Analyses of the semantic features of the lexical bundle [(VERB) PREPOSITION the NOUN of]

Siaw-Fong Chung, F. Y. August Chao, Tien-Yu Lan and Yen-Yu Lin
sfchung@nccu.edu.tw
National Chengchi University, Taiwan

ABSTRACT

This paper investigates twenty-two prepositions in two different lexical bundles – [PREPOSITION *the* NOUN *of*] (*at the point of*, *from the perspective of*, etc.) and [VERB PREPOSITION *the* NOUN *of*] (*shouted above the noise of*, *suffering from the effects of*, etc.), the only difference being that the former does not include the head verb that is present in the latter. Strings of constructions were extracted from the British National Corpus and the types of possible verbs, prepositions, and nouns in each possible combination were analyzed. The paper also details an experiment in which the types of nouns under each of the twenty-two prepositions were coded by human subjects in terms of their semantic features. Finally, a computer program was also utilized to calculate the shared meaning of the different VERBs and NOUNs. The results showed that the nouns in [(VERB) PREPOSITION *the* NOUN *of*], though they might form clusters of meanings, may not behave in the same way with and without the presence of the verbs.

Keywords: *prepositions, lexical bundles, nouns, semantic features, corpus, constructions*

I. INTRODUCTION

According to Biber et al. (2004) and Levy (2008), who investigated 'lexical bundles' in spoken versus written registers, lexical bundles, or multi-word sequences, are "the most frequent recurring lexical sequences in a register", including, but not limited to, four-word sequences such as *do you want to*, *take a look at*, *to come up with*, *I don't know what*, *one of the things*, *those of you who*, and so forth (p. 376). Their instances of bundles may or may not contain a head verb.

Most previous studies on lexical bundles focus on register-specific materials. For instance, Biber (2009) compared the most common multi-word patterns in conversation and academic writing and found that the multi-word patterns occurring in the two registers are different. Patterns in conversation tend to be fixed sequences including both function words and content words; patterns in academic writing, however, tend to

138

be formulaic frames consisting of invariable function words with an intervening variable slot that is filled by content words.

Focusing on academic prose, Biber proposed that there are numerous fillers that may occur in the frame *the * of the*. It was found that four different prepositions tend to precede *the * of* to form the four-word lexical bundles: *at the * of, on the * of, in the * of,* and *to the * of,* all of which are patterns of interest in the present paper. Among these, the most distinctive frame is *at the * of*, which co-occurs frequently with the fillers *end, time, beginning, level, expense, start, center/centre, top,* and *base*. On the other hand, *in the * of* takes several high frequency fillers that are distinctively used in this frame, namely *case, absence, form, context, course,* and *process*. Using a similar 'frame', this paper investigates the distributions of different variables (in capitals) in the pattern [(VERB) PREPOSITION *the* NOUN *of*]. The present work focuses not on any specific genre, but on material contained in the British National Corpus (BNC), a general corpus. We propose that similar clusters of nouns (and verbs) can also be found in a general corpus. Our study further hypothesizes that the VERBs and NOUNs can be measured in terms of their semantic relatedness. To answer this question, two types of methodologies were employed – one including an experimental-based analysis of semantic features, while the second involves the automatic extraction of semantically related hypernyms. The details of this will be illustrated in the next section.

In a different study, also following a genre approach, Luzón Marco (2000) investigated the collocational framework in the medical research paper. The results showed that two of the most common frameworks in the corpus are: [*the* NOUN *of*] (e.g. *the start of*), *a* NOUN *of* (e.g. *a variety of*). [*The* NOUN *of*] tends to be used in expressing the construction of nominalizations (e.g. *the cloning of*); [*a* NOUN *of*] is frequently applied to describe the process of quantifying and categorizing. Another important finding is that these two frameworks are likely to precede or follow the collocates belonging to specific semantic classes. For example, *the risk of* is always preceded by verbs with causative meanings (*related/associated with/to the risk of*). It was concluded that the selection of specific collocates for these frameworks is conditioned by the linguistic conventions of the genre. In a different study and in an attempt to improve the understanding of the function of lexical bundles in academic prose, Biber et al. (2004) compared the use of such bundles by published authors in history and biology. The most

frequent four-word lexical bundles in these genres were classified in terms of their structure groups. The findings revealed that lexical bundles in history mainly belong to two structural groups – noun phrases and prepositional phrases – while lexical bundles in biology cover a wider range of structural groups, including noun phrases, prepositional phrases, [*it* + Vbe + adjective], [Vbe + complement], and [noun phrase + V + complement] clause fragments. In general, in both history and biology genres, the majority of the bundles could be categorized into the groups containing a noun phrase with an *of* phrase fragment (e.g. *a measure of the*, *the beginning of the*) and prepositional phrases with an embedded *of* phrase (e.g. *as a function of*, *at the beginning of*, *at the university of*). From here, one can see that most of these studies in lexical bundles needed to deal with noun phrases and prepositional phrases in one way or another. For instance, Biber and Conrad (1999) found that, in academic prose, 60% of the bundles are phrasal, parts of noun phrases or prepositional phrases, as *in the case of*, *as a result of*, *on the basis of*, and *on the other hand*. Noun phrases and prepositional phrase fragments were also found as the most frequent patterns in academic prose (also found in Biber et al. 2004 and Hyland 2008a, 2008b). Similarly, scientific discourse is also characterized by very frequent occurrences of nouns, long words, prepositions, conjuncts, being agentless, and by-passives, as well as past participial adverbial clauses (Biber 1988). In a book by Silvestre (2009), he investigated the particle meanings of *in* and *on*. In his methodology, "*multi-word lexicalized expression*" was recognized as one of the criteria in extracting verb-particle constructions (VPC). Multi-word expressions were included in his VPC analysis because some uses of *in* and *on*, such as in "*to decide in favor of sb*" are "motivated by" the noun (*favor* in this example) "rather than being directly bounded to the verbal element" (p. 159). Given the above studies, we postulate that it might be useful to investigate lexical bundles by examining the nouns (and the verbs) in a given construction. This paper inspects both the nouns and the verbs in the constructions [(VERB) PREPOSITION *the* NOUN *of*], which co-occur with twenty-two different prepositions.[1]

Rather than looking at one particular preposition, this paper investigates a group of prepositions in terms of distributional patterns. As Silvestre (2009) discovered, some of the particles were more closely related to the nearby nouns than to the verbs, and this is the kind of phraseological phenomenon we inspect in this study. The foci of this study

are: (a) To compare the distributions of NOUNs and VERBs in the construction [(VERB) PREPOSITION *the* NOUN *of*] when twenty-two different prepositions are involved; and (b) To display similarities of meanings among NOUNs and VERBs in this construction. The ultimate goal is to propose a systematic way to analyze semantic features of nouns and verbs given a preposition-containing construction. Two types of methodologies were employed, namely experimental analysis of semantic features, and computational calculation of semantic meanings by measuring the common hypernym, if any, found between any two nouns or verbs. Both these methodologies complement each other and the results were cross-referred.

II. DATA FROM THE CORPUS

All data discussed in this paper were taken from the written portion of the BNC, retrieved through BNCWeb, a platform which allows access to the BNC through a search engine of its own (Hoffmann et al. 2008). Twenty-two prepositions (*about*, *above*, *across*, *after*, *against*, *among*, *around*, *as*, *at*, *beside*, *by*, *down*, *for*, *from*, *in*, *into*, *like*, *of*, *off*, *on*, *onto*, and *with*) were investigated. It was hypothesized that the groups of words that appear with a similar preposition would share some similarities in semantic features. In the following sections, the distributional patterns will first be discussed, followed by a semantic analysis by human subjects. Finally, in section III, a computational program will be discussed.

II.1. Distributional patterns

In the written portion of the BNC, 373,258 instances of [PREPOSITION *the* NOUN *of*] and 86,877 instances of [VERB PREPOSITION *the* NOUN *of*] were found. These instances were analyzed according to the different types of verbs and nouns used in them.

Table 1, below, displays the most frequent patterns for each preposition, along with their frequencies and percentages. For example, *about the <u>nature</u> of* has a frequency of 225 and the percentage of *nature* in the construction of [*about the* NOUN *of*] is 4.5%. Patterns with the same scores were all listed (as for *among* and *onto*).

Table 1. Frequencies of [(VERB) PREPOSITION *the* NOUN *of*] in the BNC.

Prep.	Four-Word Bundles [PREPOSITION *the* NOUN *of*] Most Frequent Nouns (Freq., %)	Five-Word Bundles [VERB PREPOSITION *the* NOUN *of*] Most Frequent Verb-Noun Pairings (Freq., %)
about	*about the **nature** of* (225, 4.5%)	***set** about the **task** of* (11, 0.52%)
above	*above the **level** of* (57, 10.14%)	***shouted** above the **noise** of* (3, 2.01%)
across	*across the **top** of* (49, 5.85%)	***runs** across the **front** of* (3, 1.01%)
after	*after the **death** of* (270, 7.2%)	***look** after the **interests** of* (7, 1.29%)
against	*against the **background** of* (176, 6.22%)	***seen** against the **background** of* (10, 1.02%)
among	*among the **members** of* (36, 5.15%)	***discovered** among the **remains** of* ***was** among the **members** of* ***were** among the **beneficiaries** of* ***distribute** among the **members** of **was** among the **founders** of* ***are** among the **findings** of* ***be** among the **victim**s of* (2, 1.36% each)
around	*around the **time** of* (90, 6.86%)	***was** around the **time** of* (5, 0.99%)
as	*as the **result** of* (183, 3.66%)	***used** as the **basis** of* (23, 1.24%)
at	*at the **end** of* (1086, 21.72%)	***is** at the **heart** of* (101, 0.98%)
beside	*beside the **bed** of* (4, 5.33%)	***lived** beside the **Loch** of* (2, 10%)
by	*by the **end** of* (688, 13.76%)	***completed** by the **end** of* (56, 0.45%)
down	*down the **side** of* (89, 7.09%)	***turned** down the **offer** of* (12, 1.42%)
for	*for the **rest** of* (207, 4.14%)	***called** for the **establishment** of* (24, 0.40%)
from	*from the **point** of* (143, 2.86%)	***suffering** from the **effects** of* (24, 0.44%)
in	*in the **case** of* (259, 5.18%)	***was** in the **middle** of* (88, 0.44%)
into	*into the **hands** of* (247, 4.94%)	***fall** into the **trap** of* (37, 1.36%)
like	*like the **rest** of* (158, 7.57%)	***look** like the **sort** of* (14, 3.33%)
of	*of **The House** of* (70, 1.4%)	***is** of the **order** of* (25, 1.73%)

off	*off the **coast** of* (107, 10.3%)	***fallen** off the **back** of* (6, 0.98%)
on	*on the **basis** of* (357, 7.14%)	*was on the **verge** of* (93, 0.76%)
onto	*onto the **surface** of* (15, 7.5%)	***screws** onto the **front** of* ***moves** onto the **carbon** of* ***tacked** onto the **end** of* ***built** onto the **end** of* (2, 2.56%, respectively)
with	*with the **help** of* (155, 3.1%)	***charged** with the **murder** of* (52, 0.67%)

From Table 1 it can be seen that higher percentages were generally found for the four-word bundles (without the verb) than the five-word bundles (with the verb). The percentages for the [VERB PREPOSITION *the* NOUN *of*] patterns are all lower than 5%, except *lived beside the Loch of*, although its frequency is only 2, further indicating that very few patterns were found matching this construction. For the four-word combination, higher percentages indicate that the top noun patterns are less varied (e.g. *at the end of* (21.72%), *by the end of* (13.76%), *off the coast of* (10.3%), and *above the level of* (10.14%).

From Table 1, the most frequent nouns (column 2) may not be the same as the verb-noun pairings (column 3) because the verbs added in column 3 might affect the most frequent nouns used under each combination. Interestingly, in two of the prepositions (*against* and *around*), similar nouns were found in both four- and five-word lexical bundles. This shows that *against the background of* and *around the time of* are equally frequent with or without the verbs appearing before them, further indicating the strength of the occurrences of nouns with the prepositions.[2] Some prepositions (e.g. *of*, *as*, *with*, and *about*) have a wider range of nouns, as the most frequent nouns (*The House*, *result*, *help*, and *nature*, respectively) constitute less than 5% of the total number of nouns in the [PREPOSITION *the* NOUN *of*] patterns.

Thus, Table 1 provides a general overview regarding the different prepositions when appearing in the [(VERB) PREPOSITION *the* NOUN *of*] construction. In the following section, we discuss an experiment we conducted in order to code the semantic features of the nouns.[3]

II.2. Semantic coding

Since there are twenty-two prepositions and each has its own instances to be analyzed, human subjects were trained to code the semantic features of this part of the analysis. Two Ph. D. experimenters were in charge of the experiment and the procedures that were followed are described in the following.

In this experiment, one hundred instances of each of the twenty-two prepositions were analyzed. Six English-major university subjects were paid to participate in the analysis process. Among the subjects, two senior subjects were each responsible for six prepositions, two junior subjects each took responsibility for three prepositions, and two junior subjects were each made responsible for two prepositions. The task was assigned based on a student's experience in coding the semantic features. Two of these senior students had had training in coding semantic features for over six months.

Each preposition contained one hundred noun types to be analyzed. The selection of the noun types was based on the frequency of patterns in the whole BNC, from high to low percentages. In this experiment, the singular and plural forms of the nouns were counted as one, and the duplicate one was deducted if the percentage was lower, e.g. *at the corner of* (0.32%) and *at the corners of* (0.17%), so the latter one was deducted.

The noun of the preposition was to be categorized by the subjects (e.g. *of the bank of*, *of the history of*). During the analysis process, the subjects were allowed to use dictionaries, but other documents or books, or having discussions with others were not recommended. The purpose of such restrictions was to avoid any distractions that could affect the subjects' judgment. The categorization should be based on their instinct.

The subjects were required to sort the nouns into categories based on similarity of semantic features. No exhaustive list was provided, but the generality of the category level was hinted at through the instructions. For instance, before starting, the subjects were given instructions such that *bank*, *post office*, *library*, and *cottage* should be categorized and tagged as "building". The subjects were then asked to generate the category names by themselves. All data were distributed through excel files and subjects were allowed to work at their own pace. The subjects saw the nouns in excel files, exemplified in Table 2 for the preposition *of*.

Table 2. Example of excel data used for semantic coding of nouns

	A	B	C	D	E	F	G	H
1	*of*	*the*	*	*of*	*bank*	25	0.50%	
2	*of*	*the*	*	*of*	*history*	23	0.46%	
3	*of*	*the*	*	*of*	*city*	23	0.46%	
4	*of*	*the*	*	*of*	*law*	23	0.46%	
5	of	the	*	*of*	*role*	22	0.44%	

The subjects were required to analyze the nouns in column E, which originally occupied the asterisk (*) in the phrase but were moved to the end for the sake of convenience. The result of the analysis was tagged in column H. If a noun could be categorized into more than one category (e.g. *bank*, as in (1) a financial establishment, and (2) the land alongside or sloping down to a river), all categories would be provided. Furthermore, the subjects were required to provide their own criteria for the categorization. An example of their definitions is displayed in Figure 1.[4]

Directions for categorization
1. 2D space: 某平面區域或由線構成之形狀，如field, area, circle等
2. 3D space: 包含立體空間之名詞，如kingdom, valley, nature等
3. Action: 特由人發起之短暫動作，如use, work, coming等
4. Activity: 表一活動現象且具持續性之名詞，如business, service, growth等
5. Body part: 人體部位或其他生物之身體部位，如face, hands, tail等

Figure 1. Definitions of semantic categories by subjects.

The two experimenters in charge would then collect and standardize the results from all subjects. If inconsistency was detected, the subjects were required to carry out revisions. After the coding, one of the experimenters then went through each of the instances for all 22 prepositions and checked whether consistency had been achieved. With the criteria and the revisions, the analysis process was made more systematic. Some parts of the final results of the subjects' analysis are shown in the following snapshots as examples.

◇	A	B	C	D	E	F	G	H
1	by	the	*	of	end	688	13.76%	point
2	by	the	*	of	use	84	1.68%	method
3	by	the	*	of	department	74	1.48%	part
4	by	the	*	of	time	71	1.42%	time
5	by	the	*	of	secretary	65	1.30%	profession

◇	A	B	C	D	E	F	G	H	I	J
1	onto	the	*	of	surface	15	7.50%	2D space	surface	
2	onto	the	*	of	end	14	7%	sequence	point	
3	onto	the	*	of	back	13	6.50%	body part	location	direction
4	onto	the	*	of	top	11	5.50%	location		
5	onto	the	*	of	front	9	4.50%	location		

Figure 2. Sample of completed coding.

Based on the outcomes of the semantic coding, results such as the following Tables 2 and 3 were obtained. Since the lists are long, this paper only provides selective categories. Twenty-two tables were prepared for twenty-two prepositions.

Table 2. Selective semantic features of NOUNS in [*on the* NOUN *of*].

Categories	Groups of Noun Collocates	Explanation
1	*on the (edge, verge, side) of*, *on the (top, surface, end, point, back) of*	The nouns in *on the * of* usually denote positions. The first three (*edge, verge, side*) have similar meanings. The other five (*top, surface, end, point, back*) can be used to denote different location or positions on concrete subjects; moreover, *point* and *end* can also refer metaphorically to a temporal meaning.
2	*on the (basis, grounds, floor) of*	The three nouns all refer to the base of something. However, *on the basis of* and *on the grounds of* tend to be followed by abstract nouns while *on the floor of* usually goes before concrete nouns.
3	*on the (day, night, morning) of*	The three nouns refer to different periods of the day.
4	*on the (face, outskirts, site) of*	The three are concrete (visible) nouns.
5	*on the (role, subject, eve, future, development, use, number, question, nature, issue) of*	These are abstract nouns. The phrases with *subject*, *question*, or *issue* here are usually followed by different topics or themes for discussion.

Table 3. Selective Semantic Features of NOUNS in [*at the* NOUN *of*].

Categories	Groups of Noun Collocates	Explanation
1	*at the (end, top, back, bottom, centre, edge, base, side, front) of*	These are nouns denoting locations.
2	*at the (beginning, start) of*	The nouns denoting different times also occur frequently in this construction.
3	*at the (head, hands, heart, foot) of*	The nouns found here refer to different parts of the body.

Based on the semantic coding of the nouns, we further confirm that it is possible that the nouns that share the same construction reflect certain similar clusters of meanings. In order to examine further how far these similarities can be measured, the following computational process was undertaken.

II.3. Automatic data extraction

In order to calculate all the possible verbs and nouns that might fill the [VERB] and [NOUN] slots of [(VERB) PREPOSITION *the* NOUN *of*], a program was written to measure the combination of these verbs and nouns. The program consisted in the following steps:

a) First, based on the retrieved data indicated in Table 1 above, the occurrences of each verb or noun that appears with its respective preposition were recorded. For instance, for [PREPOSITION *the* NOUN *of*], the instance *about the nature of* would mark 1 occurrence for *nature* under the preposition of *about*. For [VERB PREPOSITION *the* NOUN *of*], *seen against the background of* would mark 1 for *seen* as well as 1 for *background* for the preposition *against*.

b) For both verbs and nouns, all lemmatized forms were counted as a similar group (e.g. *seen* was grouped under *see* and so were *saw*, *sees*, *see*, and *seeing*). The lemmatization process followed Someya's (1998) e-lemma list.

c) A normalized score called the z-score was then used to measure the occurrences of verbs and nouns found in these two constructions. The z-score was selected because it reduces the problems that arise when a word is particularly high or low in frequency.[5]

As our previous hypothesis assumed that all the verbs and nouns that occur with a similar preposition might share certain similarities, our program also included a calculation of similarities. This was executed through finding out the common shared hypernym(s) for any two verbs or nouns in a lexical resource called WordNet 3.0 (cf. Fellbaum 1998). The following example shows two nouns for *among* in the [PREPOSITION *the* NOUN *of*] construction.

(a) *among the <u>group</u> of*
(b) *among the <u>world</u> of*

In WordNet, we first found many different synsets (synonymous sets) for *group* and for *world*. These synsets indicate the different meanings of *group* and *world*. *Group* has three synsets of nouns, whereas *world* has seven synsets of nouns. Each of the three synsets from *group* were paired with each of the synsets from *world* to find any common hypernyms. The number of common hypernyms was then recorded, and presented as z-scores.

The results are shown in Table 5 below. A high z-score might mean the nouns or verbs of these prepositions possess a higher number of common hypernyms. A higher number of common hypernyms usually means that the meanings among the nouns or verbs might be closer to one another. This part of the analysis thus attempted to prove our hypothesis of semantic relatedness among the nouns or verbs in the [(VERB) PREPOSITION *the* NOUN *of*] construction.

Table 5. Total Z-scores of different types of nouns.

Prep	Total Z-score		Prep	Total Z-score	
	Nouns	**Verbs**		**Nouns**	**Verbs**
as	12.01	-1.5	*off*	-0.17	1.22
with	11.72	-1.25	*onto*	-0.42	-0.51
from	11.65	-2.41	*above*	-0.73	0.14
of	10.25	-0.08	*in*	-1.19	0.83
across	1.94	-0.7	*down*	-1.8	-1.09
like	1.73	0.03	*on*	-3.25	-1.66
around	1.49	1.60	*about*	-4.36	-0.35
against	1.21	0.07	*into*	-4.86	-2.21
among	0.97	0.31	*for*	-5.35	-2.19
at	-0.01	0.87	*after*	-6.14	-0.51
beside	-0.13	0.04	*by*	-7.82	4.90

The results in Table 5 show that some prepositions (*as*, *with*, *from*, and *of*) co-appear with nouns with higher z-scores, but their verbs are not necessarily displaying higher z-scores. These controversies demonstrate that the types of nouns co-occurring with these prepositions (*as*, *with*, *from*, and *of*) are more similar than their verbs are. For instance, a closer investigation through the semantic coding in the previous section shows that the NOUNs in [*as the* NOUN *of*] display semantic groups related to amount (such as *amount*, *sum*, *majority*, *proportion*, *ratio*, etc.) and point in time or space (such as *end*, *beginning*, *center*, *start*, *last*, *first*, etc.), and so forth. The VERBs in [VERB *as the* NOUN *of*] (e.g. *regarded as*, *seen as*, *defined as*, *calculated as*, etc.) are more varied and it is harder to generate a pattern for them.[6]

Other than that, Table 5 also shows a reverse pattern, i.e., some VERBs in [VERB *by the* NOUN *of*] seem to show a higher z-score than those of NOUNs in [*by the* NOUN *of*]. This indicates that constructions such as *completed by the NOUN of*, *approved by the NOUN of*, *divided by the NOUN of*, etc. might share greater similarities than those of [*by the NOUN of*]. From this example, too, we might assume that those possessing higher scores for verbs are likely to form stronger bonds for [VERB+PREP] than those of nouns. However, this part will need further investigation, as the measurement of bonding is not the current focus of this work but will be an interesting aspect to explore.

To sum up this section, we used a computational program to calculate the similarities of meanings among the NOUNs or VERBs in the constructions [(VERB) PREPOSITION *the* NOUN *of*]. The results may help explain whether a noun behaves similarly with or without the presence of a verb in the construction [(VERB) PREPOSITION *the* NOUN *of*]. As shown in Table 5, the nouns may not behave similarly with the presence of the verbs under a similar construction.

IV. CONCLUSION

Unlike previous studies, our sequences of words contain two patterns – one with the presence of the head verb [VERB PREPOSITION *the* NOUN *of*], and one without the head verb [PREPOSITION *the* NOUN *of*]. This paper analyzes the semantic features shared by all the VERBs and NOUNs in the lexical bundle [(VERB) PREPOSITION *the* NOUN *of*]. In order to ensure that the nouns are semantically related, an experiment

was run in which subjects were asked to code the semantic features of the nouns in this construction. To compare the data, an automatic data extraction program was run to measure the shared meaning (their hypernyms) in a lexical resource.

Some limitations remain because the verbs in [VERB PREPOSITION *the* NOUN *of*], especially the copula BE (e.g. *was among the members of*), were not completely dealt with at the present stage. These copulas might cause problems as they do not possess a specific meaning, and they also tended to be dropped in the WordNet searches. The hypothesis-testing of semantic relatedness on the verbs, therefore, will need further inspection.

Notes

[1] Although more attention will be given first to the nouns.

[2] In addition, some prepositions (*among*, *around*, *at*, *in*, *of*, and *on*) appear to be less likely to form verb-particle constructions, as their most frequent patterns collocate more often with a copula BE, showing the tendency of the prepositions to become a single preposition rather than a verb-particle combination.

[3] At present, only the nouns have been discussed because the analysis of the verbs was found to entail more difficulties than expected. In addition to removing the copula BE, which contains no lexical meaning, there was also the problem of selecting suitable semantic features.

[4] Instructions and definitions were given in Mandarin to avoid misunderstanding. The results in Figure 1 might not represent the finalized code, as revisions and modification might have been undertaken.

[5] More about the z-score can also be found in McEnery and Wilson (1996) and Hunston (2002). McEnery and Wilson further mentioned that the z-score is particularly useful in "multi-word units" (p. 87).

[6] When most of the verbs fell under a general category of 'Act', this might mean a problem existed with the WordNet verb trees and it was not due to the methodology itself. However, an evaluation of the WordNet hierarchies is beyond the scope of the present work.

REFERENCES

Biber, D. 1988. *Variation across Speech and Writing.* Cambridge: Cambridge University Press.

Biber, D. 1992. "On the complexity of discourse-complexity: A multi-dimensional analysis". *Discourse Processes*, 15, 133-163.

Biber, D. 2009. "A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing". *International Journal of Corpus Linguistics,* 14 (3), 275-311.

Biber, D. and Conrad, S. 1999. "Lexical bundles in conversation and academic prose". In Hasselgard, H. and S. Oksefjell (Eds.) *Out of Corpora: Studies in Honor of Stig Johansson*. Amsterdam: Rodopi, 181–189

Biber, D., Conrad, S. and Cortes, V. 2004. "If you look at…: Lexical bundles in university teaching and textbooks". *Applied Linguistics*, 25 (3), 371-405.

Fellbaum, C. (Ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Hoffmann, S., Evert, S., Smith, N., Lee, D. and Prytz, Y.B. 2008. *Corpus Linguistics with BNCweb – A Practical Guide.* Frankfurt am Main: Peter Lang.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hyland, K. 2008a. "Academic clusters: Text patterning in published and postgraduate writing". *International Journal of Applied Linguistics*, 18 (1), 41-62.

Hyland, K. 2008b. "As can be seen: Lexical bundles and disciplinary variation". *English for Specific Purposes,* 27 (1), 4-21.

Levy, S. 2008. *Lexical Bundles in Professional and Student Writing: A Study in Linguistic Variation.* Germany: VDM Verlag Dr. Muller.

Luzón Marco, M.J. 2000. "Collocational frameworks in medical research papers: A genre-based study". *English for Specific Purposes,* 19, 63-86.

McEnery, T. and Wilson, A. 1996. *Corpus Linguistics*, 2nd ed., Edinburgh: Edinburgh University Press.

Silvestre, A.J. 2009. *Particle Semantics in English Phrasal and Prepositional Verbs: The Case of IN and ON*. Saarbrücken: VDM Verlag.

Someya, Y. 1998. *E-lemma list*. (free online software available at <http://www.lexically.net/downloads/version4/downloading%20BNC.htm>).