

Identifying Indonesian-core vocabulary for teaching English to Indonesian preschool children: a corpus-based research

Maryani

English Department, Faculty of Letters, Maranatha Christian University
Suria Sumantri 65, Bandung, 40164, Indonesia
e-mail: gepeicute@yahoo.com

Abstract: This corpus-based research focuses on building a corpus of Indonesian children's storybooks to find the frequent content words in order to identify Indonesian-core vocabulary for teaching English to Indonesian preschool children. The data was gathered from 131 Indonesian children's storybooks, which resulted in a corpus of 134,320 words. These data were run through a frequency menu in MonoConc Pro, a corpus program. Data analysis was analyzed by selecting the frequent nouns, verbs, adjectives, and adverbs before each of them was lemmatized. The result showed that the children were already exposed to both ordinary and imaginative concepts, antonym in adjective, time reference, and compound nouns. The narrative discourse clearly influenced the kind of verbs the children exposed to.

Key words: corpus, vocabulary, content words

There has been a lot of corpus research done in these past years. The availability of online corpora has greatly facilitated this kind of research. Corpus study has attracted many researchers due to the real linguistics data that appear in the corpus. Corpus can also provide guidance for finding language patterns and its usage in real-life situations. Sinclair (1996) defined a corpus as a collection of written text and spoken data which is used to find a particular linguistic phenomenon. The data chosen for a corpus can be varied such as from written register (documents, newspapers, emails, blogs, students' writings, fictions etc.) and from spoken register (telephone conversations, interviews, and daily-life conversations).

In addition, Meyer (2002) 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). Based on his principles, I noticed that a corpus can show the most frequent words in a language. Therefore, those frequent words in a corpus can contribute to what words are necessary to be taught in a particular genre. Since those words appear more frequently than the other words in a particular genre, it is more important to introduce those words first to the students.

As a matter of fact, Robbins and Ehri (1994) pointed out that young children can understand and remember the meaning of new words easily if their existing vocabulary is already developed. Based on their argument, I am certain that Indonesian preschool children will comprehend English words more easily if they have already had the concepts of the words in their first language.

There has been a growing need of teaching English to preschool children in Indonesia. However, there has not been a core vocabulary for teaching English to preschool children. Therefore, in this research, I built a written mini-corpus based on Indonesian children’s storybooks. As I am going to teach English to preschool children in near future, I want to investigate what English words are most important to teach to children of the age four to six. To be able to find out what English vocabulary would be relevant to teach to this age group, I need to know first what Indonesian vocabulary 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 accordance with Robbins and Ehri’s (1994) findings, in which kindergarten children’s vocabulary growth is improved due to listening to stories; therefore, I chose to examine the words in Indonesian children’s storybooks.

Many scholars have argued that there is a strong connection between L1 knowledge and L2 acquisition (Chen, 1992; Lotto and de Groot, 1998; Justice, 2005; Wolter, 2006); however, to my knowledge, none of the studies addressed the identification of L2 core vocabulary based on the L1 knowledge. Therefore, this research focuses on finding out the L1 children’s conceptual vocabulary to identify the L2 vocabulary to teach.

It is necessary to acknowledge what types of words that children are more exposed to in their early learning and what words that they acquire. Barrett (1995) reported that many scholars focused their research on kinds of words which were acquired during the native child’s early lexical development. The studies done by Barrett (1995) were mainly context bound. These words can refer to “classes of object; proper names of

individual objects, people, or animals; particular actions; properties, qualities, or states of objects and events” (Barrett, 1995, p. 365).

In accordance with the previous studies, Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick, and Reily (1993) pointed out that children acquired and produced about 50 to 100 nouns before developing their vocabulary with verbs and adjectives. On the contrary, Justice (2005) argued that word knowledge for children means comprehending and applying the word knowledge into production in different types of part of speech that includes nouns, verbs, adjectives, and adverbs. Justice’s statement supported the idea of children’s vocabulary knowledge is not only limited to nouns, but get expanded with verbs, adjectives, and adverbs.

Once children acquire vocabulary in their L1, there is a tendency that they will use the L1 underlying concepts of the vocabulary to comprehend the L2 words. Justice (2005) 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) supported Justice’s claim. They added that high-frequency words are the words that are not only useful but also fairly easy to acquire. However, they also agreed with Chen (1992) than children acquire a language more easily with the help of pictures or real objects, whereas adults do better by using word translation. Based on their study, the learners named the picture given faster when it represented a familiar concept in their L1. Their study supported Chen’s (1992) findings in which learners in the beginning level heavily depend on their L1 or visual representation to acquire L2 vocabulary.

In 2006, de Groot conducted another similar study showing the effect of L1 to acquiring a foreign language among 36 university students in Amsterdam. She found out that the students acquired FL words more easily if these words were paired with frequent and concrete L1 words. Moreover, their retention of the new FL words was also stronger in this situation. Dealing with word frequency, she claimed that 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) supported the preceding findings. He argued that learners’ L2 vocabulary acquisition can be influenced by their L1 conceptual knowledge. The influence gets higher especially for those

who already acquired complex L1 structures. He further explained that although learners do not necessarily use their L1 knowledge to understand new L2 concept, they are able to infer some possible combination of L2 words.

As many researchers have touched upon the relation between learning L2 with L1 vocabulary knowledge, in this section, a number of scholars pinpointed word frequency as a criterion in choosing what words to teach. Meara (1993) and Nation (1993) stated that word frequency is usually considered in designing a curriculum. They claimed that a curriculum designer usually put high-frequency words earlier in language learning. Leung (1992) also supported the idea of the importance of word frequency for children vocabulary acquisition, especially for kindergarten and first grade students as Walker, Greenwood, Hart, and Carta (1994) also addressed that children's successive school progress is influenced by the development of their early vocabulary.

Some researchers found out that the early intervention through interactive book reading and vocabulary development can boost children's literacy skills (Lonigan & Whitehurst, 1998) and inhibit them from experiencing difficulties in reading (Torgesen, 1998). Beck and McKeown (2001), and De Temple and Snow (2003) further added that reading storybooks to young children has been known as a useful way to introduce them to new vocabulary. Other researchers (Beck, McKeown, & Kucan, 2002; De Temple & Snow, 2003) pointed out that storybooks offer rich language that is not frequently heard in daily speech; therefore, teachers can introduce these new words in meaningful contexts. Justice, Pence, Beckman, Skibbe, and Wiggins (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) totally agreed with the idea that those 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.

It seems that storybooks become an effective tool to provide children with language exposure. Children know new vocabulary, even those words that are considered infrequent in daily speech, from being exposed to various words in a storybook. In contrast, Senechal (1997) claimed that young children's vocabulary grows as they are more exposed to certain words when they listen to a story more than one time. Justice (2005) also argued the usefulness of word exposure to children's vocabulary

knowledge. He pointed out that children learn vocabulary more easily if they get more exposure to it. The repeated occurrence of a word in a book helps children to acquire the word. These arguments shared the idea of word frequency as an important factor in children's vocabulary acquisition.

There have been numerous studies used storybooks to expose children to new vocabulary, either in L1 or L2. Robbins and Ehri (1994) focused their research on the children first language vocabulary acquisition. They also found that listening stories more than once and hearing repeated new vocabulary in the story can affect kindergarten children's vocabulary growth. Furthermore, they discovered that children's vocabulary size influences the way they acquire new vocabulary. Those who already know more vocabulary learn more due to their ability to use contextual clues in the stories.

In contrast with the use of storybooks in L1 vocabulary learning, Roberts and Neal (2004) based on their research on using interactive storybook reading to 43 non-native children. They discovered that the children's oral English proficiency was related to their literacy performance. In other words, these children who were exposed to words in narrative performed better in listening to English words and producing these words orally. Moreover, Silverman (2007) investigated the effectiveness of using children's literature to kindergarten students, who are both native and non-native speakers of English. She found that children's literature helped to support their literary acquisition which certainly affected their reading skill at a later age. She found that vocabulary knowledge is the most important factor for children's literacy development.

All of the above studies described the advantages of using storybooks in teaching vocabulary to children, but none of them showed what words need to put more emphasis on and what words are already in the mental lexicon. Moreover, none of the studies discussed what L2 words are most appropriate to teach in order to facilitate children's L1 conceptual knowledge and new vocabulary in L2. In order to know what L1 concepts that children are already exposed to, a collection of language used needs to be compiled.

As Hunston (2002) argued that people are not consciously aware of word, phrase, and structure frequency without any evidence from language use, she pointed out that corpus analysis is one of the ways to analyze the natural language use. Among many corpus available, Faber and Linares (2001) built a vocabulary corpus for teaching English since none of the

corpus was representative for their research. They built a corpus-based vocabulary of 800, 000 words from some European fairy tales collected by Andrew Lang in *The Red Fairy Tale Book*, *The Yellow Fairy Tale Book*, and *The Violet Fairy Tale Book*. They also added a few modern stories, such as Disney's stories with Mickey Mouse and Donald Duck, Ladybird readers, and *The Cat in the Hat Bright* and *Early Books*. They investigated how frequent the specific words appear in the text in order to find the basic vocabularies for primary school students. Their research mainly concentrated on verbs, nouns, and adjectives that frequently appeared in those electronic texts. Faber and Linares strongly believed that words become an important part in teaching language, especially for children aged four to eight who are developing their semantic. They further mentioned that this condition applies for both L1 and L2 language learning. In line with Faber and Linares' study, a corpus of Indonesian language is needed to conduct a corpus analysis of this language.

However, the research on corpus analysis is still limited to a few attempts to build corpora of different nature. Hardjadibrata (as cited in Nazief, 2000, p. 1), for example started a word analysis based on Indonesian newspapers. Nazief (2000) replicated the earlier study by conducting a research on Indonesian written corpus for adults based on Kompas, an Indonesian national newspaper.

In the area of children corpus, 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 focuses on Jakartan Indonesian spoken corpus. The other children corpus concerning Indonesian language is still in the process.

Since none of the research has been done in the area of basic vocabulary for Indonesian preschool children and as there has been a need in finding out what concepts that the Indonesian preschool children have known in their first language and what other concepts that need to be selectively addressed; therefore, I built a mini-corpus consisting Indonesian words in order to identify the frequent nouns, verbs, adjectives, and adverbs. The present research is only limited to the content words due to the existence of these words as a main part of one's language lexicon and used to convey ideas.

In order to identify a core vocabulary for teaching English to Indonesian preschool children, I need to know what concepts the children are already exposed to in the Indonesian children's storybooks. These following questions become the focus of my data analysis.

1. What nouns, verbs, adjectives, and adverbs commonly appear in the mini-corpus of Indonesian children's storybooks?
2. What English vocabulary is most appropriate to teach to Indonesian preschool children?

METHOD

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 it to them. 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.

All the data were uploaded to MonoConc Pro and ran through the frequency menu. I got a list of the most frequent till the less frequent word appeared in the corpus. The list was started from 3155 occurrences till one occurrence in the whole corpus. A word considered frequent in this corpus if this particular word appeared at least one time per story. In this corpus, a story consists of approximately 1, 000 words or tokens (since there were 131 books for the whole corpus, which was 134, 320 words). Therefore, I made a frequency cut-off point by dividing the raw number of occurrences by the number of words in the corpus. The result then was timed by the number of words per storybook. (e.g. word "kata" appeared 500 times in the whole corpus, so $(500/134000) \times 1000 = 4$, this result showed that "kata" appeared four times per storybook). As a result, only words which occurrences above 183 were considered frequent. Afterward, I selected the frequent nouns, verbs, adjectives, and adverbs from the whole frequent list.

FINDINGS AND DISCUSSION

Table 1. Frequent Nouns

Classification	Token (grouped by lemma)	Translation	Overall Frequency
family relationship	anak	a child	503
	ibu	a mother/an older woman	370
	ayah	a father	298
	kakek	a grandfather/grandpa	212
	nenek	a grandmother/grandma	247
	bunda	a mom	155
Place	rumah	a house	449
	tempat	a place	216
	laut	a sea	159
	hutan	a jungle	153
	gunung	a mountain	141
	sungai	a river	140
natural object	air	(the) water	278
	burung	a bird	213
	pohon	a tree	171
	angin	(the) wind	145
imaginative creature	putri	a princess/daughter	367
	raja	a king	246
	tuhan	god	225
time signal	hari	a day	460
	malam	night/evening	145
living creature object	orang	a person/people	291
	suara	voice	193

Table 2. Frequent Verbs

Classification	Token (grouped by lemma)	Translation	Overall Frequency
tense	sudah	have	500
	akan	will/would	500
	sedang	being	310
transitivity	ada	there is/there are /to have/to own	500
	punya	to have/ to own	156
	merasa	to feel	148
incomplete aspect	mulai	to begin	205
copula	menjadi	to become	324
	adalah	am/is/are/was/were	320
Ability	bisa	can/could	484
intention	mau	to be willing	275
	ingin	to wish	247
necessity	harus	to have to/should	254
daily life	pergi	to go	266
	makan	to eat	252
	kembali	to return	221
	datang	to come	218
	pulang	to go/to come home	201
	tidur	to sleep	165
	berjalan	to walk	161
inference	kata	say/said...	500
	tanya	ask/asked...	311
	teriak	shout/shouted...	138
natural environment	terbang	to fly	149

Table 3. Frequent Adjectives

Classification	Token (grouped by lemma)	Translation	Overall Frequency
Size	kecil	small	327
	besar	big	325
Characteristic	baik	good/kind/nice	151
Amount	banyak	many/much/a lot	192
Color	hitam	black	159
	putih	white	136
Feelings	senang	happy	153
	gembira	glad	146
	sedih	sad	143

Table 4. Frequent Adverbs

Classification	Token (grouped by lemma)	Translation	Overall Frequency
Degree	lagi	more	473
	saja	just	436
	hanya	only	269
	selalu	always	205
	pasti	certainly	173
Time	segera	immediately	227
	masih	still	216
	sekarang	now	201
	setelah	after	180
	tiba-tiba	suddenly	175

The frequent content words contributed to the lexical category order for teaching. As can be seen in the tables 1, 2, 3, and 4 for frequent content words, nouns comprise the most frequent lexical category. The result is in line with the previous findings (Fenson, et. al, 1993) that showed that children acquire nouns first before verbs and adjectives; therefore, it seems important to address the nouns first before the other content words. Since it was found that children are exposed to these frequent words in their L1, therefore, it seems likely to be easier for children to understand the English words for these concepts. Afterwards, teachers can introduce children to the frequent verbs that might lead them to create a simple sentence consisting of a subject and a predicate; for example, *A child is sleeping*. After the children are exposed to the nouns and verbs, teachers can introduce them with the frequent adjectives to accompany the nouns or the verbs. Once the children acquire the English nouns, verbs, and adjectives; for example, *A child feels happy*; the teachers can continue with the adverbs; for example, *A child always feels happy*. However, teachers can either introduce children to the frequent adjectives or adverbs. As long as children are exposed to the verbs, teachers can introduce them to the adverbs. Adequate numbers of repetitions, visual representations, and use of the language need to be addressed in the teaching and learning process.

There are several combinations of frequent nouns, verbs, adjectives, and adverbs that might be useful for teachers in introducing these English words to children.

Table 5. Combination of frequent noun, verb, and adverb

Noun	Verb	Adverb of degree	Adverb of time/place
child	go	more	now
mother	eat	just	day
father	return	only	night
grandfather	come	always	after
grandmother	sleep	certainly	sea
princess	walk	immediately	jungle
king		still	mountain
god		suddenly	river
person			
bird	go		
	eat		
	return		
	come		
	sleep		
	fly		

Based on the above table, there are several ways for teachers to introduce these concepts to children. First, they can start with the frequent nouns. Once the children learn the English words, they can get exposed to the frequent verbs. There are two different nouns (i.e. *person* or *bird*) that can be the subject of a simple sentence which consists of a subject and a predicate. As the nature of a person is walking and a bird is flying, so these words are better not to be used interchangeably if teachers want to point to the condition of human life, for example, *A person is flying* and *A bird is walking*. However, the other frequent verbs can be used both for a person and a bird. If the children already understand how to create a simple sentence, teachers can add an adverb time or place for the sentence, such as *A child is going to the jungle*. To improve the sentence, teachers can introduce the adverb of degree, such as *A child is immediately walking to the river*.

Since English verbs denote tense, it is necessary for children to learn different kinds of tense to refer to the time occurrence of an action or event. Nonetheless, teachers need to introduce the tense that is quite simple for children to understand. It seems uncomplicated to start with present continuous tense. Teachers need to address different kinds of *to be* for the subject and also the attachment of participle (i.e. *-ing*) to the verb. Whereas the other tense can follow the present continuous tense as long as there are not many variations of the verbs, simple future tense might be worth trying.

Once children get the idea of combining the nouns and the verbs, they can be exposed to the frequent adjectives as seen in the following table.

Table 6. Combination of frequent noun, verb, and adjective

Noun	Verb	Adjective
child	become	small
mother	be	big
father	feel	good/kind/nice
grandfather	have	many/much
grandmother		happy
princess		glad
king		sad
god		
person		
bird		

In the above table, it can be seen that there are different verbs that can be followed with the frequent adjectives. The nouns are the same, but in this case a bird can use the same verbs as a human, for example, *A mother feels happy* and *A bird feels sad*. There is a specific verb, *to have*, in the table that can be followed as an adjective preceded by a noun, such as *A princess has many birds*. Therefore, it might be useful to teach children a combination of noun, verb, and another noun that results in a simple sentence which consists of a subject, a predicate, and an object, for example, *A father has a child*.

Among the frequent verbs, there are several verbs that might be more difficult for children to comprehend (i.e. *ask, say, shout, and to wish*). The first three verbs need a clause preceding the verb, such as "*A mother is in the mountain,*" *said a child*, whereas the last verb needs a clause following the verb, for example, *A child wishes the jungle is small*. Therefore, these verbs are better to be the last to introduce to children. Teachers are strongly encouraged to expose children to different combinations of content words to create a simple sentence.

CONCLUSION AND SUGGESTIONS

As has been discussed in the data discussion, the frequent content words found in the corpus can be used in teaching English to Indonesian preschool children. The English equivalents will be taught to children while they are learning English. Once the children have known the concept of the words in their first language, it is easier for them to acquire the words in the second language. If the children do not know the concept, teachers need to provide supporting materials, like pictures, realias, or drawing, so the children can understand what they are learning about.

Furthermore, I found that concepts in children's world are tightly related to the structure of a language that makes teachers need to address the similarities and differences between the structure of the first and the second language. For example, in Indonesian, the adjective *baik* can refer to a person's characteristic, *Orang itu baik*, or to a condition, *Rapatnya berjalan dengan baik*. Whereas in English, the adjective *good/kind/nice* only refers to one's characteristics. To address a condition, teachers need to introduce the adverb of *good*, which is *well*, such as in *The meeting is running well*.

Although the Indonesian storybooks that are used as a data source in this research might not be available for all Indonesian children, the result of the study can be used as a guideline for preschool teachers to know what Indonesian vocabulary that preschool children are already exposed to. Teachers might apply the knowledge in English language instruction. The lexical order found can contribute to the order of instruction. Teachers can introduce the frequent nouns, followed by the frequent verbs, adjectives, and adverbs in teaching English. Besides, the study can be an insight for preschool curriculum designers to choose the most appropriate English vocabulary to teach. The result can also be used to guide teachers in using the Indonesian conceptual knowledge to use the English vocabulary creatively. In this case, teachers can create stories with English vocabulary which concepts are familiar to preschool children.

Besides, the corpus built will be useful for researchers interested in examining Indonesian children's storybooks. In addition, it might help storybooks writers to create English stories based on the concepts that Indonesian children already exposed to. Since the data is considered small, it is a good idea to develop the study with more sources from different texts, genres, and also discourse. Thus, it will enrich the corpus as well as the result.

REFERENCES

- Arka, I. W. & Simpson, J. (2007). *Challenges of developing a balanced and representative corpus for Indonesian ParGram*. Paper presented at the Eleventh International Symposium on Malay/Indonesian Linguistics, Manokwari, Indonesia.
- Barrett, M. (1995). Early lexical development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 362-392). Oxford: Basil Blackwell.
- Beck, I. L., & McKeown, M. G., (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *The Reading Teacher*, 55(1), 10-20.

- Beck, I., McKeown, M.G. & Kucan, L. (2002). *Bringing words to life: Robust vocabulary development*. New York: Guilford Press.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: 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, 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), 463-506.
- De Temple, J. & Snow, C. E. (2003). Learning words from books. In S. Stahl, A. van Kleeck & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 16-36). Mahwah: Erlbaum.
- Faber, P. & Linares, C. F. M. (2001). Design of a corpus-based vocabulary for teaching English in the primary school. In J. Raya, M., Faber, P., Gewehr, W. & Peck, A. J. (Eds.), *Effective foreign language teaching at the primary level* (pp. 25-38). New York: Oxford Press.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reily, J. S. (1993). *Technical manual for the MacArthur Communicative Development Inventory*. San Diego: Singular Press.
- 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: Springer.
- Hardjadibrata, R. R. (1969). *An Indonesian newspaper wordcount*. Cayton: Monash University.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Justice, L. M., Pence, K. L., Beckman, A. R., Skibbe, L. E., & Wiggins, A. K. (2005). *Scaffolding with storybooks*. Newark: 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.
- Lonigan, C. J. & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, 13, 263-290.

- 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, 31-69.
- Meara, P. (1993). The bilingual lexicon and the teaching of vocabulary. In R. Schreuder & B. Weltens (Eds.) *The bilingual lexicon* (pp. 279-297). Amsterdam: Benjamins.
- Meyer, C. F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge, UK: Cambridge University Press.
- Nation, P. (1993). Vocabulary size, growth, and use. In R. Schreuder & B. Weltens (Eds.) *The bilingual lexicon* (pp. 115-134). Amsterdam: Benjamins.
- Nazief, B. (2000). *Development of computational Linguistics research: A challenge for Indonesia*. Paper presented at the 15th Indonesian Scientific Conference, Tokyo, Japan.
- Robbins, C. & Ehri, L. (1994). Reading storybooks to kindergarten helps them learn new vocabulary words. *Journal of Educational Psychology*, 86(1), 54-64.
- Roberts, T. & Neal, H. (2004). Relationships among preschool English language learner's oral proficiency in English, instructional experience and literacy development. *Contemporary Educational Psychology*, 29, 283-311.
- Senechal, M. (1997). The differential effect of storybook reading on preschooler's acquisition of expressive and receptive vocabulary. *Journal of Child Language*, 24(1), 123-138.
- Silverman, R. D. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *The Elementary School Journal*, 107(4), 365-383.
- Sinclair, J. (1996). *EAGLES preliminary recommendations on corpus typology*. EAG-TCWG-CTYP/P. Pisa: ILC-CNR.
- Torgesen, J. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator*, 22, 32-39.
- Walker, D., Greenwood, C. R., Hart, B. & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, 65, 606-621.
- Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, 27(4), 741-747.
- Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergartners: What are the best predictors? *Bilingualism: Language and Cognition*, 9(1), 33-49.
- Vos, J. (2007). Can preschool children be taught a second language? Retrieved July 27, 2007 from http://www.earlychildhoodnews.com/earlychildhood/article_view.aspx?ArticleId=60.