# WHAT ENGLISH COLLOCATIONS TO TEACH FIRST TO INDONESIAN PRESCHOOL CHILDREN? A MINI-CORPUS BASED RESEARCH OF INDONESIAN CHILDREN'S STORYBOOKS

**Maryani[1]**

*Maranatha Christian University, Bandung*

gepeicute@yahoo.com

**Abstract:** Frequent content words in a mini-corpus of 131 Indonesian children's storybooks, which resulted in a corpus of 134, 320 words, were investigated. The result was used to identify what English collocations to teach first to Indonesian preschool children. The data were run through a collocation menu in MonoConc Pro, a corpus program. To identify the frequent collocations in the corpus, the preceding and following words from each frequent lemma were analyzed. All the data were calculated in terms of the whole corpus and normalized per 100,000 occurrences. The result showed that the children were already exposed to various collocations; however, it turned out that several English and Indonesian collocations were not similar in terms of syntax and morphology. Teachers cannot literally introduce those collocations to children.

**Key words**: corpus, frequent collocation, content words

**Abstrak:** Investigasi dilakukan pada kata-kata yang sering muncul dalam sebuah korpus mini dari 131 buku cerita anak-anak Indonesia yang menghasilkan sebuah korpus yang terdiri dari 134.320 kata. Hasil analisis digunakan untuk mengidentifikasi kolokasi Bahasa Inggris apa yang harus diajarkan pertama kali pada anak-anak program PAUD (Pendidikan Anak Usia Dini). Data diolah dengan menu kolokasi pada program korpus bernama *MonoConc Pro*. Untuk mengidentifikasi kolokasi-kolokasi yang sering muncul pada korpus, analisis dilakukan pada kata-kata yang ada sebelum dan sesudah *lemma*. Semua data dihitung berdasarkan keseluruhan korpus dan dinormalisasikan per 100.000 kemunculan. Hasil analisis menunjukkan bahwa anak-anak telah terpajan pada berbagai macam kolokasi; akan tetapi, beberapa kolokasi Bahasa Inggris dan bahasa Indonesia tidak sama dalam hal sintaksis dan morfologinya. Para guru secara literal tidak dapat memperkenalkan kolokasi-kolokasi tersebut pada anak-anak.

**Kata kunci:** korpus, kolokasi yang sering muncul, kata-kata konten

In recent era, the number of corpus research keeps on increasing due to the easier access of online corpora. The available data in the corpus are various as it was stated by Biber (1993), a corpus can be built on different kinds of written texts and spoken data as long as the samples

---

[1] Maryani has been teaching English since 2003 until present at English Department, Faculty of Letters, Maranatha Christian University, Bandung.

cover the range of the language usage. This condition applies to both a general and specialized corpora. Meyer (2002) later pointed out that "because corpora consist of texts (or parts of texts), they enable linguists to contextualize their analyses of language; consequently, corpora are very well suited to more functionally based discussions of language" (p. 6). As Hunston (2002) further added that corpus analysis is one of the ways to collect the natural language use because without any evidence from language use, commonly people are not consciously aware of word, phrase, structure frequency. Therefore, through the use of corpus, most frequent words in a language can be shown and those frequent words in a corpus can contribute to what collocations are frequent and important to be taught in a particular genre. Since those collocations appear more frequently than the other collocations in a particular genre, it is necessary for teachers to introduce those collocations first to the students.

Robbins and Ehri (1994) pointed out that if young children's existing vocabulary is already developed, it is easier for them to comprehend and remember the meaning of new words. Based on their view, I am certain that Indonesian preschool children will comprehend English collocations more easily if they have already had the concepts of the collocations in their first language. As the use of the English language has been spread out throughout the non-English speaking countries in the world and has increased in the past few years, the teaching and learning of English language starts as early as possible, sometimes as early as the preschool level. There is a growing demand for early second language learning in Indonesia and a need to design an English curriculum that is not solely based on teachers' intuition. One element that informs a more informed selection of materials is the assumption that children's literature might contribute to the selection of words for teaching preschool children.

Therefore, this present research focuses on what English collocations are most important to teach to children of the age four to six. To be able to find out what English collocations would be relevant to teach to this age group, first I must know what Indonesian collocations that the children have already exposed to. The first language exposure might be from songs, movies, storybooks, teachers' instructions, parents' talk, or others' children's talk. In this research, I only examine the collocations appear in Indonesian children's storybooks as it is in accordance with Robbins and Ehri's (1994) findings, in which listening to stories helps improve kindergarten children's vocabulary growth. Senechal (1997) futher added that young children's vocabulary grows as they are more exposed to certain words when they listen to a story more than one time. Leung (1992) also supported the idea of the importance of word frequency for children vocabulary acquisition, especially for kindergarten and first grade students.

This research focuses on finding out the L1 children's conceptual collocations to identify the L2 collocations to teach. Chen (1992), Lotto and de Groot (1998), Justice, et. al. (2005), and Wolter (2006) have claimed that there is a strong connection between L1 knowledge and L2 acquisition; however, to my knowledge, none of the studies addressed the identification of L2 core vocabulary based on the L1 knowledge. There is a tendency that once children have acquired collocations in their L1, they will use the L1 underlying concepts of the collocations to comprehend the L2 collocations. As it is in line with Justice, et. al. (2005) who claimed that children are able to acquire a new word when they already understand the underlying concept,

for example a child who understands the concept of bigness will be able to recognize big, large, huge, etc. Lotto and de Groot (1998) further added that high-frequency words are the words that are not only useful but also fairly easy to acquire. In 2006, de Groot found out that the students are more familiar with the concepts that they encounter more often either in written or spoken discourse. Once the students know the concepts in their L1, it is easier for them to obtain new information in L2.

In addition, Wolter (2006) argued that learners' L1 conceptual knowledge can influence L2 vocabulary acquisition. He further explained that learners are able to infer some possible combination of L2 words, but they do not necessarily use their L1 knowledge to understand new L2 concept. However, Wolter suggested that L2 learners should not rely on their L1 knowledge of collocations in constructing L2 collocations. He concluded that these learners, who greatly depend on their L1 collocation knowledge in constructing L2 word combinations, sometimes produce unusual collocations. Based on Wales' A Dictionary of Stylistics (1994), collocation "refers to the habitual or expected co-occurrence of words…its collocates ranked in order of probability of occurrence (e.g. elephant: grey, white, pink, green)" (p. 76). Halliday, Teubert, Yallop, Cermakova (2004) further explained that a word usually co-occurs with another word and these words are likely to appear together; therefore, based on the frequent content words appear in a corpus, it is also important to look at the occurrence of these content words with other content words.

Teachers can introduce new vocabulary to children in various ways, such as listening and singing songs, playing games, and listening or reading storybooks. As Justice, et. al. (2005) argued that books can help children to learn specific words that they might not learn in everyday lives, such as sprouts, seahorses, and saucers. Vos (2007) further stated that new advanced vocabularies found in the storybooks help the children prepare for comprehending the texts that they will come across in the next stages of education. Therefore, it is important to know what words are frequently appear in children's storybooks as those words will be useful to help children expand their vocabulary knowledge. At present, the use of literature in the kindergarten classrooms for offering language exposure to students has become a trend in Indonesia. The storybooks are mainly written in Indonesian, which is the official language of Indonesia. However, the research on corpus analysis is still limited to a few attempts to build corpora of a different nature. Hardjadibrata (cited in Nazief, 2000, p. 1), for example, started a word analysis based on a corpus of Indonesian newspapers. Nazief (2000) did a replication study by conducting a research on an Indonesian written corpus for adults based on Kompas, an Indonesian national newspaper. Since the Indonesian corpora available are still limited to adult language, it is essential to build a corpus of children language.

In the area of children's corpora, Gil (2006) created a corpus of 500, 000 utterances of eight Jakartan Indonesian children. He focused his study on investigating the children's acquisition of two prefixes which show active and passive markers. Whereas, Arka and Simpson (2007) proposed to build a balanced corpus based on Jakartan Indonesian spoken data. The other child based corpora of the Indonesian language is still in the building process. The Max Planck Institute (2007) also deserves a mention for its attempt to conduct a longitudinal study

of ten young children and build a Jakartan Indonesian spoken corpus, mainly focused on the dialect spoken in Jakarta, the capital city of Indonesia.

Since none of the research has been done in the area of basic collocations for Indonesian preschool children and, as there has been a need in finding out what concepts Indonesian preschool children know in their first language, I built a mini-corpus of Indonesian storybooks in order to identify the frequent collocations Indonesian children are exposed to at their early age. These following questions become the focus of my data analysis: (1) What collocations commonly appear in the mini-corpus of Indonesian children's storybooks based on the frequent contents words collected?, and (2) What English collocations are most appropriate to teach first to Indonesian preschool children?

## METHODOLOGY

### Sources and instruments
The main source for the data is 131 Indonesian children's storybooks, which resulted in a corpus of 134,320 words. A different kind of storybooks was included in the data as long as they were either used by the teachers in the classroom, available to the students at schools or at home. These storybooks were not necessarily read by the children; but either their teachers or parents read the books to them. The program which was used to upload the data was MonoConc Pro. This program provided frequency count, concordance, sort and collocation menu, which helped the data analysis.

### Procedure
This data collection was conducted in early to mid December 2007 for about 10 days in three private kindergarten schools in Bandung, West Java, Indonesia. After all data was collected, it was processed and analyzed using the corpus program, MonoConc Pro. Each file from each book was uploaded to the corpus program to find out the frequency list of the words in the whole data. Afterwards, I selected the nouns, verbs, adjectives, and adverbs that appeared in the list. I made a list for each part of speech including the frequency number per 134,320 words and the English translation for each word. I also counted the corpus frequency per 100,000 words. Then I used a concordance menu in the program to find the contexts of a particular base word. Afterwards, I used a collocate frequency data menu to find the collocation. I searched the preceding and the following of each lemma. I also counted the corpus frequency for the whole data and per 100,000 words. I also provided a gloss and a translation for each collocation. To make sure that the translation was accurate, I also rechecked the English translation of the frequent words by looking at an Indonesian English dictionary.

## FINDINGS AND DISCUSSION

### Noun collocation

There are three combinations of the noun collocation: noun + noun, noun + verb, and noun + adjective.

Table 1 shows the collocation of the frequent nouns with other frequent nouns.
Table 1. Noun + Noun Collocation

| 1st left | Token (grouped by lemma) | 1st right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| | ayah | ibunya | ayah - ibu - nya<br><br>father - mother - his/her | his/her mom and dad | 19 |
| ayah | ibu | | ayah - ibu<br><br>father - mother | mom and dad | 15 |
| ayah | ibunya | | ayah - ibu<br><br>father - mother - his/her | his/her mom and dad | 15 |
| | malam | hari | malam - hari<br><br>night - day | evening | 15 |
| malam | hari | | malam - hari<br><br>night - day | evening | 14 |
| | ayah | ibu | ayah - ibu<br><br>father - mother | mom and dad | 13 |
| raja | hutan | | raja - hutan<br><br>king - jungle | a/the king of the jungle | 11 |
| | raja | hutan | raja - hutan<br><br>king - jungle | a/the king of the jungle | 10 |

As can be seen in the above table, there are only three kinds of noun + noun collocations, namely *mom and dad*, *evening*, and *king of the jungle*. The number in brackets following each collocation shows the number of occurrences of that collocation in the corpus. The most frequent collocation appeared to be *mom and dad* (19). There were common occurrences where his or her was attached to mom and dad to address one's mom and dad (*Lina berteriak takut karena **ayah ibunya** tak ada.*) [Lina was screaming because her mom and dad were not around.], but there were two occurrences where *nya* (his/her) was also used to address *their* (*Ketika Ratih dan Dian meminta ijin **ayah ibunya** untuk memberikan bantuan uang tabungannya. Tentu saja **ayah ibunya** mengijinkannya.*) [When Ratih and Dian asked for their mom and dad's permission to give a donation from their saving, of course, their mom and dad allowed them to do that.]. Regardless of the reference of *nya*, it is obvious that *mom and dad* collocate. It is interesting to see the syntactical difference between English and Indonesian collocation for *mom and dad*. In Indonesian, this collocation becomes *dad and mom*. There

are two possible reasons for this phenomenon. First, most Indonesian families adopt the patriarchal system that respects a father more than a mother. Second, a father is commonly the one who becomes the bread-winner for the family; therefore, a father is more respectfed than a mother.

The collocation *malam hari* (evening) and *raja hutan* (king of the jungle) appeared less frequent. It is interesting to notice that *malam* itself already refers to night but it can also be used as an adjective to describe another noun (*Dia itu binatang **malam** dan memangsa binatang yang keluar malam.*) [She is a night animal and preys on the animals that go out in the night.], whereas most of the times *malam hari* functions as an adverb of time (***Malam hari** dia keluar dari persembunyiannya.*) [In the evening, she went out from her hiding place.]. In the case of *raja hutan* (the king of the jungle), it was found that the collocation represented a lion in all the occurrences. The lion was described as a strong animal that had power like a king (*Tubuhnya yang kekar dan garang membuat para hewan hutan menganggapnya raja. Orang yang ingin selamat bila melintas hutan menyebutnya Embah atau Nenek.Sebutan halus seekor singa atau si **Raja Hutan** itu.*) [His strong and firm body made all the animals in the jungle consider him a king. People who wanted to be safe if they were crossing the jungle should call him *Embah* or *Nenek*. These are the polite nicknames for a lion or the king of the jungle.]. Table 2 shows the collocation of the frequent nouns with other frequent verbs.

Table 2. Noun + Verb Collocation

| 1st left | Token (grouped by lemma) | 1st right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| kata | ibu | | kata - ibu <br><br> say - mother | says/said mother | 34 |
| kata | putri | | kata - putri <br><br> say - princess | say/said princess | 21 |
| kata | bunda | | kata - bunda <br><br> say - mom | say/said mom | 19 |
| tanya | bunda | | tanya - bunda <br><br> ask - mom | ask/asked mom | 18 |
| makan | malam | | makan - malam <br><br> to eat - night | to have dinner | 13 |
| kata | ayah | | kata - ayah <br><br> say - father | says/said father | 11 |
| kata | kakek | | kata - kakek <br><br> say - grandfather/grandpa | say/said grandfather/grandpa | 10 |

From the above table, there are seven collocations that are considered frequent in the corpus. Five collocations are the attachment of the word *kata* (say) before the nouns (i.e. *kata ibu*, *kata putri*, *kata bunda*, *kata ayah*, and *kata kakek*). The other two collocations stand by themselves (i.e. *tanya bunda* and *makan malam*). The number in brackets following each collocation shows the number of occurrences of that collocation in the corpus.

The five collocations with the same verb, *kata* (say), are worthy of note. The nouns that collocate with *say* are mother, princess, mom, father, grandpa. It is fascinating to notice that *say mother* (28) appeared more frequently than the rest. Comparing *say mother* to *say father* and *say grandpa,* the first was used commonly in giving advice (*"Lain kali hati-hati ya,"* **kata bunda**.) ["Next time please be careful," said mom.], whereas the latter was generally used in statements; for example, (*"Kami selalu ada dalam hati kalian semua. Kapan saja kalian ingin bertemu, kita akan bertemu," **kata Kakek**.) ["We will always be in your hearts. Whenever you want to meet us, we will meet each other," said grandpa.]. A mother is the one who generally cares about the children and spends more time with them than a father or a grandpa; therefore, it is probably the reason why a mother, in the children stories, is someone who usually gives advice to children. *Say princess* (21) is also quite a frequent collocation in the corpus. This is possibly due to some of the stories present the life of a princess, but not about a prince and his life. As the princess becomes the main character in some stories, it is clear that there are quite frequent utterances said by the princess.

A similar tendency is also seen in the use of *tanya bunda* (ask mom). Since a mom usually spends more time with children, it is probably the reason for the frequent occurrence of this collocation. There are many occurrences in which a mom asks her children about any situations that happened throughout the day (*"Sedang apa anak-anakku, sayang?"* **tanya bunda**.) ("What are you doing, my dear?" asked mom.). The other frequent collocation which is important to address is *makan malam* (to have dinner). There is a difference in the use of the verb following *dinner* in English and Indonesian. In English, people commonly say *to have dinner*, whereas in Indonesian, the common collocation is *to eat dinner* (*"Ayo masuk sayang. Sebentar lagi waktunya **makan malam**," seru bunda. Dani dan kak Lola masuk ke rumah.*) ["Let's come inside, my dear. It's almost time to have dinner," said mom. Dani and sister Lola came into the house.]. Therefore, teachers need to address the difference of the collocations in the two languages. Table 3 shows the collocation of the frequent nouns with other frequent adjectives.

Table 3. Noun + Adjective Collocation

| 1st left | Token (grouped by lemma) | 1st right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| | anak | kecil | anak - kecil<br><br>child - small | a small child | 18 |
| | sungai | kecil | sungai - kecil<br><br>river - small | a small river | 16 |

As can be seen in the table, there are only two frequent noun and adjective collocations in the corpus (i.e. *small child*, and *small river*). The number in brackets following each collocation shows the number of occurrences of that collocation in the corpus. Among this frequent collocation, *small child* (18) appeared most frequently followed by *small river* (16). The noun *child* and *river* usually collocate with small. There was not found any collocation of *big child* or *big river*. It is fascinating to notice that the authors of children's stories seemed to focus more on *small child* than *big child*, and *small river* than *big river*. Since the target readers of those books are children, therefore, most of characters in the stories are young children. It is presumably the reason for the frequent collocation of *small child*. The *small river* collocation is also related to the children as the characters. Since they are small, they are commonly situated with small objects; for example, it will make more sense for a child to cross a small river by herself than a big river (*Alice menyeberangi **sungai kecil**, dan muncullah seorang ksatria naik kuda.*) [Alice went across the small river, and there appeared a knight on a horse.].

**Verb collocation**
There are three combinations of the verb collocation: verb + noun, verb + verb, and verb + adverb. Table 4 shows the collocation of the frequent verbs with other frequent nouns.

Table 4. Verb + Noun Collocation

| 1st left | Token (grouped by lemma) | 1st right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| tempat | tidur | | tempat - tidur<br><br>place - to sleep | a bed | 26 |
| | tanya | bunda | tanya - bunda<br><br>ask - mom | ask/asked mom | 18 |
| | adalah | anak | adalah - anak<br><br>am/is/are - child | am/is/are a child | 16 |
| | makan | malam | makan - malam<br><br>to eat - night | to have dinner | 14 |
| | kata | ibu | kata - ibu<br><br>say - mom/mother | say/said mom/mother | 12 |

Based on the above table, there are five frequent verb and noun collocations. The most frequent one is *bed*, and the least is *say mom*. The other three in between are *ask mom*, *to be* (*am/is/are*) *a child*, and *to have dinner*. The number in brackets following each collocation shows the number of occurrences of that collocation in the corpus.

Four out of the five collocations have been discussed in the noun + verb collocation (i.e. *bed*, *ask mom*, *to have dinner*, and *say mom*). Among these four, the number of occurrences is either the same or quite similar; however, it is essential to point out the different number of

occurrence between *say mom* in noun + verb (28) and in verb + noun (12). In the discussion of noun + verb, *mom* becomes the focus. It appeared that the preceding frequent lemma usually attached to *mom* is *say*. In other words, if children are exposed to the word *mom*, they probably know that a mom is usually the one who says something to them and they need to pay attention to what she has said. Nevertheless, in the verb + noun discussion, *say* becomes the spotlight. In this case, it appeared that the following lemma that collocates with say is *mom*. Therefore, it is possible that children might not always expect that someone who commonly says something is a mom.

One collocation that has not been discussed is *to be* that is followed by *child*. In Indonesian, *adalah* (to be) can be used to address a singular subject (*Aku **adalah** anak jelek yang hanya mementingkan diri sendiri.*) [I am a bad child who only cares about myself.] or plural subjects (*Sona dan Soni **adalah anak** kembar yang sangat mirip.*) [Sona and Soni are identical twins.]. The collocation of *to be* and *child* (16) appeared quite frequently in the corpus. The possible reason for this collocation is that most of the characters in the children's storybooks are children. Besides, *to be* that collocates with child was often used to refer to the description of a child or children as the main characters in the stories. Table 5 shows the collocation of the frequent verbs with other frequent verbs.

Table 5. Verb + Verb Collocation

| 1st left | Token (grouped by lemma) | 1st right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| sedang | bermain | | sedang - bermain am/is/are being - to play | am/is/are playing | 18 |
| ingin | tahu | | ingin - tahu to wish - to know | want to know | 17 |
| | ingin | tahu | ingin - tahu to wish to know | want to know | 17 |
| akan | menjadi | | akan - menjadi will/would - become | will/would become | 17 |
| akan | pergi | | akan - pergi will/would - go | will/would go | 16 |
| | sedang | tidur | sedang - tidur am/is/are being – to sleep | am/is/are sleeping | 12 |
| sedang | tidur | | sedang - tidur am/is/are to sleep | am/is/are sleeping | 12 |
| sudah | menjadi | | sudah - menjadi has/have - become | has/have become | 10 |

As can be seen in the above table, this group consists of five collocations that occurred in a balanced proportion (i.e. *sedang bermain*, *ingin tahu*, *akan menjadi*, *akan pergi*, *sedang tidur*, and *sudah menjadi*). Among these five collocations, they can be categorized into four main groups: *am/is/are playing* and *am/is/are sleeping*, *will become* and *will go*, *have become*, and *want to know*.

From the groups above, there is a similar pattern among the first two groups that refers to the use of auxiliary verbs as a time signal. The use of *sedang* (being) shows the present continuous tense, whereas the use of *akan* (will) in the second group describes the future tense. It is interesting to view that the verbs *to play* and *to sleep* were commonly preceded by *being*, whereas the word *being* is generally followed by *to sleep*. As it is widely known that the world of young children is typically filled with playing and sleeping, it is possible that the authors use the collocation of to *be playing* and *to be sleeping* to describe the daily situation of the characters (*Suatu hari ketika dia **sedang bermain** sendirian di halaman rumahnya tiba-tiba datanglah seorang anak gadis kecil yang cantik jelita.*) [One day when he was playing by himself in his yard, suddenly a beautiful young girl came.]; (*Adik **sedang tidur** nyenyak.*) [The little brother was sleeping well.].

The reference to the future tense, *akan* (will), collocates with *to become* and *to go*. It is common in Indonesian culture to talk about one's ambition in life by asking children what they are going to become in their future. In the children's stories, it was found that the collocation *will become* was used in the same context ("*Lihat, bayi yang memakai toga ini, kelak ia **akan menjadi** guru aritmatik," kata Peri Sinar.*) ["Look, this baby with a gown, one day he will become an aritmethics teacher," said Princess Light.]. Compared to *will become*, *will go* commonly functions as a time marker to describe one's decision for the near future (*"Bu, aku **akan pergi** ke kota selama beberapa hari untuk menjual beberapa ekor babi dan ayam…," katanya.*) ["Mom, I will go to the city for a few days to sell some pigs and chickens…," said him.].

Besides the use of present continuous and future tense, the use of present perfect tense also came into view in the collocation of *sudah menjadi* (has/have become). This collocation refers to the result of a situation that seems to show children the condition at present and the effect of that condition to another situation (*Hari **sudah menjadi** gelap. "Apa yang harus kita lakukan teman-teman?" keluh Crystal bingung.*) [The day has become dark. "What should we do, friends?" complained Crystal confusedly.].The last group, *ingin tahu* (want to know), seems to be related to children's notion of being eager to know many new things around them. There is a significant difference for the use of *want to know* in English and in Indonesian. It seems awkward in English not to mention the object after *want to know*, but in Indonesian, the collocation *want to know* appears in the end of a sentence, without being followed by an object (*Noni sebenarnya tak bandel kok. Dia gadis cerdas yang selalu **ingin tahu**.*) [Actually Noni is not naughty. She is a smart girl who always wants to know.]. Therefore, it is important for teachers to address the grammatical difference between English and Indonesian use of the collocation. Table 6 shows the collocation of the frequent verbs with other frequent adverbs.

Table 6. Verb + Adverb Collocation

| 1<sup>st</sup> left | Token (grouped by lemma) | 1<sup>st</sup> right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| | mulai | sekarang | mulai - sekarang<br><br>to start - now | to start now | 12 |
| segera | pulang | | segera - pulang<br><br>immediately - go/return home | to go/return home immediately | 12 |

There were only two frequent verb and adverb collocations (i.e. *mulai sekarang* and *segera pulang*). These collocations appeared in balance. It was found that most occurrences of *mulai sekarang* (to start now) are related to the characters' promises (*"Kakek! Nenek! Saya minta maaf atas perbuatan-perbuatan saya. **Mulai sekarang**, saya akan membantu Kakek dan Nenek dengan sungguh-sungguh." Pinky meminta maaf dengan hati yang tulus.*) ["Grandpa! Grandma! I am so sorry for my actions. Starting now, I will help grandpa and grandma seriously." Pinky sincerely asked for an apology.] or commands ((**Mulai sekarang**) *jangan lagi berkelahi dengan penduduk dataran rendah.*) [Starting now, don't ever fight with the citizens of the low hill.]. It is obvious that *to start now* is a useful collocation that emphasizes the event of a promise or command should take place. Children are exposed to these collocations perhaps to let them know that a promise should be fulfilled as soon as possible. Besides, it also lets children understand that a command is usually expected to be done right away.

Different from *mulai sekarang*, *segera pulang* (to go home immediately) is used most of the time in showing the characters' urgent decision (*"Tapi aku harus **segera pulang** karena Nenek sedang sakit," kata Peter.*) ["But, I have to go home immediately because grandma is being ill," said Peter.]. The authors probably want to strengthen the urgency of going home, especially in certain conditions where there is somebody who needs attention. However, due to the tropical weather in Indonesia, the sun generally sets around 6:00 pm every single day of the year. Parents commonly ask children to go home or to go inside the house if it is already dark. This condition might be the reason why the collocation is quite frequent in the Indonesian language.

**Adjective collocation**
There is only one combination of the adjective collocation: adjective + noun.
Table 7 shows the collocation of the frequent adjectives with other frequent nouns.
Table 7. Adjective + Noun Collocation

| 1<sup>st</sup> left | Token (grouped by lemma) | 1<sup>st</sup> right | Gloss | Translation | Overall Frequency |
|---|---|---|---|---|---|
| sungai | kecil | | sungai - kecil<br><br>river - small | a small river | 16 |

| anak | kecil | | anak - kecil | a small river | 15 |
| --- | --- | --- | --- | --- | --- |
| | | | child - small | | |

Based on the above table, there is only one adjective, *small*, that can be preceded by two different nouns, *river* or *child*. The preceding nouns do not greatly influence the number of the collocation occurrence. These collocations have been discussed in the noun + adjective collocation. The main difference is that the noun + adjective collocation was sorted out from the noun as a lemma or base of the collocation, but this adjective + noun collocation was based on the adjective. However, there is no significant difference in the number of occurrences.

**Adverb collocation**
There are two combinations of the adverb collocation: adverb + verb, and adverb + adverb. Table 8 shows the collocation of the frequent adverbs with other frequent verbs. Table 8. Adverb + Verb Collocation

| 1st left | Token (grouped by lemma) | 1st right | Gloss | Translation | Overall Frequency |
| --- | --- | --- | --- | --- | --- |
| | pasti | akan | pasti - akan<br><br>certainly - will/would | will/would certainly | 21 |
| | hanya | bisa | hanya - bisa<br><br>only - can/could | can/could only | 16 |
| datang | lagi | | datang - lagi<br><br>to come - again | to come once again | 13 |
| | segera | pulang | segera - pulang<br><br>immediately - to go/return home | to go/return home immediately | 12 |

From the above table, there are four adverb + verb collocations (i.e. *will certainly*, *can only*, *to come once again*, and *to go home immediately*). The most frequent one in the corpus is *will certainly* (21). In English, these adverbs can be put either before the verb or after the verb. This cannot be applied in Indonesian. Certain adverbs can only be located before the verbs, and the others are after the verbs. Since the last adverb + verb collocation has been discussed in the verb + adverb collocation, this collocation, *to go home immediately*, will not be discussed in this section. There are three other collocations that will be discussed (i.e. *will certainly*, *can only*, and *to come once again*). The first collocation was often used to emphasize one's strong will (*"**Pasti akan** kudapatkan!" serunya penasaran sambil terus mengejar.*) ["I will certainly get it!" said him while he was continuously chasing.]. The reason of the use of this collocation is probably to show children one's strong will to achieve something. In real life, this can bring a good or bad result. It depends on what kind of will it is.

Compared to *will certainly*, *can only* commonly shows one's limitation of doing something (*Tong-tong juga ingin makan. Seandainya tadi ia membawa bekal, tentu sekarang ia tidak kelaparan. Sekarang ia **hanya bisa** memandang Cici dengan perut lapar.*) [Tong-tong also wanted to eat. If he brought his lunchbox, he would not be starving. Now he could only stare at Cici with his empty stomach.]. It is fascinating to see that children are also exposed to the collation of *can only* that shows one's limitation in a situation. The authors do not only expose children to the events in the stories that describe the winning situations, but also to some situations where a character cannot do everything but can only do something. Doing a thing is not always considered bad, as long as someone can do something one time and do something else at another time, it is perfectly fine.

*To come once again* was found to be used in addressing one's request for the future and one's action in the future. This collocation seems to show that the specific action will be repeated in the near future (*Kakek! Nanti **datang lagi** ke sini, ya.*) [Grandpa! Come again later, ok.]. It is possible that the authors want children to notice the difference between *to come* and *to come once again*. Therefore, if they want to ask someone who has already visited their house, they need to say *datang lagi* (to come once again) instead of *datang* (to come). Table 9 shows the collocation of the frequent adverbs with other frequent adverbs.

**Adeverb+Adver Collocation**
Table 9. Adverb + Adverb Collocation

| 1st left | Token<br><br>(grouped by lemma) | 1st right | Gloss | Translation | Corpus Frequency |
|---|---|---|---|---|---|
|  | tiba-tiba | saja | tiba-tiba - saja<br><br>suddenly - just | just suddenly | 16 |
| tiba-tiba | saja |  | tiba-tiba - saja<br><br>suddenly - just | just suddenly | 10 |

The above table shows that there was only one adverb + adverb collocation found in the corpus, i.e. *just suddenly*. The different location of one adverb influenced the number of occurrences, but there was not any significant difference between two of them.

The uses of *just suddenly* were specifically related to unexpected events (*Blar, blar! **Tiba-tiba saja** petir berkilat, dan hujan turun.*) [Blar, blar! Just suddenly there was lightning and rain.]. The characters did not realize something would happen, but it then happened. This collocation probably was used to give a sense of suspense to the children. The stories will be more attractive if there is some suspense, especially for children. The parents or teachers can emphasize this collocation and pause for a few seconds before continuing with the rest of the stories. This collocation can also provide a chance to ask the children what will happen in the stories.

**CONCLUSION**

In terms of introducing children with English collocations, it was found that most collocations in Indonesian and English collocations are not similar. Among the Indonesian frequent collocations, there was one noun collocation (i.e. *ayah dan ibu* (mom and dad)) and one verb collocation (i.e. *makan malam* (to have dinner)) that convey similar meanings. There is a syntactical difference between the noun collocation, in which Indonesian uses *dad and mom* whereas English uses *mom and dad*; and a morphological difference between the verb collocation, in which Indonesian uses *to eat dinner* whereas English uses *to have dinner*; however, teachers can introduce these collocations to children as well as explain the difference between English and Indonesian collocations. Related to verb collocation, there was a pattern of exposure for the Indonesian verbs: *copula be*, *to want*, and two auxiliary verbs (i.e. *will* and *has/have*). As it is similar in English, *copula be* is the most common verb, whereas *to want* is a transitive verb that can follow a noun, a pronoun, or infinitival verb. The adjective and adverb collocation are closely related to language specific, which means the Indonesian collocations that were found do not correspond to the English collocations; therefore, the frequent collocations in these categories cannot be literally introduced for teaching English to children.

The result of the study can be used as a guideline for preschool teachers to know what Indonesian collocations preschool children are already exposed to. Teachers might apply the knowledge in English language instruction and pinpoint the similarity or difference between English and Indonesian collocations. Besides, the findings can be an insight for preschool curriculum designers to choose the most appropriate English collocations to teach. The mini-corpus built will be beneficial for any researchers who are interested in examining Indonesian children's storybooks.

**REFERENCES:**

Arka, I. W. & Simpson, J. (2007). Preceedings from The Eleventh International Symposium on Malay/Indonesian Linguistics: *Challenges of developing a balanced and representative corpus for Indonesian ParGram*. Manokwari, Indonesia.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), pp.243-257.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.

Chen, H. C. (1992). Lexical processing in a non-native language: Effects of language proficiency and learning strategy. *Memory and Cognition*, 18, pp. 279-288.

de Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, *56*(3), pp. 463-506.

Gil, D. (2006). The acquisition of voice morphology in Jakartan Indonesia. In Gagarina, N. & Gulzow, I. (Eds.). *The acquisition of verbs and their grammar: The effect of particular languages*. (pp. 201-227). Dordrecht, the Netherlands : Springer.

Halliday, M. A. K., Teubert, W., Yallop, C. & Cermakova, A. (2004). *Lexicology and corpus linguistics*. New York: Continuum.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.

Justice, L. M., Pence, K. L., Beckman, A. R., Skibbe, L. E., & Wiggins, A. K. (2005). *Scaffolding with storybooks*. Newark, DE: International Reading Association.

Leung, C. (1992). Effects of word-related variables on vocabulary growth through repeated read-aloud events. In C. Kinzer & D. Leu (Eds.), *Literary research, theory, and practice: Views from many perspectives: Forty-first yearbook of the National Reading Conference* (pp. 491-498). Chicago: National Reading Conference.

Lotto, L. & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, pp. 31-69.

Meyer, C. F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge, UK: Cambridge University Press.

Nazief, B. (2000). Proceedings from 15th Indonesian Scientific Conference in Japan: *Development of computational Linguistics research: A challenge for Indonesia*. Tokyo, Japan.

Robbins, C. & Ehri, L. (1994). Reading storybooks to kindergarten helps them learn new vocabulary words. *Journal of Educational Psychology, 86*(1), pp. 54-64.

Senechal, M. (1997). The differential effect of storybook reading on preschooler's acquisition of expressive and receptive vocabulary. *Journal of Child Language*, *24*(1), pp. 123-138.

Vos, Jeannette. (2007). Can preschool children be taught a second language? Retrieved July 27, 2007 from http://www.earlychildhoodnews.com/earlychildhood/article_view.aspx?ArticleId=60.

Wales, K. (1994). *A Dictionary of Stylistics*. New York, USA: Longman Inc.

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*(4), pp. 741-747.