

An Overview of Corpus Linguistics Studies on Prepositions

Norwati Roslim

English Language Department, Academy of Language Studies

Universiti Teknologi MARA, Negeri Sembilan

Kampus Kuala Pilah, Beting

72000 Kuala Pilah, Negeri Sembilan, Malaysia

Tel: 60-648-42-355; 012-9417467 E-mail: norwatiroslim@yahoo.com

Jayakaran Mukundan (corresponding author)

Department of Language and Humanities Education, Faculty of Educational Studies

Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Tel: 60-389-46-6000 E-mail: jaya@educ.upm.edu.my

Received: December 14, 2010 Accepted: December 30, 2010 doi:10.5539/elt.v4n2p125

Abstract

Studies on prepositions have been explored in corpus linguistics. They have been studied in various perspectives mainly in relation to frequency and collocational information. In order to look at the developments of these studies, this paper will focus on the development of sequence of studies of prepositions in three decades as observed by the authors. In keeping up with the developments, this paper will also look into the scenario of English language corpus work in Malaysia. Based on these review, this paper will then further add to this body of knowledge by providing a more tangible and practical applications in dealing with prepositions from the perspectives of the teaching and learning of prepositions.

Keywords: Corpus linguistics, Prepositions

1. Developments in studies of prepositions

In the field of corpus linguistics, insights into the structure and use of language have been radically initiated by the availability of digital computers and software programmes. This allows researchers to store huge amounts of text and retrieve particular words, phrases or chunks of texts as well as sorting the linguistic items to find their typical behaviour (Kennedy, 1998; 5).

In the 1980's, one prominent corpus-based study which focused solely on prepositions was identified at a frequency level. Prepositions have been studied in a general corpus which is a corpus of many texts comprising written and/or spoken language. 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. The Brown corpus, started its compilation in 1961, which becomes a reference point for American English, was the first computer corpus compiled for linguistic research. It became available in 1964. The Lancaster-Oslo/Bergen (LOB) Corpus was compiled between 1970 and 1978. It was intended to be a British English counterpart to the Brown Corpus (Kennedy, 1998, Hunston, 2002; Baker, 2009)

Mindt and Weber (1989) studied prepositions in American and British English using the Brown Corpus and the LOB corpus. In the Brown and LOB corpora, the 14 most frequent prepositions were listed which accounted for about 90% of prepositional use.

In the 1990's, as a result of Mindt and Weber's (1989) study, another step has been taken by another researcher using six of the high frequency prepositions listed in Table1, that of the prepositions *of*, *at*, *from*, *through*, *between* and *by*. This study had moved one step ahead from the frequency level to the level of collocations.

Kennedy (1990, 1991, 1998) had undertaken studies on prepositions which initially based on findings of the 14 most frequent prepositions in the Brown and LOB (Lancaster-Oslo/Bergen) corpora by Mindt and Weber (1989) on the basis that 'when the high frequency and difficulty of acquisition of the English prepositional system is considered, it is somewhat surprising that there have not been more corpus-based studies of how the system is used' (1998: 139).

Kennedy (1990) provided results of prepositions *at* and *from* based on the LOB (Lancaster-Oslo/Bergen) corpus. It showed that the prepositions tend to collocate with particular words. The most common word classes that occur immediately before *at* and *from* were nouns and pronouns with 42% of tokens for *at* and 45% of tokens for *from* and verbs with 32% of tokens for *at* and 29% of tokens for *from*.

Kennedy (1991) further revealed a detailed analysis of *between* and *through* by exploring their linguistic ecology using the one-million-word LOB (Lancaster-Oslo/Bergen) corpus of adult written British English, which is made up of 500 representative 2000-word samples from a wide variety of genres. The Oxford Concordance Program – OCP2 (Hockey & Martin, 1988) was used in the study to reveal collocational information about *between* and *through* in the LOB corpus. *Between* occurred 867 times and *through* occurred 776 times in the corpus and they clearly occurred in different contexts. The findings were discussed in three aspects; collocations with preceding word, collocations with following word and semantic functions of *between* and *through* in context:

Collocations with preceding word

There was a striking difference in the words which most frequently came directly before *between* and *through*. 65.7% (570 tokens) of both nouns and pronouns typically preceded *between*, whereas 43.2% (335 tokens) of verbs were the most common word class preceding *through*.

Collocations with following word

There were fewer recurring collocations with following rather than preceding words. However, by treating different personal pronouns, numbers and place names, 282 of the tokens (33%) following *between* occurred in combinations appearing four or more times. Similarly, 217 of the tokens (28%) following *through* recur. *Between* and *through* tend to collocate more strongly with preceding words.

In the same study, Kennedy then moved from the collocational information to semantic functions of prepositions.

Semantic functions of between

Major semantic functions in which *between* occurred include 31.5% (273 tokens) for location, 4.4% (38 tokens) for movement, 11.6% (84 tokens) for time, 4.8% (42 tokens) for dividing and sharing and 49.7% (430 tokens) for other relationships.

Semantic functions of through

Major semantic functions in which *through* occurred include 36.8% (285 tokens) for unimpeded motion, 17.7% (138 tokens) penetration of a barrier or obstruction, 7.2% (56 tokens) for perception through an obstruction, 8.1% (63 tokens) for time, 25.3% (196 tokens) for agent-intermediary-instrument and 4.9% (38 tokens) for causation.

The analysis of the semantic functions in which *between* and *through* occurred in the LOB corpus suggest why these words may be difficult to learn or use. For example both are associated with movement, time and other relationships have complex semantic structures involving abstractness.

Kennedy (1998) also provided results of major semantic functions *at* (n=5951) and *by* (n=5386) which occurred in the LOB (Lancaster-Oslo/Bergen) corpus. Major semantic functions of *at* are location (49%), time (43%), event/activity (6%), quantity/degree (16%), state/manner (1%), causation (1%) and miscellaneous (4%). Whereas major semantic functions of *by* include agent marker (64%), means or manner (16%), location (2%), time (4%), measurement (3%) and miscellaneous (11%).

For all Kennedy studies, all of them went beyond the linguistic description by providing a statistical dimension in terms of frequency of collocations and semantic functions of prepositions based on use in context. Even though it is not easy to predict what particular quantitative information can be of pedagogical significance, such empirical information may account for uncertainty in our intuitions or difficulties in learning and thus may contribute to improvements in pedagogical practices.

The most frequent of all English prepositions, *of*, was considered by Kennedy on the basis of Sinclair's (1991) pilot analysis of the large Birmingham corpus. The preposition *of* is not normally used in the dominant prepositional structure, namely the prepositional phrase. However, it is noted that *of* tends to collocate with preceding items rather than following.

Renouf & Sinclair (1991) studied multi-word 'collocational frameworks' which are pairings of function words with a variable lexical slot, for instance, *a + ? + of*, *be + ? + to*, *many + ? + of*. They used a one-million-word corpus of spoken British English and a 10-million-word corpus of written British English from the Birmingham Corpus. They analyzed collocational framework as another way to present and explain language patterns.

In 2000's, Hunston and Francis presents their studies on major open classes (verb, noun, adjective) in *A Corpus-Driven Approach to the Lexical Grammar of English* (Hunston & Francis 2000). They provided a 'pattern grammar' framework as used in Collins Cobuild English Dictionary (1995) which includes 75, 000 most frequent words in the Bank of English. In general, the pattern of a word consists of the elements that follow it, but it may also include elements which precede it. With reference to *A Corpus-Driven Approach to the Lexical Grammar of English* (Hunston & Francis 2000), the following shows the patterns of these classes which take prepositions as part of the patterns.

The patterns of verbs

1. The verb is followed by a prepositional phrase or adverb group:

V prep/adv : She chewed *on* her pencil.

V *about* n : He was grumbling *about* the weather.

In other cases, the verb is followed by a noun group, adjective group, 'ing' clause or wh-clause introduced by a specific preposition. This pattern is V *about*, V *at* n, V *as* adj, V *by* -ing etc., depending on the preposition. Examples include:

He was grumbling *about* the weather.

The rivals shouted *at* each other.

The prepositions which are used in patterns like this are *about*, *across*, *after*, *against*, *around/round*, *as*, *as to*, *at*, *between*, *by*, *for*, *from*, *in*, *in favour of*, *into*, *like*, *of*, *off*, *on*, *onto*, *out of*, *over*, *through*, *to*, *towards*, *under*, *with*.

2. The verb is followed by a noun group and a prepositional phrase or adverb group:

V n prep/adv : Andrew chained the boat *to* the bridge.

Stir the sugar *in*.

Sometimes the pattern is formed with the word way and a prepositional phrase or adverb group:

V way prep/adv : She ate her way *through* a pound of chocolate.

In other cases, the verb is followed by a noun group and another noun group, adjective group or wh-clause introduced by a specific preposition. This pattern is V n *about* n, V n *at* n, V n *as* adj etc., depending on the preposition. Examples include:

I warned him *about* the danger.

I saw the question *as* crucial.

The prepositions which are used in patterns like this are *about*, *against*, *as*, *as to*, *at*, *between*, *by*, *for*, *from*, *in*, *into*, *of*, *off*, *on*, *onto*, *out of*, *over*, *towards*, *with*.

3. The verb pattern contains the word it. The main patterns are as follows.

Introductory it: it V prep clause

General it: it V adj prep/adv

The prepositions most frequently used in patterns like this are as follows: *at*, *by*, *from*, *in*, *into*, *on*, *out of*, *under*, *with*.

The patterns of nouns

1. The noun is preceded by a specific preposition.

from N : I've been blind in my right eye *from* birth.

on N : The film was shot *on* location in Washington.

to N : They went *to* school together every day.

2. The noun is followed by a prepositional phrase introduced by a wide range of prepositions.

N prep

3. The noun is followed by a prepositional phrase introduced by a specific preposition: N *of* n, N *for* n, N *from* n, etc., Examples include:

It was the latest in a series of acts *of* violence.

Their hatred *for* one another is legendary.

The threat *from* terrorists is at its highest for two years.

The prepositions most frequently used in patterns like this are as follows: *about, against, among, as, at, behind, between, for, from, in favour of, in, into, of, on, over, to, towards, with*.

The patterns of adjectives

1. The adjective is followed by a prepositional phrase introduced by a wide range of prepositions.

ADJ prep

2. The adjective is followed by a prepositional phrase introduced by a specific preposition: ADJ *as* n, ADJ *of* n, ADJ *on* n, etc., Examples include:

We felt inadequate *as* parents.

I think he's fully aware *of* those dangers.

He's always been very dependent *on* me.

Biber et. al (2000: 91-93) studied prepositions 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 conversation. 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, 2000: 91-93).

All these studies which were based on various corpora to search for frequency levels and collocational information were observed by the authors as a basis for the developments of English language corpus work in Malaysia.

2. Developments of English language corpus work in Malaysia

There are not many corpora on the English language used in Malaysia. If they have been created, they are not easily available to the public (Menon, 2009). As to date, there are four notable corpora developed by different universities in Malaysia.

A corpus of English language of Malaysian School Students which is known as the EMAS Corpus was created by the Universiti Putra Malaysia (Arshad Abd. Samad et al., 2002). This corpus contains written essays and oral data of 872 students from year 5, Form 1 and Form 4 from selected primary and secondary schools in three states in Malaysia. Another corpus created by the same university was a textbook corpus of the Form One to Form Five Malaysian English language textbooks (Mukundan and Anealka Aziz, 2007).

The next two corpora were developed by the Universiti Malaya. The first is the MACLE Corpus or the Malaysian Corpus of Learner English (Knowles and Zuraidah, 2004) based on students essays. The second is the Corpus of Malaysian English (COMEL), a spoken corpus project which is still in the process of development.

Another learner corpus which has not made public is the Corpus Archive of Learner English Sabah-Sarawak or known as the CALES Corpus (Botley et al., 2005). It is an ongoing project which started in 2003. The corpus consists of argumentative essays written by students taking English proficiency courses at the Sabah and Sarawak campuses of the Universiti Teknologi MARA, Universiti Malaysia Sarawak (UNIMAS) and Universiti Malaysia Sabah (UMS).

With regard to studies of prepositions using the corpora available in Malaysia, Norwati Roslim (2004, 2009) have utilized the EMAS Corpus and the textbook corpus developed by the Universiti Putra Malaysia.

Using the EMAS Corpus, Norwati Roslim (2004) studied the use, mastery and developmental patterns of English prepositions of place, *in, on* and *at* used by the year 5, Form 1 and Form 4 Malaysian students in their written picture essays. Based on all the concordance lines where *in, on* and *at* had appeared as prepositions of place, grammatical collocations of these prepositions seemed to be the students' problem which had affected their writing ability.

Mukundan & Norwati Roslim (2009) then conducted a corpus-based investigation on *to, of, in, on, from, at, by, after, before, between, near, under, behind* and *in front of* which are presented in three English language textbooks used by lower secondary schools in Malaysia. The findings showed that there was a difference between the textbook corpus and the British National Corpus (BNC) in terms of the order of these grammatical items. Another finding was the similarities and differences of the use of these items as prepositions in textbooks in terms of their co-occurrence with other parts of speech.

Both studies provide recommendations in the teaching and learning of English as a second language (ESL) in a Malaysian context. Based on the EMAS Corpus, recommendations were made to include collocations in the teaching and learning activities while the study on the textbook corpus provided guidelines for teachers in deciding how best to supplement the text with activities that will give learners exposure to target grammar items which were not sufficiently presented in the textbook.

3. Implications on the teaching and learning of English

Generally, much has been said about contributions of corpus linguistics in the teaching and learning of English and the authors have observed this in relation to findings gained in these review of corpus-based research on prepositions.

The teaching of grammar has always been an important concern in Malaysia. The syllabus of the Kurikulum Bersepadu Sekolah Menengah (KBSM) or the Integrated Secondary School Curriculum for English language as outlined by the Curriculum Development Centre (CDC) of the Ministry Of Education Malaysia (MoE) emphasizes its importance and the grammar items to be taught which also include English prepositions are listed in the syllabus. Therefore, these review of studies are very significant as they could benefit curriculum planners, textbook writers and teachers in the Malaysian ESL context.

3.1 Curriculum planners

Curriculum planners would benefit from these studies as they would be able to determine the contents of the English language syllabus for schools in Malaysia. Mindt and Weber (1989) and Kennedy (1998) studies have shown the most frequent and common prepositions in a general corpora, the LOB and the Brown Corpus. The curriculum planners may want to decide on selecting items for linguistic features to be included in the syllabus based on word frequency. They would be able to consider which prepositions and how many prepositions need to be introduced and retained in every level of school years. They can further check whether the present curriculum is at a level appropriate to the needs of students with respect to prepositions specifically.

Additionally, findings on collocation studies (Sinclair and Renouf, 1991; Kennedy, 1998; Norwati Roslim, 2004) and pattern grammar (Hunston and Francis, 2000), may help curriculum planners to consider the addition of including prepositions in the syllabus within its collocational and grammatical framework and not only prepositions per say.

3.2 Textbook writers

Corpus linguistics has its impacts also on the content of language teaching (Kennedy, 1998; Biber & Conrad, 1999; Hunston, 2002) Findings from studies on prepositions will benefit the textbook writers in choosing texts to illustrate prepositions. For instance, descriptions in *The Longman Grammar of Spoken and Written English* (LGSWE) (Biber et. al, 2000) are helpful because they allow new factors to be considered in decisions about materials development. The frequency of use of features and the way that people actually use features in conversation, fiction, newspaper language, and academic prose provide information for teaching materials. By providing information about the frequency of use of linguistic features, corpus linguistics can inform decisions about priorities in ESL teaching materials (Conrad 2000). With regard to prepositions, they are the most frequent function word class in news and academic prose, however, they are much less common in conversation (Biber et. al, 2000: 91-93).

Furthermore, studies on the English language textbooks (Mukundan & Norwati Roslim, 2009) would also be beneficial for textbook writers. It provides information to textbook writers on decisions of the types of materials and activities that they can look into while preparing for the next cycle of English language textbooks. With the help of suitable materials and activities developed by textbook writers for the textbook content, the presentation of prepositions would be sufficient and teachers would be able to draw students' attention in order to help them in the learning of prepositions.

Such important role of corpora goes beyond simply providing more realistic examples of language usage but also looking critically at existing language teaching materials (Mc Enery & Wilson, 2001: 120).

3.3 Teachers

These reviews of studies would also have implications on the teachers themselves. The findings from Mukundan & Norwati Roslim (2009) on the present school textbooks provide teachers with additional knowledge on the content of the textbooks with regard to the teaching of prepositions in grammar sections as well as how prepositions appear throughout these textbooks. Teachers could recognize which prepositions should be given more attention based on their appearance as a whole throughout the textbook. In fact, teachers can prepare themselves with strategies and provide supplementary activities for prepositions which appear less. Therefore, the strengths and weaknesses on the

content of the textbooks with regard to the presentation of prepositions provide some kind of indicators to teachers to come out with activities which will be useful for learners during their English periods.

4. Conclusion

Studies on prepositions in the field of corpus linguistics have revealed some of the greatest contributions of corpus linguistics for the teaching and learning of English in the ESL context. However, more studies need to be conducted to add to the body of knowledge as prepositions are constant source of difficulty for the ESL/EFL learners and therefore must be taken seriously and studied more systematically.

References

- Arshad et al. (2002). *The English of Malaysian School Students (EMAS) corpus*. Serdang: Universiti Putra Malaysia Press.
- Biber, D. et. al. (2000). *The Longman Grammar of Spoken and Written English*. Essex: Pearson Education Limited.
- Biber, S.P., De Alwis, C., Metom, L., and Izza, I. (2005). *CALES: A corpus-based Archive of Learner English in Sarawak. Final Project*. Unit for Research, Development and Commercialization. Universiti Teknologi MARA.
- Hunston, S. and Francis, G. (2002). *Corpora in Applied linguistics*. Cambridge: Cambridge University.
- Kennedy, G. (1991). Between and through: The company they keep and the functions they serve. In K. Aijmer and B. Altenberg (eds). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. Longman. 95-110.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Kennedy, G. (2007). Between and through: The company they keep and the functions they serve. In Wolfgang Teubert & Ramesh Krishnamurthy (eds.). *Corpus Linguistics : Critical Concepts in Linguistics*. Vol IV. Oxford : Routledge. 26-40.
- Knowles, G and Zuraidah Mohd Don. (2004). Introducing MACLE: The Malaysian Corpus of Learner English. *1st National Symposium of Corpus Linguistics and Foreign Language Education*, South China Normal University, Guangzhou, China, 10-14 October 2004.
- Mc Enery, T. & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Menon, S. (2009). Corpus-based analysis of lexical patterns in Malaysian secondary school Science and English for Science and technology textbooks. Universiti Putra Malaysia.
- Mindt, D. and Weber, C. (1989). Prepositions in American and British English. *World Englishes*, 8: 229-238.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Mukundan, J. and Norwati Roslim. (2009). Textbook Representation of Prepositions. *English Language Teaching Journal: Canadian Center of Science and Education*. 2(4). 13-24.
- Mukundan, J. & Anealka, A. H. (2007). A forensic study of vocabulary load and distribution in five Malaysian Secondary School Textbooks (Forms 1-5). *Pertanika Journal of Social Science and Humanities*. 15(2): 59-74.
- Norwati Roslim. (2004). Corpus Linguistics: Dealing with a Learner Corpus. Paper presented at International Language Learning Conference [1st: 16-18 December 2004: Penang]
- Norwati Roslim. (2004). A corpus-based study on an English preposition of place, *at*. In Mukundan, J. et al. (Ed). *ELT Matters 2: Developments in English Language Learning and Teaching*. Serdang: Universiti Putra Malaysia Press.
- Renouf, A. and Sinclair, J.M. (1991). Collocational frameworks in English. In Aijmer, K. & Altenberg, B. (1992). *English Corpus Linguistics*. New York: Longman Inc.

Table 1. The 14 most frequent prepositions in the Brown and LOB corpora (from Kennedy, 1998: 139)

	Brown Corpus (%)	LOB Corpus (%)
of	29.8	28.6
in	17.1	16.4
to	9.1	8.8
for	7.4	7.1
with	6.0	5.8
on	5.1	5.1
at	4.4	4.4
by	4.3	4.6
from	3.6	3.8
into	1.5	1.3
about	1.0	1.0
through	0.7	0.6
over	0.7	0.6
between	0.6	0.7