meaning
1	*bumiputera*	*bumi* (earth) + *putera* (son/prince)	native
2	*warganegara*	*warga* (people) + *negara* (country)	citizen
3	*tandatangan*	*tanda* (mark) + *tangan* (hand)	signature
4	*olahraga*	*olah* (exercises) + *raga* (body)	sports/athletics
5	*matahari*	*mata* (eye) + *hari* (day)	sun
6	*kakitangan*	*kaki* (foot) + *tangan* (hand)	staff
7	*tanggungjawab*	*tanggung* (to bear) + *jawab* (to response)	responsibility
8	*sukarela*	*suka* (like) + *rela* (willing)	voluntariness
9	*antarabangsa*	*antara* (among) + *bangsa* (race)	international

10	*kerjasama*	*kerja* (work) + *sama* (together)	co-operation
11	*setiausaha*	*setia* (loyal) + *usaha* (effort)	secretary
12	*suruhanjaya*	*suruhan* (order) + *jaya* (success)	commission
13	*pesuruhjaya*	*pesuruh* (emissary) + *jaya* (success)	commissioner
14	*jawatankuasa*	*jawatan* (position) + *kuasa* (authority)	committee

If we accept the structures above as constructions consisting of two conjoined components, we can proceed to analyse them through the compoundhood checklist like we have done so far.

We begin by analysing the structure and headedness of the compounds. All three compounding schema can be found within this group of compounds. The most common one is the [X Y] (X) structure. For instance, the noun head *tanda* 'mark' is modified by the noun modifier *tangan* 'hand', resulting in the endocentric nominal compound *tandatangan* 'signature' ($[N_1\ N_2]\ (N_1)$). There are also a few [X Y] (Z) structures where none of the components can be accounted for as either a semantic or a categorical head. For instance, *setiausaha* is neither a state of being *setia* 'loyal', nor it is a type of *usaha* 'effort', (but) resulting in the exocentric nominal compound *setiausaha* 'secretary' ($[A_1\ N_1]\ (N)$). There is only one example of an [X Y] (Y) structure. In the compound *bumiputera*, the head is the right-hand component of *putera* 'son/prince', which is modified by the left-hand component *bumi* 'earth' (i.e. literally 'son of earth'), resulting in the left-headed endocentric nominal compound *bumiputra* 'native' ($[N_1\ N_2]\ (N_2)$).

Further modification to the compounds will not affect their components separately. For instance, adding *lelaki* (man) to *tanggungjawab* 'responsibility' produces [*tanggungjawab*] [*lelaki*] 'man's responsibility', not *[*tanggung*] [*jawab lelaki*] *'response male, to bear'.

Similarly, the components cannot be separated by insertion of other elements, e.g. insertion of *lelaki* (man) to *warganegara* 'citizen' ungrammatically produces either *[*warga*] [*lelaki negara*] *'people/denizen, country-man', or *[*warga lelaki*] [*negara*] *'male denizen, country'. With regard to component switching, swapping, for instance, *tandatangan* 'signature' to *tangan tanda* *'mark(ed) hand' does not bring about the intended meaning of 'signature'. The compounds that are able to be circumfixed will undergo a process similar to any other single words in the language (e.g. *olahraga* 'athletics' results in *keolahragaan* '(everything) to do with athletics'). With regard to reduplication, there are no arguments on whether reduplication will only take place on the head component or otherwise. The whole compound will be reduplicated (just like any other single words) if it needs to be (e.g. *setiausaha-setiausaha* 'secretaries'). To a great extent, these 'established' compounds truly adhere to the syntactic criteria of compoundhood, arguing that the components of the compounds are strongly bound.

Phonological criteria are not applicable to the compounds. Although the components are conjointly spelled, there is no evidence of assimilation between them (e.g. *suka* /suka/ + *rela* /rela/ maintains as *sukarela* /sukarela/). There is also no evidence of stress on a particular component within the compounds because they are all treated (hence pronounced) as a single unit (instead of two separate components).

There is strong support for the compounds to be considered as lexicalised structures. Lexicalisation is especially true for those with non-transparent components, e.g. *setia* (loyal) + *usaha* (effort) produces *setiausaha* 'secretary', *kaki* (foot) + *tangan* (hand) produces *kakitangan* 'staff', *mata* (eye) + *hari* (day) produces *matahari* 'sun', etc. The meaning of these compounds cannot be derived from the literal meaning of their components, and thus they are necessarily lexicalised. Nevertheless, there are some that are transparently compositional, e.g. *kerja* (work) + *sama* (together) produces *kerjasama* 'co-operation', *antara* (among) + *bangsa*

(race) produces *antarabangsa* 'international', etc. In other words, the overall meaning of the compound can be derived from the meaning of its components. Even so, the components are always treated as single units, conjointly spelled and limited in number. Furthermore, they also adhere to the criteria of compoundhood (i.e. disallow modification, separation and switching), which supports the idea of component unification. Taking these points into account, it is safe to assume that even the more transparent of the established compounds are more likely to be lexicalised than they are to be compositional constructions.

There are two kinds of relationship observable between the components of the compounds. First is the attributive type, where there is an interaction between a head component (H) and a modifying component (M). For instance in *tandatangan*, it is the right-hand component of *tangan* (M) that is further modifying the left-hand component *tanda* (H) to express the meaning 'mark of hand', i.e. 'signature'. Conversely, in *bumiputera*, it is the left-hand component of *bumi* (M) that is further modifying the right-hand component of *putera* (H) to express the meaning 'son of land', i.e. 'native'.

The second type of relationship observable is the coordinative type. Examples of this type include endocentric ones such as *tanggungjawab* 'responsibility' (from *tanggung* 'to bear' + *jawab* 'to response'), and exocentric ones such as *kakitangan* 'staff' (from *kaki* 'feet' + *tangan* 'hands'). In this type of compound, neither of the components modifies the other. They are both equally important in conveying the intended meaning of the compounds.

The analysis above illustrates that it is possible to tease out and clarify the internal structure of these so-called 'established' compounds. All three compounding schema can be found within this group of compounds, i.e. the [X Y] (X), [X Y] (Y) and [X Y] (Z), along with the division of endocentric and exocentric headedness. It is also clear that the compounds conform to the proposed criteria of compoundhood. The compounds disallow modification,

insertion and switching of their components, and they are sensitive to circumfixation and reduplication accordingly. The compounds also have strong arguments for lexicalisation, further asserting their status as compounded structures in the language. Finally, we have also established that there are at least two kinds of relationship observable within these compounds, i.e. the attributive and the coordinative.

Having said this, it is important to understand that, in a real-world context, these compounds are not necessarily perceived as combinations of two different components. It is true that speakers can decompose the structures into their individual components (if they want or have to), but under normal circumstances these compounds are treated and perceived as single concrete units. In other words, although the combinations in (61) above can be classified as NN/VV/AA (ATR/CRD) (endo/exo) nominal compounds as we have demonstrated above, it also makes perfect sense to simply classify them as 'established' compounds in the language (as has been done by Malay linguists in general).

5.3.2 Verbal compounds

In general, there are fewer examples of verbal compounds in the corpus in comparison to nominal ones. Nevertheless, several different types of verbal compounds are evident, namely the Verb-Noun (VN), Verb-Verb (VV), and Verb-Adjective (VA) compounds.

5.3.2.1 The VN verbal compounds

There are some examples of Verb-Noun combinations that can be considered as verbal compounds.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Verb-Noun (VN) attributive (ATR) endocentric (endo) verbal compounds.

(62) VN (ATR) (endo) verbal compounds

	Compounds	Gloss	Meaning
1	*bawa diri*	(to take/carry + self)	to sulk
2	*layan diri*	(to serve + self)	to serve oneself
3	*buang negeri*	(to banish + state)	to expatriate

The combinations above have the structure of $[V_1 \ N_1] \ (V_1)$, where (V_1) is the endocentric verbal head of the compounds. The verb (V_1) is arguably the semantic head of the combinations, i.e. *bawa diri* is to *bawa* (to take/carry) oneself away, *layan diri* is to *layan* (to serve) oneself and *buang negeri* is to *buang* (to banish) from a state. Accordingly, we can assume that (V_1) is the categorical head as it projects its verb category onto the compounds as a whole. The compounds disallow separation, switching and further modification of their components as these will disrupt the meaning of the compounds. The compounds cannot undergo circumfixation and reduplication. The placement of stress on the components also has no significance, nor is there any evidence of assimilation between them. In terms of lexicalisation, we cannot say with certainty that they are fully stored structures. This is because their meaning is both transparent and compositional via their components. To a fair degree, the relationship between the verb component (V_1) and the noun component (N_1) can be interpreted as having a head + modifier relationship, i.e. attributive. All in all, the structures in (62) above can be classified as examples of VN (ATR) (endo) verbal compounds in the language.

(ii) Idiomatic

There are a few examples from the corpus that have been analysed as Verb-Noun (VN) exocentric (exo) idiomatic verbal compounds.

(63) VN (exo) idiomatic verbal compounds

	Compounds	Gloss	Meaning
1	*tumbuk rusuk*	(to pound/punch + flank)	to bribe
2	*cakar ayam*	(to scratch + chicken)	scribbling/bad handwriting
3	*cuci mata*	(to clean + eyes)	to look at woman (for pleasure)

These compounds have the exocentric structure of [V_1 N_1] (V). The overall verbal category (V) and the meaning of the compounds cannot be derived from the inner verb (V_1) (or the noun (N_1)) component. For instance, *tumbuk rusuk* (to punch + flank) is not the action of literally punching the side (ribs) of someone. Similarly, *cakar ayam* (to scratch + chicken) has nothing to do with 'scratching a chicken'.[120] For this reason, the compounds above are truly exocentric compounds, in both the categorical and the semantic sense.

The compounds also disallow component modification, separation and switching, which further supports the idea of the components being intact. Reduplication and circumfixation are not applicable to the compounds. There is also no significance in stress placement and/or any proof of assimilation between the components to further support compoundhood.

---

[120] The word *cakar* can also mean the noun 'claw' and '(claw) marks'. In this case, the compound would be an NN combination instead of VN like it is categorised here. The meaning still remains the same (i.e. scribbling/bad handwriting); it only differs in the categorisation. In other words, the compound *cakar ayam* can be placed in NN (exo) idiomatic nominal compounds (i.e. the *tangkai jering* types). *Cakar ayam* is also the name for a traditional sweet that looks like a messy scribbling.

As mentioned, the meaning of the compounds cannot be literality derived from the meaning of their components (i.e. the components are opaque). Therefore, they can be considered as 'true' idiomatic compounds with a strong claim for lexicalisation. However, we once again find it difficult to exactly pin down the type of relationship between the components.[121] The issue is again due to lack of a concrete headedness. Therefore, the combinations in (63) above are classified only as examples of VN (exo) idiomatic verbal compounds in the language.

5.3.2.2 The VV verbal compounds

From the corpus, there are a few Verb-Verb combinations that can be considered as verbal compounds.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Verb-Verb (VV) attributive (ATR) endocentric (endo) verbal compounds.

(64) VV (ATR) (endo) verbal compounds

	Compounds	Gloss	Meaning
1	*cari gali*	(to search + to dig)	to explore
2	*ambil alih*	(to take + to move)	to replace
3	*kahwin lari*	(to marry + to run)	to elope

---

[121] Although there is an inclination to argue that the components are in a subordinative relation.

4	*ambil kira*	(to take + to count)	to consider
5	*temu duga*	(to meet + to question)	to interview
6	*lipat ganda*	(to fold + to fold/double)	to multiply

The compounds above follow the schema of [V₁ V₂] (V₁). The verb (V₁) can be considered as their semantic head, e.g. *cari gali* is an act of *cari* (to search), *ambil alih* is an act of *ambil* (to take), *temu duga* is an act of *temu* (to meet), etc. Accordingly, (V₁) is also the categorical head of the compounds (as they are all verbal compounds).

Tests of modification, insertion and component switching disrupt the meaning of the compounds. For instance, adding the word *minyak* 'oil' to the compound *cari gali* produces [*cari gali*][*minyak*] 'oil exploration', not *[*cari*][*gali minyak*] *'to search, oil-digging'; insertion ungrammatically produces *[*cari minyak*][*gali*] *'to search oil, digging' or *[*cari*][*minyak gali*] *'to search, oil-dig'; and switching from *cari gali* to *gali cari* *'dig - search' disrupts the intended meaning of the compound. Similar outcomes can be seen on all of the other compounds, indicating that their components act as unified combinations. Most of the compounds can be circumfixed (e.g. *ambil alih* becomes *pengambilalihan* 'the process of replacing'), while some cannot (e.g. *berkahwin larikan* *'to make someone to commit in an elopement'). None of the compounds are subjected to reduplication.

In terms of phonology, there is no evidence of specific component stressing to establish compoundhood, nor is there any evidence of assimilation between the components. In terms of semantic criteria, there is no concrete argument for lexicalisation. This is mainly due to the components being fairly literal and transparent (e.g. *kahwin* 'to marry' + *lari* 'to run' is 'to elope', *temu* 'to meet' + *duga* 'to question' is 'to interview', etc.).

The relation between their components can be described as having an attributive (head + modifier) relationship (e.g. *temu duga* is an act of *temu* 'to meet' and to *duga* 'to question' during the meeting). Overall, the combinations in (64) above can be classified as examples of VV (ATR) (endo) verbal compounds in the language.

(ii) Coordinative (CRD)

There are two combinations in the corpus that can be analysed as Verb-Verb (VV) coordinative (CRD) endocentric (endo) verbal compounds.

(65) VV (CRD) (endo) verbal compounds

	Compounds	Gloss	Meaning
1	*tunjuk ajar*	(to show + to teach)	to guide (to advise)
2	*campur aduk*	(to mix + to mix)	to mix (completely/thoroughly)

The compounds above follow the schema of $[V_1\ V_2]$ (V). The categorical and semantic heads of the compounds cannot be taken as an exclusive projection of either (V_1) or (V_2), but instead as an equal projection of both components. In other words, (V_1) is as equally important as (V_2) (e.g. the meaning of to *campur* 'to mix' is equally as important as the meaning of to *aduk* 'to mix' in order to produces the compound meaning of to *campur aduk* 'to mix (completely/thoroughly)').

In general, tests of modification, insertion and component switching disrupt the meaning of the compounds. Insertion of the conjunction *dan* (and) fairly maintains the meaning of the compounds (e.g. *tunjuk dan ajar* (show and teach) 'to guide'), arguably proving that the combinations are phrases rather than compounds. However, as we have discussed before, such

an argument is redundant because it is only confirming that the components of coordinated compounds are indeed in a 'coordinated' relationship. The compounds can be circumfixed (e.g. *campur aduk* becomes *mencampuradukkan* 'to make (something) thoroughly mix'). All these observations provide a degree of evidence for the argument that the components are in a unified state. There is no evidence of stress patterning to establish compoundhood, nor is there any evidence of assimilation between the components. There is also no concrete argument for lexicalisation as the components have a literal and transparent meaning. Finally, as stated, the components can be described as having a coordinative relationship, where both components are equally important in deriving the intended meaning of the compound. Overall, the combinations in (65) above can be classified as examples of VV (CRD) (endo) verbal compounds in the language.

5.3.2.3 The VA verbal compounds

There are a few Verb-Adjective combinations that can be considered as verbal compounds.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Verb-Adjective (VA) attributive (ATR) endocentric (endo) verbal compounds.

(66) VA (ATR) (endo) verbal compounds

	Compounds	Gloss	Meaning
1	*temu ramah*[122]	(to meet + friendly/chatty)	to have a talk

---

[122] *Temu ramah* can also be a nominal compound as in 'a talk'.

2	*kenal pasti*	(to recognise + sure)	to identify
3	*tolak tepi*	(to push + side)	to put aside
4	*pandang rendah*	(to look + low)	to underestimate
5	*tolak ansur*	(to push + gradual)	to tolerate

The compounds above follow the structure of [V$_1$ A$_1$] (V$_1$). To a certain degree, the verb (V$_1$) can be considered as the semantic head of the compounds (i.e. *temu ramah* is an act of *temu* 'to meet', *kenal pasti* is the act of *kenal* 'to recognise', etc.).[123] Accordingly, the verb (V$_1$) can also be considered as the categorical head projecting the verbal category to the compounds as a whole.

The compounds are sensitive to modification, insertion and component switching. For instance, modifying *temu ramah* with *radio* (radio) results in [*temu ramah*][*radio*] 'a radio talk/interview', not *[*temu*][*ramah radio*] *'to meet, radio-friendly'; insertion as in *[*temu radio*][*ramah*] *'to meet radio, friendly' or *[*temu*][*radio ramah*] *'to meet, friendly radio'; and switching as in *_ramah temu_ *'friendly-meeting', will all disrupt the meaning of the compound. A similar outcome is apparent for the rest of the compounds in (66). This observation can be taken as evidence that the components of the compounds are in union.

In general, the compounds are not subjected to reduplication and/or circumfixation.[124] The compounds are also not justifiable by stress patterning and/or assimilation. In terms of semantic criteria, there is no strong argument for lexicalisation given that the meaning of the compounds can be derived from the meaning of the components (e.g. *temu* 'to meet' + *ramah*

---

[123] An exception can perhaps be made for *tolak ansur* 'to compromise', where neither of the compound's components, (V$_1$) (*tolak* 'to push') nor (A$_1$) (*ansur* 'gradual'), are truly its semantic head.

[124] Only *kenal pasti* can be circumfixed (e.g. *mengenalpastikan* 'to identify for').

'friendly/chatty' is to 'meet and have a friendly talk').[125] Only *tolak ansur* has a stronger case for lexicalisation as the components of *tolak* 'to push' + *ansur* 'gradual' seem to be more opaque and less transparent to convey the intended meaning 'to tolerate'.

The components are in an attributive modified - modifying relationship (e.g. *temu ramah* is an act of *temu* 'to meet' and to be *ramah* 'friendly/chatty' during the meeting). Overall, the combinations in (66) above can be classified as examples of VA (ATR) (endo) verbal compounds in the language.

5.3.3 Adjectival compounds

There are only two types of adjectival compounds available from the corpus; that is, the Adjective-Noun (AN) and the Adjective-Adjective (AA) combinations.

5.3.3.1 The AN adjectival compounds

There are several Adjective-Noun combinations from the corpus that can be analysed as adjectival compounds.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Adjective-Noun (AN) attributive (ATR) endocentric (endo) adjectival compounds.

---

[125] The compound is also more compositional in the sense that there are other analogous compounds with a similar structure to *temu ramah*, e.g. *temu bual* (to meet + to talk) 'talk/interview', *temu duga* (to meet + to question) 'interview', etc.

(67) AN (ATR) (endo) adjectival compounds

	Compounds	Gloss	Meaning
1	*biru laut*	(blue + sea)	ocean blue
2	*merah jambu*	(red + red guava)	pink
3	*hijau daun*	(green + leaf)	leafy green
4	*kuning langsat*	(yellow + a local fruit)	yellowish
5	*putih telur*	(white + egg)	egg white - like
6	*kelabu asap*	(grey + smoke)	smoky grey
7	*lut sinar*	(penetrating + ray)	transparent
8	*telap air*	(permeable + water)	penetrable
9	*celik huruf*	(aware/able + letters)	literate
10	*celik akal*	(aware/able + mind)	intelligent

In general, the combinations above have the prototypical compound structure of $[A_1 N_1]$ (A_1), where the (A_1) is the endocentric adjectival head of the combinations. The adjective (A_1) is the semantic head of the combinations, i.e. they are the superordinate of the whole compound. For instance, *biru laut* is a type of *biru* (blue), in this case ocean blue, and *celik huruf* is to be *celik* (aware/able), in this case to read. The adjective (A_1) can thus be accepted as the categorical head as well, since a compound's overall category is the same as that of its head (i.e. all of the compounds above are adjectives).

The compounds disallow separation of their components. Insertion of other elements disrupts the meaning of the combinations, e.g. insertion of *tua* (old) between the components of *biru laut* (blue + sea) produces [*biru tua*] [*laut*] 'dark blue, sea', or [*biru*] [*tua laut*] 'blue, old sea'. Similarly, switching the components will also disrupt the meaning of the compounds,

e.g. *biru laut* becomes *laut biru* 'blue sea'. The two tests of insertion and switching provide the general argument for the combinations above having unified components. However, the test of modification does not necessarily yield a clear-cut outcome. For example, adding *kereta* (car) to *kelabu asap* (grey + smoke) produces [*kereta*] [*kelabu asap*] 'smokey-grey car', not *[*kereta kelabu*] [*asap*] *'grey car, smoke'. Having said that, adding *kerata* (car) as in [*kelabu*] [*asap kereta*] can actually mean 'car-smoke grey', as in 'the shade of grey of car smoke'. Perhaps a better exemplification would be with *hijau daun* (green + leaf) 'leafy green'. One can imagine two different shades of leafy green, e.g. [*hijau*] [*daun mangga*] 'mango-leaf green' vs. [*hijau*] [*daun pandan*] 'screw pine-leaf green'. This illustrates that the modifying component (N_1) *daun* (leaf) might not be as 'fused' to the head component as it initially seems.[126]

There is no evidence that any of the combinations above can be shown to not be compounds via stress patterning. Also, there is no occurrence (or potential occurrence) of assimilation between the components of the compounds. Similarly, there is no concrete proof that the combinations are lexicalised forms.

In comparison to the other types of compounds analysed at so far, the AN compounds show the least adherence to the compoundhood criteria (i.e. only adhering to the syntactic criteria of separation and switching). This might be a sign that the components involved are not strictly bounded, hence not necessarily lexicalised. Furthermore, the combinations are compositional in the sense that the meanings can be derived from their components, e.g. *biru laut* (blue + sea) is the shade of blue similar to the sea's colour, *telap air* (permeable + water) is the ability to be penetrated by water, etc. Having such transparency arguably implies that the compounds are not necessarily stored as a whole in the lexicon.

---

[126] A note on the other two syntactic tests, i.e. circumfixation and reduplication: taken as a whole, none of the combinations listed above can undergo circumfixation and reduplication in the language.

Nonetheless, the adjective component (A₁) and the noun component (N₁) of the combinations can be interpreted as having a head + modifier relationship, i.e. attributive. All in all, the structures in (67) above can be classified as AN (ATR) (endo) adjectival compounds in the language.

(ii) Idiomatic

From the corpus, there are also several combinations that can be analysed as Adjective-Noun (AN) exocentric (exo) adjectival compounds.

(68) AN (exo) idiomatic adjectival compounds

	Compounds	Gloss	Meaning
1	*berat tulang*	(heavy + bones)	lazy
2	*ringan tulang*	(light + bones)	hard working
3	*panjang tangan*	(long + hand)	likes to steal
4	*terang hati*	(bright + heart)	easy to gain knowledge
5	*manis mulut*	(sweet + mouth)	smooth-tongued
6	*besar kepala*	(large + head)	stubborn/arrogant

All of the compounds above have the structure of [A₁ N₁] (A). The overall adjectival (A) category of the compounds is not the projection of the inner adjectival (A₁) component. For instance, *berat tulang* (heavy + bones) is not the state of being *berat* (heavy), nor is it a kind of *tulang* (bones). *Berat tulang* has an exocentric meaning of 'being lazy'. Similarly, *besar kepala* (large + head) is not the state of being *besar* (large), nor is it a kind of *kepala* (head), but instead means to be 'stubborn/arrogant'. In this sense, neither the categorical nor the

semantic heads of the compounds in (68) can be derived from their components. Therefore, they can be categorised as exocentric compounds.

With regard to the syntactic test of separation, the insertion of other elements between the components interrupts the meaning of the compounds. For instance, insertion of *seperti* (like) between *berat tulang* results in *berat seperti tulang* '(literally) heavy like (a) bone'. Further modification of the compounds also disrupts the unity of the components, e.g. adding *sikap* (attitude) to *ringan tulang* 'hard working' results in [*sikap*][*ringan tulang*] 'hard-working attitude', and not *[*sikap ringan*][*tulang*] *'light attitude, bones'. Likewise, the switching of components will bring a totally different meaning to the compounds, e.g. *ringan tulang* 'hard working' becomes *tulang ringan* 'bones that are (literally) light', *besar kepala* 'arrogant' becomes *kepala besar* 'a (literally) big head', etc. These observations can be taken as proof that the components of the compounds are unified. However, the other two tests of circumfixation and reduplication are not applicable to the compounds. The compounds cannot be grammatically circumfixed (e.g. *meringantulangkan* *'to make (oneself) industrious'), nor can they be grammatically reduplicated (e.g. *ringan-ringan tulang, *ringan tulang - ringan tulang, *ringan tulang - tulang).

Phonological criteria are also not applicable to the compounds. Applying different stress patterns to the components will not distinguish them from being phrases. Similarly, there is no evidence of assimilation between the components of the compounds. However, the compounds have a strong case for lexicalisation. This is mainly because the meanings of the combinations are not transparent. In other words, it is (almost) impossible to derive the meaning of these compounds from the meaning of the components and the meaning of their combinations.

In terms of the relationship between the adjective component (A$_1$) and the noun component (N$_1$) of the combinations, we cannot be entirely sure whether it is accurate to interpret them as having a head + modifier relation, i.e. an attributive relationship.[127] This is mainly due to the fact that they are exocentric compounds, and identifying the 'exact' head component here is not as clear-cut as it is for endocentric compounds. For this reason, the structures in (68) above will be classified only as AN (exo) adjectival compounds in the language.

5.3.3.2 The AA adjectival compounds

There are several Adjective-Adjective combinations from the data that can be analysed as adjectival compounds. Similar to the Adjective-Noun compounds, most of the Adjective-Adjective compounds denotes colour schemes as well.

(i) Attributive (ATR)

Below is the list of combinations that have been analysed as Adjective-Adjective (AA) attributive (ATR) endocentric (endo) adjectival compounds.

(69) AA (ATR) (endo) adjectival compounds

	Compounds	Gloss	Meaning
1	*merah muda*	(red + young)	rosy
2	*biru tua*	(blue + old)	dark blue

---

[127] However, we are sure that the relationships between the components are not of the subordinate or the coordinate type.

3	*kuning tua*	(yellow + old)	dark yellow
4	*merah cerah*	(red + bright)	light red
5	*merah tua*	(red + old)	crimson

In general, the AA adjectival compounds here are very similar to the earlier discussed AN adjectival compounds. Obviously, the main difference between them is the category of the modifying component, i.e. an adjective here (instead of a noun before) with the compounding structure of [A$_1$ A$_2$] (A$_1$). The adjective (A$_1$) is the semantic head of the combinations being that the whole compound is the hyponym of the head (A$_1$), e.g. *merah muda* is a hue of *merah* (red), *biru tua* is a hue of *biru* (blue), etc. Accordingly, a compound's overall category is the same as that of its head (A$_1$), i.e. all of the compounds above are adjectives.

With regard to the tests of syntactic criteria, the compounds disallow modification, separation and switching of their components.[128] For example, adding *baju* (shirt) to *merah muda* (red + young) results in [*baju*] [*merah muda*] 'rosy (coloured) shirt', not *[*baju merah*] [*muda*] *'young red shirt'. Inserting *baju* (shirt) ungrammatically results in *[*merah baju*] [*muda*] *'shirt-red, young', or *[*merah*] [*baju muda*] *'young-shirt, red'. Similarly, switching the components of *merah muda* to *muda merah* ungrammatical results in *'young-red'. These observations demonstrate the intactness of the components within the compounds.

With regard to phonological criteria, none of the compounds are subjected to stress patterning, nor are their components subjected to any sort of assimilation. There is also no concrete proof that the compounds are lexicalised. Although the modifying components (e.g. *cerah* 'bright', *muda* 'young' and *tua* 'old' results in slightly opaque – for instance, *tua* (old)

---

[128] The remaining syntactic tests of compoundhood are not applicable as none of the compounds can undergo circumfixation (**kemerahmudaan* *'to be pinkish') or reduplication (e.g. **merah-merah muda*, **merah muda - merah muda*, **merah muda-muda*).

in *biru tua* does not imply an 'old'-blue, but a 'dark' shade of blue – nonetheless, the association of *tua* 'old' being 'dark' is not too far-fetched. Taken as a whole, the compounds can be considered as compositional and transparent (i.e. the meanings can be derived from their components), hence not necessarily lexicalised.

The relationship between the head component (A_1) and the modifying component (A_2) is that of the head + modifier relationship, i.e. attributive. All in all, the structures in (69) above can be classified as AA (ATR) (endo) adjectival compounds in the language.[129]

5.4 Chapter summary

From the previous analysis and discussion, it is clear that the Malay language has several different kinds of compounded structures. Guided by the devised Malay compoundhood checklist, we have managed to identify and separate the compounds into their respective subcategories under the umbrella of nominal, verbal and adjectival compounding. Below is the summary list of the different types of compounds available from the corpus of this study.

---

[129] There is an AA adjectival combination from the corpus that is between being an attributive and/or a coordinative compound, i.e. *muda remaja* (young + teenage) 'youth (state of being young)'. On one hand, the adjective *remaja* (teenage) seems to be modifying the adjective head *muda* (young) by further elaborating the quality of being 'young'. In this sense, we can say that the two components have the head-modifier relationship. Conversely, the state of being a *remaja* (teenage) is also being 'young', and thus we can also say that the two components are in a coordinative relationship.

(70) Types of Malay compounds[130]

(i) Nominal compounds

The NN nominal compounds	
(5.3.1.1 (i))	NN (ATR) (endo) nominal compounds
	e.g. the *jam tangan* 'wristwatch' types.
(5.3.1.1 (ii))	NN (CRD) (endo) nominal compounds
	e.g. the *ibu bapa* 'parents' types.
(5.3.1.1 (iii))	NN (ATR) (endo) partially idiomatic nominal compounds
	e.g. the *duit kopi* 'tips' types.
	NN (exo) idiomatic nominal compounds
	e.g. the *tangkai jering* 'miser' types.
The NV nominal compounds	
(5.3.1.2 (i))	NV (ATR) (endo) nominal compounds
	e.g. the *padang tembak* 'shooting range' types.
The NA nominal compounds	
(5.3.1.3 (i))	NA (ATR) (endo) nominal compounds
	e.g. the *kerusi malas* 'lounge chair' types.
Other nominal compounds	
(5.3.1.4 (1) (i))	VN (SUB) (exo) nominal compounds
	e.g. the *sapu tangan* 'handkerchief' types.
(5.3.1.4 (1) (ii))	VN (exo) idiomatic nominal compounds
	e.g. the *hirup darah* 'extortionist' types.

---

[130] Types of Malay compounds available from the corpus of this study.

(5.3.1.4 (2) (i))	AN (ATR) (endo) nominal compounds
	e.g. the *perdana menteri* 'prime minister' types.
(5.3.1.4 (3)) Established compounds	NN/VV/AA (ATR/CRD) (endo/exo) nominal compounds
	e.g.: the *tandatangan* 'signature' types, (ATR) (endo). the *tanggungjawab* 'responsibility' types, (CRD) (endo). the *kakitangan* 'staff' types, (CDR) (exo).

(ii) Verbal compounds

The VN verbal compounds	
(5.3.2.1 (i))	VN (ATR) (endo) verbal compounds
	e.g. the *bawa diri* 'to sulk' types.
(5.3.2.1 (ii))	VN (exo) idiomatic verbal compounds.
	e.g. the *tumbuk rusuk* 'to bribe' types.
The VV verbal compounds	
(5.3.2.2 (i))	VV (ATR) (endo) verbal compounds
	e.g. the *cari gali* 'to explore' types.
(5.3.2.2 (ii))	VV (CRD) (endo) verbal compounds
	e.g. the *tunjuk ajar* 'to guide' types.
The VA verbal compounds	
(5.3.2.3 (i))	VA (ATR) (endo) verbal compounds
	e.g. the *temu ramah* 'to have a talk' types.

(iii) Adjectival compounds

The AN adjectival compounds	
(5.3.3.1 (i))	AN (ATR) (endo) adjectival compounds.
	e.g. the *biru laut* 'ocean blue' types.
(5.3.3.1 (ii))	AN (exo) idiomatic adjectival compounds.
	e.g. the *berat tulang* 'lazy' types.
The AA adjectival compounds	
(5.3.3.2 (i))	AA (ATR) (endo) adjectival compounds.
	e.g. the *merah muda* 'rosy' types.

The list above can also be projected as a diagrammatic illustration, as shown below.

(71) Diagrammatic representation of the types of Malay compounds

Malay compounds

Subordinative　　　Attributive　　　Coordinative　　　Idiomatic

compounds

Exo.　　　　　Endo.　　　　Endo.　　　Exo.　　　Exo.

(VN) *sapu tangan*　　(NN) *jam tangan*　　(NN) *ibu bapa*　　(NN) *kakitangan*

(NV) *padang tembak*　　(VV) *tunjuk ajar*

(NA) *kerusi malas*　　(AA) *sukarela*

(VN) *bawa diri*

(VV) *cari gali*

(VA) *temu ramah*

(AN) *biru laut*

(AA) *merah muda*

Right-headed:

(NN) *bumiputra*

(AN) *perdana menteri*

(NN) *duit kopi*
(partially idiomatic)

Nominal:

(NN) *tangkai jering.*
(VN) *hirup darah.*

Verbal:

(VN) *tumbuk rusuk.*

Adjectival:

(AN) *berat tulang*

As a conclusion, we will once again go through the main topics of the checklist, summarising the general observations, along with addressing a number of issues which are apparent from the outcome of the analysis.

Firstly, we have seen that the language can have at least three types of compounding structures, namely the [X Y] (X) structure, the [X Y] (Y) structure and the [X Y] (Z) structure. The most common among them is the [X Y] (X) structure, where the head can be derived from the left-hand component (X) of the compound. The structure of [X Y] (Z) seems to be fairly common, where the head (Z) is not necessarily the projection of either of the components involved. Finally comes the very limited structure of [X Y] (Y), where the head is the projection of the right-hand (Y) component of the compound.

The issue of headedness establishes whether a given compound is an endocentric or an exocentric compound. The analysis observed two types of headedness, namely the categorical head and the semantic head. The most common type of headedness is the one where both categorical and semantic heads are derivable from a single (same) component within the compound. When this is the case, the compound can be clearly labelled as an endocentric compound. From the analysis, this type of compound predominantly has the left-hand component as (both) the semantic and the categorical heads, i.e. the [X Y] (X) structures (e.g. *jam tangan* is a type of *jam*, and the whole compound is also a noun as projected by the noun head *jam*). There are also a few examples where the head is on the right-hand component. Nonetheless, both categorical and semantic heads in this type of compound are still projected from the same right-hand component, i.e. the [X Y] (Y) structures (e.g. *perdana menteri* is a *menteri*, and the whole compound is also a noun as projected by the noun head *menteri*).

The other kind of headedness can be categorised as the [X Y] (Z) structures. One type of headedness under this structure is when a component projects only as the semantic head (or

only as the categorical head). Recall the compound *sapu tangan* (handkerchief) has a semantic head *sapu* (to wipe), but the *sapu* does not project its verbal category to the whole nominal compound. In this case, the compound has an endocentric semantic head, but an exocentric categorical head.[131] Another type of headedness under the [X Y] (Z) structure is when both of the components, or neither of the components, represent the categorical and/or the semantic heads of the compound. The former is the *ibu bapa* (parents) type, where both *ibu* (mother) and *bapa* (father) are equally the semantic and the categorical heads of the compound. In this sense, the compounds have endocentric semantic and categorical heads.[132] The latter is the *kakitangan* (staff) type, where neither *kaki* (foot) nor *tangan* (hand) is the semantic head of the compound. Although the components are both nouns, they cannot be truly attributed as the reason why the whole compound is nominal. In this sense, the compounds have exocentric semantic and categorical heads.[133]

The next main issue on the checklist concerns the proposed criteria of compoundhood, i.e. the syntactic and phonological criteria. With regard to syntactic criteria, there is not much debate on whether the compounds comply to the tests of modification, insertion and switching as proof of compoundhood. The analysis has shown to a very large extent that all of the compounds disfavour further modification of components, disallow insertion of other elements between components, and also disallow switching between the components. However, the same cannot be said with the tests of reduplication and circumfixation. Recall how it has been suggested that, upon reduplication, both components of a given compound should merge and

---

[131] Recall that the noun component *tangan* (hand) cannot be attributed to the overall nominal category of the compound.

[132] Technically, the *ibu bapa* type of compound can also be represented as the [X Y] (XY) structure. For ease of representation, we will keep using the [X Y] (Z) schema, with the (Z) representing the idea of exocentricity of either the semantic, the categorical, or of both/neither semantic and/or categorical headedness.

[133] Compounds with both exocentric semantic and categorical heads are clearer in the idiomatic types, e.g. in *hirup darah* (extortionist) neither *hirup* (to suck) nor *darah* (blood) is the semantic and/or the categorical head of the compound.

reduplicate as a whole.[134] The analysis, however, has shown that not all compounds can be reduplicated, and the ones that can be will only reduplicate the head component and not the whole compound (cf. *jam-jam tangan* 'wristwatches' (not **jam tangan - jam tangan*), *kerusi-kerusi malas* 'lounge chairs' (not **kerusi malas - kerusi malas*), etc. Similarly, it has been suggested that the components of a compound should merge upon circumfixation.[135] However, the analysis has also shown that not all compounds can undergo circumfixation as well (cf. **menjamtangankan* *'to make (something) into a wristwatch', **berjamtangankan* *'to wear (something) as a wristwatch', **mengerusimalaskan* *'to make (something) into a lounge chair', etc.).

On the other hand, the proposed phonological criteria of stress and assimilation are arguably specified occurrences within the Malay compounding phenomenon. Throughout the analysis, we are only confident to argue for the compound *ibu bapa* 'parents' as conveying some sort of a different meaning due to different stress patterning.[136] As for the rest of the analysed compounds, there is no concrete evidence that different stress placement can contribute to a compound having different meanings. A separate study focusing on this matter of Malay compound stress patterns is definitely needed. Similarly, assimilation is also a highly restricted occurrence within Malay compounding. This is clear throughout the analysis, as even the compounds with potential components for assimilation still do not assimilate (e.g. *gambar*

---

[134] Cf. Hassan (2006) in his argument that reduplication can prove compoundhood; e.g. *uji kaji* 'experiment' is reduplicated as *ujikaji-ujikaji* 'experiments', proving that the components are in essence unified and compounded, which is why they merge and reduplicate as a whole upon reduplication.

[135] Cf. Hassan (2006) in his argument that circumfixation can prove compoundhood; e.g. *uji kaji* 'experiment' can be circumfixed into *pengujikajian* 'the process of experimenting', once again proving that the components are unified and compounded which is why they are merged upon circumfixation.

[136] As mentioned, placing stress on the first component /ˋibu bapa/ can convey the meaning of 'father's mother', instead of the neutral stress /ibu bapa/ to mean 'parents'. Nevertheless, it is a forced occurrence as the better way to say it is *ibu kepada bapa* 'mother of father' or 'father's mother'.

rajah (/gambar + radʒah/) does not assimilate to */gambaradʒah/, *alat tulis* (/alat/ + /tulis/) does not assimilate to */alatulis/, etc.

The question is then, if a given combination is unable to fulfil all of the proposed criteria in the checklist, does it mean that the combination is not a compound? Clearly, not all of the proposed criteria are effective (or applicable) as tests of compoundhood. As we have seen throughout the analysis, it can be difficult for a given combination to necessarily meet all the proposed criteria in order to be considered a compound. If this was to be the case, Malay would virtually only have about 15 compounds in the language (i.e. mainly the established compounds), because only the established compounds arguably adhere to all of the proposed criteria of compoundhood in the checklist (cf. 5.3.1.4 (4) on established compounds). Perhaps a better way to look at this matter is to divide the proposed criteria into primary and secondary tests. Obviously, a given combination needs to fulfil certain criteria in order for it to be considered as a compound to begin with. To this end, we suggest that the syntactic tests of insertion, modification and switching, remain the primary test of Malay compoundhood. This suggestion is based on the effectiveness of these tests as proof of compoundhood as seen throughout the analysis. Complying with them would indicate that the given combination is essentially a compounded form. On the other hand, the remaining tests of reduplication, circumfixation, stress patterning and assimilation are administered as a secondary test. This is based on the fact that these tests are less applicable (and can be highly specific or isolating occurrences) tests of compoundhood. If a given combination also fulfils the secondary test, it is therefore further asserting its status as a compounded structure.

In terms of lexicalisation of compounds, the general idea held throughout the analysis on whether a given compound is compositional or lexicalised is that it is dependent on the transparency and/or opaqueness of the components. In other words, a compound with transparent components is more likely to be compositional (hence non-lexicalised), in

comparison to a compound with non-transparent components (hence lexicalised). However, one must understand that it is not the case that transparent compounds cannot be lexicalised. It is perhaps better to understand this as: compounds with transparent components can be composed without necessarily having the need to be lexicalised (cf. the attributive compound types). On the other hand, it is necessary for compounds with opaque components to be lexicalised because the intended meaning of the compounds will never be understood through the meaning of the components (cf. the idiomatic compound types).[137]

The final concern of the checklist is to establish the relationship between the compounded components. In general, the most common type of relationship visible from the data is the attributive (ATR) type. In this type of compound, the non-head component acts as a modifier, further expressing a property (or quality) of the head component. This type of relation is apparent in compounds with endocentric headedness, i.e. the (NN) *jam tangan*, (NV) *padang tembak*, (NA) *kerusi malas*, (NN-right head) *bumiputra*, (AN-right head), *perdana menteri* (VN) *bawa diri*, (VV) *cari gali*, (AA) *merah muda*, (AN) *biru laut*, and the (VA) *temu ramah* types.

The other relationship which is apparent in this study is the coordinative (CDR) type. In this type of compound, both components are important in terms of contributing to the whole meaning of the compound. The (CRD) compounds are also mostly of the endocentric types, i.e. the (NN) *ibu bapa*, (AA) *sukarela* and (VV) *campur aduk* types, with one example of the exocentric type, i.e. (NN) the *kakitangan* compound. In this study, the term coordinative is used in a broad sense which includes both synonymous coordinative compounding and antonymous coordinative compounding. For instance, the *ibu bapa* (mother + father) types

---

[137] Nevertheless, this study still (at best) assumes the degree of composition and/or lexicalisation of the analysed compounds. Perhaps a much more specific test is needed if one is to be certain of the lexicalisation status of a given compound.

have antonymous components, while the *campur aduk* (to mix + to mix) types have synonymous components. The *kakitangan* (foot + hand) compound is an example of a synonymous coordinative exocentric compound.

The coordinative relation (hence coordinative compounding) is not properly acknowledged in the language. Based on the analysis of this study, we strongly disagree with the idea that Malay does not have coordinative compounding. As a matter of fact, this type of compounding is ever-present in the so-called established compounds (i.e. *kakitangan*, *tanggungjawab*, *sukarela*). Given that the established compounds have long been held as the epitome of Malay compounding, it is perplexing that Malay scholars have denied the existence of coordinative compounding in the language.

The least common relationship that can be seen in this study is the subordinate (SUB) relation. As mentioned, the constituents of subordinate compounds have the grammatical relations of 'complementation', i.e. head-complement relation. In addition to synthetic compounds like *truck driver* having the head-complement relation (i.e. *truck* is the complement of the deverbal head *driver*), Bissetto and Scalise (2005) also consider compounds such as *apron string* and *cat food* as having a head-complement relation as well. They explain that the argument relation between the components is related to the semantic *R*-relation/argument (as mentioned in the [X *R* Y] schema), where the *R* is seen as linking the X and Y through a grammatical or semantical relationship. In this case, the semantic relation between *apron* and *string* is that of an 'of' relation (i.e. string of apron), and the relation between *cat* and *food* is that of a 'for' relation (i.e. food for cats). Thus, *string* and *food* are considered as the head components taking *apron* and *cat* as their complements respectively.

Lieber (2009) disagrees with this view, arguing that compounds like *apron sting* and *cat food* have more of an attributive relation than a subordinative one. This is because there is

not enough argument structure in the compound to be a head-complement structure. Spencer (2011) also finds it difficult to clearly distinguish between a compound with a head-complement relation and a compound with a head-adjunct relation (through Bissetto and Scalise's (2005) perspective). If we accept Bissetto and Scalise's (2005) terms of subordinate relation, this study would have to re-evaluate compounds such as *jam tangan* 'wristwatch' as having a head-complement relation. In other words, we would have to accept that the head noun *jam* (watch) takes the non-head noun *tangan* (hand) as its complement (i.e. *jam untuk tangan* 'watch for hand'). This, however, is obviously an inaccurate analysis, as it is difficult to understand how *tangan* can be the complement of *jam*.

In general, this study has shown that most Malay compounds do not bear the head-complement relation between their elements (i.e. the head-modifier relation is more prominent in general). Nevertheless, the corpus does provide us with very limited examples of subordinative compounding, i.e. the *sapu tangan* 'handkerchief' types. In this type of compounding, the semantic head verb *sapu* (to wipe) can be argued as taking the noun component *tangan* (hands) as its complement (cf. 5.3.1.4 (1) (i)). This is the only evidence that is available from the analysis of a subordinative relation within Malay compounding.

On the other hand, we also find it difficult to best represent the relationship between the components of idiomatic/metaphoric compounds, i.e. the *tangkai jering* types. The question is, can their components truly be assigned to have either the subordinate, attributive or coordinative relationship? This uncertainty is partly caused by the fact that they are essentially headless compounds. This lack of headedness makes it difficult to assign a particular relationship between the two components. Although one can argue that the combination of *tangkai jering* can have an attributive relation in the sense that the word *jering* (jering fruit) is specifying the type of *tangkai* (stalk), i.e. 'jering fruit stalk', assigning this sort of relation means that the combination is read literally. This is similar to other compositional *tangkai*

constructions, e.g. *tangkai rambutan* is literally 'rambutan stalk', *tangkai mangga* is literally 'mango stalk', etc. An attributive relation does not hold between the components of *tangkai jering* when it is in a compounded context (i.e. to mean a miser). The compound *tangkai jering* must necessarily be lexicalised as a whole because a compositional reading of the combination will never bring about the intended meaning of 'a miser'. In this sense, we cannot truly say that *tangkai jering* (as a compound) can have an attributive relationship between its components. As such, this study suggests that these types of compounds need to be separated, i.e. to be categorised as idiomatic or metaphoric compounds without having any affiliation to the subordinative, attributive and coordinative distinction.

Clearly, the analysis (guided by the Malay compoundhood checklist) has managed to identify and separate the many different types of Malay compounds, as discussed. The analysis has also allowed this study to establish several core concepts and characteristics of Malay compounding, as discussed above. The outcome of the analysis and the study as a whole has definitely further clarified and contributed to a better understanding of the Malay compounding phenomenon.[138]

---

[138] It is crucial that we also mention some of the shortcomings of the analysis. The analysis did not take into account several types of combinations that can be found in the corpus due to several reasons. Among them is the fact that some combinations were inaccurately combined or simply combinations that are non-existent in the language, e.g. *arah pandu* (direction + drive) should be *pandu arah* (navigation), *denyut dada* (throb + chest) should be *denyut jantung* (heartbeat), and *beri diri* (to give + self) probably should be *serah diri* (surrender). Another reason for dismissing some of the combinations is because the type or class of the combined components is not under the consideration of this study. Among others, the combination of *ayah tiri* (father + step-) and *anak tiri* (child + step-) is not analysed because the component *tiri* (step-) is more of an exclusive combining form. Similarly, in the combination of *gotong royong* (cooperate + royong), the form *royong* is specifically bound only to this particular combination, making the combination more likely a rhyming reduplication. The term *anakanda* (child + you/your) is an archaic fossilised pronoun usually used within the high and royal settings of the language. Other combinations such as *ikat mati* (to tie + die), *ikut dekat* (to follow + near), *usaha sama* (effort + same) and *kemas kini* (to tidy + now) were also left out of the analysis because they consist of an adverbial component (which is not under the consideration of this study). This study also did not analyse combinations containing derived components, examples of which are given here with the affixations in bold: *salur**an** runcit* (channel + retail), *graf **ter**arah* (graph + directed), *medan sinar**an*** (field + radiation), *mata **pel**ajar**an*** (focus + studies), *tunjuk **per**asaan* (show + feelings) and *jambat**an** cerucuk* (bridge + pile). We believe that these combinations are beyond the scope of this study, which definitely needs a separate focus in order to analyse these issues.

Chapter 6: Summary

In this final chapter, we present a chapter-by-chapter summary of this thesis.

This study opened its first chapter by presenting some of the many reasons why compounding is a universally interesting and important linguistic phenomenon. The focus of this study is of course on Malay compounding. A quick history of the Malay language was first introduced in terms of its regional influence in South East Asia, and the two most influential varieties of Malay, namely *Bahasa Melayu* (Malaysian Malay) and *Bahasa Indonesia* (Indonesian Malay), were distinguished. It was important to do this as the two varieties are not exactly mutually intelligible, especially in terms of their vocabularies. This study thus dismisses Indonesian Malay compounding, focusing only on the Malaysian Malay compounding phenomenon. The first chapter ended by presenting the objective and aims of this study, i.e. to (i) identify the pertinent issues in relation to the concept of Malay compounding, (ii) establish an organised approach to account for Malay compounding and (iii) recognise all the possible constructs that can be qualified as compounds in Malay.

Chapter 2 is divided into two sections, one concerning the issues of complex word formation and the other concentrating on the universals of compounding. In the first section, the question of where in the grammar does complex word formation take place was entertained. The issue is whether complex words are handled by the morphological module or by the syntactical module of the grammar. We looked at some arguments on how syntax has been justified as being able to represent and handle morphological formations. Thus, a separate morphological module for handling word formation was deemed redundant, uneconomical and unnecessary, i.e. morphology should simply be subsumed under syntax. However, this claim is problematic. The basic principles of morphology are distinct to those of syntactical

principles, and some properties are exclusive only to morphology. Arguments were given in support of morphology as an autonomous creative and generative module in the grammar (the lexicalist approach, among others). We agree that it is better to assume both morphology and syntax as independent modules with the ability to interact with each other. To this end, the study adopted Ackema and Neeleman's (2004) competition model, which supports the idea that morphology and syntax are equally capable of generating complex structures. We have argued that this model can sufficiently account for complex word formation such as that of compounding.

In the second part of the chapter, we looked at the universals of compounding. On the topic of definition, we addressed the challenges of coming up with a universally acceptable definition of compounding. This is also related to the issue of the types of components that are suitable to be used as the building blocks of compounds. The problem lies in the fact that defining and identifying the components of compounds is highly language-specific. To counter this problem, this study took into account Guevara and Scalise's (2009) proposal of the [X *R* Y] (X)(Y)(Z) structure as the prototypical compounding schema, as it is broad enough to function as a universal compounding schema.

The next topic examined was headedness. We explained that there are various head positions within compounds, i.e. the left, right, headless, and double-headed compounds. However, identifying the head is not the simple act of observing the position of an item inside a given compound. First, we looked at how the head is the component responsible for the overall category of compounds. Next, we discussed how the head is the core meaning of the whole compound. Finally, we looked at how the head is the grammatical locus of compounds. These can be translated as the features or properties of headedness, which are not necessarily transferred from one single component. The headedness features in compounding must be

understood as independent from one another, i.e. a tripartite definition of the categorical, semantic and morphological heads.

On the topic of compoundhood criteria, we looked at the matters of orthographic, phonological, morphological, syntactic and semantic criteria of compounding. It has been argued that a compound should abide by these criteria in order to be considered a compound. However, as we have seen, some criteria are more relevant than others, while some are simply language-specific. However, some are general enough to be considered as universally acceptable criteria of compoundhood. In essence, it is insufficient to rely on one single criterion, nor is it practical to account for all of the criteria, only for a given combination to be accepted as a compounded form. The appropriate criteria for compoundhood are thus language-specific.

The final discussion on compound classification focused on Bisetto and Scalise (2005) and Scalise and Bisetto's (2009) suggestion of classifying compounds under the headings of subordinative, attributive and coordinative compounding. The headings represent the correlation between the components of a given compound. Classification of this manner has been argued as being able to universally represent the many types of different compounds in the world's languages.

Chapter 3 began with an introduction to the orthography and sound system of the Malay language. This provided the basics for the structure, spelling and pronunciation of Malay words. The next focus was on the processes of affixation and reduplication of the three major lexical categories in the language, i.e. nouns, verbs and adjectives. We first looked at the different types of affixation (i.e. prefixes, suffixes, infixes and circumfixes), their meaning and their grammatical functions in relation to the lexical roots/words. We then looked at the different types of reduplication process (i.e. the full, partial and rhyming), along with their

meaning and grammatical functions. The chapter closed by bringing together both processes, affixation and reduplication, in relation to the process of compounding.

In Chapter 4, we reviewed eight works/studies of Malay compounding in the literature, discussing and commenting on each and every one of them respectively. Although the topics of discussion vary from one linguist to another, there are nevertheless certain recurring ones seen throughout. Among others, discussions on Malay compounding generally commence with a definition of the subject matter. This usually concerns the types of compounding elements and how the combinations of these elements are seen as unified entities. Another topic of discussion is the headedness of compounding, i.e. Malay compounds are commonly recognised as having a head + modifier relation. This also relates to the issue of compounds being either endocentric or exocentric types. Yet another common concern is the criteria of compounding. Some linguists focus on the orthographic criteria, while others are more concerned with the syntactic or semantic ones. Finally, Malay linguists usually discuss the types of compounds, categorising them as free forms (nominal, verbal, adjectival types) and idiomatic compounds.

A table summary of the compounding topics that have been presented/discussed by the respective linguists can be found in section 4.9. We highlight here some of the prominent topics raised by each linguist. In Hassan's (1974, 1986, 2006) works, some of the main discussion includes the issue of headedness (endocentric/exocentric distinction), compounds as single unified entities (disallowing separation between the components) and testing compoundhood via reduplication (i.e. compounds should reduplicate as a whole). Hassan (1986) is also one of the few who acknowledge that compounding in Malay is an unclear phenomenon, which is a controversial idea. Musa (1993) also discusses the concept of head-modifier relation, hence recognising the endocentric/exocentric distinction. However, more importantly, he puts forward the idea that, in terms of meaning, compounds are structures that are in between structures with a literal meaning and structures with a totally opaque meaning. In this sense, he

is able to accept that a given combined structure can be both a compound and a phrasal structure. Musa (1993) also believes that compounding in Malay is perhaps not clearly understood as a morphological phenomenon in the language. The work of Karim (1995) presents some basics on the topic of compound definition, types of compounds and spelling convention. Similarly, Karim et al. (2008) also describe Malay compounding through the matters of definition, compound types (free forms, established forms, specific terms and idiomatic forms), spelling convention (effects of affixation and reduplication) and compounds vs. phrases (via insertion test). Both Karim (1995) and Karim et al. (2008) put forward the idea of testing compoundhood via circumfixation (i.e. components of a compound must merge upon circumfixation). Similarly, they also maintained the idea that in general only the head (i.e. left-hand) component of a compound should undergo reduplication. Sew's (2007) work is interesting in the sense that he recognises more than just the head-modifier and endocentric/exocentric distinction, also recognising the idea of appositional and coordinate compounds in Malay. This is one of the few works acknowledging that the language might have more than just the typical free, established and idiomatic forms of compounding. Lastly, Ismail and Jalaluddin's (2008) study attempts to identify compounded structures through a five-point criteria test of compoundhood, i.e. components cannot be separated, components cannot switch places, compounds should not be able to further generate other compounds alike, compounds should be able to be circumfixed, and the meaning of a compound must be entirely different from the meaning of its components. As mentioned, we have discussed and commented on the merits and the weakness of each of these works/studies respectively. The aim was to draw an overview of how the compounding phenomenon has been addressed and understood by different Malay linguists over the years.

Chapter 5 is divided into three subsections. The first two revisited the previous chapters, to reiterate the important issues and topics that were needed to build the foundations of Malay compoundhood and which were used as analysis tools in the last section.

The initial concern questioned whether compounds in Malay are the products of morphological merger or of syntactic merger. This question can be difficult to answer, given that compounds and phrases in Malay are structurally identical objects. The principles of the competition model argue that both morphological and syntactical modules will compete to generate the merger. The preference, however, will be for the unmarked structure, and we have argued that syntax is the unmarked structure in Malay. However, the module also argues that compounds can occur when there is no syntactic competitor. For instance, if the syntactic merger expresses transparent semantics, the generating of the morphological merger will be blocked. However, if the morphological merger expresses semantics that cannot be expressed by the syntactic merger, then the morphological merger can be generated as a compound. To this end, the competition model is accountable enough to explain compounding as a distinct object from that of a structurally identical phrasal object in the language. However, distinguishing Malay compounds only on semantic grounds can be misleading, and thus a proper definition of compounding on formal grounds was still needed, which was the undertaking of the subsequent section.

Accordingly, the discussion in the second subsection concerned the issues of definition, components, headedness, criteria and classification of Malay compoundhood. With regard to definition and components of Malay compounding, we came to agree that Malay compounding can take roots, bases, morphemes and words as the primary units of compounding, with the prototypical structures of [X R Y] (X)(Y) for endocentric and [X R Y] (Z) for exocentric compounds. On the issue of heads, we observed that the compounds can have three head positions, i.e. the default left-head, headless compounds and some right-head compounds. We

agreed that the relevant concept of headedness in Malay compounding are the categorical head (i.e. the component that projects its category onto the overall compound) and the semantic head (i.e. the head is the hyponym of the compound). On the topic of criteria, we have seen how the degree of relevance of a criterion varies from language to language, and this is no exception in Malay. We came to agree that the suitable criteria for Malay compoundhood are the syntactical criteria (inseparability, modification, component switching, circumfixation and reduplication), phonological criteria (stress and assimilation) and semantic criteria (lexicalisation). Finally, we considered Bisetto and Scalise's (2005) classification of subordinate, attributive and coordinate compounding as sufficient in classifying the types of Malay compounds in this study. All of the information and discussions gained up to this point of the study enabled us to devise a checklist of Malay compoundhood in 5.2.5 (47), reiterated below.

Malay compoundhood checklist

	Malay compoundhood	Comments
1.	Definition - components of compounding: - roots, bases, morphemes and words	Prototypical compounding schema for Malay: [X R Y] (X)(Y)(Z). Where, X and Y are the components, R represents the relationship, and (X), (Y) and (Z) represents the overall head/category.

2.	Headedness
	
	- Categorical head
	
	    (The head is the part that determines
	
	    the whole category of the compound)
		Identify the head of a given compound, distinguish between

	- Semantic head (The head is the hyponym of the compound)	endocentric and exocentric compounds.

3.	Criteria
	
	- Syntactic criteria
	
	    - insertion/modification
	
	    - inability to switch constituents
	
	    - circumfixation/reduplication
	
	- Semantic criteria (when applicable)
	
	    - lexicalisation
	
	- Phonological criteria (when applicable)
	
	    - stress and/or assimilation
		Use these tests to identify whether a given concatenated structure can be considered as a compounded structure.

4.	Classification - Subordinative (SUB) - Attributive (ATR) - Coordinative (CRD)	Classify compounds according to their types in accordance with Bisetto and Scalise's (2005) compound classification.

The final part of the chapter utilised the compoundhood checklist to analyse the combined structures (exemplified as compounds) in the selected Malay works and studies. An overall claim can be made from the outcome of the analysis. In general, it is evident that the most common compounding structure in the corpus is the [X Y] (X) structure, followed by the [X Y] (Z), with the [X Y] (Y) being the least common. The most common type of headedness is the endocentric type, where both categorical and semantic heads are projected from the same head component within the compound. The test of modification, test of insertion and test of

switching seems to be the most applicable and reliable test of Malay compoundhood. On the other hand, the test of reduplication, test of circumfixation, test of stress patterning and test of assimilation seems to be applicable only to a limited degree, as they are selective and restrictive test of Malay compoundhood. To this end, a division of the criteria into primary and secondary tests has been suggested. The analysis has also shown how the compositional or lexicalised status of Malay compounds is not a clear-cut issue as it is dependent on the transparency and/or opaqueness of the components.

Finally, through the assistance of the checklist, we managed to clarify and identify the different types of compounds available from the corpus. Among the major ones are the NN, NV and NA nominal compounds, the VN, VV and VA verbal compounds, the AN and AA adjectival compounds, and also idiomatic compounds. In terms of the relationship, the most common ones available from the corpus are the attributive (ATR) endocentric types. There is also evidence of the coordinative (CRD) types of compounding, which have been denied recognition in Malay literature. There are only a few limited examples of the subordinate (SUB), which might suggest that Malay compounds do not favour the head-complement relationship. In addition to these three categories of compounding, this study also suggests that it is perhaps best that idiomatic compounding in Malay be categorised separately from the ATR, CRD and SUB types of compounds. The summary list of the different types of compounds available from the corpus is detailed in 5.3.4 (70). The simplified diagrammatic illustration presented in 5.3.4 (71) is reiterated below.

Diagrammatic representation of the types of Malay compounds

Malay compounds

Subordinative   Attributive   Coordinative   Idiomatic compounds

Exo.   Endo.   Endo.   Exo.   Exo.

(VN) *sapu tangan*

(NN) *jam tangan*

(NV) *padang tembak*

(NA) *kerusi malas*

(VN) *bawa diri*

(VV) *cari gali*

(VA) *temu ramah*

(AN) *biru laut*

(AA) *merah muda*

(NN) *ibu bapa*

(VV) *tunjuk ajar*

(AA) *sukarela*

(NN) *kakitangan*

Right-headed:

(NN) *bumiputra*

(AN) *perdana menteri*

(NN) *duit kopi*
(partially idiomatic)

Nominal:

(NN) *tangkai jering.*
(VN) *hirup darah.*

Verbal:

(VN) *tumbuk rusuk.*

Adjectival:

(AN) *berat tulang*

In the course of this study, some topics were only briefly touched upon or left otherwise unresolved. Case in point, the concept of synthetic compounding in Malay definitely deserves further research from the introductory proposal that we have made here. Yet another concern is that this study only looked at the compounding of nominal, verbal and adjectival components. An analysis of other lexical categories must be taken into consideration in the future. Also, as it is beyond the scope of this study, a separate focus on compounds with derived components and compounds with more than two word components is needed in future research. Additionally, a further study is perhaps needed to provide a more detail account (or an alternative account) of the idiomatic type of compounding in the language that we have mentioned. These are a few topics born out of this study that are definitely worthy of future research.

As a final remark, we believe that this study has managed to achieve what it was set out to do. It has managed to identify the important issues concerning Malay compounding, managed to resolve them by coming up with a better-organised conception of Malay compounding, and also managed to identify and categorise the different types of compounds in the Malay language. With this, we believe that this study has contributed to a more systematic and comprehensive understanding of the topic in Malay linguistics specifically, and enriches the larger body of compounding knowledge generally.

References

Ackema, P., & Neeleman, A. (2004). *Beyond morphology: Interface condition on word formation*. Oxford: Oxford University Press.

Ahmad, Z. A. (1965). *Pelita bahasa Melayu. penggal 1, permulaan nahu Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Anderson, S. R. (1988). Morphological theory. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey: Volume 1, linguistic theory: Foundations* (pp. 146-191). Cambridge: Cambridge University Press.

Anderson, S. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, Massachusetts: MIT Press.

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. Cambridge, Massachusetts: MIT Press.

Aronoff, M. (2012). Morphological stems: what William of Ockham really said. *Word Structure, 5*(1), 28-51.

Aronoff, M., & Fudeman, K. (2005). *What is morphology?*. Oxford: Blackwell Publishing.

Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.

Bauer, L. (1988). *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.

Bauer, L. (1998). When is a sequence of two nouns a compound in English?. *English Language and Linguistics*, *2*(1), 65-86.

Bauer, L. (2000). Word. In G. Booij, C. Lehmann & J. Mugdan (Eds.), *Morphology: An international handbook on inflection and word-formation* (pp. 247-257). Berlin and New York: Walter de Gruyter.

Bauer, L. (2001). Compounding. In M. Haspelmath, E. Konig, W. Oesterreicher & W. Raible (Eds.), *Language typology and language universals* (pp. 695-707). Berlin and New York: Walter De Gruyter.

Bauer, L. (2006). Compound. In K. Brown (Ed.), *Encyclopaedia of language and linguistics, vol. 2* (pp. 719-726). Amsterdam: Elsevier.

Berent, I., & Pinker, S. (2007). The dislike of regular plurals in compounds: Phonological familiarity or morphological constraint?. *The Mental Lexicon, 2*, 129-181.

Bisetto, A., & Scalise, S. (2005). Classification of compounds. *Lingue e Linguaggio*, *2*, 319-332.

Booij, G. (2005). *The grammar of words. An introduction to linguistics morphology (first edition)*. Oxford: Oxford University Press.

Booij, G. (2010). Morphological analysis. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 507-529). Oxford: Oxford University Press.

Booij, G. (2012). *The grammar of words. An introduction to linguistics morphology (third edition)*. Oxford: Oxford University Press.

Borer, H. (1998). Morphology and syntax. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 151-190). Oxford: Blackwell.

Botha, R. (1983). *Morphological mechanisms*. Oxford: Pergamon Press.

Bybee, J. (1995). Diachronic and typological properties of morphology and their implications for representation. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 225-246). New Jersey: Lawrence Erlbaum Associates.

Ceccagno, A., & Basciano, B. (2007). Compound headedness in Chinese: An analysis of neologisms. *Morphology*, *17*(2), 207-231.

Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184-221). Waltham: Blaisdell.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*(6), 991-1013.

Clark, E.V., Gleman, S. A., & Lane, N. M. (1985). Compound nouns and category structure in yong children. *Child Development*, *56*, 84-94.

Clark, R. (2009). Austronesian languages. In B. Comrie (Ed.) *The world's major languages* (pp. 781-790). London: Routledge.

Coates, R. (1999). *Word structure*. London: Routledge.

Cunnings, I., & Clahsen, H. (2007). The time-course of morphological constraints: Evidence from eye-movements during reading. *Cognition*, *104*, 476-494.

Donalies, E. (2004). *Grammatik des Deutschen im europäischen vergleich: Kombinatorische begriffsbildung, teil 1: Substantivkomposition*. Mannheim: Institut für Deutsche Sprache.

Dressler, W. U. (2006). Compound types. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound word* (pp. 23-44). Oxford: Oxford University Press.

Embick, D., & Noyer, R. (2007). Distributed morphology and the syntax-morphology interface. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (pp. 289-324). Oxford: Oxford University Press.

Fabb, N. (1998). Compounding. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 66-83). Oxford: Blackwell.

Fabregas, A., & Scalise, S. (2012). *Morphology: From data to theories*. Edinburgh: Edinburgh University Press.

Guevara, E., & Scalise, S. (2009). Searching for universals in compounding. In S. Scalise, E. Magni & A. Bisetto (Eds.), *Universals of language today* (pp. 101-128). Netherlands: Springer.

Halim, A. (1974). *Intonation in relation to syntax in bahasa Indonesia Jakarta*. Djambatan.

Halle, M. (1997). Distributed morphology: Impoverishment and fission. In B. Bruening, Y. Kang & M. McGinnis (Eds.), *MITWPL 30: Papers at the interface* (pp. 425-449). Cambridge, Massachusetts: MITWPL.

Halle, M., & Marantz, A. (1994). Some key features of Distributed Morphology. In A. Carnie & H. Harley (Eds.), *MITWPL 21: Papers on phonology and morphology* (pp. 275-288). Cambridge, Massachusetts: MITWPL.

Harahap, D. (1991) *Persoalan bahasa*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Harley, H. (2009). Compounding in Distributed Morphology. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 129-144). Oxford: Oxford University Press.

Harley, H., & Noyer, R. (1999). State-of-the-article: Distributed morphology. *Glot International*, *4*(4), 3-9.

Haskell, T., MacDonald, M., & Seidenberg, M. (2003). Language learning and innateness: Some implications of compounds research. *Cognitive Psychology*, *47*, 119-163.

Haspelmath, M., & Sims, A. D. (2010). *Understanding morphology*. London: Hodder Education.

Hassan, A. (1974). *The morphology of Malay*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Hassan, A. (1986). *Penerbitan kata dalam bahasa Malaysia*. Petaling Jaya: Penerbit Fajar Bakti.

Hassan, A. (1987). *Isu-isu perancangan bahasa: Pengintelektualan bahasa Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Hassan, A. (1997). *Bahasa Melayu sebagai bahasa ilmu dan esei-esei lain*. Shah Alam: Penerbit Fajar Bakti.

Hassan, A. (2006). *Morfologi*. Kuala Lumpur: PTS Professional Publishing.

Heah, L. H. C. (1989). *The influence of English on the lexical expansion of bahasa Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Hoeksema, J. (1992). The head parameter in morphology and syntax. *Language and Cognition*, *2*, 119-132.

Ismail, Z., & Jalaluddin, N. H. (2008). Kata majmuk dalam perkamusan: Satu analisis semantik dan pragmatik. In N. H. Jalaluddin & N. Baharom (Eds.), *Leksikologi dan leksikografi Melayu* (pp. 335-371). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Jackendoff, R. (2002) *Foundations of knowledge*. Oxford: Oxford University Press.

Jackendoff, R. (2009). Compounding in the Parallel Architecture and Conceptual Semantics. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 105-128). Oxford: Oxford University Press.

Jalaluddin, N. H., & Syah, A. H. (2009). Penelitian makna imbuhan peN- dalam bahasa Melayu: Satu kajian rangka rujuk silang. *GEMA Online Journal of Language Studies*, *9*(2), 57-72.

Karim, N. S. (1978). *Bahasa Malaysia syntax – some aspects of its standardization*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Karim, N. S. (1995) *Malay grammar for academics and professionals*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Karim, N. S. (2010). *Panorama bahasa Melayu sepanjang zaman*. Kuala Lumpur: Penerbit Universiti Malaya.

Karim, N. S., Onn, F. M., Musa, H. H., & Mahmood, A. H. (2008). *Tatabahasa Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Katamba, F. (1993). *Morphology*. London: MacMillan Press.

Katamba, F., & Stonham, J. (2006). *Morphology*. London: Palgrave MacMillan.

Kornfeld, L. M. (2009). IE, Romance: Spanish. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 436-452). Oxford: Oxford University Press.

Lees, R. B. (1960). *The grammar of English nominalization*. Bloomington: Indiana University Press.

Libben, G. (2006). Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1-22). Oxford: Oxford University Press.

Lieber, R. (2004). *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Lieber, R. (2009). A lexical semantic approach to compounding. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 78-104). Oxford: Oxford University Press.

Lieber, R. (2010). *Introducing morphology*. Cambridge: Cambridge University Press.

Lieber, R., & Scalise, S. (2006). The lexical integrity hypothesis in a new theoretical Universe. *Lingue e Linguaggio*, *6*, 7-32.

Lieber, R., & Stekauer, P. (2009). Introduction: Status and definition of compounding. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 3-18). Oxford: Oxford University Press.

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics*, *4*(2), 201-225.

Maris, Y. (1980). *The Malay sound system*. Kuala Lumpur: Penerbit Fajar Bakti.

Matthews, P. H. (1974). *Morphology: An introduction to the theory of word-structure*. Cambridge: Cambridge University Press.

McIntyre, A. (2014). Constraining argument structure in nominalizations: The case of English-er. *Lingua*, *141*, 121-138.

Md Salleh, R. (1993). Sistem pengimbuhan bahasa Melayu. *Jurnal Dewan Bahasa*, 921-927.

Mintz, M. W. (1994). *A student's grammar to Malay and Indonesian*. Singapore: EPB Publishers.

Mohd Don, Z., Knowles, G., & Yong, J. (2008). How words can be misleading: A study of syllable timing and "stress" in Malay. *The Linguistics Journal*, *3*(2), 66-81.

Montermini, F. (2010). Units in compounding. In S. Scalise & I. Vogel (Eds.), *Cross-disciplinary issues in compounding* (pp. 77-92). Amsterdam: John Benjamins.

Musa, H. (1993). *Binaan dan fungsi perkataan dalam bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (1975). *Essays on Malaysian linguistics*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (1980). *Nahu Melayu mutakhir*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (1988). *Susur galur bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (1991). *Kepelbagaian fonologi dialek-dialek Malayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (1992a). *The linguistic scenery in Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (1992b). Malay as a pluricentric language. In M. Clyne (Ed.), *Pluricentric languages: Differing norms in different nations* (pp. 401-420). Berlin: Mouton de Gruyter.

Omar, A. H. (1993). *Language and society in Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Omar, A. H. (2010). *Kajian dan perkembangan bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Onn, F. M. (1980). *Aspects of Malay phonology and morphology: A generative approach*. Kuala Lumpur: Universiti Kebangsaan Malaysia.

Onn, F. M. (2014). *Saya dan fonologi bahasa Melayu: Mendekati tatabahasa sejagat*. Kuala Lumpur: Universiti Kebangsaan Malaysia.

Othman, A. (1984). *Tatabahasa bahasa Malaysia*. Kuala Lumpur: Penerbitan Sarjana.

Pesetsky, D. (1985). Morphology and logical form. *Linguistic Inquiry*, *16*(2), 193-246.

Pinker, S. (1994). *The language instinct*. London: Penguin Books.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456-463.

Plag, I. (2003). *Word formation in English.* Cambridge: Cambridge University Press.

Prentice, D. J. (1987). Malay (Indonesian and Malaysian). In B. Comrie (Ed.), *The world's major languages* (pp. 913-935). London: Croom Helm.

Ralli, A. (2009). IE, Hellenic: Modern Greek. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 453-463). Oxford: Oxford University Press.

Rosenberg, M. (2007). Classification, headedness and pluralization: Corpus evidence from French compounds. *Acta Linguistica Hungarica, 54*(3), 341–360.

Rubino, C. (2005). Reduplication: Form, function and distribution. In H. Bernhard (Ed.), *Studies on Reduplication* (pp. 11-30). Berlin: Mouton de Gruyter.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Foundations, vol. 1*. Cambridge, Massachusetts: MIT Press.

Scalise, S., & Bisetto, A. (2009). The classification of compounds. In R. Lieber & P. Stekauer (Eds.), *The Oxford handbook of compounding* (pp. 34-53). Oxford: Oxford University Press.

Scalise, S., & Fabregas, A. (2010). The head in compounding. In S. Scalise & I. Vogel (Eds.), *Cross-disciplinary issues in compounding* (pp 109-126). Amsterdam: John Benjamins.

Scalise, S., & Guevara, E. (2005) The lexicalist approach to word-formation and the notion of the lexicon. In P. Stekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp.174-187)*. The Netherlands: Springer.

Scalise, S., & Vogel, I. (2010). Why compounding?. In S. Scalise & I. Vogel (Eds.), *Cross-disciplinary issues in compounding* (pp. 1-18). Amsterdam: John Benjamins.

Selkirk, E. (1982). *The syntax of words*. Cambridge, Massachusetts: MIT Press.

Sew, J. W. (2007). *Reduplicating nouns and verbs in Malay: A conceptual analysis*. Kuala Lumpur: University Malaya Press.

Spencer, A. (1988). Bracketing paradoxes and the English lexicon. *Language, 64*(4), 663-682.

Spencer, A. (1991). Morphological theory. Oxford: Blackwell.

Spencer, A. (2003). Does English have productive compounding. In G. Booij, J. De Cesaris, A. Ralli, & S. Scalise (Eds.), *Topics in morphology: Selected papers from the third Mediterranean morphology meeting* (pp. 329- 341). Barcelona: IULA.

Spencer, A. (2005). Word formation and syntax. In P. Stekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp. 73-98). Netherlands: Springer.

Spencer, A. (2006). Morphological universals. In R. Mairal & J. Gil (Eds.), *Linguistics universals* (pp. 101-129). Cambridge: Cambridge University Press.

Spencer, A. (2011). What's in a compound?. *Journal of Linguistics*, *47*, 481-507.

Syed Jamaluddin, S. Z. (2005). *The evolution of Malay affixes*. Pulau Pinang: Penerbit Universiti Sains Malaysia.

Tadmor, U. (2009). Malay-Indonesian. In B. Comrie (Ed.), *The world's major languages* (pp. 791-818). London and New York: Routledge.

Taft, M. (1988). A morphological-decomposition model of lexical representation. *Linguistics*, *26*, 657-667.

Ten Hacken, P. (1994*). Defining morphology: A principled approach to determining the boundaries of compounding, derivation, and inflection*. Hildesheim: Georg Olms Verlag.

Van Zanten, E., & Van Heuven, V. J. (1998). Word-level stress in Indonesian: Its communicative relevance. *Bijdragen tot de Taal-, Land-, en Volkenkunde*, *154*(1), 129-144.

Williams, E. (1981). On the Notions "Lexically Related" and "Head of a Word". *Linguistic Inquiry*, *12*(2), 245-274.

Appendices

The combinations considered as compounds by the respective linguists (in their respective work/study) are listed below.

Appendix (1)

Abdullah Hassan's (1974) *The morphology of Malay.*

Hassan categorises the compounds in his work according to the lexical categories (nominal, verbal, adjectival) and the endocentric/exocentric division.

(i) Nominal (endocentric)

No.	Combination/Compound	Meaning
1	*abu rokok* (ash + cigarette)	cigarette ash
2	*padang tembak* (field + shoot)	shooting range
3	*kapal terbang* (ship + fly)	aeroplane
4	*rumah sakit* (house + sick)	hospital
5	*kelab malam* (club + night)	night club
6	*orang luar* (person + outside)	outsider/alien
7	*anakanda* (child + you)	I/me (polite)
8	*kata dua* (word + two)	ultimatum
9	*ayah tiri* (father + step-)	stepfather
10	*baja butir* (fertiliser + granule)	granulated fertiliser
11	*menteri besar* (minister + big)	chief minister
12	*surat khabar* (letter + news)	newspaper
13	*perdana menteri* (prime + minister)	prime minister
14	*mahaguru* (great + teacher)	great teacher/professor

(ii) Nominal (exocentric)

No.	Combination/Compound	Meaning
1	*cucuk sanggul* (to pierce + hair bun)	hair pin
2	*kaji bumi* (study + earth)	geology
3	*sapu tangan* (to wipe + hand)	handkerchief

(iii) Nominal (semantically exocentric)

No.	Combination/Compound	Meaning
1	*panjang tangan* (long + hand)	likes to steal (thief)
2	*tangkai jering* (stalk + a kind of fruit)	miser

(iv) Verbal (endocentric)

No.	Combination/Compound	Meaning
1	*ikat mati* (to tie + (die) completely)	(to tie) a dead knot
2	*pandang rendah* (to look + low)	to underestimate
3	*tolak tepi* (to push + side)	to put aside
4	*ikut dekat* (to follow + near)	take a short cut
5	*temu duga* (to meet + to anticipate/guess)	interview

(v) Verbal (exocentric)

No.	Combination/Compound	Meaning
1	*hirup darah* (to suck + blood)	to extort/extortionist
2	*cuci mata* (to clean + eyes)	to look at woman (for pleasure)
3	*buang negeri* (to banish + state)	to expatriate
4	*tunjuk perasaan* (to show + feeling)	to demonstrate
5	*bawa diri* (to carry + self)	to sulk
6	*gulung tikar* (to roll + mat)	to go bankrupt

(vi) Adjectival (endocentric)

No.	Combination/Compound	Meaning
1	*biru tua* (blue + old)	dark blue
2	*biru laut* (blue + sea)	ocean blue

(vi) Adjectival (semantically exocentric)

No.	Combination/Compound	Meaning
1	*tinggi hidung* (high + nose)	arrogant
2	*berat tulang* (heavy + bone)	lazy

Appendix (2)

Hashim Musa's (1993) *Binaan dan fungsi perkataan dalam bahasa Melayu*.

Musa's work categorises the examples straightforwardly as either nominal, verbal or adjectival compounds.

(i) Nominal compounds

No.	Combination/Compound	Meaning
1	*ibu bapa* (mother + father)	parents
2	*papan hitam* (board + black)	blackboard
3	*purba kala* (ancient + period)	ancient times
4	*aneka warna* (variety + colour)	colour assortments
5	*tanah air* (land/soil + water)	homeland
6	*matahari* (eye + day)	sun
7	*kerjasa sama* (work + together)	cooperate
8	*usaha sama* (effort + together)	collaboration
9	*angka ganda* (figure + multiple)	multiplication figure
10	*alat tulis* (tool + write)	stationery
11	*hak cipta* (rights + create)	copyright
12	*gambar rajah* (picture + diagram)	diagrammatic picture
13	*jana kuasa* (to generate + power)	generator
14	*garis pusat* (line + centre)	diameter
15	*beri diri* (to give + self)	surrender
16	*timbang cara* (weigh + method)	arbitration
17	*kaji bumi* (study + earth)	geology
18	*temu duga* (to meet + to question)	interview
19	*cari gali*	explore

	(to search + to dig)	
20	*urus niaga* (to manage + business)	transaction
21	*had laju* (limit + speed)	speed limit
22	*isi padu* (content + solid)	volume
23	*urus setia* (to manage + loyal)	secretariat
24	*setia kawan* (loyal + friend)	solidarity
25	*setia usaha* (loyal + effort)	secretary
26	*air tanih* (water + soil)	soil water
27	*garis arus* (line + current)	streamlines
28	*roda tenaga* (wheel + energy)	flywheel
29	*arah pandu* (direction + to guide)	direction
30	*daya cipta* (ability + create)	creativity
31	*takat beku* (point + freeze)	freezing point
32	*had kenyal* (limit + elastic)	elastic limit
33	*denyut dada* (pulse + chest)	heartbeat
34	*cetak contoh* (to print + example)	samples
35	*tukar servis* (to change + service)	exchange service
36	*perdana menteri* (prime + minister)	prime minister
37	*naib canselor* (vice + chancellor)	vice-chancellor
38	*timbalan menteri* (deputy + minister)	deputy minister
39	*guru besar* (teacher + large/big)	head teacher
40	*lembu daging* (cow + meat)	beef producing cows
41	*ayam telur* (chicken + egg)	egg producing hens
42	*kuih kering* (cake + dry)	cookies/biscuits
43	*benda alir* (NV) (object + flow)	liquid-like (things)

(ii) Verbal compounds

No.	Combination/Compound	Meaning
1	*kahwin lari* (to marry + to run)	to elope
2	*kenal pasti* (to recognise + sure)	identify
3	*titik berat* (point + concern)	emphasise
4	*gotong royong* (cooperate + redup. form)	cooperation
5	*ambil kira* (to take + to count)	to consider
6	*ikat mati* (to tie + (die) completely)	to tie a dead knot
7	*beritahu* (to give + to know)	to tell
8	*layan diri* (to serve + self)	to serve oneself

(iii) Adjectival compounds

No.	Combination/Compound	Meaning
1	*lut sinar* (penetrating + ray)	transparent
2	*celik akal* (aware/able + mind)	able to think
3	*celik huruf* (aware/able + letters)	able to read
4	*kelabu asap* (grey + smoke)	smoky grey
5	*telap air* (permeable + water)	penetrable
6	*merah muda* (red + young)	rosy
7	*kuning tua* (yellow + old)	dark yellow
8	*merah cerah* (red + bright)	light red

Appendix (3)

Nik Safiah Karim's (1995) *Malay grammar for academics and professionals*.

Karim (1995) categorises the compounds in her book under the umbrella of nominal, verbal and adjectival compounding. The compounds are then subcategorised as either free forms, established forms, idiomatic forms, or academic/scientific terminologies.

(i) Nominal

(a) Free forms

No.	Combination/Compound	Meaning
1	*guru besar* (teacher + big/large)	head teacher
2	*tengah hari* (middle + day)	midday
3	*temu ramah* (to meet + friendly)	to have a talk
4	*surat khabar* (letter + news)	newspaper
5	*ibu bapa* (mother + father)	parents

(b) Established forms

No.	Combination/Compound	Meaning
1	*antarabangsa* (among + race)	international
2	*bumiputera* (earth + son/prince)	native
3	*jawatankuasa* (position + authority)	committee
4	*kakitangan* (foot + hand)	staff
5	*kerjasama* (work + together)	cooperation
6	*warganegara* (people + country)	citizen
7	*olahraga* (exercise + body)	athletics
8	*pesuruhjaya* (emissary + success)	commissioner
9	*setiausaha*	secretary

	(loyal + effort)	
10	*suruhanjaya* (order + success)	commission
11	*tandatangan* (mark + hand)	signature
12	*tanggungjawab* (to bear + to response)	responsibility
13	*matahari* (eye + day)	sun

(c) Academic or scientific terminologies

No.	Combination/Compound	Meaning
1	*analisis teks* (analysis + text)	text analysis
2	*keluarga asas* (family + basic)	nuclear family
3	*saluran runcit* (channel + retail)	retail outlet
4	*jambatan cerucuk* (bridge + pile)	pile bridge

(d) Idiomatic forms

No.	Combination/Compound	Meaning
1	*anak emas* (child + gold)	favourite person
2	*berat tulang* (heavy + bones)	lazy
3	*kaki botol* (leg/foot + bottle)	drunkard
4	*tumbuk rusuk* (to pound + flank)	to bribe
5	*cakar ayam* (to scratch + chicken)	scribbling

(ii) Verbal

(a) Free forms

No.	Combination/Compound	Meaning
1	*terima kasih* (receive + love)	to thank
2	*kenal pasti* (to recognise + sure)	to identify
3	*ambil alih* (to take + to remove)	to replace

4	*tunjuk ajar* (to show + to teach)	to guide
5	*campur aduk* (to mix + to mix/stir)	to mix
6	*lipat ganda* (to fold + to double)	to multiply
7	*tolak ansur* (to push + gradual)	to tolerate
8	*daya serap* (strength + absorb)	absorbance

(iii) Adjectival

(a) Free forms

No.	Combination/Compound	Meaning
1	*hijau daun* (green + leaf)	leafy green
2	*muda remaja* (young + teenage)	youth
3	*merah tua* (red + old)	crimson

(b) Idiomatic forms

No.	Combination/Compound	Meaning
1	*panjang tangan* (long + hand)	likes to steal (thief)
2	*ringan tulang* (light + bones)	hard working
3	*terang hati* (bright + heart)	easy to gain knowledge

Appendix (4)

Jyh Wee Sew's (2007) *Reduplicating nouns and verbs in Malay.*

Sew's (2007) study only exemplifies nominal compounds. The compounds are subdivided into 3 types labelled as types A, B and C respectively.

(i) Nominal Compound type A - The (NN, NV, NA) head-modifier compounds.

No.	Combination/Compound	Meaning
1	*air mata* (water + eye)	tears
2	*ibu kota* (mother + city)	capital
3	*kereta api* (car/cart + fire)	train
4	*nasi goreng* (rice + fry)	fried rice
5	*tukang masak* (artisan + cook)	chef /cook
6	*pisau cukur* (knife/blade + shave)	razor
7	*sekolah rendah* (school + low)	primary school
8	*papan hitam* (board + black)	blackboard
9	*kerusi malas* (chair + lazy)	lounge chair

(ii) Nominal Compound type B - Also known as appositional compounds.

No.	Combination/Compound	Meaning
1	*hutan rimba* (woods + forest)	large/dense jungle
2	*jiran tetangga* (neighbour + neighbour)	neighbours
3	*bala tentera* (troops + soldier)	army

(iii) Nominal Compound type C - Also known as coordinate compounds.

No.	Combination/Compound	Meaning
1	*ibu bapa* (mother + father)	parents
2	*rumah tangga* (house + stairs)	household
3	*periuk belanga* (pan+ pot)	pots and pans

Appendix (5)

Nik Safiah Karim et al. (2008) *Tatabahasa Dewan*.

Karim et al. (2008) exemplifies the compounds in their book according to the main distinction of nominal, verbal and adjectival compounds. This is followed by the subcategories of free forms, established forms, specific terms, and idiomatic forms.

(i) Nominal compounds

(a) Free forms

No.	Combination/Compound	Meaning
1	*air mata* (water + eye)	tears
2	*air hujan* (water + rain)	rain water
3	*meja makan* (table + eat)	dining table
4	*meja tulis* (table + write)	desk
5	*jam tangan* (watch + hand)	wristwatch
6	*alat tulis* (tool + write)	stationery
7	*jalan raya* (road + large/great)	main road
8	*kerusi malas* (chair + lazy)	lounge chair
9	*kerani kanan* (clerk + right)	senior clerk
10	*tunjuk ajar* (to show + to teach)	guide/advice
11	*kata kunci* (word + key)	keyword
12	*nasi minyak* (rice + oil)	oily rice
13	*bom tangan* (bomb + hand)	grenade
14	*guru besar* (teacher + large)	head teacher
15	*tengah hari* (middle + day)	midday
16	*rukun tetangga*	neighbourhood committee

	(principal + neighbour)	
17	*rumah tangga* (house + stairs)	household
18	*anak tiri* (child + step-)	stepchild
19	*ibu bapa* (mother + father)	parents
20	*rumah sakit* (house + sick)	hospital
21	*terima kasih* (to receive + love)	to thank
22	*kertas kerja* (paper + work)	paperwork
23	*urusetia* (to manage + loyal)	secretariat
24	*balai raya* (hall + large)	community hall
25	*atur cara* (to arrange + manner)	program
26	*surat khabar* (letter + news)	newspaper
27	*kapal terbang* (ship/vessel + fly)	aeroplane
28	*bandar raya* (town + large)	city

(b) Established forms

No.	Combination/Compound	Meaning
1	*bumiputera* (earth + son/prince)	native
2	*warganegara* (people + country)	citizen
3	*antarabangsa* (among + race)	international
4	*jawatankuasa* (position + authority)	committee
5	*kakitangan* (foot + hand)	staff
6	*tandatangan* (mark + hand)	signature
7	*tanggungjawab* (to bear + to response)	responsibility
8	*pesuruhjaya* (emissary + success)	commission
9	*kerjasama* (work + together)	cooperate
10	*olahraga*	athletics

	(exercises + body)	
11	*sukarela* (like + willing)	voluntariness
12	*suruhanjaya* (order + success)	commission
13	*matahari* (eye + day)	sun

(c) Specific terms

No.	Combination/Compound	Meaning
1	*analisis teks* (analysis + text)	text analysis
2	*mogok umum* (strike + general)	general strike
3	*titik buta* (spot + blind)	blind spot
4	*cari gali* (to search + to dig)	explore
5	*deria rasa* (sense + taste)	sense of taste
6	*graf terarah* (graph + directed)	directed graph
7	*hukum alamiah* (law + natural)	natural law
8	*medan sinaran* (field + radiation)	radiation field
9	*mata pelajaran* (focus/centre + studies)	subject
10	*roda tenaga* (wheel + energy)	flywheel
11	*gambar rajah* (picture + diagram)	diagrammatic picture
12	*kertas kerja* (paper + work)	paperwork
13	*darah panas* (blood + hot)	mammal
14	*model linear* (model + linear)	linear model
15	*garis pusat* (line + centre)	diameter
16	*kanta tangan* (lens + hand)	magnifier
17	*segi tiga* (angle + three)	triangle
18	*reka bentuk* (to create + shape)	design
19	*lut sinar*	transparent

	(penetrating + ray)	
20	*hak milik* (rights + belonging)	ownership
21	*pita suara* (tape + voice)	vocal cords
22	*batu kapur* (stone + lime)	lime stone
23	*kemas kini* (tidy + present)	update
24	*daya serap* (NV) (strength + absorb)	absorbance

(d) Idiomatic forms

No.	Combination/Compound	Meaning
1	*kaki ayam* (foot + chicken)	barefoot
2	*anak emas* (child + gold)	favourite person
3	*buah hati* (fruit + heart)	lover (sweetheart)
4	*tumbuk rusuk* (to pound + flank)	bribe
5	*duit kopi* (money + coffee)	tips (money)
6	*makan angin* (to eat + wind)	sightseeing
7	*pilih kasih* (select + love)	favouritism
8	*manis mulut* (sweet + mouth)	smooth-tongued
9	*berat tangan* (heavy + hand)	lazy
10	*ringan tulang* (light + bones)	hard working
11	*buku lima* (fist + five)	fist/brass knuckles
12	*berat tulang* (heavy + bones)	lazy
13	*lipas kudung* (cockroach + stump)	speedy person
14	*cakar ayam* (to scratch + chicken)	scribbling

(ii) Verbal compounds

(a) *Free forms*

No.	Combination/Compound	Meaning
1	*kenal pasti* (to recognise + sure)	to identify
2	*ambil alih* (to take + to move)	to replace
3	*tunjuk ajar* (to show + to teach)	guide
4	*terima kasih* (to receive + love)	to thank
5	*campur aduk* (to mix + to mix/stir)	completely mixed
6	*lipat ganda* (to fold + to fold/double)	multiply
7	*tolak ansur* (to push + gradual)	tolerate
8	*beritahu* (to give + to know)	to tell

(iii) Adjectival compounds

(a) *Free forms*

No.	Combination/Compound	Meaning
1	*merah jambu* (red + red guava)	pink/rosy
2	*hijau daun* (green + leaf)	leafy green
3	*biru laut* (blue + sea)	ocean blue
4	*kuning langsat* (yellow + a local fruit)	yellowish
5	*putih telur* (white + egg)	egg white - like
6	*tengah hari* (middle + day)	midday
7	*luar biasa* (outside + ordinary)	unusual

(b) Idiomatic forms

No.	Combination/Compound	Meaning
1	*panjang tangan* (long + hand)	likes to steal (thief)
2	*besar kepala* (large + head)	stubborn/arrogant
3	*ringan tulang* (light + bones)	hard working
4	*terang hati* (bright + heart)	easy to gain knowledge
5	*rabun ayam* (short/long sighted + chicken)	short/long sighted during dusk