Formulaic Language Patterns in Social Science and Natural Science Research Papers: A Corpus-Based Study

Yen-Yu Lin, Siaw-Fong Chung
National Chengchi University, Taiwan

0608

The Asian Conference on Education 2013

Official Conference Proceedings 2013

iafor
The International Academic Forum
www.iafor.org

Phraseology is a general term used to describe the tendency of words, and groups of words, to co-select frequently in some environment than in others to achieve a certain meaning. (Hunston, 2011, p. 5). A great deal of research has focused on the recurrent continuous sequence of words using a wide range of terminology such as semantic sequences (Hunston 2008), n-grams (Cheng et al. 2006, Forchini & Murphy 2008), and lexical bundles (Biber et al. 1999, Biber, 2009, Biber & Barbieri 2007, Biber et al. 2004, Nesi & Basturkmen 2006). In EAP (English for Academic Purposes), these studies are mostly discipline-based and have revealed the existence of linguistic differences across genres. For instance, Hyland (2008), by exploring the forms, structures and functions of 4-word bundles in a 3.5 million corpus of four disciplines, found that there was disciplinary variation. Some other scholars examined the use of phraseological patterns in specific section of research articles: Hsieh (2005) compared two corpora of 50 published journal articles and 50 non-native conference papers to shed light on the differences of the use of lexical bundles and functional moves in the abstracts; Lin (2006), in a similar vein, explored frequent continuous sequences of words and their rhetorical functions in the introduction section of research articles in the field of computer science.

There are also other studies on phraseology that have paid attention to discontinuous sequences of words. These studies focused on the recurrent sequences of words that form a frame containing several slots (e.g. Biber 2009; Cheng et al, 2011; Marco 2000; Römer 2010; Renouf & Sinclair 1991, Stubbs 2007); the choice of the words enclosed within the frame is determined by the elements of this frame (Eeg-Olofsson & Altenberg, 1994). Renouf and Sinclair (1991) examined seven specific collocational frames (e.g. *a * of)* investigating the common fillers that occur in each framework. Biber (2009) and Marco (2000) followed a similar approach to investigate the formulaic frames in academic writing: Biber (2009) identified four high-frequency frameworks with similar constructions (*in the * of, to the * of, on the * of, at the * of*) selecting different sets of fillers. For example, *at the * of*, as the most distinctive frame, co-occurs frequently with the fillers *end, time, beginning, level, expense, start, center/centre, top,* and *base*, while *in the * of* takes several high frequency fillers, such as *case, absence, form, context* and *process*. Marco considered three frames in a corpus of medical research papers. The results showed that two of the most common frames are: [*the* NOUN *of*] and [*a* NOUN *of*]. [*The* NOUN *of*] tends to be used in expressing the construction of nominalizations (e.g. *the administration of, the measurement of*); [*a* NOUN *of*] is frequently applied to describe the process of quantifying and categorizing (e.g *a number of*, *a period of*). Another important finding is that the frames are likely to precede or follow the collocates belonging to

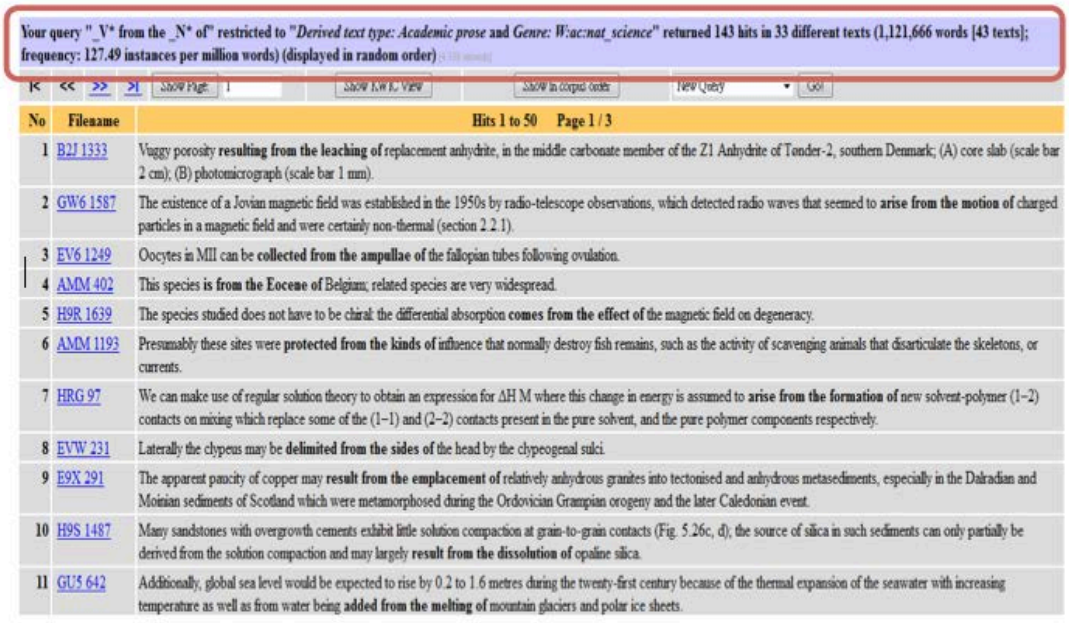
specific semantic classes. For example, *the effect/results of* are followed by a noun naming a drug or by a nominalization expressing an action in the treatment. This study echoed Halliday and Martin's (1993) analysis on scientific texts which suggested that nominalization is crucial in structuring scientific knowledge in a static, synoptic representation of reality. In addition to those mentioned above, a number of studies also indicated that the formulaic sequences in academic writing are parts of noun or prepositional phrases and end with prepositions (e.g. *the nature of the, the end of the, in the case of, on the basis of*) (e.g. Biber, 2009; Biber et al., 1999; Corte, 2004; Hyland, 2008). In sum, the studies reviewed demonstrated the prevalence and significance of the discontinuous sequence -- [PREPOSITION *the* NOUN *of*] -- and exemplify that it is selective in terms of the words as fillers. Marco's study on medical research papers further reflected the fact that each discipline may prioritize different phraseological patterns and thus shows variety in the use of the discontinuous sequences. We can anticipate that each discipline might have its own distinct phraseological profile. To the best of our knowledge, the cross-discipline comparison of the realization of a specific formulaic sequence like [PREPOSITION *the* NOUN *of*] has rarely been done. Lin and Chung (2012), in her study on [VERB *from the* NOUN *of*] in the field of commerce elucidated that the verbs preceding this prepositional framework may also have impact on what nouns can co-occur. Accordingly, the present study targeted at the formulaic sequence [VERB *from the* NOUN *of*] and investigates the distributions of different lexical items in the slots (in capitals) based on the assumption that different disciplines are distinctive in their preference for the filler terms in this common formulaic pattern.

Methodology

British National Corpus (BNC) was adopted for the present study. BNC is a 100 million word collection of text of written and spoken English from a wide range of sources and genres. In this study, the data was retrieved through the *BNCweb*, a web-based interface for searching data from BNC. Strings of constructions were extracted under the restriction of written – academic prose. A search string was applied to query for the target construction, as illustrated in (1):

(1) _V* *from the* _N* *of*

After keying in the syntactic pattern as our search string and restricting the query in written text, the interface returned the instances along with the pattern. Figure 1 below illustrates the query result page of natural science as an example.



Your query "_V* from the _N* of" restricted to "*Derived text type: Academic prose* and *Genre: Wac:nat_science*" returned 143 hits in 33 different texts (1,121,666 words [43 texts]; frequency: 127.49 instances per million words) (displayed in random order)

No	Filename	Hits 1 to 50 Page 1 / 3
1	B2J 1333	Vuggy porosity **resulting from the leaching of** replacement anhydrite, in the middle carbonate member of the Z1 Anhydrite of Tønder-2, southern Denmark; (A) core slab (scale bar 2 cm); (B) photomicrograph (scale bar 1 mm).
2	GW6 1587	The existence of a Jovian magnetic field was established in the 1950s by radio-telescope observations, which detected radio waves that seemed to **arise from the motion of** charged particles in a magnetic field and were certainly non-thermal (section 2.2.1).
3	EV6 1249	Oocytes in MII can be **collected from the ampullae of** the fallopian tubes following ovulation.
4	AMM 402	This species **is from the Eocene of** Belgium; related species are very widespread.
5	H9R 1639	The species studied does not have to be chiral: the differential absorption **comes from the effect of** the magnetic field on degeneracy.
6	AMM 1193	Presumably these sites were **protected from the kinds of** influence that normally destroy fish remains, such as the activity of scavenging animals that disarticulate the skeletons, or currents.
7	HRG 97	We can make use of regular solution theory to obtain an expression for ΔH M where this change in energy is assumed to **arise from the formation of** new solvent-polymer (1–2) contacts on mixing which replace some of the (1–1) and (2–2) contacts present in the pure solvent, and the pure polymer components respectively.
8	EVW 231	Laterally the clypeus may be **delimited from the sides of** the head by the clypeogenal sulci.
9	E9X 291	The apparent paucity of copper may **result from the emplacement of** relatively anhydrous granites into tectonised and anhydrous metasediments, especially in the Dalradian and Moinian sediments of Scotland which were metamorphosed during the Ordovician Grampian orogeny and the later Caledonian event.
10	H9S 1487	Many sandstones with overgrowth cements exhibit little solution compaction at grain-to-grain contacts (Fig. 5.26c, d); the source of silica in such sediments can only partially be derived from the solution compaction and may largely **result from the dissolution of** opaline silica.
11	GU5 642	Additionally, global sea level would be expected to rise by 0.2 to 1.6 metres during the twenty-first century because of the thermal expansion of the seawater with increasing temperature as well as from water being **added from the melting of** mountain glaciers and polar ice sheets.

Figure 1 The Query Result Page of *BNCweb*

We searched in two disciplines, namely, natural science and social science. The top row shows the number of hits returned from this query. The total number of words tokens, and the number of instances per million words could also be found. Table 1 summarizes the quantitative results of the BNC search results.

Table 1 Summary of the BNC Search Results

The Target Pattern	VERB *from the* NOUN *of*	
The disciplines	Social Science	Natural Science
No. of hits	431	143
No. of texts	114	44
Frequency (per million)	91.17	127.49

As can be seen, the results of [VERB *from the* NOUN *of*] in nature science returned 143 hits in 33 different texts, amounting to 127.49 instances per million words**.** Although the total number of [VERB *from the* NOUN *of*] is higher in social science, the frequency per millions words is a bit lower than that of natural science.

These retrieved results were then exported to text files and opened in MS Excel for the follow-up semantic feature analysis (see Figure 2).

	A	B	C	D	E	F	G	H	I
1	No.	Instance				Frequency	Percenta	Semantic Coding	
2	111	derived	from	the	activities	1	0.70%	activity	
3	30	follows	from	the	associativity	1	0.70%	association	
4	3	collected	from	the	ampullae	2	1.40%	body; organ	
5	40	emerging	from	the	mouth	1	0.70%	body	
6	120	formed	from	the	aftershafts	1	0.70%	body	
7	43	innervated	from	the	ganglia	1	0.70%	body; organ	
8	118	move	from	the	lacuna	1	0.70%	body; organ	
9	20	moved	from	the	foot	1	0.70%	body	
10	45	radiating	from	the	ganglia	1	0.70%	body; organ	
11	26	transferred	from	the	cloaca	1	0.70%	body; organ	
12	107	determined	from	the	structure	1	0.70%	building	
13	116	result	from	the	alteration	1	0.70%	change	
14	31	originated	from	the	transposition	1	0.70%	change	
15	127	mapped	from	the	ability	1	0.70%	competence	
16	39	come	from	the	influence	1	0.70%	effect	
17	27	comes	from	the	effect	1	0.70%	effect	
18	12	dying	from	the	effects	1	0.70%	effect	
19	95	obtained	from	the	Doppler-shift	1	0.70%	effect; process	
20	112	calculated	from	the	enthalpy	1	0.70%	element	
21	60	derived	from	the	occurrence	1	0.70%	existence	
22	16	form	from	the	introduction	1	0.70%	existence	

Figure 2 Encoding Format in the MS Excel File

The nouns were categorized based on their similarity of semantic features. The aim of a semantic feature analysis was to decide whether an item is a member of a particular meaning group. It was hypothesized that the words that can appear in the same slot of the target pattern would share some similarities in semantic features. Most of the category names were generated by the researchers but some were borrowed from Marco's study. Next, each category was counted and converted into percentage.

Results and Discussion

Based on the outcomes of the semantic coding, results such as the following Tables 2 and 3 were obtained. The two tables below display the semantic categories for the nouns in this pattern in the two disciplines, along with their frequencies and percentages. It should be noted that, since social science has much more noun categories in [VERN *from the* NOUN *of*], Table 3 shows only the distribution pattern of its top twenty semantic categories, which account for over the 62% of the total frequency. The finding shows that there are striking differences between the two disciplines. Table 2 first shows the results of natural science.

Table 2 The Distribution of Semantic Groups of NOUNS in [VERB *from the* NOUN *of*] in Natural Science

Semantic Categories	Raw Frequency	Percentage
1. 'location' (e.g. *surface, side, centre*)	34	23.78%
2. Nominalization expressing procedures (e.g. *distortion, rotation, melting, mixing*)	30	20.98%
3. 'object' (e.g. *ring, milk, pen*)	11	7.69%
4. 'a part of the body ; organ' (e.g. *mouth, foot, ganglia*)	9	6.29%
5. 'quantity' (e.g. *ratio, amount, sum*)	7	4.90%
6. 'lefts' (e.g. *rest, remainder*)	6	4.20%
7. 'existence' (e.g. *occurrence, introduction, presence*)	3	2.10%
8. 'time; period' (e.g. *history, Eocene*)	3	2.10%
9. 'relation' (e.g. *correlation, interaction*)	4	2.80%
10. 'path' (e.g. *orbit*)	3	2.10%
11. 'features' (e.g. *properties, characters*)	3	2.10%
12. 'effect' (e.g. *effect, influence*)	8	5.59%
13. 'investigation' (e.g. *study, analysis*)	2	1.40%
14. Others (e.g. s*peech, trouble, factor*)	20	13.99%
Total	143	100.00%

From Table 2, we can see that, in natural science, the two most frequent semantic categories reach around forty percent of total, which represents a large proportion. The top group, which denotes different positions or locations on subjects (e.g. *end, surface, back*), accounts for 23.78% of the total frequency; the second biggest group, with a rate of approximately 21%, mainly deals with nominalization of natural processes (e.g. *erosion, dissolution, recrystallization*). The next two categories are concerned with something visible and parts of the body. The predominance of the above-mentioned four categories can be attributed to the essence of this discipline which tends to focus on concrete objects, the natural process of evolution, and the procedures of empirical studies. The result also shows that the quadruplets (e.g. from *the center of*, *from the distortion of)* including the items from the top two categories, which denote locations and nominalizations, tend to be preceded by the verbs expressing 'developing or coming from something else' (e.g. *grow, derive, arise*) in order to describe either where something originates or the process of formation. Example (1) and (2) illustrate this.

(1) *The thermal tide may also be responsible for dim ultraviolet radiation observed to* ***come from the middle of*** *the night hemisphere.* (GW6 288)
(2) *However, it is now considered to be an ironstone* ***derived from the weathering of*** *an exhalative, spessartine-bearing protore (Nicholson, 1987).* (E9X 984)

Example (1) shows a specific object starting to appear from a certain location *(the night hemisphere*); similarly, Example (2) displays that an ironstone, which is discolored and disintegrated, comes to exist through the process of exposing to the weather.

In contrast to the results of natural science, what was discovered in social science shows that the nouns in [VERB *from the* NOUN *of*] spread out over a wider and different range of semantic categories as the most frequent semantic category constitute merely 9.51% of the total number of nouns in this pattern (see Table 3).

Table 3 The Distribution of Semantic Groups of NOUNS in [VERB *from the* NOUN *of*] in Social Science

Semantic Categories	Raw Frequency	Percentage
1. 'view' (e.g. *idea, notion, perspective, viewpoint*)	41	9.51%
2. 'location' (e.g. *base, border, University, County*)	36	8.35%
3. 'something negative' (e.g. *limitation, failure, question, difficulties*)	25	5.80%
4. 'existence/non-existence' (e.g. *introduction, presence, absence*)	19	4.41%
5. 'use' (e.g. *use, application, implementation*)	14	3.25%
6. 'effect' (e.g. *effect, influence, impact*)	14	3.25%
7. 'standard' (e.g. *pattern, model, level*)	13	3.02%
8. 'lefts' (e.g. *rest, ashes, remainders*)	12	2.78%
9. 'object' (e.g. *ring, object, work*)	12	2.78%
10. 'investigation' (e.g. *study, analysis, research*)	11	2.55%
11. Nominalization expressing an activity (e.g. *identification, reordering, measurement*)	9	2.09%
12. 'record' (e.g. *history, record*)	9	2.09%
13. 'type' (e.g. *kind, type, category*)	9	2.09%
14. 'combination' (e.g. *integration, merging, composition*)	8	1.86%
15. 'creation' (e.g. *creation, invention, production*)	8	1.86%

16. 'action' (e.g. *action*)	7	1.62%
17. 'sharing' (e.g. *distribution, redistribution*)	5	1.16%
18. 'experience' (e.g. *experience*)	6	1.39%
19. 'field' (e.g. *domain, field*)	6	1.39%
20. 'feeling/desire' (e.g. *frustration, gloom, ambivalence*)	6	1.39%
Total	224	62.65%

As can be seen, the frequent collocates from the top four groups refer to people's point of view (9.51%) (e.g. *ideas, perspective, assumption*), followed by those denoting different location or institutions (8.35%) (e.g. *edge, end, University, Gulf*), bad qualities or unpleasant situations (5.80%) (e.g. *decline, problem, difficulty*), and the existence/non-existence of something (4.41%) (*e.g. presence, introduction, absence*). Except for the category denoting locations and nominalization expressing an activity, other noun groups in Table 3 are different from those occurring in natural science. Overall, these nouns seem to be mainly associated with people's perception/feelings and experience (see Categories 1, 18 and 20), the application/cause and effect of something (see Categories 4, 5 and 6), an average/normal quality or requirement (see Categories 7 and 13). This suggests that social science focuses more on the discussion about people's thoughts and feelings, the power to produce results, as well as the importance of principles and rules.

Regarding the question about what verbs precede the quadruplets (e.g. *from the view of*, *from the base of*, *from the limitation of)* with the items from the top four noun categories, our observation are as follows. It was found that the verbs such as *emerge, derive,* and *stem*, co-occurred frequently with the nouns from the 'view' group for the author to express where certain perspectives originate and the way in which someone understands information. Examples (3) and (4) show this.

(3) *The second point of similarity* ***emerges from the point of*** *his analysis where he talks of changing codes of conduct.* (CRU 245)

(4) *Our understanding of the vertical (as opposed to the horizontal) movements of the lithosphere during continental rupture is largely* ***derived from the interpretation of*** *the sediments laid down in passive margin basins.* (JOT 938)

These two examples show that the opinion (*the second point*) and the knowledge about how something works (*our understanding*) are established from a careful examination and observation of something.

The second biggest group of nouns, the 'location' noun category, which refers to positions in time or space, in this frame was found to follow the verbs related to movement or the action of starting and finishing as exemplified by Examples (5) and (6).

(5) *The responsibility for primary and secondary education (and related school matters) in Wales was* ***transferred from the Secretary of*** *State for Education and Science to the Secretary of State for Wales. (H8D 598)*

(6) *The coalowners' response in most areas was to give notice that their workers' contracts were* ***terminated from the end of*** *the month… (FAW 676)*

Example (5) shows that the responsibility for education was moved from one institution to another; Example (6), on the other hand, displays an abstract movement in time – the contracts was coming to the end of a specific month.
In some other conditions, the verbs (e.g. *arise, derive, result, suffer, escape*) commonly combine with the quadruplets including nouns denoting something unpleasant or unwanted to indicate cause-effect relations or the action of avoiding difficulties/hardships. (see Examples (7) and (8)).

(7) *After death there are two sets of emotions to be handled, one concerning the mourning* ***resulting from the loss of*** *a loved one, and the other arising from the unconscious hostility felt towards the dead person,…* (ECY 1019)

(8) *The theory of the global system based on transnational practices is an attempt to* ***escape from the limitations of*** *state-centrism.* (HTC 83)

Through the use of *resulting from the loss of*, Example (7) illustrates that, due to the deprivation from failure to keeping something, the great sadness occurred; Example (8) emphasizes the action of getting away from a bad situation.

Following that, as shown in Examples (9) and (10), the noun category referring to the (non-)existence of something has a high probability to co-occur with the verb *aris*e when the author presents the effects of applying something or being deficient in certain aspects.

(9) *Further serious environmental consequences have* ***arisen from the introduction of*** *grazing herbivores (such as cattle, sheep, goats, horses, donkeys and mules) to which the Andean environment is not naturally adapted but which are the mainstay of modern Andean populations.* (BIE 487)

(10) *Consensus* ***arises from the absence of*** *any fundamental challenge for change in society; individuals judge what is in their best interests within a basic framework of laws.* (ALM 359)

More specifically, Example (9) describes that serious problems concerning environment destruction originate from breeding grazing herbivores. Example (10) expresses that the general agreement usually appears from the lack of intention of making changes in society.

In summary, it is interesting to find that, while the distribution pattern of nouns in [VERB *from the* NOUN *of*] are different in natural science and social science, the two disciplines share a trend, that is, the predominant verbs preceding this formulaic sequence are usually *derive, arise, result* and so on.

Table 4 displays the results of semantic classification of the verbs, some of which are in line with the finding presented above. Again, due to the diversity in terms of the verb categories in social science, only the top 12 categories, which constitute 67% of the total frequency, are shown in the table below.

Table 4 The Distribution of Semantic Groups of VERBS in [VERB *from the* NOUN *of*] in Natural Science and Social Science

Natural Science			**Social Science**		
Semantic Categories of the Verbs	Raw Freq.	PCT	Semantic Categories of the Verbs	Raw Freq.	PCT
1. 'appear' (e.g. *arise, come, emerge*)	37	26%	1. 'appear' (e.g. *arise, come, emerge*)	72	17%
2. 'get' (e.g. *obtain, acquire, derive*)	30	21%	2. 'get' (e.g. *obtain, receive, gain*)	70	17%
3. 'cause' (e.g. *result*)	18	13%	3. 'cause' (e.g. *result*)	32	8%
4. 'find' (e.g. *find, determine*)	7	5%	4. 'delete/reduce' (e.g. *detract, remove, eliminate*)	24	5%
5. 'separate' (e.g. *separate, isolate*)	7	5%	5. 'separate' (e.g. *separate, isolate, alienate*)	15	4%
6. 'study' (e.g. *calculate, collect*)	7	5%	6. 'bring/ take' (e.g. *draw, drag*)	12	3%
7. 'measure' (e.g. *estimate, compute,*	6	4%	7. 'move' (e.g. *move, shift, transfer*)	10	2%

measure)					
8. 'grow' (e.g. *grow, accrete, accrue*)	5	3%	8. 'investigate' (e.g. *view, see*)	10	2%
9. 'move' (e.g. *move, transfer*)	5	3%	9. 'undertake' (e.g. *undertake, suffer*)	8	2%
10. 'release' (e.g. *emit, radiate, emanate*)	4	3%	10. 'disappear' (e.g. *disappear, escape, flee*)	7	2%
11. 'mental processes' (e.g. *know, learn, infer*)	3	2%	11. 'grow' (e.g. *grow, develop, accrue*)	7	2%
12. others (e.g. *speech, trouble*)	14	10%	12. 'back' (e.g. *retreat, withdraw, recoil*)	6	1%
Total	143	100%	**Total**	279	67%

As can be seen, the top three verb categories of [VERB *from the* NOUN *of*] in both disciplines are the same: the verb group with the sense of appearing is ranked as the most frequent group, followed by 'get' verbs and 'cause' verbs. The prevalence of the verbs from these three groups suggests that [VERB *from the* NOUN *of*] is mainly concerned with tracing origins/sources. The origins can refer to either a concrete location (e.g. institutions, positions on subjects) or an abstract position (e.g. causes, effects).

Compared to the three main verb groups, the percentage of the other categories in the two disciplines are much lower – each of them constitutes no more than 5% of the total. The distribution of these verb groups shows that social science and natural science have distinct preferences. This confirms that the nature of discipline would influence word selection in a common formulaic frame (Groom, 2009). We can see in Table 4 from the column on the left hand that the verbs in natural science center on the activity of evaluating, researching, and getting to know something ('study', 'measure', and 'mental process' verbs). Some of them deal with physical movements and changes ('grow', 'move', and 'release' verbs) .

In social science, some verbs are different from those in natural science. For example, certain items are associated with the action of getting rid of something (e.g. *eliminate, exclude*); some others relate to the tolerance of something unpleasant (e.g. *suffer, undertake*). On the other hand, similar verb groups in natural science, such as 'grow' and 'move' verbs can also be found.

In summary, in natural science, there is a strong tendency for nouns denoting locations and concrete subjects to occur in the frame; by contrast, the nouns referring to people's perception, unpleasant events, and the use of something are likely to appear in social science. The verbs such as *derive, arise, come, emerge*, which commonly occur with the preposition *from*, are also found to spread out in both disciplines. Our semantic analysis of nouns and verbs shows distinctive differences

across the two disciplines. However, some difficulties for categorization were encountered. It was sometimes difficult to define a clear boundary between categories. For example, in social science, the 'use' and 'combination' noun categories appear to overlap with nominalization expressing an activity. A more stringent criterion is therefore necessary for future work.

Conclusion

Phraseology has been one of the most rapidly growing areas in corpus linguistics recently. Our study reveals that formulaic patterns or strings of words are pervasive in natural language use. The aim of this study was to explore the extent to which formulaic patterns contribute to the uniqueness of each discipline in academic writing. Through semantic analysis of NOUNs and VERBs in [VERB *from the* NOUN *of*], the findings show considerable variations across natural science and social science in the categorization of these words and their frequency. The two disciplines are characterized by their distinct preferences for the nouns and verbs. Natural science is concerned with empirical analysis and processes of growing and moving, while social science discusses much about the acceptance or rejection of point of view, the suffering and releasing from limitations and problems, and the effect of doing/not doing something. In other words, natural science investigates a variety of natural phenomenon; social science deals with the complex interaction among people, between gain and loss, as well as between separation and cooperation.

The finding has clear implications for pedagogic practice It alerts us to pay attention to what disciplines our students are from and what kinds of text our students need to write. Moreover, the design of teaching materials can benefit from the findings of formulaic sequences in different disciplines, allowing instructors to focus on the specific ways of creating meanings appropriate to particular kinds of writing.

Acknowledge:

This work was supported in part by the National Science Council under the Grants NSC101-2410-H-004-176-MY2.

References:

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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Biber, D. & Barbieri, F. 2007. Lexical bundles in university spoken and written

registers. *English for Specific Purposes, 26* (3), 263-286.

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

Cheng, W., Greaves, C. & Warren, M. 2006. From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics, 11* (4),411-433.

Chung, Siaw-Fong, F. Y. August Chao, Tien-Yu Lan, and Yen-Yu Lin. (2011). Analyses of the Semantic Features of the Lexical Bundle [(VERB) PREPOSITION the NOUN of]. *Language Value, 3*(1), 138-152. Also presented at the Third International Seminar on Metaphor and Discourse: Verb-particle Constructions and their Underlying Semantic Systems. Universitat Jaume I, Castelló de la Plana, Spain. October 19-20.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397–423.

Forchini, P. & Murphy, A. (2008). N-grams in comparable specialized corpora: Perspectives on phraseology, translation, and pedagogy. *International Journal of Corpus Lingustics, 13* (3), 351-367.

Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics, 13*(3), 271–295.

Hunston, S. (2011). *Corpus Approaches to Evaluation*. London/New York: Routledge.

Hsieh, W. H. (2005). *Text analysis and online material development for EAP graduate courses: Teaching of abstract writing in the field of applied linguistics discipline.* Unpublished master thesis, Tsing Hua University, Hsin Chu, Taiwan.

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

Lin, Yen-Yu and Siaw-Fong Chung. (2012). The Use of "VERB from the NOUN of" in Published Writing in Commerce. Poster presented at the *6th Conference on Language, Discourse, and Cognition (CLDC2012)*. National Taiwan University, Taiwan. May 4-6.

Marco, M. (2000). "Collocational frameworks in medical research papers: A genre-based study". *English for Specific Purposes, 19* (1), 63-86.

Nesi, H. 8c Basturkmen, H. 2006. Lexical hundles and discourse signaling in academic lectures". *International Journal of Corpus Linguistics, 11* (3), 283-304

Renouf, A. 8c Sinclair, J. McH. 1991. Collocational frameworks in English. In K. Aijmer & B. Altenherg (Eds.), *English Corpus Linguistics*. London: Longman, 128-143.

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic hook reviews". *English Text Construction, 3* (1), 95-119.

iafor