MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology

Vijayakumar Chintalapalli
Assistant Professor
University of Dammam, KSA

Mirza Mohammed Ali Baig
Lecturer in English
University of Dammam, KSA

Dr C Vijayakumar is currently an Assistant Professor at University of Dammam, Saudi Arabia. He completed his M.Phil. and PhD in English Language Education from the English and Foreign Languages University (EFLU), Hyderabad. His areas of interest include ESP, EAP, CALL, Corpus based Language Learning and Pedagogical Lexicography. Email: vijayakumar.chintalapalli@gmail.com

Mr. Mirza Mohammed Ali Baig is a Lecturer in English at University of Dammam, Saudi Arabia. He has 25 years of teaching experience in the EFL contexts. His areas of interest include EFL teaching, Academic Writing, and Language through Literature.

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology

by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

Abstract

The discourses of science and technology, in the process of accounting for innovations, either invent new 'words' or reinvent the existing words in new contexts of language use. However, the process of upgrading the communicative conventions takes place within the constraints of discourse communities. This is reflected in the quotation that forms the first part of the title of the paper which Michael Stubbs pointed at (Stubbs, 1996). Thus, in the case of lexis, every piece of text the learner encounters either introduces the target word in an unfamiliar co-text and context or reinforces the existing combinations and contexts in novel ways. As the genres or domains become more focussed or specialized, the occurrences of 'primed'(Hoey, 2005) words become obvious. Such primed combinations, *phraseologisms* (Meunier & Granger, 2008), are said to affect learner comprehension and use of lexis within the domains. In this paper, we consider suggestions offered by experts from the field of English for Academic Purposes, and make an attempt to show how we can find phraseologisms by exploring the 120 million word Academic sub-corpus of Corpus of Contenporary American English (COCA) created by Mark Davies(Davies, 1998-). While the major objective of the paper is to provide a coherent pedagogic model for instruction, it will also demonstrate with examples how the model can be applied in teaching specific purpose vocabulary.

Key words: *academic corpus, phraseology, corpus-based analysis, concordance*

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

Introduction

Computer mediation in language instruction is a rather challenging task. For instance, the design and use of blogs and other social media require a special commitment from both teachers and learners. When used appropriately with traditional methods of instruction they have the potential to take us beyond our common understanding of what we can do in a non-mediated class. The present paper is situated in a context where teacher's knowledge requires the support of an additional tool such a dictionary or a corpus. It is the teaching of meaning units that are specific to certain discourses. These units are significant aspects that constitute a large part of spoken and written language use. A good command of these chunks or prefabs or sequences allows us to process the input faster and faciliates text comprehension. A few of such constructions include *visible to the naked eye, a significant amount of, strong tea,* and *please find attached.*

Contexts and Referents

While these chunks when memorized and processed as single units help us with fluency, there are other kinds of 'speech habits', such as the sentence patterns Hornby(1946) identified, that help us create new models for language teaching and learning. Now the challenge ahead of the teacher is not so much with the number of meanings the learner has to memorize but with the process of enabling the learner transfer his learning to new contexts of understanding and use. Which means the learners need to be aware of the use of the chunk or word in its context. Let us look at the table Hornby provided.

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

1	2	3	4
This That It	is	(made of) (a piece of) (a bit of) (a lump of) (a sheet of)	Glass Wood Paper Stone Leather Sugar Furniture

The above table when used as an exercise will provide the learner with a structure in which the word 'lump' can be used. But the examples in the table can serve a limited purpose that of teaching a specific pattern, and not the word and any strategies of learning the word knowledge. They do not also provide any clues as to what words generally go together. Moreover, they propose to see both *form* and *meaning* as two different independent dimensions of language.

Form and Meaning

The views discussed in the subsequent sections mainly discuss the following views: The unit of meaning is not a word, but a lexical item. The size of this unit is anything between a word and a sentence. Each lexical item has constituents and each constituent heavily depends on the other in meaning making. No constituent is less or more important. Each word's/unit's meaning needs to be studied in relation to the textual environment; and the process of meaning making needs to be text to item rather than item to text. While trying to identify the patterns of these units we need to study nodes both syntagmatically and paradigmatically.

According to Oxford Advanced Learner's Dictionary, for instance, *a lump* is something that refer to "hard or solid, usually without a particular shape" "a heavy, lazy or stupid person" and "to a swelling under the skin". These three meaning senses are provisional and independent of their contexts of use. The learner who uses the learner's dictionary as a learning tool can remember the three senses associated with the word, and apply that

MJAL8:2Summer 2016 *ISSN 0974-8741*

knowledge—bottem up—when he meets the word in contexts. Let us now look at the following instance from Corpus of Contemporary American English (COCA).

> At this stage the fry are totally helpless and they can be seen as a solid mass, quivering like **a lump of** jelly, and difficult to distinguish as fish at all.

In the above instance the immediate collocate to the right of the phrase 'a lump of' is 'jelly'; and jelly in this context metaphorically refers to 'fry', a small fish. According to the meanings given in the dictionary a lump usually refers to either hard or solid. But when we describe a living creature generally we do not use 'lump'. However, in the above context the word lump refers to 'fry' and not 'jelly'. This discussion refers to the fact that meaning of a word is not a concatenation of a sequence of residual meanings of individual words; rather it is deduced from the combination of units that constitute the larger context. The meaning is more like a mixture of idiolects of each individual which form the larger *dialect* of the society. Although there are distinct differences between each of the idiolects, put together, each individual of the dialect understands the other without a problem (Sinclair, 2005). Within this scheme individual eccentricities that are markedly different from a common understanding of many are taken into account. Only the general agreements and common area are considered as 'provisional' meanings of the units. And some of these meaning units are codified in the dictionaries. If we observe the instance the other words that equally match the phrase 'a lump of jelly' is 'a solid mass'. In this context *lump* and *jelly, solid* and *mass* are collocates that eventually refer to 'fry'. And all these words are context dependent, that is, 'Practical Fishkeeping'.

The points we can derive from the above discussion are that:

- form and meaning cannot be seperated;
- the use of a specific form with a specific meaning largely depends on the co-text and the context;
- meaning(s) are provisional and context dependent;
- concept development is a crucial step in vocabulary development;

MJAL8:2Summer 2016 *ISSN 0974-8741*

- the larger focus of vocabulary instruction needs to aim at enabling the learner use or transfer the knowledge of word across a range of possible contexts;

The Idiom Principle and the Open-choice Principle

The argument that 'there is no distinction between form and meaning' (Sinclair, 1991: 7) and meanings are made in chunks of language led Sinclair to articulate a theory of phraseology: *The Idiom principle*. According to this theory, the language user has access to him or her large number pre-constructed chunks that are stored and retrieved in language performance. (1991: 110). And the language user can describe and interpret his understanding or use of language with the help of this theory. The following example as given in (Hunston & Francis, 2000)should be in place. If we interpret, for example the utterance *I must confess* applying the 'idiom principle' its meaning would be 'I am going to tell you something you may find unpleasant…'; whereas if the same sequence is interpreted in terms of its segments by applying the *open-choice principle* its meaning would be 'I am under an obligation to admit to a wrong doing'. In the case of *idiom principle* it is not possible to substitute, for instance, the subject *I* with another noun such as *he* or *she* because the first interpretation is associated only with *I must confess* and does not agree with its variants. The application of any one principle in analysis depends on many aspects as he says,

> For normal texts, we can put forward the proposal that the first mode to be applied is the idiom principle, since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretive process switches to the open-choice principle, and quickly back again. Lexical choices which are unexpected in their environment will presumably occasion a switch (*emphasis added*); choices which, if grammatically interpreted, would be unusual are an affirmation of the operation of the idiom principle (Sinclair, 1991:114).

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

Interpretation of a text largely depends on the environment in which a particular lexical word occurs. As we interpret a text we usually apply the idiom principle by default. However, if our interpretation of the text, as interpreted according to the idiom principle, is ambiguous and does not match with the rest of the text's interpretation, we re-interpret the text (phraseology) according to the open-choice principle.

The argument that the meaning of *a* lexical item is distributed across a string of words; and 'if the language is analysed according to the idiom principle, the meaning unit would be the primary unit of analysis' (Hunston and Francis, 2000: 25). For instance, in the phrase *have a bath, invisible to the naked eye, as accurate as the naked eye…* the meaning of *have or the naked eye* can only be interpreted with the help of words surrounded by them and not in isolation. It means that meanings of the lexical items actually spread across the phrases. Hence, lexis and grammar should not be treated as two separate domains in language description.

Francis, (1995), cited in Hunston and Francis, (2000), also echoes the same view that lexis cannot be separated from grammar. She further argues that lexical items tend to co-occur with certain syntactic structures and syntactic structures tend to co-occur with certain lexical items. And their co-occurrence is not random but systematic. She also says that in the process of meaning making an individual's choice and selection of lexis and grammar that adequately captures the communicative intent or the concepts to be communicated begins with lexis and not syntax.

> …lexis is communicatively prior to syntax. As communicators we do not proceed by selecting syntactic structures and independently choosing lexical items to slot into them. Instead, we have concepts to convey and communicative choices to make which require central lexical items, and these choices find themselves syntactic structures in which they can be said comfortably and grammatically.
>
> (Hunston & Francis, 2000)

MJAL8:2Summer 2016 *ISSN 0974-8741*

Such a description of language, where the distinction between lexis and grammar is blurred, refuses to provide separate descriptions for lexis and grammar. It means that 'the learner's reference collection will no longer be divided into 'dictionary', 'grammar' and 'usage book'' (Sinclair 1987: 107 as quoted in Hunston and Francis, 2000). It also means that identifying the grammatical patterns of lexical items is an integral part of lexicographer's task. This proposal of a lexico-grammar of English was fully realized in two volumes of 'pattern grammar' (Francis, 1995); and was published by Cobuild. As part of methodology each lexical item (both words and phrases) for around 75,000 words and phrases was examined for its 'environment' in the *Bank of English Corpus*. Sinclair and his colleagues from the Cobuild project believed that a comprehensive description of words with their grammatical patterns would help learners with their composition skills. It is at this point Hornby's patterns and Sinclair and Francis's models of language descriptions converge. Both the descriptions aim to provide the learners adequate help with their composition in addition to providing help with decoding skills.

a) the meanings of lexical items depend on the surrounding words;
b) both the idiom principle and the open-choice principles are ways of interpreting the texts. Language user faced with an utterance can use either of the ways to interpret it;
c) lexis and grammar should not be treated as two separate domains in language descriptions;
d) A lexico-grammatical description of language tends to blur the distinction between lexis and grammar; and gives primacy to phraseology
e) The major pedagogical implication of the lexico-grammatical approach aims to help learners with composition.

A Pedagogic Model: Corpus-based Exploration

Corpus query of any kind usually involves the observation of patterning from an "unbiased selection of lines from the whole concordance." As Sinclair(2003: x) suggests this unbias is

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

not only an important condition at the level of selecting data for analysis, but also at the level of analysis. He argues, "…many linguists argue that if you find a single example that does not fit a generalization you should discard the generalization. This is just not valid in corpus work,…because corpora contain all sorts of variation and even error… so *you have to hunt for general patterns among a amass of varying and sometimes even conflicting evidence*. (emphasis added)"

Although there are linguists who wish to account for each instance at the level of example itself and not compare the word/phrase with other instances of language use for purposes of generalization, Sinclair's views are different from them. Sinclair believed that when a large quantum of organized data was observed it was possible to observe a pattern in it; and lexis and every other meaningful particle has a tendency to throw up such patterns.

According to Johns T. (1991), the computer is an informant and does not have to be 'intelligent'. On the contrary, the learner needs to be intelligent to find answers to question from data. The most usual format of organizing data on a computer is called KWIC (Key Word In Context) in which the node or the key word is arranged 'one below the other down the centre of the page with a fixed number of characters to the left and to the right' (p: 4). The exploration of concordances,

> …can have a considerable influence on the process of language learning, stimulating enquiry and speculation on the part of the learner, and helping the learner to also develop the ability to see patterning in the target language and to form generalizations to account for that patterning (p: 5).

clarity , " O my God , " and withdrew	a lump of	**amberish** ear wax that he then wiped on his thumb to
be powered by conventional fuels ? If you start with	a lump of	**anthracite** coal (the best fuel known at the time)
International Trade . # " The Chinese people are not	a lump of	**bean** curd , " Mr. Zheng added . # Some Western
a circle Inside were several handfuls of coca , and	a lump of	**black** rock . # Vic took a few leaves into his
, which is highly regarded in art circles , includes	a lump of	**Blu-Tak** stuck to a wall , and a crumpled-up sheet of
. He even ...." " Even what ? "	A lump of	**bread** stuck in his throat . Max swallowed . It was
game will keep it sweet for many days ; and	a lump of	**charcoal** inserted into the cavity of a drawn bird is

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

ssary in the weak acidic solution , the planisher drew	a lump of	**charcoal** -- specially prepared from willow , carefully
the pathway , is a large mousetrap on which sits	a lump of	**cheese** . The cat tiptoes over to the refrigerator , vanis
at home , you can , sort of , put	a lump of	**clay** at the end of a yardstick and sit in the
very tired after making Adam and Eve but still had	a lump of	**clay** left over . He tried to make a third figure
er-filled plastic container like those at salad bars , plus	a lump of	**clay** to shape into something that floats . At Hermosa
. " But to us , the roofing business was	a lump of	**clay** waiting to be molded . " # ONE OF THESE
it moist . The potter begins the pot by taking	a lump of	**clay** , tossing it in the air , twisting and shaping
, that in this case , instead of yardstick and	a lump of	**clay** , you 've got little two-micron-thick and

Figure: Concordances for 'a lump of' from COCA-A

Usually the scheme of corpus exploration follows a cyclical process wherein each 'research question' the learner wants to answer (examples: how does a so-and-so word is used in the discourse of popular magazines? Does this word occur frequently in academics?...) is examined in the light of the corpus evidence. Sinclair (2003) summarized this cyclical process in seven steps. The first step is called initiative: it involves the learners or users to observe the words to the left or to the right of the node to decide on the strongest patterning. The second step is called interpret: the users having observed the patterning forms a hypothesis. In the third step the user tests the hypothesis and it is referred to as consolidate. Here the users observe the surrounding co-text in relation to the node to go beyond the phrase boundaries. In the next step the user having understood the way the patterning has taken place re-examines the hypothesis and reinterprets the data. The next step is to report the observations. The next three steps are crucial: recycle, result and repeat. At the step recycle the user can concentrate on other patterns in the concordances, while the other two steps primarily meant to formulate generalizations which can either lead the user to repeat the process or use the formulation.

In this approach the corpus data are used either to confirm the codified knowledge about language or is used to extend our understanding of the phenomena beyond the codified knowledge. In either case corpus can be mainly used to support, supplement and validate the theory/knowledge already documented. Hence the learner can take off from a given rule and validate it in the light of evidence. It is said in corpus literature that such an exploration usually leads the users to formulate yet another hypothesis to be examined. This model of

MJAL8:2Summer 2016 *ISSN 0974-8741*

exploration—known to the unknown—best suits with learners who have less familiarity with the corpus explorative procedures. At a later stage, when the learners have acquired the corpus analysis skills, they can be allowed to explore the unknown zones of language phenomena on their own.

Phraseological Network of 'Significant'

Collocational significance can be used by language teachers to identify specific combinations that frequently co-occur. A random search of COCA (Corpus of Contemporary American English) for the most frequently used collocates of the word 'significant' shows us more than a hundred hits across a range of contexts. Every context either uses specific collocates in novel ways or introduces the words in new contexts. Reading through every instance is certainly an enriching experience. However, when offering specific purpose courses, reading random instances will prove to be a rather tiresome and futile task. Hence a careful selection needs to be made to study specific instances that have direct relevance to learners needs.

In the table given below, the word 'significant' collocates with a range of words. However, in academics, the word has a strong bonding with 'statistically' which is revealed in the form of MI (Mutual Information) value while it goes frequently with 'portion' in the spoken genres. MI value counts the statistical significance of a specific combination across the corpus and reports the success of combination in terms of a value. It is believed, if the MI value is more than 3.0 the combinations tend to be stronger and require attention.

General Collocates	MI	Academic Collocates	MI	Fiction	MI	Spoken	MI
differences	6.49	no	3.17	most	3.24	number	3.23
between	3.59	differences	5.18	number	3.58	amount	4.99
statistically	9.14	statistically	7.61	difference	4.95	impact	4.56
differences	5.61	differences	5.29	event	5.35	progress	4.98
found	3.19	found	4.80	amount	5.46	difference	3.60
effect	4.59	effect	3.89	events	5.12	changes	4.19
effects	4.34	effects	3.23	portion	6.65	numbers	3.50
Role	3.41	revealed	4.50	changes	5.11	role	3.09

MJAL8:2Summer 2016 *ISSN 0974-8741*

changes	3.79	showed	4.22	role	5.04	damage	4.10
showed	4.09	interaction	4.22	statistically	9.85	increase	3.75
impact	4.07	main	3.63	progress	4.72	risk	3.33
=	3.56	gender	3.15	damage	4.61	reduction	5.54
revealed	5.22	f	3.36	numbers	3.96	cuts	3.71
results	3.22	correlation	4.64	percentage	7.55	portion	5.89
relationship	3.36	correlations	5.07	particularly	3.88	fairly	4.22
amount	3.91	indicated	3.11	glance	3.84	step	3.09
P	3.38	predictor	5.68	differences	5.87	development	4.16
interaction	5.65	predictors	5.28	fortune	4.30	event	3.63
Main	3.51	improvement	3.67	glances	3.84	factor	4.20
increase	3.32	contribution	4.00	loss	3.77	differences	4.45

In the table above a few words that frequently collocate with 'significant' are given. Depending on learning goals, teachers can direct the focus to specific combinations rather than all.

Frequency and Discipline

The study of collocations can become refined if attention is paid to specific disciplines. For instance the following table provides the most frequently used collocates of the most frequently used academic vocabulary of Science and Technology in COCA.

Ratio	word	Total	Science and Technology	Frequent Collocate	Frequency	Frequency Overall	MI Vlaue
2.14	SPECIES	23861	19236	ENDANGERED	556	916	6.01
1.66	SCIENTIST	19311	12804	SOIL	149	4903	5.48
1.5	SURFACE	18168	11143	EARTH	232	4677	4.02
2.02	SOFTWARE	14774	8945	HARDWARE	233	1214	6.19
1.58	PLANET	11656	9759	EARTH	64	4677	3.7
1.52	TEMPERATURE	10924	8218	DEGREES	164	2217	5.15
2.83	TELESCOPE	10776	10438	SPACE	367	4939	6.01
1.59	SOIL	10101	6670	WATER	399	16035	3.16
2.1	ENGINEERING	9967	7159	MECHANICAL	121	2777	5.75

MJAL8:2Summer 2016 *ISSN 0974-8741*

1.87	UNIVERSE	9555	6207	EARLY	70	5593	3.4
2.83	GALAXY	9532	9359	MILKY	84	543	4.58
2.15	LABORATORY	9405	5406	NATIONAL	545	9795	4.87
2.34	EMISSION	8624	5110	REDUCTION	66	2083	5.75
1.77	DIGITAL (J)	8454	5856	IMAGING	133	1123	6.28
2.93	ASTRONOMER	7196	6986	AMATEUR	52	391	3.87
1.87	SOLAR (J)	7085	6986	SYSTEM	602	18582	4.37
1.99	CONSERVATION	7008	4271	BIOLOGY	332	2571	5.5
1.8	GENETIC(J)	6701	3860	DIVERSITY	206	2712	5.78
1.82	CARBON	6535	4774	DIOXIDE	860	1181	8.56

This list is provisional and represents the collection of texts used to compile the database. Each word selected here, in addition to being a strong collocate of the word given, can partner with a range of others to produce various meaning senses. Every combination represents a specific context; and every context represents a larger text, which, as mentioned, is part of the collection texts used to compile the corpus. The co-occurrence of any two words as the most frequently used combination thus largely depends on the texts we choose.

MI values of the identified strong collocates provide an accurate representation of the contexts selected to build the database. It is possible that a specific combination of words may not co-occur as frequently in a general database as they occur in a specific database. To avoid any misinterpretation, one needs to increase the size of the database to make it as 'large' as possible. For instance the word 'genetic' in the general academic database frequently collocates with 'diversity'. However, in the spoken database of COCA the word 'predisposition' with an MI value of 11.22 shows a strong bonding with genetic. It is followed by words such as *markers* (9.33), *defects* (8.98), while *diversity* is positioned as not so frequent. This particular difference in terms of co-occurrence across genres and disciplines pose a challenge to learners as well as teachers. It is at this stage the choice needs to be made from the boundless variety of possibilities.

Combinations and Constraints

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

A large part of language instruction has been directed to identifying language units that frequently occur in specific areas of study. However, little emphasis has been laid on disciplinary constraints in using the learnt units. It has been observed that students tend to use the already available units of meaning without considering their relevance to specific areas of study. Often their use of a word restricts its phraseological purview to a set of hackneyed chunks that seem to find a common ground across areas of study. Combinations that markedly typify specific discourse communities rarely get the attention of many users. What makes the learning of such discipline typified chunks obvious is a constant exposure to those units in a range of discipline specific contexts. If the users, however, wish to explore the general tendency of words across genres their approach to corpus study is generally recorded in the following way.

SIGNIFICANT

There are 63691instances available for the word **Significant** in the COCA. The word spread across various registers is most often found in the Academic register, where it occupies a sizable frequency of 41639 examples. The following table illustrates the proportion of frequency the given word shares in the different registers.

Spoken	Fiction	Magazine	Newspaper	Academic
6161	1183	7723	6985	41639

This particular word occurs most in the academic register than in any other register. Its frequency in the remaining registers after adding up amounts to roughly about half of its frequency in the academic register. With reference to the concordances, it is observed that the word Significant most frequently occurs as an adjective, as in the cases –
significant amounts of natural gas; significant amounts of capital; report significant additional difficulties; significant artillery support etc.
Moreover, this word means-- being important in effect or meaning; fairly large or rich in implication. The common synonyms of this word are considerable, hefty, substantial, large, noteworthy, important etc.
With reference to COCA, it is also noted that the collocations related to the word are-
Noun: advances, support, amounts, attention, change, result, difference etc
Miscellaneous: there, no, find, show etc

Figure: A student exploration of the word 'significant' in COCA

MJAL8:2Summer 2016 *ISSN 0974-8741*

'Speakers are free, but only within constraints': A Corpus-based Study of Academic *Phraseologisms* in Science and Technology by 1.Vijayakumar Chintalapalli and 2. Mirza Mohammed Ali Baig

The above figure explicates a student's exploration of the word 'significant' in COCA. The student here has paid attention to word's meanings, collocations as well as patterns of use. She observed that, the word 'significant' was more frequently used in 'academic registers' compared to other. Furthermore she found a few phraseologisms such as *significant amounts of natural gas, significant amounts of capital, report significant additional difficulties* and *significant artillery support* that seem to appear more frequently in the corpus.

If the student wishes to research the database for specific combinations within disciplines she could refine and limit the search to fit her purpose. For instance, in the following instances the search option is confined to the field of 'medicine'. These concordances here expose students to patterns that are specific to the field of study.

from 27 countries . Symptomatic heart failure patients **with**	significant	left ventricular dysfunction (ejection fraction <35%) in
possibly accelerates the development of SCC in patients **with**	significant	risk factors , presumably by impairing normal immune
. An open surgery is used more frequently in patients **with**	significant	trapezial subluxation because the wings of the Artelon mplant
116 patients during their recovery at home after TKR surgery **.**	Significant	pain (i.e. , score >40 on a visual analog scale ranging
was confirmed on ASSR audiometry by abnormal tympanograms **.**	Significant	changes in the predicted auditory behavioral thresholds were
to reanalyze what they had been doing and resulted in **a**	significant	self-study and improvement in policies and procedures
ecological regime , increased mosquito populations are not **a**	significant	side effect of a small wetland construction project such as the
11 patients dog allergen immunotherapy for 1 year and noted **a**	significant	decrease in specific IgE , conjunctival and skin sensitivity ,
) . Patients in the 12 mg/m 2 group showed **a**	significant	advantage in the analysis of time to confirmed EDSS
in obstetric ultrasound in the United States have led to **a**	significant	increase in the prenatal diagnosis of clubfoot . The 16- to
# The results of the study showed that there was **a**	significant	increase in the inspiratory muscle strength of the
.073) were moderate . In addition , there was **a**	significant	gender by ethnicity interaction , F(1,3993) = 5.63 , p = .018
pressures was compared statically to body weight , indicated **no**	significant	change in sensor sensitivity over a short period of inactivity
trend of observations of unlocked guns was not **statistically**	significant	in this subsample . # No adverse health events associated
the value zero . This means that they were not **statistically**	significant	at 5% level (two-sided test) . For all outcome measures
indicated that , although improvements were not **statistically**	significant	, the improvements that did take place were maintained .
osteoporosis . Although chi-squares were not **statistically**	significant	, and there was little difference for the comparable and the

Figure: concordances for the word 'significant' in the medical sub-genre of COCA

The word significant appear to form a stong bonding with a few words such as 'statistically' and 'patients' it shows a general tendency to collocate with adjectives such as *changes, decrease, increase,* and *advantages* that signify 'a shift from one point or condition to the other'. The phrase '*patients with significant* (condition)' such as *dysfunction, risk factors, pain,* and *subluxation* are specific to the field of medicine. This phraseologism shows a

MJAL8:2Summer 2016 *ISSN 0974-8741*

negative tendency in terms of its 'semantic prosody': negative or positive meaning sense. In addition, he would also notice the ways the reporting verbs play their role in conveying the message.

The following table charts the use of three most frequent *verbs* used in science and technology.

Rank Order	Word	Strong Collocations	A frequently used pattern
1	**Provide**	**To the right: ~ adj**(dedicated, better, interactive, new, rich, significant, higher, strong, special, additional) **noun** (insights, incentives opportunity, evidence)	Be-form + verb (expected, intended) + to + **~** + adjectives with positive semantic prosody + noun (mostly uncountable) Ex: ...infrared thermocouple probes has been designed to **provide** low-cost emissivity-independent contact
2	**Include**	**To the left: adj** (typical, other (2.43), optional); **noun** (specialities, features, examples, applications, challenges, accomplishments, designs, statements) **~** **To the right: noun** (information, topics) with less than 3 MI Value No nouns to the right side of the node occurred with more than 3.0 MI value.	Noun + that/which + **~** + (ajective) + noun Ex: ...factory-supplied toolkit that includes a chain break
3	**Develop**	**To the right**: **adj** (customized, requisite, instructional, analytical, innovative, sophisticated, effective) **noun** (skills, understanding, plans, strategies, hypotheses, abilities, protocols, prototype(s) In the field of agriculture 'develop germplasm' is found as a strong collocate	Noun + to + **~** + adjective + noun Ex: a) Three-year multidisciplinary study to develop GHG mitigation strategies b) data to develop a full-scale version

MJAL8:2Summer 2016 *ISSN 0974-8741*

The table above provides a glimpse of bonding as revealed by COCA for Science and Technology. These three words are used in a range of combinations with many other words as well. We have, however, considered here a set of parameters to opt out many words. The parameters include

a. Words with less than 3 occurrences in science and technology were not considered even though their MI values were more than 3.0;
b. Words with high frequency and less than 3.0 MI values; and
c. Words that belonged to one specific speciality within science and technology were not considered for inclusion.

Selection of words for instructional purposes can consider the criteria above. However, when dealing with individual texts where specific combinations of words appear to co-occur frequently, teachers can explicitly direct learner attention to those features. Teaching units of meaning promotes better comprehension and text production. All teachers need to do is to channel learner attention to meaning(s) as conveyed by chunks and not just individual words.

Conclusion

This paper highlighted the need to study word combinations as they occur in specific academic fields. It also showed, with examples, how our search options determine our observations, which, in turn, affect our learning of discipline specific phraseologisms. If the students study concordances with a focus on specific aspects of word knowledge, every attempt leads to a discovery of new pieces of information. In other words, corpus study becomes an enriching experience when approached with a purpose; and exploring word combinations enables learners to enrich their vocabulary knowledge.

MJAL8:2Summer 2016 *ISSN 0974-8741*

References

Davies, M. (1998-). *The Corpus of Contemporary American English: 450 million words, 1990-present.* Utah: Brigham Young University.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger, & F. Meunier, *Phraseology: An interdisciplinary perspective* (pp. 3-26). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Hoey, M. (2005). *Lexical Priming: A new theory of words and language.* New York: Routledge.

Hornby, A. S. (1946). Sentence patterns and substitution tables (1). *English Language Teaching , 1* (1), 16-23.

Hunston, S., & Francis, G. (2000). *Pattern Grammar.* Amsterdam: John Benjamins.

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English.* Philadelphia: John Benjamins B.V.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. J. (eds), *In Classroom Concordancing [ELR Journal 4]* (pp. 1-16). Birmingham: CELS University of Birmingham.

Meunier, F., & Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching.* John Benjamins Publishing Company: John Benjamins Publishing Company.

Sinclair, J. M. (1991). *Corpus, Concordance and Collocation.* Oxford: Oxford University Press.

Sinclair, J. (2005). Meaning in the framework of corpus linguistics. In W. T. (ed.), *Lexicographica* (pp. 20-32). Tubingen: Max Niemeyer.

Stubbs, M. (1996). *Text and Corpus analysis.* Oxford: Blackwell.