



## **A Comparative Analysis of Word Structures in Malay and English Children's Stories**

**Lay Wah Lee\*, Hui Min Low and Abdul Rashid Mohamed**

*School of Educational Studies, Universiti Sains Malaysia, 11800 Penang, Malaysia*

### **ABSTRACT**

Malay is described as an alphabetic language with salient syllabic structures. In our attempt to develop a reading intervention program for early Malay struggling readers, word analysis of Malay children's stories was conducted. Additionally, in order to have a better understanding of Malay word structures, a cross-linguistic comparison with English was attempted. The results indicate significant cross-language differences for Malay and English words in terms of phoneme-grapheme correspondences, syllabic structure and types of inflectional morphemes. Malay is empirically shown to be a transparent language but with multiple syllabic structures and inflectional morphemes. The analysis also revealed that the most frequent occurring word structures in the Malay texts were bi-syllabic, with CV+CVC, CV+CV, V+CVC, and CVC+CVC word structures. This suggests that unlike English, the major set of word stimuli in early Malay reading intervention programs have to be bi-syllabic, which implies that additional syllabic decoding skills have to be taught in early reading intervention.

*Keywords:* Word analysis, cross-linguistics, English, Malay, reading intervention

### **INTRODUCTION**

The mappings between orthography, phonology and semantics of languages give rise to the differences in writing scripts which normally fall into three main

types: alphabetic, syllabic and logographic (DeFrancis, 1989). Standard Malay is an alphabetic-syllabic writing script used in Malaysian schools. Malay is an agglutinative language. The meaning of words can be changed by adding inflectional morphemes such as prefixes, suffixes, and circumfixes to the root words. For example, the verb 'makan' (eat), when added with the suffix '-an', becomes 'makanan' (food)

#### **ARTICLE INFO**

##### *Article history:*

Received: 5 January 2011

Accepted: 8 July 2011

##### *E-mail addresses:*

lwah@usm.my (Lay Wah Lee), halowmin@hotmail.com (Hui Min Low), rich@usm.my (Abdul Rashid Mohamed)

\* Corresponding author

(refer to Benjamin, 2009, for a review on Malay affixes). Due to this agglutinative feature, poly- or multisyllabic words are common in this language.

In this paper, we aimed to analyse the structure of Malay words in order to inform reading intervention programme for Malay-learning struggling readers. This analysis is crucial since many researchers have provided evidence that the incidence and severity of reading disabilities are influenced by the orthography and phonology of a language (e.g., Ellis *et al.*, 2004; Katz & Frost, 1992; Miles, 2000; Ziegler & Goswami, 2005). For languages with alphabetic and syllabic scripts, many researchers have differentiated between languages with transparent and opaque letter-sound correspondences. Transparent languages, such as German, Greek, Spanish, Turkish and Welsh, have letters and sounds with an almost one to one relationship, while opaque languages such as French and English have inconsistent letter-sound correspondences (see Ellis *et al.*, 2004 for a review). This distinction is believed to contribute to processing differences in word naming, where transparent orthography promotes direct phonological decoding while opaque orthography requires additional visual and onset-rime processing, on top of phonological decoding (Ellis *et al.*, 2004; Katz & Frost, 1992; Miles, 2000; Ziegler & Goswami, 2005).

To date, many reading intervention programs have been designed for English-speaking children. The efficacy of these English reading intervention programs has

been widely studied and reported (Vellutino *et al.*, 2004). To our knowledge, there is no evidence-based treatment programme available yet for children who are struggling to learn Malay. If we are going to utilize and to base our intervention programme on the treatment frameworks that are originally designed for English-speaking children, we need to first understand the cross-linguistic differences between Malay and English. This has been made explicit to us by reviews of the first author's earlier works (Lee & Wheldall 2010; Lee *et al.*, 2009). In line with the orthographic depth hypothesis (Katz & Frost 1992), we hold the same view as the reviewers of our earlier work that Malay language has specific phonological and orthographic features that potentially entail the struggling readers of this language to be remediated with instruction different from that for the English-learning readers. However, such a view is hard to justify since there are almost no empirical data that have systematically revealed the cross-linguistic differences between English and Malay word structures. As a result, it remains unjustifiable which aspects of treatment used for the English-learning readers are transferable, and which are not.

Few researchers have contrasted the cross-linguistic differences between Malay and English, especially in relation to the analysis of word structure. Earlier works by Lee (2008), and Lee and Wheldall (2009) provided some descriptive similarities and differences between these two languages. Most of the descriptions remained hypothetical, which we have sought to

improve in this study. The aspects of similarities include the fact that the Malay writing system is phoneme-based and it uses the same Latin alphabetic script, i.e. the 26 letters as in English (Lee, 2008). On the other hand, Lee (2008) identified two major features in Malay which are distinctive from the widely studied English language.

First, Malay has a more transparent orthography as compared to English. This means that the grapheme-phoneme correspondences in Malay are more direct as compared to English. For example, the Malay word *malam* (night) consists of five letters (i.e. 'm' + 'a' + 'l' + 'a' + 'm') and five phonemes (i.e. /m/ + /a/ + /l/ + /a/ + /m/). This results in a 100% direct grapheme-phoneme correspondence. In comparison, the English word *night* has five letters ('n' + 'i' + 'g' + 'h' + 't') but only three phonemes (/n/ + /aɪ/ + /t/). The grapheme-phoneme correspondence is not as direct as it is in Malay. Lee (2008) previously has indicated that this feature places Malay on a similar granularity-transparency dimension as German or Italian. Researchers, such as Wydell and Butterworth (1999), Miles (2000) and Ziegler and Goswami (2005), have found that children who are learning a transparent language such as Malay acquire phoneme awareness more rapidly than those who are learning a less transparent language such as English. However, a recent study involving a Malay-speaking child, with severe reading impairments in a trial reading intervention programme (Lee 2010), indicated severe difficulties in decoding Malay word stimuli. The child continued

to struggle with phoneme-grapheme decoding despite an intensive intervention programme. According to Miles (2000), even in transparent orthographies, there are inconsistencies in phoneme-grapheme correspondences that dyslexic children still have difficulty with. Therefore, it is important to continue to investigate the features of the phonemes and graphemes in the Malay language context.

Second, as reviewed in Lee (2008), native Malay words are based on four distinct syllable structures, i.e. V, VC, CV and CVC (Hamdan, 1988). These basic structures can be combined in multiple ways to form two syllabic words like *bola* (ball), with the structure of CV+CV or more complex words like *mesyuarat* (meeting), with the structure of CV+CV+V+CVC. Malay words are mostly formed by two or more distinct syllables (Karim *et al.*, 2004), with very few mono-syllabic words. The dominance of multi-syllabic words in Malay imposes a challenge in the design of a reading intervention programme for the early Malay poor readers since most of the existing English intervention programmes use mono-syllabic words as the primary stimulus set (e.g. Gillon, 2008; Lindamood & Lindamood, 2007). Mono-syllabic words that rhyme, such as 'fat' and 'bat', are the desired stimulus set in those programmes which focus on facilitating phoneme awareness at the word level in English. The differences in English and Malay word structures discussed so far would imply that some aspects of reading intervention in Malay need to be different

from the conventional English programmes. However, as proposed earlier, there is a lack of empirical data that has systematically revealed the nature and the scale of these differences. Therefore, there is a lack of evidence-based reference that can be used to design a systematic remedial reading instruction programme in Malay language.

So far, we have descriptively reviewed two primary Malay word characteristics which are different from English. As compared to English, Malay has (1) more direct grapheme-phoneme correspondences and (2) more multi-syllabic words. In this study, we aimed to systematically examine these cross-linguistic differences by analyzing children's translated stories (labelled as E and M to refer to the works written in English and in Malay, respectively). This analysis allows us to contrast the orthography and phonology of the two languages by controlling the type of texts used. We also aimed to further gather differences between the languages, and second, to gain greater insights into the selection of word stimuli for a Malay reading intervention programme. In order to guide our investigation, we focused on two research questions:

1. What are the linguistic differences in the Malay and English word structures?
2. How can these differences be used to inform the selection of word stimuli for a Malay reading intervention programme?

## METHOD

Since we aimed at focusing on analyzing word structures, we specifically looked for reading materials with a close match in translation at the sentential level. Children's translated story books fulfilled this requirement. We randomly chose two stories for our analysis. The first story is *Tom Thumb* (E1 and M1) and the second story is *Thumbelina* (E2 and M2). Both stories are classic fairy tales. The stories were retold by Jeff K.L. Lay and published in Malaysia in the year 2001. The stories and language levels were assessed by a trained teacher and a qualified translator in terms of their suitability for readers in early elementary school.

### *Translation*

The translator also commented on the quality of the translation. The stories contained simple vocabularies and short sentences (*e.g., He could ride a horse*). Literal translations were used to translate from English into Malay. It is a translation strategy whereby text is translated "word-by-word" abiding by the target language grammar. This translation strategy is considered acceptable since sentences in English and Malay both adhere to the "subject + predicate" word order rule. Overall, the meanings of the target text are rather well-maintained. Some examples of the translation are provided to illustrate this:

**Source text 1 (English):** Tom Thumb was the son of a woodcutter and his wife.

**Target text 1 (Malay):** Tom Thumb ialah anak kepada seorang pemotong kayu dan isterinya.

**Back translation 1:** *Tom Thumb is/was the son for one woodcutter and his wife.*

**Source text 2 (English):** A mouse often came to the mole's house for dinner.

**Target text 2 (Malay):** Seekor tikus tanah sering datang ke rumah tikus mondok itu untuk makan malam.

**Back translation 2:** *One mouse always comes to the house of the mole for dinner.*

Since both the stories originated from the West, the source text contained some cultural terms, such as *snow, fairy, winter*. However, the illustrations in the book served to give the meaning of those cultural terms. As a whole, the translated Malay texts were judged by the professional translator as a good translation as it conveyed the same message as the source text, with the original meaning maintained.

### *Analysis*

We used a general term 'text' to refer to all the stories, i.e., E1, M1, E2 and M2. For each text, we first converted it into an electronic format by typing it into the Clan transcription processing programme (MacWhinny, 2000). We then used the commands in the Clan programme to calculate the number of word-types and word-tokens. The number of word-types refers to the count of different words in the text. This count functions to provide a compressed view of word variety. Hence,

the count excludes the re-occurrence of the same word in the text. By contrast, the count of word-tokens considers both word variety and any repetitions. It functions to provide a fuller view of word distributions in the text.

Next, we exported the word-type and word-token results from the analysis in the Clan program to a word processor. For both sets of result, we first screened the data for non-words (e.g., oh, hmm) and proper nouns (e.g., Tom Thumb, Thumbelina). Our preliminary analysis showed that these items formed less than 10% of the total word distribution. Despite its relatively small percentage, we excluded these items in order to set up clearer language boundaries for analysis purposes. We then coded the data for (1) word length (grapheme) – the number of letters in the word, (2) phoneme combination – the number of phonemes in the word, (3) number of syllables – the number of syllables identifiable with their syllabic boundaries in a word, and (4) types of inflectional morphemes – prefixes, suffixes, circumfixes, reduplications and irregular words.

## **RESULT**

### *Word frequency*

The word frequency analysis via the Clan programme provided us with the word-type and word-token results. The results are presented in Table 1. These data function to provide an overview of the word size and variety.

The data in Table 1 showed that the English texts contained 213 and 215 word-types while the Malay texts contained 216

TABLE 1  
Results of word-type and word-token analysis\*

Language	Texts	Number of word-types	Number of word-tokens	Type-token ratios
English	E1	215	482	0.446
	E2	213	485	0.439
Malay	M1	215	413	0.521
	M2	224	450	0.498

\* The calculation generated by the Clan program did not exclude non-words and proper nouns. The non-words and proper nouns were excluded in the later analyses.

and 224 word-types. Even though the number of word-types in Malay and its corresponding English texts was almost comparable, there were slightly more word-tokens in the English texts (482 and 485) than in the Malay texts (413 and 450). As a result, a higher type-token ratio was found in the Malay texts as compared to the English. This could be an English to Malay translation effect, where words such as 'woodcutter' and 'swallow' as presented in the methodology section, have multi-lexical forms in Malay, i.e., 'pemotong kayu' and 'burung layang-layang'.

#### *Reliability*

For the subsequent coding for English and Malay word items, the inter-coder reliability was checked for the segmentation of phonemes and syllables before we continued with the analysis. Both the adult coders in this study have almost equivalent bilingual literacy in English and Malay, and the first coder has previous research experience in analyzing Malay and English texts. The inter-coder reliability was checked for 351 English word items and 363 Malay word-types in the texts. The analysis excluded

duplicating word-types that occurred in both stories. For example, the word 'day' occurs in both E1 and E2. Hence, one duplicate was removed from this analysis. In total, there were 72 duplicates that were excluded from the English and Malay lists, respectively. The results revealed a high consistency between the two coders, with alpha values above .7 for all the comparison datasets (Pallant, 2007). The consistency levels of 86.0% (n=351) and 95.0% (n=363) were achieved for the segmentations of English and Malay phonemes. This produced high alpha values of .98 and .99, respectively, when tested with Cronbach's alpha coefficient. The consistency levels of 91.2% (n=351) and 98.9% (n=363) were achieved for the segmentations of English and Malay syllables, which in turn produced high alpha values of .95 and .99. A higher inter-coder consistency was found for Malay than English coding. Thus, the finding suggests that the Malay phonemes and syllables are easier to be coded compared to English. This is likely due to the greater transparency of the Malay language compared to English.

*Grapheme-phoneme Correspondence*

The number of letters and the number of phonemes were calculated as a way to investigate the grapheme-phoneme correspondences in English and Malay words. In order to illustrate the grapheme-phoneme relationships in our data, the English and Malay words were grouped into two categories: (1) words with a direct match of letter and phoneme (e.g. the word 'cut' in English consisted of three letters and three phonemes, /kʌt/), and (2) words without a direct match of letter and phoneme (e.g. the word 'come' in English consisted of four letters, but only three phonemes /kʌm/). The results are summarized in Table 2.

As indicated by the results in Table 2, about 70% of the words in the English texts did not have a direct letter and phoneme match. In contrast, the opposite was observed in the Malay data. In more specific, 72.9% and 76.0% of the words in the Malay texts had a direct match of letter and phoneme. The examples are: *aku* (/aku/), *sini* (/sini/) and, *besar* (/besar/). Meanwhile, the words without a direct match in Malay are words consisting of diphthongs (/ai/, /au/, /oi/) and digraphs (/gh/, /kh/, /ŋ/, / /, /

sy/) (Awang, 2004). The examples of these words contained in the texts are *engkau* (/eŋkau/) and *orang* (/oraŋ/). The result shows that these words occur in about one quarter of the total distribution of word-types in each Malay text analyzed. An independent t-test was conducted to assess the statistical relationship of the differences observed. The Malay texts were found to have significantly more words with a direct phoneme and grapheme match ( $M=0.74$ ,  $SD=0.44$ ) than the English texts ( $M=0.30$ ,  $SD=0.46$ ),  $t(850,2)=14.35$ ,  $p<.001$ . The dominance of words with a direct grapheme-phoneme correspondence in the Malay texts provided the evidence for the extent of grapheme-phoneme transparency in Malay, relative to English.

Fig.1 and Fig.2 portray the pictorial representations of the analysis results. An example is given to help readers understand the grapheme-phoneme correspondence graphs. As shown in Fig.1, which portrays the grapheme-phoneme correspondence in the English words, the word example A has five graphemes (represented by the point A1) but only two phonemes (represented by the point A2). Meanwhile, the review

TABLE 2  
Grapheme-phoneme correspondence in English and Malay words

Language	Text (Story)	Grapheme-phoneme correspondence		Number of word-types analyzed (exclude non-words and proper nouns)
		Direct match (e.g. three letters and three phonemes)	No direct match (e.g. three letters but only two phonemes)	
English	E1	31.0%	69.0%	213
	E2	30.0%	70.0%	210
Malay	M1	72.9%	27.1%	214
	M2	76.0%	24.0%	221

of the English data suggests that A could either be ‘where’ which is composed by the phonemes /w/ and /eə/ or ‘there’ which is composed by /ð/ and /eə/. Hence, the wider the difference in the corresponding points in the graph, the less perfect the grapheme-phoneme correspondence will be.

A comparison of Fig.1 and Fig.2 suggests that the number of phonemes in Malay words more directly corresponds with the number of letters in the words, as compared to the English data. This is represented by a narrower spread of the line that represents the ‘number of phonemes’ away from the line that represents the ‘number of letters’ in Fig.2, compared to the pattern seen in Fig.1. The patterns in both graphs showed that there are more words

in English with inconsistent grapheme-phoneme correspondences as compared to Malay. The pattern also reveals that the inconsistency between the number of grapheme and phoneme of a word varies more in English compared to the Malay words.

Taken together, the graphs and the calculations in this section provide empirical data which indicate that the proportion of direct grapheme-phoneme correspondences in Malay is higher than in English. In addition, the data also revealed that despite the transparency of the Malay language, as asserted by Miles (2000), there are still inconsistencies in certain grapheme-phoneme correspondences. These inconsistencies may cause difficulty

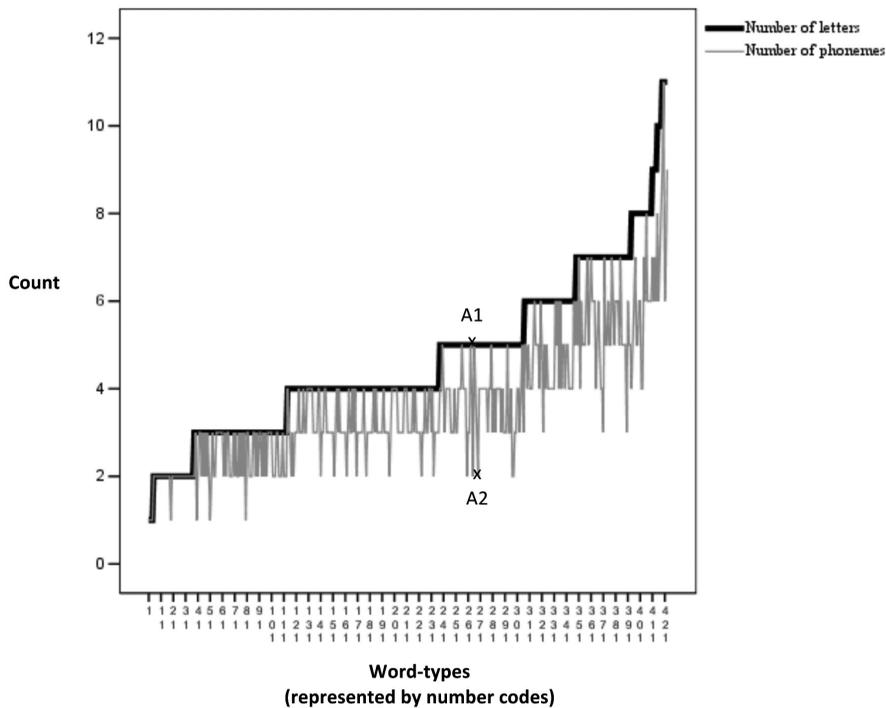


Fig.1: The grapheme-phoneme correspondence in the English words



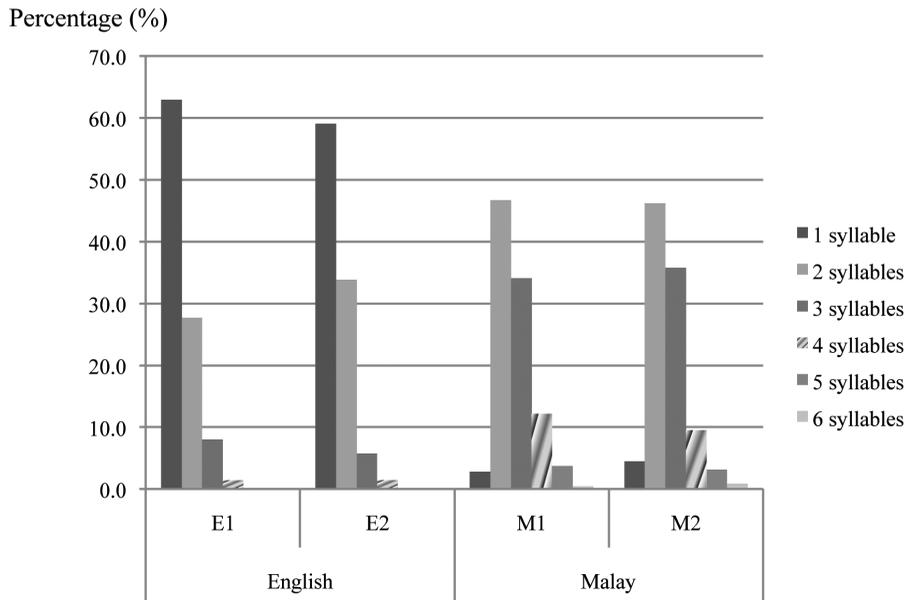


Fig.3: The number of syllables in the words

types used in Malay than in English. The inflectional morphemes used in Malay included prefix (e.g. ber+cakap), suffix (e.g. arah+an), circumfix (e.g. men+dapat+kan), reduplication (e.g. ikan-ikan) and affixed reduplication (men+jerit-jerit). These words involved manipulations at word (e.g. ikan-ikan) and syllable (e.g., ber+cakap) levels. In addition, the English morpheme rules are less transparent and less predictable than Malay since about 10% of the inflected words in English are in irregular forms (refer to E1 and E2 in Fig.4). Words with irregular forms involve manipulations with the smaller linguistic units, i.e., the phonemes. Some examples of these irregular words are might (may), sold (sell) and men (man). We referred to the phonemic variations in these irregular words as manifestations of phoneme-level manipulations. Nonetheless,

this type of phoneme-level manipulations is totally absent in the Malay data.

The overall findings confirmed our initial descriptions that there are more words in Malay with direct grapheme-phoneme correspondence and with multi-syllabic structures relative to English. The analysis of inflectional morphemes added to our understanding that as compared to English, there is more manipulation at the syllabic level than at the phonemic level for the Malay inflected words.

Given this set of findings, it seems crucial to consider the characteristics of the syllable in the selection of word stimuli in a reading intervention program. For this purpose, we investigated the distribution of words with different syllabic structures for the root words used in the Malay texts. Table 3 presents the data on the

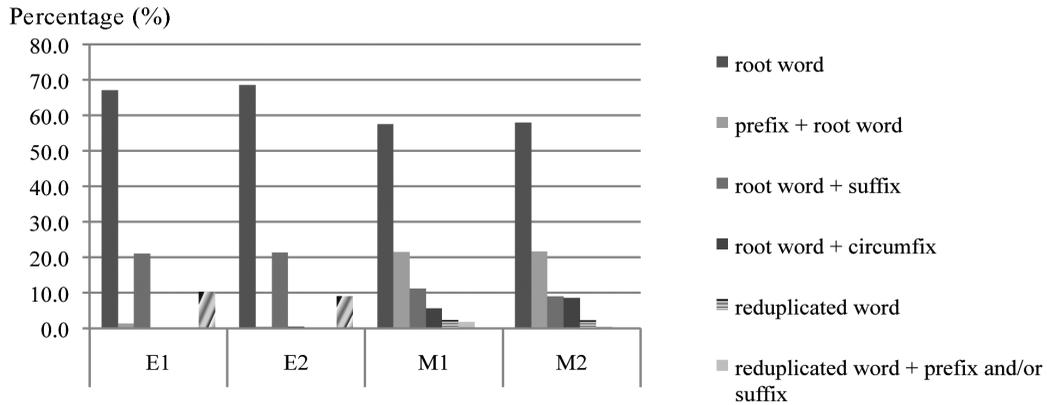


Fig.4: Inflectional morphemes in the words

most frequent occurring word structures in the Malay texts. Then, the most frequent occurring syllabic structures were deduced from this dataset.

Based on the data in Table 3, the most frequently occurring syllabic structures deduced are CV (highlighted in grey) and CVC (highlighted in black). These two syllabic structures formed the most frequently occurring word structures in mono-, bi-, tri- and quadri-syllabic words analyzed. For example, as shown in Table 3, 34.2% of bi-syllabic words in texts M1 and M2 are formed by the structure of CV+CVC (n=54) and 17.7% are formed by the structure of CV+CV (n=28). In fact, the most frequently occurring words in the texts are also these two types of bi-syllabic words. We listed these words in the appendix, together with words with the structure of V+CVC (n=14) and CVC+CVC (n=12).

## DISCUSSION

In this study, a contrastive word analysis was conducted to two English and Malay

translated children's stories. The results indicated significant cross-language differences for the texts analyzed (E1, E2, M1 and M2). The cross-linguistic differences were found in the grapheme-phoneme correspondence, syllabic structures and inflectional morphemes. Consistent with our initial presumptions, there were more words in the Malay texts with direct grapheme-phoneme correspondences than the English texts. Second, there were very few words in the Malay texts with mono-syllabic structure as compared to the words in the English texts. These two differences were proven to be statistically significant. Third, relative to English, more manipulations take place at the syllable level of Malay words than at the phoneme level.

In regard to the second finding, it is important to make explicit that the dominance of English monosyllabic word that we found in this study is partly determined by the nature of the texts analyzed. Our finding corresponds with the finding from the English word corpus built by Masterson, Stuart, Dixon and Lovejoy

TABLE 3  
Syllabic structures in the words found in texts M1 and M2

	Syllable structures	Examples in the Malay texts	Number of items	Percent in each category	Total percent (n=197)
1 syllable	CVC	baik, lain, daun, jauh	4	36.4	2.0
	CV	di, ke, si	3	27.3	1.5
	<b>CVC</b>	dan, pun	2	18.2	1.0
	CVC	yang	1	9.1	0.5
	CV	kau	1	9.1	0.5
			11	100.0	
2 syllables	CV+ <b>CVC</b>	bebas, besar etc.	54	34.2	27.4
	CV+CV	beli, cuba etc.	28	17.7	14.2
	V+ <b>CVC</b>	amat, akan etc.	14	8.9	7.1
	<b>CVC</b> + <b>CVC</b>	lembut, nampak etc.	12	7.6	6.1
	CV+CVC	datang, kering etc.	10	6.3	5.1
	CV+CVC	sangat, jangan etc.	8	5.1	4.1
	<b>CVC</b> +CV	lembu, pergi etc.	5	3.2	2.5
	V+CV	ini, itu etc.	5	3.2	2.5
	<b>CVC</b> +CVC	terbang, kandang etc.	3	1.9	1.5
	CVC+ <b>CVC</b>	sungguh, pengsan etc.	3	1.9	1.5
	CV+CV	walau, kalau	2	1.3	1.0
	CV+V	dia, dua	2	1.3	1.0
	CV+VC	luar, tuan	2	1.3	1.0
	CVC+CV	bangga, sangka	2	1.3	1.0
	VC+ <b>CVC</b>	untuk	1	0.6	0.5
	V+CVC	ingin	1	0.6	0.5
	CV+CV	bunga	1	0.6	0.5
	CV+CVC	kenyang	1	0.6	0.5
	<b>CVC</b> +CV	sampai	1	0.6	0.5
	V+CVC	orang	1	0.6	0.5
	V+V	ia	1	0.6	0.5
	VC+CV	engkau	1	0.6	0.5
			158	100.0	

cont'd Table 3

3 syllables	CV+CV+CV	lelaki, berasa etc.	12	54.5	6.1
	CV+V+CV	suara, suatu	2	9.1	1.0
	CVC+CV+CV	gembira, kembali	2	9.1	1.0
4 syllables	V+V+CVC	ialah	1	4.5	0.5
	CV+CV+CV	telinga	1	4.5	0.5
	CV+CV+CV <u>C</u>	belakang	1	4.5	0.5
	CV+CV+CV <u>V</u>	teratai	1	4.5	0.5
	CV+CV+V	semua	1	4.5	0.5
	CV+CV+VC	keluar	1	4.5	0.5
			22	100.0	
4 syllables	CV+CV+CV+CV	matahari, daripada etc.	3	50.0	1.5
	CV+CV+VC+CV	keluarga	1	16.7	0.5
	CV+CVC+CV+VC	perempuan	1	16.7	0.5
	V+CV+CV+CV	apabila	1	16.7	0.5
		6	100.0		

C=consonant, V=vowel, C=digraph (/ng/, /ny/ and less frequently /gh/, /kh/, /sy/), V=diphthong (/ai/, /au/, /oi/).

(2010) who found that 2/3 of the words in the early reading materials for British children are monosyllabic words. Therefore, it seems that our claim pertaining to the distributional differences of mono-syllabic and multi-syllabic words in Malay and English texts is valid, particularly in relation to children's early reading materials. Correspondingly, these observed cross-linguistic differences are believed to have an impact on reading interventions, which typically make use of the common vocabulary that beginning readers are experiencing with.

The prominence of the syllabic structures in Malay words suggests that this needs to be incorporated into the design of reading instructional program in Malay. However, though it might be useful to refer

to the instructions of teaching English-learning children to read multi-syllabic words (Duncan & Seymour 2003), those instructions might not be appropriate. The reason is that prosody or stress pattern was found to influence the reading of multi-syllabic words in young English learners (Duncan & Seymour, 2003). On the other hand, Malay is not a stress-timed language as compared to English (refer to Aris, 2010, for a review). The syllables in Malay are generally produced with equal intervals of time (Platt & Weber, 1980), with stressed syllables produced with equal length and loudness (Juliah 1993). Hence, other than the quantity difference as we found in this study, there seems to be also other aspects of cross-linguistic differences that we need to

recognize when we design a Malay reading remedial program based on an English framework. Other syllable-timed languages, such as Spanish (Mathes, Linan-Thompson, Pollard-Durodola, Hagan & Vaughn 2001) form complementary materials that can be referred to.

Cross-language interaction and bilingualism are becoming increasingly common. This cross-linguistic analysis is important in relation to language acquisition for the bi- or multilingual individuals. In the Malaysian context, we aim to use the current findings to inform our future work on Malay-English bilingual children who are struggling to read in schools that have bi- and trilingual policies. Despite both languages having similar alphabetic scripts, there are significant word structure differences to indicate that cross-language transfer may not be symmetrical across both languages as was found by previous researchers such as Mishra and Stainthorp (2007). The results from this comparative analysis also provide a reference on cross-linguistic differences between Malay and English to help non-Malay readers to understand Malay better.

This study provides empirical data revealing the prominence of CV and CVC structures in Malay texts. The most frequent occurring word structures in the Malay texts were found to be bi-syllabic word structures, including the structures of CV+CVC, CV+CV, V+CVC and CVC+CVC. This indicates that bi-syllabic word structures are the word structures that young Malay readers are most likely to encounter. This evidence set the rationale for us to use bi-

syllabic words as the major set of word stimuli in our early reading intervention programme. This contrasts with reading intervention programs designed for English-speaking children, which typically used mono-syllabic words as the primary word stimuli in tasks to promote phonological awareness (e.g. Gillon, 2008; Lindamood & Lindamood, 2007).

There are two major implications when bi-syllabic word stimuli are used in early reading intervention programmes. First, syllabic manipulation skills have to be taught in addition to phonemic manipulation skills when learning to decode at an early word reading level. Second, cognitive load increases when two types of grain-sizes (syllable and phoneme) are manipulated at the word-level. These two implications provide additional challenges that need to be considered when developing early reading intervention programs for transparent languages such as Malay. In English, it is the opacity of grapheme-phoneme correspondences that impose the challenge for reading intervention. Different aspects of languages present different challenges to learners. Therefore, despite the transparency of Malay, there are still a sizable group of children who are facing difficulties in reading. For example, a study conducted in the state of Penang identified that 9.4% of Grade 1 students in elementary schools had learning disabilities. Amongst these students, 92.3% of them were found to have severe reading disabilities (Socio-economic & Environmental Research Institute Penang, 2003).

## LIMITATION OF RESEARCH

A limitation of our research is that we conducted our analysis based only on two English texts and two Malay texts. However, despite the smaller number of texts used, there was a high inter-text consistency for the results obtained. Based on this, we felt that as a preliminary study, the dataset that we analyzed was adequate to give some baseline data. With the direction provided by this preliminary study, we will now proceed to include more texts in our database. A word corpus will be built from this database so that word stimuli can be selected to form the basis for the Malay reading intervention programme.

## CONCLUSION

In conclusion, before developing a reading intervention programme for poor readers in a language, we suggest that it is important to first identify the orthographic and phonological components of the word structures in the language. In addition, we advocate that cross-language comparisons need to be conducted, if English intervention programmes are to be used as a basis for developing reading intervention programmes for other languages.

## REFERENCES

- Aris, M. (2010). *Stress and intonation between Malay and English* [Electronic Version]. Retrieved on Sept 2, 2010, from <http://linguallyspeaking.wordpress.com/2009/11/04/stress/>
- Awang, S. (2004). *Teras pendidikan bahasa Melayu: Asas pegangan guru* [Core of Malay language education: Teachers' foundation beliefs]. Bentong: PTS Publications Sdn Bhd.
- Benjamin, G. (2009). Affixes, Austronesian and iconicity in Malay. *Bijdragen tot de Taal-, Land- en Volkenkunde*, 165(2/3), 291-323.
- DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.
- Duncan, L. G., & Seymour, P. H. K. (2003). How do children read multisyllabic words? Some preliminary observations. *Journal of Research in Reading*, 26(2), 101-120.
- Ellis, N. C., Natsume, I., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H. P., Polyzoë, N., Tsipa, M-L, & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, 39(4), 438-468.
- Furnes, B., & Samuelsson, S. (2009). Preschool cognitive and language skills predicting Kindergarten and Grade 1 reading and spelling: a cross-linguistic comparison. *Journal of Research in Reading*, 32(3), 275-292.
- Gillon, G. (2008). *The Gillon phonological awareness training program: an intervention program for children at risk for reading disorder* (Program handbook). Christchurch: University of Canterbury.
- Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia*, 6, 133-151.
- Hamdan, H. A. R. (1988). Dasar pendeskripsian sistem fonologi bahasa melayu [Basic description of Malay phonology]. In Farid, M. O. (Ed.), *Bunga rampai fonologi bahasa melayu*. Petaling Jaya: Penerbit Fajar Bakti Sdn. Bhd.
- Juliah, M. B. (1993). Stress and meaning: Malay UKM students' ability to apply English word stress. *Academic exercise*. Universiti Kebangsaan Malaysia.

- Kamus Dewan Edisi Keempat. (2005). Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Karim, N. S., Onn, F. M., Haji Musa, H., & Mahmood, A. H. (2004). *Tatabahasa Dewan*. Ed. baharu. [Malay grammar. New edition]. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In Frost, R., & Katz, L. (Eds.), *Orthography, phonology, morphology and meaning* (pp. 67-84). Amsterdam: New Holland.
- Lay, J. K. L. (2005a). *Tom Thumb and Thumbelina. 2 stories*. Penang: Monosetia Sdn. Bhd.
- Lay, J. K. L. (2005b). *Tom Thumb dan Thumbelina. 2 cerita*. Penang: Monosetia Sdn. Bhd.
- Lee, L. W. (2008). Development and validation of a reading-related assessment battery in Malay for the purpose of dyslexia assessment. *Annals of Dyslexia*, 58, 37-57.
- Lee, L. W., Wheldall, K., Beaman, R., & Madelaine, A. (2009). A conceptual framework for remedial reading instruction for low-progress early readers in Malaysia. *Asia Pacific Journal of Speech, Language and Hearing*, 12(1), 57-70.
- Lee, L. W. (2010). Severe dyslexia: A treatment case study. Research monograph. Universiti Sains Malaysia.
- Lee, L. W., & Wheldall, K. (2010). Acquisition of Malay language word reading skills: Lessons from low progress early readers. *Dyslexia*.
- Lindamood, C., & Lindamood, P. (2007). *The Lindamood phoneme sequencing (LiPS) program*. Austin: Pro-Ed, Inc.
- Mishra, R., & Stainthorp, R. (2007). The relationship between phonological awareness and word reading accuracy in Oriya and English: A study of Oriya-speaking fifth-graders. *Journal of Research in Reading*, 20(1), 23-37.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3<sup>rd</sup> ed.). Mahwah: Lawrence Erlbaum Associates.
- Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology*, 101, 221-242.
- Mathes, P. G., Linan-Thompson, S., Pollard-Durodola, S. D., Hagan, E. C., & Vaughn, S. (2001). *Lectura proactiva para principiantes: Intensive small group instruction for Spanish speaking readers*. Austin: Vaughn Gross Center for Reading and Language Arts, University of Texas.
- Miles, E. (2000). Dyslexia may show a different face in different languages. *Dyslexia*, 6, 193-201.
- Oxford English Dictionary. (1989) (2<sup>nd</sup> ed.). Oxford: Oxford University Press.
- Pallant, J. (2007). *SPSS Survival Manual: A step-by-step guide to data analysis using SPSS for Windows* (version 15) (3<sup>rd</sup> ed.). New South Wales: Allen & Unwin.
- Platt, J., & Weber, H. (1980). *English in Singapore and Malaysia: Status, features and functions*. Kuala Lumpur: Oxford University Press.
- Socio-economic & Environmental Research Institute Penang. (2003). *Learning difficulties among Standard 1 pupils in Penang, 2002*. Penang: Socio-economic & Environmental Research Institute Penang.
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2-40.
- Wydell, T. N., & Butterworth, B. (1999). A case study of an English-Japanese bilingual with monolingual dyslexia. *Cognition*, 70, 273-305.

- Xuereb, R. (2009). Literacy skills in Maltese–English bilingual children. *Journal of Research in Reading, 32*(3), 327-348.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 2-29.

**APPENDIX**

Words with the most frequently occurring word-structures in M1 and M2.

CV+CVC (n=54)		CV+CV (n=28)	V+CVC (n=14)	CVC+CVC (n=12)
bawah	musim	beli	akan	bandar
bebas	pasar	budi	amat	hendak
benih	penuh	cuba	anak	kuntum
besar	perut	dari	arah	lembut
bijak	pokok	diri	asal	mondok
boleh	rumah	hari	atas	muncul
bukan	salah	hati	awak	nampak
cepat	sayap	jadi	ayah	rumput
comel	sedar	jari	ayam	sambil
cukup	sejuk	juga	ikan	sarkas
dalam	sihat	kaki	ikut	sempat
dapat	tanah	kami	itik	tempat
dapur	tebal	kata	izin	
dunia	telah	kaya	oleh	
gadis	tetap	kayu		
gagal	tidak	kuda		
halus	tidur	lagi		