

Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development

Norwati Roslim¹, Muhammad Hakimi Tew Abdullah², Anealka Aziz³, Vahid Nimehchisalem⁴, and Azhani Almuddin⁵

¹ Akademi Pengajian Bahasa, Universiti Teknologi MARA, Negeri Sembilan, Rembau Campus, Malaysia
norwati@uitm.edu.my

² Faculty of Communication and Media Studies, Universiti Teknologi MARA, Negeri Sembilan, Rembau Campus, Malaysia
muhhammad_hakimi@uitm.edu.my

³ Akademi Pengajian Bahasa, Universiti Teknologi MARA, Shah Alam, Malaysia
anealka@uitm.edu.my

⁴ Faculty of Educational Studies, Universiti Putra, Malaysia
vahid@upm.edu.my; nimechie@gmail.com

⁵ Akademi Pengajian Bahasa, Universiti Teknologi MARA, Negeri Sembilan, Rembau Campus, Malaysia
azhaniazhani@yahoo.com

Article history

Received: 19 May 2020

Accepted: 8 June 2020

Published: 10 June 2020

Abstract

Numerous corpus studies have suggested that teaching materials design could greatly benefit from the empirical information about language use provided by corpus linguistics. In spite of the awareness that corpus-based research can offer valuable insights for materials development, still relatively small number of studies report on the practical applications of corpus data for teaching materials development. There is no clear guideline or framework on how corpora and corpus studies could assist in developing teaching materials. Hence, this study focusses on one grammatical item which poses problems to Malaysian learners, that is, prepositions. The objectives are (i) to identify prepositions in the British National Corpus as a reference corpus and the descriptions offered by linguists and grammarians as a reference grammar, and (ii) to provide a framework to use reference corpus, reference grammar and corpus-based research, as a resource for developing materials in the teaching of prepositions. In order to meet the objectives, content analysis was used as the methodology throughout this study. The findings showed that reference corpus, reference grammar and corpus-based research could be used systematically as guidance to develop corpus-informed materials. It is hoped that this contribution of knowledge could have an impact on second language learning-teaching.

Keywords: *corpus research, corpus-informed materials, second language teaching, prepositions*

Introduction

In the context of the teaching of English as a second language, the compilation and application of corpora have extended from the creation of dictionaries, grammatical description of English to other varieties of ap-
Universiti Teknologi MARA, Vol. 4, No. 2, 2020

Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development applications, such as, syllabus design, materials development and methodology which are relevant to language teaching.

The developers of dictionaries (Partington, 2001) and learner dictionaries (Aston, 2001) often use varieties of corpora for the construction of entries. Corpus-based dictionary projects which were published from the 1980s include Collins Cobuild English Language Dictionary whose first edition (Cobuild, 1987) used a 20-million-word corpus of (mainly written) contemporary English, and whose second edition (Cobuild, 1995) drew on 200 million words from the Bank of English and the Longman Dictionary of Contemporary English (LDOCE) which exploited the 100-million-word British National Corpus (Aston, 2001).

The description of English has also used corpora to complement grammatical descriptions in grammar reference books. The emergence of corpus linguistics has led to studies on grammar descriptions that ‘derive evidence directly from texts’ (Kennedy, 1998, p. 7) rather than based on intuition or introspection. For instance, according to Biber, Johansson, Leech, Conrad and Finegan (1999), *The Longman Grammar of Spoken and Written English* (LGSWE) adopts a corpus-based approach investigating the patterns of structure and use to provide grammatical descriptions in which it serves as an aid to complement grammatical descriptions in *A Comprehensive Grammar of English language* (Quirk, Greenbaum, Leech, & Svartvik, 1985).

Other applications of corpora which are relevant to language teaching have also been mentioned by many corpus linguists in studies on frequency and collocation information. The role of frequency in corpus linguistics is essential. Even though the role of frequency has been criticized, for instance, in Thornbury (2004) who noted that in terms of frequency, it is not always the case that the most frequently occurring items are the most useful and in terms of an item’s teachability, Mahlberg (2007) argues that frequency is relative, and an important factor in corpus work is comparison. Frequencies of individual words have to be seen in relation to words in their contexts. Even though function words are very frequent and can co-occur in a variety of texts, we can still identify collocation tendencies and see how frequent words have their own patterns.

Also, it has been said that corpus linguistics has its impacts on the content of language teaching and also language teaching methodology (Hunston, 2002; Kennedy, 1998). It is greatly known that the applications of corpora have spread from dictionaries to the grammatical description of English, as well as other applications, that is, curriculum design, the production of reference tools and classroom grammar teaching (Meunier, 2002; Cheng, 2010; Hughes, 2010; Oghigian & Chujo, 2012; Tribble & Wingate, 2013; Dilay & Dilai, 2016). Corpora can enlighten the syllabus, and most teaching materials can be used around corpus data and as a result, many studies have attempted in using corpora or a published corpus study as a resource or reference guide for materials development (Barlow, 2002, as cited in Krajka, 2009, Reinhardt, 2010; Chang & Kuo, 2011; Timmis, 2013; Ha, 2016).

While many studies have highlighted that applications of corpora have spread and have gained interests in the teaching of English language, there is still lack of available literature on corpus-informed materials development. Hence, this paper intends to fill in the gap by bringing into focus some insights from a reference corpus, reference grammar and corpus research for the development of teaching materials.

Literature Review

Corpus linguistics has been described as the study of language on the basis of text corpora (Aijmer & Altenberg, 1992). It has also been further described that corpus linguistics is based on bodies of text as the domain of study and as the source of evidence for linguistic description and argumentation (Kennedy, 1998). It is also perhaps best described in simple terms as the study of language based on examples of 'real-life' language use (McEnery & Wilson, 2001). Linguists have always used the word corpus to describe a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings which have been collected for linguistic study. More recently, the word has been reserved for collections of texts (or parts of text) that are accessed electronically (Hunston, 2002).

Corpora are compiled for many different purposes. As Tribble and Jones (1990) describe, a corpus can be made up of individual text or several texts. A corpus which consists of several texts is categorized as specialist corpora and general corpora. Hunston (2002) categorized corpus types in specialized corpora, general corpora, comparable corpora, parallel corpora, learner corpora, pedagogic corpora, historical or diachronic corpora, and monitor corpora.

Corpus linguistics deals with the principles and practice of using corpora in a language study (Crystal, 1992). The primary question in any usage of corpora is what do descriptions and analyses of language from large corpora of texts reveal about the lexical patterns and grammatical structures that people use. The goal of corpus linguistics, is "to identify and analyze the grammatical and lexical patterns that speakers and writers use" (Chapelle, 2001, p. 42) and the aim of corpus linguistics can also be seen as "the analysis and the description of language use, as realized in text(s)" (Tognini-Bonelli, 2001, p. 2). Reppen, Biber and Fitzmaurice (2002) propose the main research questions that can be investigated through corpus analysis focus on the use of a particular linguistic feature: a single word, a set of related words, a grammatical construction, or the interaction between particular words and grammatical structures.

Corpus-based studies on prepositions

Studies on prepositions have been explored in the field of corpus linguistics and they have been conducted in various perspectives mainly in relation to frequency and collocation information. Roslim and Mukundan
Universiti Teknologi MARA, Vol. 4, No. 2, 2020

Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development (2011) have presented an overview of corpus linguistics studies on prepositions which could have contributed to the teaching and learning of prepositions.

Prepositions have been studied for dictionaries such as Collins Cobuild English Dictionary for Advanced Learners (2001) which gives information on the frequency of words using five frequency bands. The words in the five frequency bands are of immense importance to learners because they make up 95% of all spoken and written English. Many of the words in frequency band 5 are the common grammar words such as the, and, of, and to, which are an essential part of the way we put words together. The words in the top two bands, bands 4 and 5, account for about 75% of all English usage which mostly include common prepositions which shows their obvious importance.

Prepositions have also been studied in different varieties of English mainly conversation, fiction, newspaper language, and academic prose. Although it is often said that function words (in this case prepositions), as opposed to individual lexical words, are frequent in any text, there are wide differences among registers. Prepositions are the most frequent function word class in news and academic prose, however, they are much less common in conversations. This means prepositions appear more in written than spoken discourse. Academic prose and news reportage have the highest frequency of nouns and also the highest frequency of prepositions which serve as extensions or specifications of nouns (Biber et al., 1999, pp. 91-93).

Prepositions have been studied in a general corpus which is a corpus of many texts comprising written and/or spoken language. It is often used as a baseline in corpus studies. Well-known general corpora include the British National Corpus (100 million words) and the Bank of English (400 million words in January 2001). Much earlier general corpora were the Lancaster-Oslo/Bergen (LOB) corpus, consisting of written British English, and the Brown corpus, consisting of written American English. They were both compiled in the 1960's and comprising a million words each (Hunston, 2002, pp. 14-15).

Studies on prepositions have also utilized pedagogic corpora, a type of corpus which “consists of all the language a learner has been exposed to, for instance, course books, readers, plus any tapes” (Hunston, 2002, p. 14). In a study based on a pedagogic corpus, Mukundan and Roslim (2009) conducted a study on prepositions used in the Malaysian Secondary School English language textbooks of Forms One, Two and Three. This study looked at the prepositions which were outlined in the syllabus and presented in the textbooks. Talif, Wong, Abdullah, Rashid and Jan (2007) took a different perspective in their presentation of collocation list of preposition **in** also in a Form One English textbook (Abdullah, Wee & Bien, 2002). The study showed the type of words that collocate with preposition **in** and the list also showed the environment in which it occurred. They highlighted collocation of such items would allow teachers to introduce to the learners the contexts the word may appear in.

Studies on prepositions have also been carried out using a learner corpus. This is a collection of texts produced by learners of a language (Hunston, 2002, p. 15). In Malaysia, for instance, the data in the EMAS

(English of Malaysian School Students) corpus has revealed that the words *to, in, of, for, at, on, into* and *with* were among the top 25 most frequent words used in the students' written and oral data (Samad, Hassan, Mukundan, Kamarudin, Rahman, Rashid & Vethamani, 2002). A few studies related to prepositions have been conducted using the EMAS Corpus. Roslim (2004a,b) conducted a study on the use of prepositions of place, *in, on* and *at* using the EMAS corpus and the findings revealed the problems faced by learners in using these prepositions.

The abovementioned empirical research have supported the evidence of how corpora and corpus studies can benefit English language teaching as found in Hunston's (2002) similar viewpoints with Kennedy (1998, p. 288):

Such analysis can be used not only for syllabus design and the sequencing of pedagogical materials, (but they) can also be used for the consciousness-raising of classroom teachers about the dimensions of the learners' task and for prioritizing the emphasis given to various learning goals in the classroom...the quantitative evidence on the distribution of the component parts of the language becomes a yardstick against which to evaluate subjective judgements about the goals and content of instruction.

In summary, the literature has shown a variety of corpus research utilizing different types of corpora such as general, pedagogic and learner corpora. With reference to English prepositions, most studies are concerned with identifying prepositions in terms of frequency and collocational information as well as problems faced by learners in using prepositions in both speaking and writing. However, to date, none of the studies we reviewed has provided any comprehensive framework to use corpora and corpus-based studies to develop teaching and learning materials.

The Study

It is the concern of this study to focus on prepositions as it is important to master the use of prepositions as they are "grammatical words or function words that mainly contribute to the grammatical structure of the sentence" (Thornbury, 2002, p. 14). Having difficulties in using the English prepositions will hinder one to be clearly understood by others in both speaking and writing. For instance, in the ESL/EFL situations, such difficulties will become an obstacle for the learners to be able to communicate effectively.

Despite the importance of prepositions on communication, studies as those found in Darus and Subramaniam (2009), Jalaluddin, Awal and Bakar (2008), Mohameda, Goh and Eliza (2004), Roslim (2004a, 2004b),

Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development and Talif and Vethamani (1989) have shown that Malaysian students have problems in the use of prepositions. The problems continue even after their school years. This is shown in studies by Ahour and Mukundan (2012), Hong, Rahim, Hua and Salehuddin (2011), Yunus and Awab (2011), Ting, Mahadhir and Chang (2010), Sudhakaran (2008) and Darus, Maasum, Stapa, Omar and Aziz (2007) on the problems Malaysian students face in using prepositions correctly at university levels. It is emphasized by Noordin and Samad (2005) that “Malaysians will need to be proficient in English as a medium for communicating with people from around the globe” (p. 1). Since “communication is an integral part of society, the mechanics of language has to be understood in terms of how it facilitates communication. This includes the knowledge of grammar as without it, communication will fail as a structure will be lacking” (Manaf, 2008, p. 1). “Attention to grammar has a long, if somewhat controversial, history in second language (L2) teaching” (Loewen, 2004, p. 1). However, the status of grammar “has been assumed to be a pre-requisite of true proficiency” and “has been even treated as an obstacle to successful language learning” (Hoey, 2009, p. 143).

This study is carried out based on the premise that corpus linguistics has become quite influential in English language teaching in recent years (Juchem-Grundmann, 2010; Furniss, 2013; Ishihara & Paller, 2016), however, the actual use of corpora (large, searchable collections of real language, electronic texts) still seems to be fairly limited. Besides, little knowledge is known on how to use corpora although the availability of corpora has helped teachers to prepare materials. With regard to grammatical items, it is essential to use a reference grammar. Additionally, corpus-based research that have been conducted provide findings that could assist in materials development (Carloni, 2017; Vyatkina, 2020). As such, there is a need to provide a framework on how a reference corpus, a reference grammar and corpus-based studies can be used as a resource to develop materials for classroom use. This study aims to fill the gap between corpora and corpus-based research and materials development as a way of assisting material developers in understanding potential uses of corpus linguistics for English language teaching. Based on this aim, the following research questions were addressed:

1. What are the prepositions presented in the British National Corpus and how do they compare with the descriptions offered by linguists and grammarians?
2. In what way can the presentation of prepositions in the British National Corpus and Quirk et al. (1985) corpus be used to develop corpus-informed materials?

Methodology

This study examined the British National Corpus and the pedagogic corpus through content analysis. Krippendorff (2004, p. 3) describes “content analysis entails a systematic reading of a body of texts, images, and symbolic matter, not necessary from an author’s or user’s perspective”.

Content analysis is described as “a research method applied to written or visual materials for the purpose of identified specified characteristics of the material” (Ary, Jacobs & Sorensen, 2010, p. 457). Content analysis could also be used for analyzing and tabulating the frequency of occurrence of topics, ideas, opinions and other aspects of the content of written and spoken communication in a set of language teaching materials. A “frequency” is a type of content indicator. A “frequency” content indicator is one in which the number of times one or more content characteristics occur (George, 1959, as cited in Krippendorff & Bock, 2009). Even though, originally it refers to frequencies of words per text or surface unit, the concept has extended to include quantifying by means of a frequency survey of classifications of “syntactic, semantic or pragmatic categories” (Titscher, Wodak & Vetter, 2000, p. 55). In this study, the materials analyzed are the British National Corpus and the Cambridge Learner Corpus. It serves the purpose of content analysis to examine the frequency and the predetermined categories of prepositions for corpus-informed materials.

Content analysis provides a systematic approach. An important requirement of a systematic research is that it is carried out in such a way that its results can be verified by other investigators who follow the same steps as the original researcher. Content analysis is a replicable research technique (Krippendorff, 2004). Since in this study, a coding system will be applied by a human coder which is the first author, at least one more coder is needed to provide validation of a coding scheme (Neuendorf, 2002).

Reliability

Reliability is “the degree to which members of a designated community agree on readings, interpretations, responses to, or uses of given texts or data” (Krippendorff, 2004, p. 212). Menon (2009) establishes criterion for selection of an independent coder as follows:

1. Individual who has had experience teaching English language for the past 5 years as familiarity with grammatical word classes was essential.
2. Individual who was willing to undergo coding training sessions with the first author.

In this study, the coder underwent a training session on the coding system and to achieve uniformity, disagreements were discussed until both the first author and coder came to a conclusive agreement.

Corpus size

This study involved the analysis of a corpus of Malaysian English language textbooks of Form One to Five secondary levels and this applies to the characteristics of a corpus-based analysis which is described by Biber et al. (1998, p. 4) as ‘a principled collection’. One principle to consider is the size of the corpus. Gen-

Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development erally, “a very large corpus will be richer in information, with more interesting output and more likely to represent accurately features of a language” (Tribble & Jones, 1990, p. 15). However, Oostdijk and de Haan (1994) conclude that there is no such thing as the best optimum sample size and that the suitability of the sample depends on the specific study undertaken. The convenience of computer software has made all kinds of analysis possible, and at the same time corpus-based method is only a complementary approach as opposed to the more traditional linguistic approach. Bauer and Renouf (2001) maintain that corpus-based studies will force the linguists to face phenomena that may be overlooked by the traditional methodologies.

Population and Sampling

For the purpose of this study, the population for the English corpus is defined as the Malaysian English language textbooks used by the Form One, Two, Three, Four and Five secondary school students. This corpus consisted of 311,214 running words.

The books used for this corpus-based analysis were:

1. Khairi Izwan bin Abdullah, Doris Wee and Toe Hwee Bien, (2002). *English Form 1*. Carina Lee Yu Ching (Editor), Pustaka Alhas (M) Sdn Bhd.: Kuala Lumpur.
2. Rohini Shanta, Uma Kurap and Shalani Lorenz, (2003). *English Form 2*. Datin June Yogeswaran and Choo Kar Yik (Editors). Penerbit Mega Setia Emas Sdn. Bhd.: Petaling Jaya.
3. Ahmad Salihin bin Mat Saat, Mohd Salleh bin Salehin and Wan Zuraidi bin Wan Zaid, (2003). *English Form 3*. Ee Ewe Hock (Editor). Sutrapadu (M) Sdn Bhd.: Petaling Jaya.
4. Tan Phiak Lee and Angelina Ng Kim Leng, (2003). *English Form 4*. Irene Leow (Editor). Mutiara Cemerlang Sdn. Bhd.: Petaling Jaya.
5. Rani Parasuraman, Carole Chung Mei Choo and Shadhana Popatlal, (2003). *English Form 5*. Judith Ann Arokiasamy and Christine Tan Bee Lian (Editors). Sutrapadu (M) Sdn Bhd.: Petaling Jaya.

Instrumentation

This study utilized WordSmith Tools 4.0. WordSmith Tools are integrated programmes that look at how words behave in a text. Oxford University Press uses the tools for the lexicographic work in preparing dictionaries. These tools are also useful for language teachers, students and researchers in investigating language patterns.

The reliability and capability of WordSmith Tools were reviewed by Biber and Reppen (2001) in which they compared WordSmith to MonoConc Pro and found that WordSmith was more capable in corpus-based research. The capability of this software has also been verified by Mukundan (2004) in the exploration for

suitable software to analyze prescribed textbooks compared with other seven concordance software programme (Concordance 3.0. HAMLET for Windows, TextQuest 1.37, HyperResearch 2.0.3, EZText 3.06c, MicroConcord (DOS), MonoConc Pro 2.0 and WordSmith Tools. There are three analysis tools available in the software, that of, WordList, Concord and Keyword.

Data Collection Procedures

It is fundamental to acknowledge that the data used in this study was taken from a pedagogic corpus compiled by Mukundan and Anealka (2007). Each page of the textbooks was photocopied and scanned and converted into a Tagged Image File (TIF) Format. This was then saved and put through the Optical Character Recognition (OCR) software which converted all TIF files into text files (.txt). The text files were then checked for errors before saving and renaming them according to the respective units of the textbook. These text files were analyzed in this study using WordSmith Tools 4.0 software programme created by Scott (2004).

This corpus consisted of five Malaysian secondary school English language textbooks and comprises 311,214 running words. This textbook corpus has been used in many significant corpus-based studies in Malaysia (Mukundan & Leong, 2012; Philip, Mukundan & Nimehchisalem, 2012; Khojasteh & Kafipour, 2012; Mukundan & Khojasteh, 2011; Mukundan & Roslim, 2009; Mukundan & Anealka, 2007, 2009). These studies provide illuminating findings on vocabulary loading and distributions and grammatical structures presented in the textbooks. However, it should also be noted that the study on the grammar parts of this textbook corpus was first attempted by Mukundan and Roslim (2009).

A reference corpus used in this study was The British National Corpus (BNC). The BNC is a 100 million word collection of samples of written and spoken language from a wide range of sources. It is designed to represent a wide cross-section of British English from the later part of the 20th century. The written part of the BNC (90%) includes extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The spoken part (10%) consists of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins (<http://www.natcorp.ox.ac.uk/>). In this study the list of prepositions outlined in the Companion Website for word frequencies in written and spoken English which is based on the British National Corpus (Leech, Rayson & Wilson, 2001) was used as a point of reference to identify the list of prepositions.

According to Leech (2017), a reference corpus is used by scholars and researchers, lexicographers and grammarians. Besides, language materials developers will want to use the reference corpus as an aid to develop teaching materials of various kinds.

The reference grammar used in this study was *A Comprehensive Grammar of the English Language* (Quirk et al., 1985). This study follows the possible meanings of prepositions describes in Quirk et al. (1985) which have become the most used reference for grammar studies.

Data Analysis

The analysis of data for research question 1 involved the retrieval of prepositions from the British National Corpus frequency list of prepositions and the common simple and complex prepositions appeared in Quirk et al. (1985) as well as studies on prepositions in Talif and Vethamani (1989), Arshad et al. (2002), Mohameda et al. (2004), Roslim (2004a, 2004b), Darus et al. (2007), Sudhakaran (2008), Darus and Subramaniam (2009), Jalaluddin et al. (2008), Ting et al. (2010), Hong et al. (2011), Yunus and Awab (2011), and Ahour and Mukundan (2012). The list of prepositions from the BNC and Quirk et al. (1985) were then compared in terms of the ranking order (BNC) and prepositional meanings (Quirk et al., 1985). Each word examined was tagged and coded as preposition. In order to enable the researcher to study prepositions in context and come up with a reliable list of all the prepositions understudied, this was done manually and intuitively with the aid of an independent coder. This procedure of employing an independent rater or coder is to ensure that forgery is avoided during the analysis. This procedure has also been used in other corpus-based studies by Mukundan et al. (2012), Philips et al. (2012), Khojasteh and Kafipour, (2012), Mukundan and Khojasteh (2011), Mukundan and Roslim (2009), Menon (2009), Akbari (2009), and Manaf (2008).

In order to address the second research question, selection of prepositions were taken from the British National Corpus and Quirk et al. (1985). A qualitative page-by-page content analysis of textbook materials was then applied to analyse the ways in which the presentation of prepositions can be used to develop corpus-informed materials. The presentation of prepositions and the various types of exercises in the textbooks would provide a platform in which corpus-based research could lead to corpus-informed materials and eventually, the development of a framework for corpus-informed materials.

Findings

We report our findings following the order of the research questions.

1. What are the prepositions presented in the British National Corpus and how do they compare with the descriptions offered by linguists and grammarians?

This study outlined the prepositions presented in the British National Corpus and compared with the descriptions of prepositions in Quirk et al. (1985). The prepositions in the British National Corpus were compared to Quirk et al. (1985) prepositional meanings: place, time, the cause/purpose spectrum, the means/agentive spectrum, accompaniment, support and opposition and other prepositional meanings.

Table 1 Prepositions in the BNC and their prepositional meanings in Quirk et al. (1985)

List of prepositions	List of frequency order in the BNC	Prepositional meanings in grammar descriptions (Quirk et al., 1985)						
		Place	Time	Cause/Purpose	Means/Agentive	Accompaniment	Support and opposition	Others
in	2	√	√					
on	6	√	√	√	√			√
near	42	√						
under	19	√						
behind	33							
in front of	62	√						
by	7	√	√		√			
at	8	√	√	√	√			
between	15	√	√					
before	24		√					
after	14		√					
since	34		√					
during	23		√					
until	38		√					
within	22	√	√					
to	3	√	√	√				
from	9	√	√	√				
across	31	√						
towards	26	√						
out of	20	√						√
away	46	√						
from								
of	1							√
for	4		√	√			√	√
among	30	√	√					
between	15	√						

Table 1 shows prepositions outlined in the BNC in terms of their frequency order. Preposition *in* was ranked 2nd in the BNC, *on*, the 6th rank, *near*, the 42nd, *under*, the 19th, *behind*, the 33rd, *in front of*, the 62nd, Universiti Teknologi MARA, Vol. 4, No. 2, 2020

Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development by, the 7th, *at*, the 8th, *before*, 24th, *after*, the 14th, *to*, the 3rd, and *from*, the 9th. *Between*, was ranked 15th and *of*, was ranked 1st in the BNC. Prepositions *since* (34th), *during* (23rd), *until* (38th), *within* (22nd), *across* (31st), *towards* (26th), *for* (4th) and *among* (30th). *Out of* was on the 20th rank and *away from*, 46th rank of the BNC. Prepositions rank order in the BNC indicates the degree of importance in terms of which prepositions need to be introduced to learners.

Table 1 also shows clearly, prepositions have several different functions and different prepositions can have very similar uses. This further contributes to the difficulties in learning to use prepositions correctly among Malaysian learners apart from other anticipated problems. In studies on prepositions in Malaysian learners' writing, Talif and Vethamani (1989) had categorized the errors into three: wrong choice of prepositions, omission of prepositions and redundant use of prepositions. This also appeared in other studies by Ahour and Mukundan (2012), Hong et al. (2011), Yunus and Awab (2011), Ting et al. (2010), Darus and Subramaniam (2009), Jalaluddin et al. (2008), Sudhakaran (2008), Mohamed et al. (2004), Roslim (2004a,b) and Yong (2001). These categorizations were described in both Wong's (1983) simplification features and Oxford's (1990) compensation strategies. These problems also appeared in other grammar studies such as modals (Manaf, 2008) and phrasal verbs (Akbari, 2009). The spatial and temporal meanings of *in*, *on* and *at* are among the several groups of prepositions that are frequently confused by learners (Celce-Murcia & Larsen-Freeman, 1999). There appear to be instances where these different prepositions can have very similar uses such as *in* the morning, *on* Monday morning or *at* night (Swan, 2001).

2. In what way the presentation of prepositions in the British National Corpus and Quirk et al. (1985) can be used to develop corpus-informed materials?

In terms of the descriptions of prepositions, preposition of place is the chief prepositional meaning as in Quirk et al. (1985) and Celce-Murcia and Larsen-Freeman (1999). Hence, this should be given priorities before other prepositional meaning of prepositions.

Table 2 Presentation of prepositions in the grammar teaching sections in the textbook corpus

Preposi- tions	Form 1 (18 Chapters)	Form 2 (15 Chap- ters)	Form 3 (15 Chap- ters)	Form 4 (14 Chap- ters)	Form 5 (15 Chap- ters)
Preposition of place	Chapter 2		Chapter 10		
Preposition of time	Chapter 5	Chapter 8	Chapter 10	Chapter 11	
Preposition of direction	Chapter 10	Chapter 8	Chapter 11	Chapter 10	Chapter 13
Preposition		Chapter 10	Chapter 13		

of quantity					
Preposition		Chapter 10			
of purpose					
Preposition		Chapter 11			
of association					
Total	3	3	4	4	1

Table 2 shows the distribution of the teaching of prepositions with regard to their prepositional meanings or functions in the Forms 1-5 textbooks. The analysis showed there were three chapters on the teaching of prepositions in the Form 1 textbook, three chapters in the Form 2 textbook, four chapters in the Form 3 textbook, four chapters in the Form 4 textbook and one chapter in the Form 5 textbook. The most central prepositional meanings, that is place, was presented twice throughout Forms 1-5, which is in Chapter 2 for Form 1 and Chapter 10 for Form 3. Prepositions of time appeared throughout in the Forms 1-4 textbooks; Chapter 5 (Form 1), Chapter 8 (Form 2), Chapter 10 (Form 3) and Chapter 11 (Form 4). However, prepositions of directions appeared in all the textbooks: Chapter 10 (Form 1), Chapter 8 (Form 2), Chapter 11 (Form 3), Chapter 10 (Form 4) and Chapter 13 (Form 5). Preposition of quantity appeared in two chapters: Chapter 10 (Form 2) and Chapter 13 (Form 3). Prepositions of purpose appeared in Chapter 10 and prepositions of association appeared in Chapter 11 in the Form 4 textbook.

The findings in Table 2 has shown that the textbook materials have not fully given priorities to the teaching of preposition of place as the chief prepositional meaning. Priorities were mainly given to prepositions of time and directions. As a result, it was found in studies on prepositions of place in the EMAS (The English of Malaysian School Students) corpus that Malaysian learners in Year 5 (primary school), Form 1 (secondary school) and Form 4 (secondary school) had problems in using prepositions of place in, on and at correctly. The problems occurred among students in primary schools as well as secondary schools (Roslim, 2004a,b). The problems with prepositions of place also appeared in samples of errors in a study conducted by Darus and Subramaniam (2009) as shown in the following example, ‘So many dirty plates and glasses can be seen everywhere *at* (Preposition) the school canteen’ in which the correct preposition is ‘*around*’ rather than ‘*at*’.

In the present study, prepositions of place were not presented in all the textbooks, instead prepositions of directions and prepositions of time have taken the central roles. Hence, this could contribute to the problems in using prepositions of place correctly among Malaysian learners in which material developers should be aware of.

The frequency order of prepositions also play an important role in terms of which prepositions need to be introduced accordingly.

The prepositions which were introduced explicitly to the Form 1 students in the Form 1 textbook were *on*, *near*, *in*, *under*, *in front of*, *next to/by*, *behind* and *between*. They were introduced in a grammar section called ‘Language Focus’ and the subheading of ‘Talking about position’ presented the semantic functions of these prepositions. Pictures and examples of sentences were used to present these prepositions on page 18.

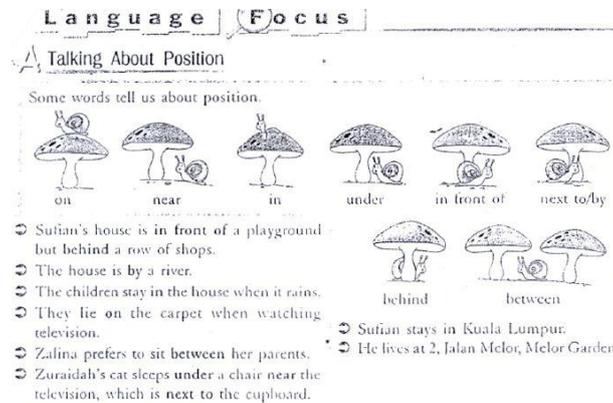


Figure 1 Presentation of *on*, *near*, *in*, *under*, *in front of*, *next to/by*, *behind* and *between* in the Form 1 textbook

Figure 1 shows the use of the subheading ‘Talking about position’ in presenting the functions of the words that tell about position. Pictures were used to illustrate these words, *on*, *near*, *in*, *under*, *in front of*, *next to/by*, *behind* and *between* which were used as prepositions.

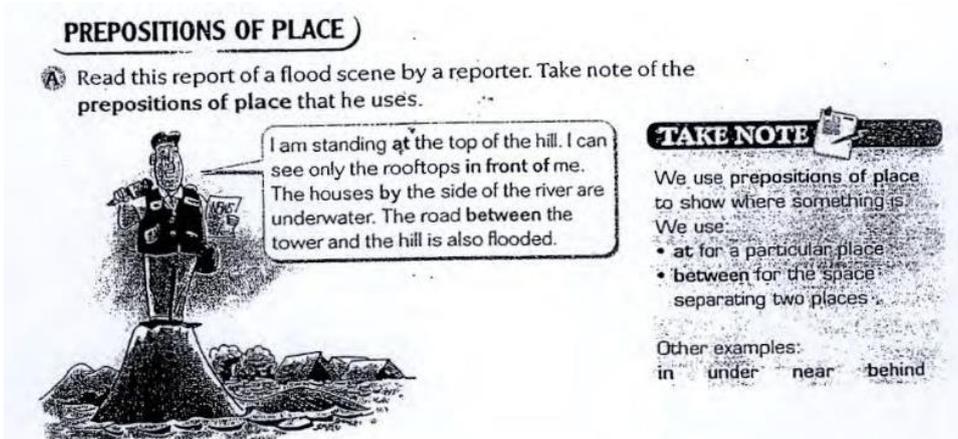


Figure 2 Presentation of *at*, *in front of*, *by* and *between* in the Form 3 textbook

Figure 2 shows prepositions *at*, *in front of*, *by* and *between* were presented in a grammar section of Chapter 10 called “Grammar Web” and the subheading of “Prepositions of place” in the Form 3 textbook. They were presented in a form of a news report with a picture of a reporter on page 136. The explanations of these prepositions were given in the “Take Note” with other examples of prepositions of place which were given unequal treatment.

Table 3 Presentation of prepositions in the textbook corpus, the BNC and suggested corpus-informed prepositions

Textbook corpus	British National Corpus (BNC)	Suggested Corpus-informed Prepositions for Material Developers
on	6	in
near	42	on
in	2	by
under	19	at
in front of	62	between
by	7	under
behind	33	behind
between	15	near
at	8	in front of

The presentation order of prepositions in the textbook corpus does not follow the order of frequently used prepositions in the BNC. Hence, there is a need for material developers to refer to a reference corpus to assist them in selecting and sequencing prepositions.

Hence, having analyzed the textbook materials, it is important for material developers to consider which prepositional meanings to be taught first based on the BNC and descriptions of linguists and grammarians in reference grammar.

The findings from this study supports the evidence in terms of the complexity of prepositions which were described by many scholars such as Baldwin et al. (2009), Kosur (2008), Hoffman (2004), Huddleston and Pullum (2002), Tan (2002), Hunston (2002), Swan (2001), Celce-Murcia and Larsen-Freeman (1999), Biber et al. (1999), Lindstromberg (1998), Hughes (1992), Kaplan (1989) and ESL researchers namely Mukundan and Roslim (2009), Darus and Subramaniam (2009), Sudhakaran (2008), and Rosli and Edwin (1989) to name a few. The regularities and irregularities with regard to prepositions found in the KBSM syllabus as well as textbooks provide evidence leading to problems in understanding prepositions among ESL learners in Malaysia. In addition to this, this finding also supports criticisms on textbooks in other Malaysian Secondary School textbook studies such as Mukundan and Roslim (2009, 2011), Mukundan and Khojasteh (2011), Mukundan and Anealka (2009, 2007) and Menon (2009).

Based on the findings from Research Questions 1 and 2, a framework is designed to develop corpus-informed materials for teaching material developers. This is shown in the following figure:

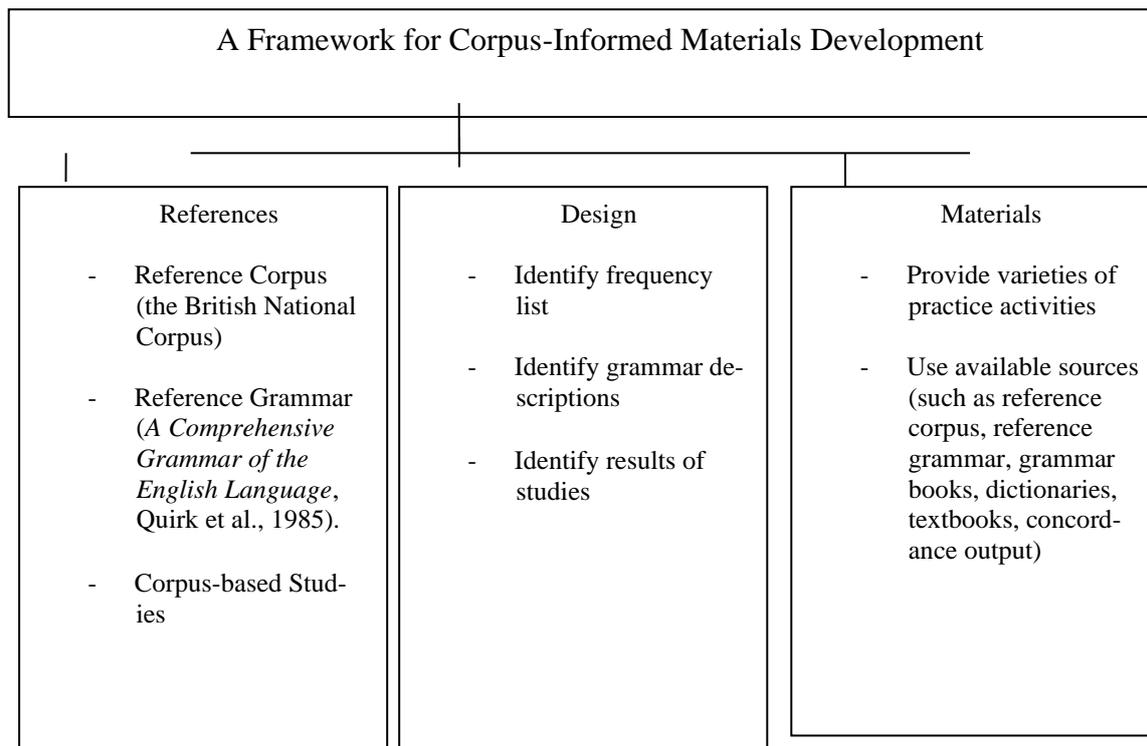


Figure 3 A Framework for Corpus-Informed Materials Development

Discussions

The findings from this study are very significant for material developers in the Malaysian ESL context. Material developers should reconsider the selection and the sequence of prepositions in the teaching materials. A list of preposition based on the frequency order of prepositions outlined in the British National Corpus (BNC) is encouraged to be referred to throughout the development of teaching materials. Frequency of occurrence is one of the factors that needs to be considered when one wants to decide on the inclusion of grammar items or vocabulary items in developing teaching materials even though it is not the only criterion. The frequency data from corpora can indicate whether a particular feature is likely to be worth learning and thus makes it relevant to material designers. This view is emphasized by linguists such as Evison (2010), Biber and Reppen (2002), Tognini-Bonelli (2001) and Aston (2001) to name a few.

The BNC is the most suitable reference point because it is an established corpus cited by many scholarly researchers in their corpus-based studies. In this study the list of prepositions provided in the BNC is very informative to aid the analysis of prepositions. However, comparisons can be made with other corpora namely the Brown and Lancaster-Oslo/Bergen (LOB) corpora used in studies by Mindt and Weber (1989, as

cited in Kennedy, 1998). Mindt and Weber listed the 14 most frequent prepositions which accounted for about 90% of prepositional use. They were *of, in, to, for, with, on, at, by, from, into, about, through, over* and *between*. Hence, in spite of the BNC as the reference corpus was the most suitable for this study, findings from other corpora studies could be an aid to support the evidence.

Based on the findings, material developers could develop teaching based on frequently occurring prepositions as presented in the reference corpus, the BNC. The selection of prepositions would be more appropriate if it is selected based on the frequency of use in line with what one might reasonably expect from corpus data of spoken and written English. Even though Tsui (2004) does highlight the need to consider the infrequent ones as a basis for selecting and sequencing items, this could be practiced in enrichment activities. Therefore, the findings from the present study could contribute information to material developers in making language learning more successful.

Based on the premise that corpus-based research could inform material and textbook writers on linguistic contents, there is a need to evaluate the selection and sequence of prepositions through findings of corpus-based research as there are more than 100 single and multi-word (complex) English prepositions. Material writers could create a material or a textbook mapping on the list of prepositions to be taught to students. This idea of creating a material and textbook mapping if taken seriously could tackle criticisms on textbooks posed by many such as Mukundan and Roslim (2011), Mukundan and Khojasteh (2011), Mukundan and Anealka (2007, 2009), and Menon (2009) who have claimed that textbooks are prepared based on intuition and assumptions rather than on corpus-based research.

It has been noted in the literature that linguists and grammarians such as Baldwin, Kordini and Villavicencio (2009), Kosur (2008), Hoffman (2004), Huddleston and Pullum (2002), Tan (2002), Hunston (2002), Swan (2001), Celce-Murcia and Larsen-Freeman (1999), Biber et al. (1999), Lindstromberg (1998), Hughes (1992) and Kaplan (1989) do offer descriptions of prepositions, However, Quirk et al. (1985) was used in this study because it is the most established reference as the basis for grammar descriptions and has been used in numerous studies. In the Malaysian context too, studies on vocabulary and grammar items such as those by Mukundan and Khojasteh (2011), Mukundan and Roslim (2009), Menon (2009), Manaf (2008) and Sudhakaran (2008) to name a few have used Quirk et al. (1985) as a point of reference. In fact, it has been declared by Manaf (2008) that the Malaysian schools are encouraged to follow descriptions of grammar by Quirk et al. (1985). According to Wong (1983), standard British English has been selected in Malaysia as the criterion because this is the variety of English which has been instructed in the country ever since the language was first introduced into Malaysia.

It is essential to restate that the present study has added to the literature that reference corpus, reference grammar and corpus-based research could be used systematically as a guide to develop corpus-informed ma-

Norwati Roslim, Muhammad Hakimi Tew Abdullah, Anealka Aziz, Vahid Nimehchisalem, and Azhani Almuiddin
Corpus Research Applications in Second Language Teaching: A Look into Corpus-informed Materials Development
terials. This study is important as it has prepared the platform for other studies to be carried out in the hope of improving the teaching and learning English as a second language in the Malaysian context. With the framework developed from this study, it is hoped that it could provide clearer insights for material and textbook writers to develop corpus-informed materials.

Acknowledgements

The authors wish to thank Universiti Teknologi MARA Negeri Sembilan, Rembau Campus, Malaysia and Universiti Teknologi MARA, Shah Alam, Malaysia for all the support. Thanks are also due to Professor Dr Jayakaran Mukundan for his guidance and never ending support throughout.

References

- Abdul Rashid Mohameda, Goh Li Lian & Wan Rose Eliza (2004). English Errors and Chinese learners. *Sunway College Journal*. 1: 83-97. Retrieved from <http://eprints.sunway.edu.my/>
- Ahour, T. & Mukundan, J. (2012). Errors and variations of TESL students' written description. *Pertanika Journal Social Science and Humanities*. 20 (1): 55-64.
- Aijmer, K. & Altenberg, B. (1992). *English Corpus Linguistics*. New York: Longman Inc.
- Akbari, O. (2009). *A corpus-based study on Malaysian ESL Learners' use of phrasal verbs in narrative compositions*. (Unpublished Doctoral Thesis). Serdang: Universiti Putra Malaysia.
- Arshad Abd. Samad, Fauziah Hassan, Jayakaran Mukundan, Ghazali Kamarudin, Sharifah Zainab Syed Abd. Rahman, Juridah Md. Rashid, & Malachi Edwin Vethamani. (2002). *The English of Malaysian School Students (EMAS) Corpus*. Serdang: Universiti Putra Malaysia.
- Ary, D., Jacobs, L. C. & Sorensen, C. (2010). *Introduction to Research in Education* (8th ed.). California: Wadsworth Cengage Learning.
- Aston, G. (2001). *Learning with Corpora*. Bologna: CLUEB and Houston, TX: Athelstan.
- Baldwin, T., Kordini, V. & Villaviencio, A. (2009). Prepositions in Applications: A survey and introduction to the Special Issue. *Computational Linguistics*. 35(2): 119-149.
- Bauer, L. & Renouf, A. (2001). A corpus-based study of compounding in English. *Journal of English Linguistics*. 29(2): 101-123.
- Biber, D. & Reppen, R. (2002). What does a frequency have to do with grammar teaching? *Studies in Second Language Acquisition*. 24: 199-208.

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education. British National Corpus. Retrieved from <http://www.natcorp.ox.ac.uk/>
- Carloni, G. (2017). Chapter Eight Clil Materials Development In Higher Education: Corpus Linguistics And Digital Tool Giovanna Carloni. *A Journey through the Content and Language Integrated Learning Landscape: Problems and Prospects*, 105.
- Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The Grammar book: an ESL/EFL teacher's course*. 2nd ed. Boston: Heinle & Heinle.
- Chang, C. F., & Kuo, C. H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, 30(3), 222-234.
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.
- Cheng, W. (2010). What can a corpus tell us about language teaching? In A. O'Keefe & M. McCarthy (Eds.). *The Routledge handbook of corpus linguistics* (pp. 319-332). USA/Canada: Routledge.
- Collins COBUILD advanced learner's English dictionary*. (2001). Glasgow: Harper Collins Publishers.
- Dilay, I. P., & Dilai, M. P. (2016). Designing a corpus-informed EFL textbook. *Проблеми сучасного підручника*, (17), 123-131.
- Evison, J. (2010). What are the basics of analyzing a corpus? In A. O'Keefe & M. McCarthy (Eds.). *The Routledge handbook of corpus linguistics* (pp. 122-135). USA/Canada: Routledge.
- Furniss, E. (2013). Using a corpus-based approach to Russian as a foreign language materials development. *Russian Language Journal/Русский язык*, 63, 195-212.
- Ha, M. J. (2016). Linking adverbials in first-year Korean university EFL learners' writing: A corpus-informed analysis. *Computer Assisted Language Learning*, 29(6), 1090-1101.
- Hoey, M. (2009). Suppose grammars were not true, what would we teach? In J. Mukundan (Ed.). *Readings on ELT materials* (pp. 143-164). Petaling Jaya: Pearson Malaysia.
- Hoffman, S. (2004). Are low-frequency complex prepositions grammaticalized? On the limits of corpus data – and the importance of intuition. In H. Lindquist & C. Mair (Eds.). *Corpus approaches to grammaticalization in English* (pp. 171-210). Amsterdam/Philadelphia: John Benjamins
- Hong, A. L., Hajar Abdul Rahim, Hua, T. K. & Khazriyati Salehuddin (2011). Collocations in Malaysian English Learners' Writing: A corpus-based error analysis. *3L: The Southeast Asian Journal of English Language Studies*. 17: 31-44.
- Huddleston, R. & Pullum, G. K. (2002). *The Cambridge Grammar of the English language*. Cambridge: Cambridge University Press.

- Hughes, R. (2010). What a corpus tells us about grammar teaching materials. In A. O'Keefe & M. McCarthy (Eds.). *The Routledge handbook of corpus linguistics* (pp. 401-411). USA/Canada: Routledge.
- Hughes, R. (1992). *English grammar made simple*. Kuala Lumpur: Fajar Bakti Sdn. Bhd.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Ishihara, N., & Paller, D. L. (2016). Informed Materials for Teaching Pragmatics: The Case of Agreement and Disagreement in English. In *SLA research and materials development for language learning* (pp. 103-118). Routledge.
- Juchem-Grundmann, C., & Krennmayr, T. (2010). Corpus-informed integration of metaphor in materials for the Business English classroom. *Fostering language teaching efficiency through cognitive linguistics*, 317-336.
- Kamariah Yunus & Su'ad Awab. (2011). Collocational competence among Malaysian Undergraduate law students. *Malaysian Journal of ELT Research*. 7(1): 151-202.
- Kaplan, J. P. (1989). *English Grammar: Principles and Facts*. New Jersey: Prentice Hall.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman Publishing.
- Khojasteh, L. & Kafipour, R. (2012). Are Modal Auxiliaries in Malaysian English language Textbooks in line with their usage in Real Language? *English Language Teaching*. 5(2): 68-77.
- Kosur, H. M. (2008). *Forms and Functions of Prepositions*. Retrieved from <http://www.rockpicklepublishing.com/essays/formandfunctionsofprepositions.html>.
- Krajka, J. (2010). Dialoguing in curriculum development: On designing Informtion and Communication Technology Training in Modern Philology Setting. : In M. Kusiak (Ed.). *Dialogue in Foreign Language Education* (pp. 111-121). Wydawnictu Uniwersytetu Jagiellonskiego
- Krippendorff, K. & Bock, M. A. (2009). *The Content Analysis Reader*. Los Angeles: SAGE Publications.
- Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*. Beverly Hills, CA: SAGE.
- Leech, G. (2017). The Importance of Reference Corpora. Retrieved from <https://www.uzei.eus/wp-content/uploads/2017/06/06-Geoffrey-LEECH.pdf>
- Leech, G., Rayson, P., & Wilson, A. (2001). Word Frequencies in Written and Spoken English: based on the British National Corpus. Retrieved from ucrel.lanc.ac.uk/bncfreq/flist.html
- Lindstromberg, S. (1997). *English Prepositions Explained*. Amsterdam: John Benjamins Publishing Co.
- Loewen, S. (2004). Attention to grammar in incidental focus on form. In Arshad Abd. Samad (Ed.) *Perspectives on Grammar in Language Education* (pp. 1-17) Serdang: Universiti Putra Malaysia Press.
- Mahlberg, M. (2007). Lexical items in discourse: identifying local textual functions of sustainable development. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.) *Text, Discourse and Corpora: Theory and analysis*. (pp. 191-218). London: Continuum.

- McEnery, T. & Wilson, A. (2001). *Corpus Linguistics: An Introduction*. (2nd Edition). Edinburgh: Edinburgh University.
- Menon, S. (2009). *Corpus-based analysis of lexical patterns in Malaysian Secondary School Science and English for Science and Technology Textbooks*. (Unpublished Doctoral Thesis). Serdang: Universiti Putra Malaysia.
- Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung & S. Petch-Tyson. (Eds.). *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 119-141). Amsterdam/Philadelphia: John Benjamins.
- Mukundan, J. (2004). *A composite framework for ESL textbook evaluation* [Unpublished doctoral dissertation]. Universiti Putra Malaysia, Serdang.
- Mukundan, J. & Anealka Aziz (2009). Loading and distribution of the 2000 high frequency words in Malaysian English Language Textbooks for Form 1 to Form 5. *Pertanika Journal Social Science and Humanities*. 17(2): 141-152.
- Mukundan, J. & Anealka Aziz Hussin. (2007) A forensic study of vocabulary loading and distribution in five Malaysian secondary school English textbooks (Forms 1 – 5). *Pertanika Journal of Social Sciences and Humanities*. 15(2): 59-74.
- Mukundan, J. & Khojasteh, L. (2011). Modal Auxiliary Verbs in Prescribed Malaysian English Textbooks. *English Language Teaching*. 4(1): 79-89.
- Mukundan, J. & Roslim, N. (2009). Textbook Representation of Prepositions. *English Language Teaching*. 2(4): 13-24.
- Mukundan, J. & Leong, C. H. (2012). Distribution of Articles in Malaysian Secondary School English Language Textbooks. *English Language and Literature Studies*. 2(2): 62-70.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, California: Sage Publications.
- Nooreen Noordin & Arshad Abdul Samad. (2005). Examining the importance of EST and ESL textbooks and materials: Objectives, content and form. *English for Specific Purposes World Online Journal for Teachers*, 9(4). Retrieved from http://www.esp-world.info/Articles_9/textbooks.htm
- Nor Hashimah Jalaluddin, Norsimah Mat Awal & Kesumawati Abu Bakar. (2008). The Mastery of English language among lower secondary school students in Malaysia: A linguistic analysis. *European Journal of Social Sciences*. 7(2): 106-119.
- Oghigian, K., & Chujo, K. (2012). Corpus informed writing for science and engineering. *Journal of the College of Industrial Technology, Nihon University*, 45, 1-13.
- Oostdijk, N. & de Haan, P. (1994). *Corpus-based research into language*. Amsterdam: Rodopi.

Partington, A. (2001). Corpora and their uses in language research. In G. Aston (Ed.). *Learning with Corpora*. Houston: Athelstan.

Philip, A., Mukundan, J. & Nimehchisalem, V. (2012). Conjunctions in Malaysian Secondary School English Language Textbooks. *International Journal of Applied Linguistics & English Literature*. 1(1): 1-10.

Quirk, R., Greenbaum S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of English language*. Harlow: Longman.

Reinhardt, J. (2010). The Potential of Corpus-Informed L2 Pedagogy. *Studies in Hispanic & Lusophone Linguistics*, 3(1).

Reppen, R. Biber, D. & Fitzmaurice, S. (Eds.). (2002). *Using corpora to explore linguistic variation*. Amsterdam: Benjamins.

Rosli Talif, Bee Eng, W. Ain Nadzimah Abdullah, Sabariah Md Rashid & Jariah Mohd Jan. (2007). Corpora analysis: Benefits towards ESL learners' vocabulary competence. In T. Bee Hoon, Washima Che Dan, Mar-dziah Hayati Abdullah, Noritah Omar, Faiz Sathi Abdullah & Rosli Talif (Eds.). *Theoretical and practical orientations in Language and Literature Studies* (pp. 114-131). Petaling Jaya: Pearson Malaysia.

Rosli Talif & Vethamani, E. M. (1989). *Error analysis of Form Four English compositions*. The English Teacher. XVIII: 110-124.

Roslim, N. (2004a). Corpus linguistics: Dealing with a learner corpus. In Tengku Sepora Tengku Mahadi (Ed.). *Proceedings of the first International Language Learning Conference* (pp. 128-145). Kuala Lumpur: Perpustakaan Negara Malaysia.

Roslim, N. (2004b). A corpus-based study on an English prepositions of place, *at*. In J. Mukundan, Dzeelfa Zainal Abidin & D. Singh (Eds.). *ELT Matters 2: Developments in English Language Learning and Teaching* (pp. 147-159). Serdang: Universiti Putra Malaysia Press.

Roslim, N. & Mukundan, J. (2011a). An overview of corpus linguistics studies on prepositions. *English Language Teaching*. 4(2): 125-131.

Saadiah Darus and Subramaniam, K. (2009). Error analysis of the written English essays of secondary school students in Malaysia: A case study. *European Journal of Social Sciences*. 8(3): 483-495.

Saadiah Darus, Tg Nor Rizan Tg Mohd Maasum, Siti Hanim Stapa, Nazlia Omar & Mohd Juzaidin Abd Aziz. (2007). Developing an error analysis marking tool for learners. *Proceedings of the 7th International Conference on Applied Computer Science*. 7: 356-359. Retrieved from

<http://www.academia.edu/263348/Developing> An Error Analysis Marking Tool for ESL Learners

Scott, M. (2001). Comparing corpora and identifying key words, collocations and Frequency distributions through the WordSmith Tools Suite of computer programs. In M. Ghadessy, A. Henry & R. L. Roseberry, (Eds.) *Small Corpus Studies and ELT* (pp. 47-67). Amsterdam/Philadelphia: John Benjamins Publishing Co.

Swan, M. (2001). *Practical English Usage*. Oxford: Oxford University Press.

- Sudhakaran, B. (2008). *The use of prepositions among Malay adult ESL learners*. (Unpublished Doctoral Thesis). Serdang: Universiti Putra Malaysia.
- Timmis, I. (2013). Corpora and materials: Towards a working relationship. *Developing materials for language teaching*, 461-474.
- Thornbury, S. (2002). *How to teach vocabulary*. England: Pearson Education Limited.
- Thornbury, S. (2004). *How to teach grammar*. Malaysia: Pearson Education Limited.
- Ting, S. H., Mahanita Mahadhir and Chang, S. L. (2010). Grammatical errors in spoken English of university students in oral communication course. *GEMA OnlineTM Journal of Language Studies*. 10 (1): 53-70. Retrieved from <http://journalarticle.ukm.my/>
- Titscher, S., Meyer, M., Wodak, R. & Vetter, E. (2000). *Methods of text and discourse analysis*. London: Sage.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Philadelphia: John Benjamins.
- Tribble, C. & Jones, G. (1990). *Concordances in the classroom: A resource book for teachers*. London: Longman.
- Tribble, C., & Wingate, U. (2013). From text to corpus—A genre-based approach to academic literacy instruction. *System*, 41(2), 307-321.
- Tsui, A. B. M. (2004). What teachers have always wanted to know – and how corpora can help. In J. M. Sinclair (Ed.). *How to use corpora in language teaching* (pp.39-66). Amsterdam/ Philadelphia: John Benjamins.
- Umi Kalthum Abdul Manaf. (2008). *The Use of Modals in Malaysian ESL Learners' Writing*. (Unpublished Doctoral Thesis). Serdang: Universiti Putra Malaysia.
- Vyatkina, N. (2020). Corpus-Informed Pedagogy in a Language Course: Design, Implementation, and Evaluation. In *New Technological Applications for Foreign and Second Language Learning and Teaching* (pp. 306-335). IGI Global.
- Wong, I. (1983). *Simplification features in the structure of colloquial Malaysian English*. Singapore: Singapore University.
- Yong, J. Y. (2001). Malay/Indonesian Speakers. In M. Swan and B. Smith (Eds.). *Learner English: A Teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.