EXTRACTION OF COMPOUND NOUNS IN MALAY NOUN PHRASES USING A NOUN PHRASE FRAME STRUCTURE

SUHAIMI AB RAHMAN
NAZLIA OMAR
MOHD JUZAIDIN AB AZIZ

ABSTRACT

This paper addresses the process of extracting compound nouns in Malay noun phrases using a noun phrase frame structure. Studying in a compound noun area is very important to see the dependency of the word that can produce a correct meaning of the sentence. Each complete sentence in Malay must have their respective compound nouns. The compound nouns can be extracted from a subject and predicate of the sentences. The subject must consist of a noun phrase, while the predicate may have a verb phrase, an adjective phrase or a prepositional phrase. In a compound noun structure, we may know the head and modifier of the sentences. The compound noun has been discussed in detail by our native language experts. Many issues were highlighted to further strengthen the concept of compound nouns. Based on the issues discussed, we attempted to use a noun frame structure technique using a computer to extract compound nouns.
Keywords: Compound Noun, Noun Phrase Frame Structure, Malay Noun Modifier Tagger Tool, Malay Compound Noun Analyser.

ABSTRAK

Kertas ini akan membincangkan proses untuk mengekstrak kata majmuk yang terdapat di dalam sesuatu frasa nama dalam Bahasa Melayu. Kaedah untuk mengekstrak kata majmuk ini ialah dengan menggunakan struktur kerangka frasa nama. Kajian ke atas kata majmuk ini amatlah penting bagi menentukan kedudukan hubungan antara perkataan-perkataan untuk menghasilkan susunan yang betul dari segi sintak dan makna. Setiap ayat dalam Bahasa Melayu seharusnya mempunyai kata majmuk. Kata majmuk boleh didapati daripada pecahan frasa subjek dan frasa predikat. Setiap frasa tersebut akan mengandungi kata majmuk yang terkandung inti dan penerang di dalam nya. Setiap ayat akan dimulai dengan frasa nama dan diikuti dengan salah satu daripada frasa kerja, frasa sifat atau frasa sendi. Isu-isu kata majmuk ini telah dibincangkan dengan lebih lanjut oleh pakar-pakar bahasa atau penyelidik. Justeru itu, melalui perkara-perkara yang telah dibincangkan, kami bersedia akan menjalankan satu penyelidikan dan pembangunan untuk mengkaji serta mengekstrak kata majmuk dari frasa-frasa nama Bahasa Melayu berpandukan kepada struktur kerangka frasa nama.

Kata kunci: Kata Majmuk, Struktur Kerangka Frasa Nama, Perisian Penglabelan Penerang Kata Nama Bahasa Melayu, Penganalisa Kata Majmuk Bahasa Melayu.

INTRODUCTION

As sighted in Rencana Pilihan (2009), Malay language is a language spoken and written by 40 million people across the Malacca Strait, Peninsula of Malaysia, Southern Thailand, Singapore, Eastern Coast of Sumatra, Riau Islands in Indonesia, Western Coastal Sarawak, Brunei and West Kalimantan in Borneo. In Malaysia, a public agency known as Dewan Bahasa dan Pustaka (DBP[1]) is responsible for coordinating the use of Malay language in Malaysia and Brunei. Various studies were conducted to strengthen usage of Malay language in various domains (reference). The grammar issues in Malay language have been discussed by (Addullah Hassan

---

[1]Dewan Bahasa dan Pustaka (DBP) atau nama asalnya Balai Pustaka ditubuhkan pada 22 Jun 1956 di Johor Bahru. Penubuhan DBP adalah untuk mengembangkan bahasa Melayu sebagai bahasa kebangsaan dan bahasa rasmi negara. (http://http://www.dbp.gov.my/)

2004; Addullah Hassan 1992; Fazal Mohamed 2009; Fazal Mohamed 2008; Nik Sapiah et al. 2010; Ong Ching Guan 2009).

One of the important issues addressed in studying Natural Language Processing for Malay language is a compound noun. Researchers such as (Ken Barker 1998; L. Barrett 2001; Takaaki Tanaka and Yoshihiro Matsuo 2001; Vivi Nastase and Stan Szpakowicz 2003) discussed the fundamental concept of compound nouns in English. Most compound nouns in English are constructed by nouns that are modified by other nouns or adjectives.

As explained by (Addullah Hassan 1992; Nik Sapiahet al.2010), the compound noun in Malay is made up of two categories such as i) noun and noun modifier; and ii) noun and non-noun modifier. In the first category-noun and noun modifier, all the compound nouns are grouped according to their respective class names. The words are combined with nouns followed by other nouns. It has 13 different types of class names. In the second category-noun and non-noun modifier, a compound noun is formed based on 6 different class names. The words are combined with nouns followed by other non-nouns.

Andrew, H et al. (2008) explained the usage of semantic relations such as hypernym and meronym used in their research work. Both of these semantic relations are used to show the linkages between nouns. They give a few words, for example *boat, houseboat,* and *speedboat*. The word *boat* is a super class to the sub-class *houseboat* and *speedboat.* The compound *houseboat* and *speedboat* is therefore, a hyponym of *boat*, i.e. *"..is a kind of boat"*. A semantic relation hypernym is used to describe the relation *"..is a kind of"* between the words. They also used meronym semantic relation to elicit "*attribute"* or *"property"* of the sub-class words. The *speedboat* is a super-class to its *length, size* and *engine.* It means that the meronym of word *speedboat* is *length, size* and *engine*. Andrew, H et al. (2008) referred to the WordNet to identify the nouns using both hypernym and meronym relations.

(H.Sundblad 2002; Vivi Nastase and Stan Szpakowicz 2003) discussed the bracketing technique process to find the possibility of the existence of a compound noun in a sentence. They use a left and right branching format to place a group of words. They gave the example for left-branching and right-branching techniques used in a sentence. For example, phrase 1 :*laser printer manual*, after the left-branching applied, the following bracketing *((laser printer) manual)* will be produced, while for the right-branching, the following bracketing *(desktop (laser printer))* for the phrase *desktop laser printer* will be created. However, there is still a lack of empirical research in solving ambiguity in pairing the word.

The rest of the paper is organized as follows. In section 2 we will discuss the compound nouns in a Malay sentence and section 3 gives a brief conclusion.

COMPOUND NOUNS IN MALAY SENTENCE

The main activity in our research work is to observe and find an acceptable technique to extract a pair of compound nouns in Malay noun phrases. Referring to the compound noun construction process in Figure 1, we compiled nearly 107 pairs of words that fit with 13 categories of noun and noun modifier defined in our research requirement (Suhaimi Ab Rahman et al. 2011; Suhaimi Ab Rahman et al. 2012).

In order to label Malay word's Part-of-Speech (POS), Table 1 shows several Malay POS taken from Arbak Othman (2006) with their corresponding English POS. We will use the Malay POS labels shown in Table 1 for the examples and discussion outlined in our studies.

TABLE 1. Malay POS label

Description in Malay (English)	Malay POS Label
Kata Nama(Noun)	*KN*
Kata Kerja(Verb)	*KK*
Kata Adjektif(Adjective)	*KAdj*
Kata Adverba(Adverb)	*KAdv*
Kata Hubung(Conjunction)	*KH*
Kata Sendi Nama(Preposition)	*KSN*
Kata Penentu(Determiner)	*Pent*
Kata Penguat(Intensifier)	*KPeng*
Kata Bantu(Auxiliary)	*KB*
Kata Nama Inti (Head Noun)	*KNInt*
Bilangan (Numeric)	*KBil*
Penjodoh Bilangan (Classifier)	*Penbil*
Gelaran (Title)	*Gel*
Ayat (Sentence)	*A*
Frasa Adjektif (Adjective Phrase)	*FA*
Frasa Kerja (Verb Phrase)	*FK*
Frasa Nama (Noun Phrase)	*FN*
Frasa Sendi (Prepositional Phrase)	*FS*
Subjek (Subject)	*S*
Predikat (Predicate)	*P*
Inti (Head)	*H*
Penerang (Modifier)	*Mod*

In general, compound nouns in Malay sentences can be constructed using the following three steps:
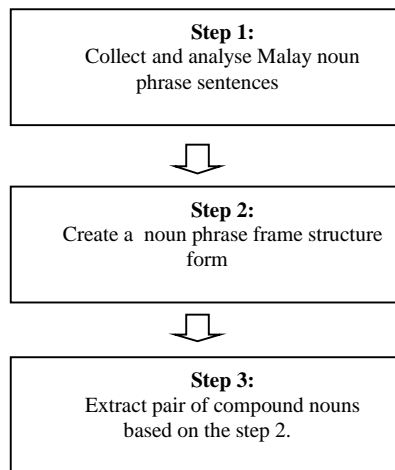

FIGURE 1. Compound noun construction process

Step1: Collect and analyse sentences in Malay noun phrase sentences

In our data preparation process, we currently focus on noun phrase sentences to identify the compound nouns that exist in the sentence. In order to get the noun phrase sentences, we gathered 600 examples of Malay sentences. Out of 600 examples, we may know a number of compound nouns placed in the sentences. This is done using Malay compound nouns analyser tool. All the 600 examples are gained from a dictionary, children story book, internet, and magazine. The collection of examples is done manually. The prototype Malay compound nouns analyser tool is developed to detect the possibility of the existence of a pair of compound noun in a Malay noun phrase sentence. This tool uses a technique named a noun phrase frame structure to extract compound noun from a noun phrase sentence but it does not detect a compound noun in verb phrases, adjective phrases and prepositional phrases.

At present, the method of splitting the sentence into two sub phrases, a subject and a predicate is done manually to become as the input to our prototype system. The words in each

subject and predicate will be assigned with a respective POS. The POS for the words are done automatically using our prototype Malay language noun modifier tagger tool.

Table 2 shows the examples of Malay noun phrases collected from our data preparation process.

TABLE 2. Examples of Malay noun phrase sentences

Sentence Examples	Subject	Predicate
Ahmad guru	*Ahmad[KN] + guru[KN]*	-
Daging lembu	*Daging[KN] + lembu[KN]*	-
Tahun naga	*Tahun[KN] + naga[KN]*	-
Ahli jawatankuasa surau	*Ahli[KN]+ jawatankuasa[KN]*	*surau[KN]*
Kereta baru proton	*Kereta[KN]+baru[KN]*	*proton[KN]*
Buku cerita kanak-kanak	*Buku[KN]+cerita[KN]*	*kanak-kanak[KN]*
Buku cerita dongeng	*Buku[KN]+cerita[KN]*	*dongeng[KN]*
Kulit buah koko	*kulit[KN]*	*buah[KN]+koko[KN]*
Bidang kuasa kerajaan negeri	*bidang[KN]+kuasa[KN]*	*kerajaan[KN]+ negeri[KN]*
Datuk Bandar kuala lumpur	*datuk[KN]+bandar[KN]*	*Kuala[KN]+ Lumpur[KN]*
Tin susu itu sudah saya buka	*tin[KN]+susu[KN]+itu[Pent]*	*sudah[KB] + saya[KN] + buka[KN]*

Based on the examples shown in Table 2, we divided ach sentence into a subject and a predicate. Identifying a correct group of words as subjects and predicates is very important to find the existence of a compound noun in the sentence. The correctness of group of words as subjects and predicates was verified by the linguist.

Step 2: Analyse and create a noun phrase frame structure.

To find a compound noun in a Malay sentence, we use a noun phrase frame structure depicted in Table 3.This frame structure contains a list of sequence word categories starting from numeric, classifier until determiner. Not all sentences must fulfil all these categories. The rules are necessary in determining the appropriate position for each word from input sentence to be placed in a noun phrase frame structure. Other than the words that are assigned with their respective POS, an additional label known as the noun modifier is needed for the rule to be more precise and recognised as a compound noun. (Addullah Hassan 2004; Addullah Hassan 1992; Nik Sapiah et al. 2010) also discussed these issues by showing the importance of a noun phrase frame structure in constructing a noun phrase or Malay sentence.

The examples of noun phrase rules are discussed below:
Sentence 1:   Malay: *"Gunung itu amat tinggi."*
              English: The mountain is very high.
Sentence 2:   Malay: "*Mr Ahmad seorang guru matematik sekolah saya.*"
              English: Mr Ahmad is a mathematics teacher at my school.
Sentence3:   Malay: "*Dua orang kanak-kanak perempuan sedang bermain di taman permainan.*"
              English: Two girls are playing at the playground.

Based on the examples in Sentence 1, 2 and 3, we can use a noun phrase frame structure to arrange the position of words.

TABLE 3. Noun phrase frame structure for Malay sentence 1, 2, and 3

	Noun Phrase							
	Noun modifier category			Compound Noun				Noun modifier category
	Numeric	Classifier	Title	Head Noun	Head Noun	Modifier	Modifier	Determiner
POS and noun modifier category name	*[KN]* *[KBil]*	*[KN]* *[PenBil]*	*[KN]* *[Gel]*	*[KN]*	*[KN]*	*[KN]*		*[KN]* *[Pent]*
Row 1	-	-	-	*gunung*	-	-		*itu*
Row 2	-	-	*Mr*	*Ahmad*	-	-		-
Row 3	*Seorang*	-	-	*guru*	-	*Matematik*	*sekolah saya*	-
Row 4	*dua*	*orang*	-	*kanak-kanak*	-	*Perem-puan*	*sedang bermain di taman permainan*	-

Referring to the noun phrase frame structure in Table 3, we can summarize the discussion as follows:
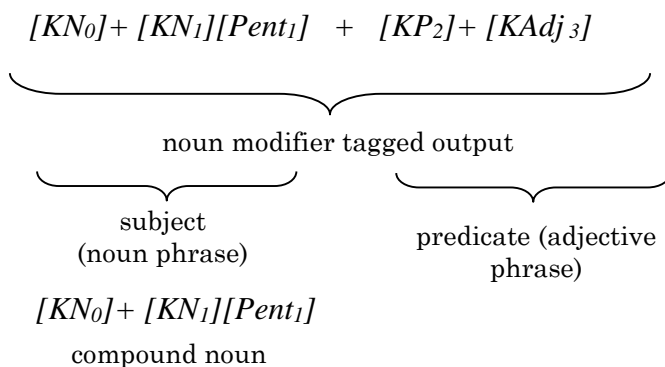
Sentence 1:
Malay: *"Gunung itu amat tinggi."*
English: The mountain is very high.

The noun modifier labels for the words in Sentence 1 are:
Gunung[KN$_0$]+itu [KN$_1$][Pent$_1$]+amat [KP$_2$]+ tinggi[KAdj$_3$]   (syntax 1)

In syntax 1, only the first two words in Sentence 1 comply with the rule in Row 1 :*[KN$_0$]+ [KN$_1$][Pent$_1$]*. The remaining words are grouped as adjective phrases. In this rule, a compound noun for a noun phrase is nominated as follows:

[KN$_0$]+ [KN$_1$][Pent$_1$]   +   [KP$_2$]+ [KAdj$_3$]

noun modifier tagged output

subject
(noun phrase)

predicate (adjective phrase)

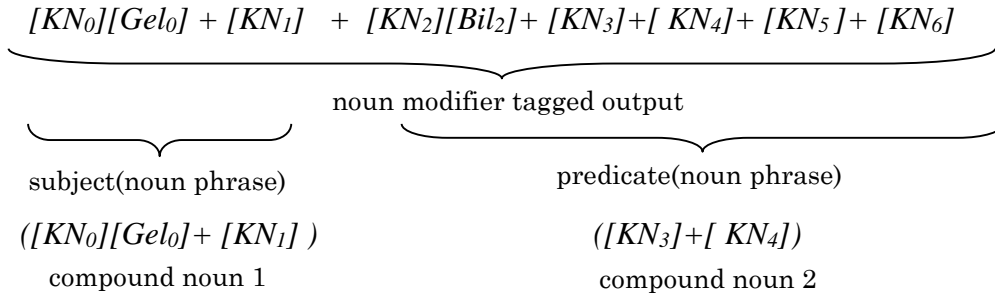*[KN$_0$]+ [KN$_1$][Pent$_1$]*

compound noun

Sentence 2:
Malay: "*Mr Ahmad seorang guru matematik sekolah saya.*"
English: Mr Ahmad is a mathematics teacher at my school.

The noun modifier labels for the words in Sentence 2 are:
"*Mr[KN_0][Gel_0] Ahmad[KN_1] seorang[KN_2][Bil_2] guru [KN_3] matematik[KN_4] sekolah[KN_5] saya[KN_6]."* (syntax2)

In syntax2, the first two words in Sentence 1 comply with the rule in Row 2 :$[KN_0]$+ $[KN_1][Pent_1]$, while remaining of words are in Row 3: $[KN_2][Bil_2]$+ $[KN_3]$+$[KN_4]$+ $[KN_5]$+ $[KN_6]$. In this rule, a compound noun for a noun phrase is nominated as follows:

$$[KN_0][Gel_0] + [KN_1] \quad + \quad [KN_2][Bil_2]+ [KN_3]+[KN_4]+ [KN_5]+ [KN_6]$$

noun modifier tagged output

subject(noun phrase)   predicate(noun phrase)

$([KN_0][Gel_0]+ [KN_1])$ $([KN_3]+[KN_4])$

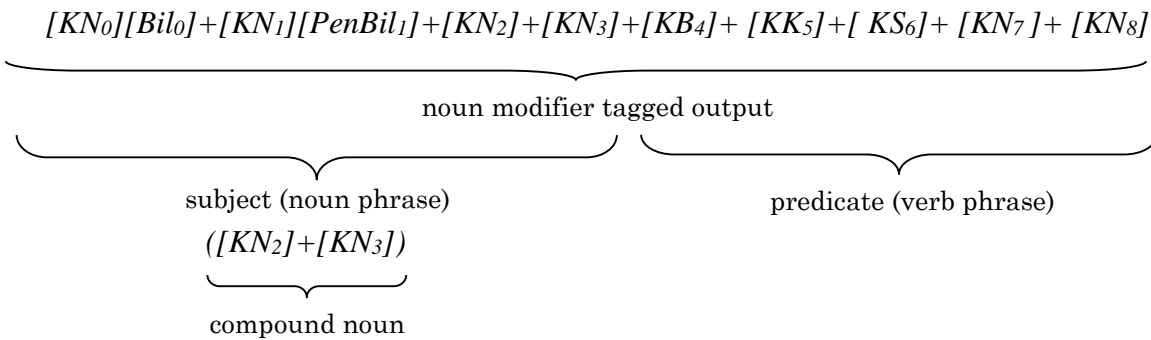compound noun 1   compound noun 2

Sentence 3:
Malay: "*Dua orang kanak-kanak perempuan sedang bermain di taman permainan.*"
English: Two girls are playing on the playground.

The noun modifier labels for the words in Sentence 3 are:
"*Dua[KN_0][Bil_0] orang[KN_1][PenBil_1] kanak-kanak[KN_2] perempuan[KN_3] sedang[KB_4] bermain[KK_5] di[KS_6] taman[KN_7] permainan[KN_8].*" (syntax 3)

In syntax3, the first two words in Sentence 3 satisfy with the rule in Row 4 :$[KN_0][Bil_0]$+$[KN_1][PenBil_1]$+$[KN_2]$+$[KN_3]$ while remaining of words are in verb phrases, such as $[KB_4]$+ $[KK_5]$+$[KS_6]$+ $[KN_7]$+ $[KN_8]$. In this rule, a compound noun for a noun phrase is nominated as follows:

$$[KN_0][Bil_0]+[KN_1][PenBil_1]+[KN_2]+[KN_3]+[KB_4]+ [KK_5]+[KS_6]+ [KN_7]+ [KN_8]$$

noun modifier tagged output

subject (noun phrase)   predicate (verb phrase)

$([KN_2]+[KN_3])$

compound noun

All formulated rules will be kept in a Malay noun phrase structure rule database for reference. The more examples of Malay sentences collected, the more noun phrase rules can be constructed. The rules are also discussed in [Michael, N 2002].

Step 3: Extract compound nouns based on step 2.

We developed a prototype of a Malay language compound noun analyser to identify and recognize which part of the words in Malay noun phrase sentences are in compound nouns. The Malay language compound noun analyser will use a technique described in Step 2. Below are examples of the output produced using a Malay language compound noun analyser.

We tested the system using 100 simple Malay sentences. The sentences are from general domain. A few examples of the sentences that were tested are as follows:

TABLE 4. Sentence examples to be tested

No.	Example	Sentence Structure	Result
1	Bapa budak itu guru sekolah agama. *(The boy's father is a religious school teacher.)*	S: Bapa budak itu P: guru sekolah agama	Both subject and predicate can be detected from our system
2	Kain kapas buatan Malaysia. *(Cotton cloth made in Malaysia.)*	S: Kain kapas P: buatan Malaysia	Both subject and predicate can be detected from our system
3.	Pada malam ini bulan sungguh terang. *(Moon looks so bright tonight.)*	S: Pada malam ini P: bulan sungguh terang	System did not detect the adjective phrase in a predicate
4.	Lelaki itu sedang makan di atas meja. *(The man was eating on the table.)*	S: Lelaki itu P: sedang makan di atas meja.	System did not detect the verb phrase in a predicate
5.	Blok asrama itu untuk pelajar-pelajar lelaki sahaja. *(The hostel block is for male students only.)*	S: Blok asrama itu P: untuk pelajar-pelajar lelaki sahaja	System did not detect the preposition phrase in a predicate

Based on 100 sentences tested, we found that a number of compound nouns exist in the sentences as shown in Table 5:

TABLE 5. Number of compound noun

Type of sentences	No. of sentences	No. of compound nouns detected
Noun phrase and noun phrase	55	95
Noun phrase and verb phrase	20	20 from a noun phrase.
Noun phrase and adjective phrase	15	15 from a noun phrase.
Noun phrase and preposition phrase	10	10 from a noun phrase

Referring to Table 5, the total number of compound nouns in a noun phrase was detected using rules in Step 2. The calculation of the total of compound nouns that exist as subjects and predicates is based on the following summation formula:

$$Sum_{subject} = \sum_{i=1}^{n=100} CN_i \qquad (1)$$

$$Sum_{predicate} = \sum_{i=1}^{n=100} CN_i \qquad (2)$$

$$\text{Total CN} = Sum_{subject} + Sum_{predicate} \qquad (3)$$

Where n=100 represents the total number of Malay sentences that need to be tested while the variables Compound Noun (CN), $sum_{subject,}$ and $sum_{predicate}$ represent the total number of compound nouns obtained by using a Step 2. The summation is based on the number of

compound nouns discovered from subjects and predicates of the sentences, represented as Total CN. Referring to Table 5, the Total CN will be assigned to value 140. It means that there are 140 compound nouns detected from 100 examples tested in Malay sentences.

Table 6 shows a few examples of compound nouns obtained from the testing of 100 sentences in Malay sentences using the summation formula discussed earlier.

TABLE 6. Example of compound nouns

Compound Noun (CN)	Compound Noun (CN)
CN (Orang[kn], Melayu[kn])	CN (Penduduk[kn], Bandar[kn])
CN (Pelajar[kn], Amerika[kn])	CN (Anak[kn], Johor[kn])
CN (Askar[kn], British[kn])	CN (Orang[kn], Tempatan[kn])
CN (Bahasa[kn], Melayu[kn])	CN (Tepi[kn], Pantai[kn])
CN (Masyarakat[kn], Cina[kn])	CN (Anak[kn], Desa[kn])
CN (Kaum[kn], Banjar[kn])	CN (Cina[kn], Melaka[kn])
CN (Keturunan[kn], Jawa[kn])	CN (Pinggir[kn], Kota[kn])
CN (Harimau[kn], Bintang[kn])	CN (Kawasan[kn], Barat[kn])
CN (Kelapa[kn], Sawit[kn])	CN (Wilayah[kn], Selatan[kn])
CN (Pokok[kn], Getah[kn])	CN (Angin[kn], Timur[kn])
CN (Air[kn], Limau[kn])	CN (Jalan[kn], Tengah[kn])
CN (Buah[kn], Mangga[kn])	CN (Simpang[kn], Kanan[kn])
CN (Baju[kn], Kemeja[kn])	CN (Ruang[kn], Dalam[kn])
CN (Kerusi[kn], Rotan[kn])	CN (Pintu[kn], Hadapan[kn])
CN (Gelang[kn], Emas[kn])	CN (Kulit[kn], Luar[kn])
CN (Kasut[kn], Getah[kn])	CN (Tingkat[kn], Atas[kn])
CN (Kipas[kn], Kain[kn])	CN (Rumah[kn], Belakang[kn])
CN (Beg[kn], Kulit[kn])	CN (Jam[kn], Tangan[kn])
CN (Pingat[kn], Gangsa[kn])	CN (Gelang[kn], Kaki[kn])
CN (Penyendal[kn], Pintu[kn])	CN (Tali[kn], Pinggang[kn])
CN (Penjaga[kn], Stor[kn])	CN (Tudung[kn], Kepala[kn])
CN (Pam[kn], Air[kn])	CN (Sarung[kn], Kaki[kn])
CN (Gunting[kn], Kain[kn])	CN (Rantai[kn], Leher[kn])
CN (Penyelenggara[kn], Pejabat[kn])	CN (Barah[kn], Kulit[kn])
CN (Pemandu[kn], Teksi[kn])	CN (Penyakit[kn], Jantung[kn])
CN (Pengurus[kn], Syarikat[kn])	CN (Kepala[kn], Lutut[kn])
CN (Pelayan[kn], Tetamu[kn])	CN (Batang[kn], Hidung[kn])
CN (Alat[kn], Pemadam[kn])	CN (Lampu[kn], Minyak[kn])
CN (Sabun[kn], Pencuci[kn])	CN (Dapur[kn], Gas[kn])
CN (Kayu[kn], Pengukur[kn])	CN (Seterika[kn], Arang[kn])
CN (Minyak[kn], Pelincir[kn])	CN (Kipas[kn], Elektrik[kn])
CN (Tali[kn], Penyidai[kn])	CN (Kereta[kn], Motor[kn])
CN (Getah[kn], Pemadam[kn])	CN (Kapal[kn], Layar[kn])
CN (Jamu[kn], Pengawet[kn])	CN (Kapal[kn], Api[kn])
CN (Batu[kn], Penanda[kn])	CN (Mesin[kn], Kaki[kn])
CN (Kayu[kn], Pemukul[kn])	CN (Perahu[kn], Galah[kn])
CN (Ayam[kn], Jantan[kn])	CN (Lesen[kn], Perniagaan[kn])
CN (lembu[kn], Betina[kn])	CN (Surat[kn], Kebenaran[kn])
CN (Orang[kn], Lelaki[kn])	CN (Borang[kn], Pendaftaran[kn])
CN (Kaum[kn], Perempuan[kn])	CN (Hari[kn], Perayaan[kn])
CN (Budak[kn], Lelaki[kn])	CN (Garis[kn], Permulaan[kn])
CN (Anak[kn], Perempuan[kn])	CN (Surat[kn], Perjanjian[kn])
CN (Harta[kn], Kerajaan[kn])	CN (Majlis[kn], Perasmian[kn])

The list of compound nouns produced from our prototype Malay language compound noun analyser is important to be used as a data sample in evaluating the accuracy of the method used in our prototype system. The accuracy of compound noun results was evaluated by the linguist or the language expert to verify the correctness of the compound noun produced from the prototype system. The language expert will analyse the sentences to identify compound nouns in the sentences. The results from the language expert will be compared with the result generated from our system to find the percentage of similarity results. Therefore, we can measure the percentage of accuracy method used in our prototype system.

The result of our study will be used in our next research work to detect head modifier in a noun phrase sentence based on the compound nouns method.

CONCLUSION

In this paper, we described fundamental process on extracting compound noun from Malay sentences. As discussed, the Malay language compound noun analyser tool only detected a noun phrase, and not detecting a verb phrase, an adjective phrase and a prepositional phrase. The rules to detect compound noun can be constructed using a noun phrase frame structure rule. This frame structure rule also has been explained in Section 2. In creating a noun phrase frame structure rule, we also plan to create another two different frames to handle problems such as solving issues on appearing word verb and adjective in a noun phrase. The accuracy of compound noun detected in a noun phrase is based on the number of formulated rules. The more rules created means more possibility of getting compound nouns in a noun phrase to be improved. The final result such as a list of compound nouns will be verified from a language expert, so that we know the correctness of the process, particularly in using the technique to extract compound noun in a noun phrase noun phrase for Malay sentences.

ACKNOWLEDGMENT

REFERENCES

Addullah Hassan.2004. *Tatabahasa Bahasa Melayu, 4th edition,* PTS Publications & Distributors Sdn. Bhd.

Abdullah Hassan.1992. *Linguistik Am, 10 edition*, PTS Profesional Publishing Sdn.Bhd.

Andrew, H, David Cheng, Khurshid Ahmad. 2008. The head-modifier principle and multilingual term extraction. Universityof Surrey, UK.

Arbak Othman, Nik Safiah Karim.2006. *Kamus komprehensif Bahasa Melayu, Cetakan kedua*, Penerbit Fajar Bakti Sdn.Bhd.

Fazal Mohamed Mohamed Sultan. 2009. Struktur sintaksis frasa nama bahasa Bateq. *GEMA Online Journal of Language Studies*,9(1), 47-61, Malaysia.

Fazal Mohamed Mohamed Sultan. 2008. Struktur sintaksis frasa nama (FN) Bahasa Melayu. *Jurnal Bahasa,8*(2), 204-219, Malaysia.

H.Sundblad. 2002. Automatic Acquisition of Hyponyms and Meronyms from Question Corpora.*ECAI'2002*, Lyon,France.

Ken Barker. 1998. A trainable bracketer for noun modifiers. University of Ottawa, Canada.

L. Barrett, A. R. Davis, and B. J. Dorr. 2001. Interpretation of compound nominals using wordnet. In: *CICLing*, pages 169-181.

Rencana pilihan 2009. Memastikan kepentingan Bahasa Melayu kekal sepanjang zaman. http://abihulwa.blogspot.com/2009/10/ memastikan-kepentingan-bahasa-melayu.html. [31 Oktober 2009]

Michael, N.2002. Artificial *Intelligence: A guide to Intelligent Systems, Second edition,* PearsonEducation Limited.

Nik Sapiah Karim, Farid M.Onn, Hashim Haji Musa, Abdul Hamid Mahmood,2010. *Tatabahasa Dewan , 3rd edition (cetakan kelima),* Dewan Bahasa dan Pustaka (DBP), Malaysia.

Ong Ching Guan.2009. *Kuasai Struktur Ayat Bahasa Melayu,*Dewan Bahasa dan Pusataka (DBP), Malaysia.

Suhaimi Ab Rahman, Nazlia Bt Omar, Noor Baizura Che Hassan. 2012. Construction of Compound Nouns (CNs) for Noun Phrase in Malay Sentence.*International conference on Information*

*Retrieval and Knowledge Management, CAMP'12 (Capaian Maklumat dan Pengurusan Pengetahuan),*Mines Wellness Hotel, Malaysia.

Suhaimi Ab Rahman, Nazlia Omar and Mohd Juzaidin Ab Aziz. 2011. Transformation of Malay Head Modifier Noun Phrase into a Thematic Relation Structure. *International Conference on Electrical Engineering and Informatics, (Volume 3),* PP 1775-1779, Bandung-Indonesia.

Takaaki Tanaka and Yoshihiro Matsuo. 2001. Ex-traction of compound noun translations from non-parallel corpora (in Japanese). *Trans. of theInstitute of Electronics, Information and Communication Engineers,* 84-D-II(12):2605-2614.

Vivi Nastase and Stan Szpakowicz.2003. Exploring noun-modifier semantic relations. *5th International Workshop on Computational Semantics (IWCS-03),*Tilburg, The Netherlands.

Suhaimi Ab Rahman
Software Engineering Department,
Universiti Tenaga Nasional (UNITEN),
Jalan IKRAM-UNITEN,
43000 Kajang, Selangor, Malaysia.
smie@uniten.edu.my

Nazlia Omar,
Mohd Juzaidin Ab Aziz
School of Computer Science,
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia.
no@ftsm.ukm.my, din@ftsm.ukm.my