Exploring phraseological variations by concgramming: The realisation of complete patterns of variations*

Winnie Cheng · Maggie Leung
(The Hong Kong Polytechnic University)

Cheng, Winnie and Maggie Leung. 2012. Exploring phraseological variations by concgramming: The realization of complete patterns of variations. *Linguistic Research* 29(3), 617-638. The significance of studying the co-selection of words has long been recognized. More traditional corpus linguistic approaches or tools help to find co-selections in the form of contiguous words (i.e. n-gram, or lexical bundles and clusters) or non-contiguous patterns. However, phraseologies in the form of non-contiguous co-occurrence with positional variations are rarely examined. Here it is argued that they are worth examining and have significance for better understanding language use and meaning. One reason for the rare discussion of these phraseologies is that they are not easily discovered with more traditional approaches and tools. This paper describes the realisation of different patterns of phraseological variations, exemplified with five concgrams (i.e. co-occurrence of words) extracted from two profession-specific corpora. With the use of an innovative corpus linguistic software, ConcGram 1.0 (Greaves, 2009), the frequencies and patterns of all of the possible phraseological variations (constituency and positional variations) of the concgrams are analysed. The illustration and analysis have implications on the application values of studying phraseological variations using concgramming. *(The Hong Kong Polytechnic University)*

**Keywords** phraseology, ConcGram 1.0, word co-selection, constituency and positional variations, phraseological variations

1. Introduction

Since the 1960s, an important area of study in corpus linguistics is to uncover the

---

* We are thankful to two anonymous reviewers for their helpful comments and criticism. All remaining errors are of course ours. The work described in this paper was substantially supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region (GRF Project No.: PolyU 5459/08H).
extent of word co-selections (Sinclair, Jones, & Daley, 1970). The idiom principle (Sinclair, 1987) suggests that “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair, 1991, p. 110). The principle is based on phraseological tendency in language use, meaning that words are not randomly selected, but rather they are co-selected by writers and speakers to convey meanings. However, anybody interested in the full range of these word co-selections faces the problem of extracting them from texts or corpora.

Traditionally, researchers focus on the co-selection of contiguous words because they are able to extract them by generating ‘n-grams’, i.e., the recurrent contiguous words that constitute a phrase or a pattern of use in texts or a corpus. N-grams are also termed ‘lexical bundles’, ‘word clusters’, or ‘lexical clusters’ (see, for example, Biber, Conrad, & Cortes, 2004; Carter & McCarthy, 2006; Hyland, 2008; Nesi & Basturkmen, 2009; Biber, Kim, & Tracy-Ventura, 2010; Adel & Erman, 2012; Csomay, 2012). Linguistic realizations of n-grams are based on the number of words in the sequence, for example, bi-grams (e.g. ‘interest rates’), tri-grams (e.g. ‘assets and liabilities’), and so on. Previous studies have investigated n-grams in different genres, for example, academic prose (Biber, Johansson, Lecch, & Conrad, 1999; Cortes, 2004; Biber, 2006, 2009; Hyland, 2008; Chen & Baker, 2010; Byrd & Coxhead, 2010; Adel & Erman, 2012), classroom talk (Biber et al., 2004; Biber & Barbieri, 2007; Csomay & Cortes, 2009; Neely & Cortes, 2009; Herbel-Eisenmann, Wagner, & Cortes, 2010; Csomay, 2012), conversation (Biber et al., 1999; Biber, Conrad, Reppen, Byrd, & Helt, 2002; Biber, 2009; Crossley & Salsbury 2011), and European Union documents (Jablonkai, 2010). Biber et al. (2004), for instance, investigated the use of lexical bundles in two university registers, one spoken (university teaching) and one written (textbooks). The lexical bundles identified from university teaching and textbooks are analysed in terms of types and their frequency distribution, and then classified based on their discourse functions. In another study, Hyland (2008) examined 4-word bundles in a corpus of academic writing across four disciplines, including electrical engineering, biology, business studies and applied linguistics. Having examining the forms and functions of the 4-word bundles, Hyland (2008) found considerable variations across types of academic writing as well as across disciplines, indicating that “writers in different fields draw on different resources to develop their arguments, establish their creditability and persuade their
Phrases such as ‘current assets’ can be generated using n-gram, but instances realised in the form of, for example, ‘current financial assets’ or ‘current tax assets’ would be missed. ‘Skipgram’ (Wilks, 2005; Guthrie, Guthrie, & Wilks, 2009) was later developed to deal with the limitations of n-grams. It is used to find non-contiguous word co-occurrence, in other words, “gapped n-grams” (Cheng, Greaves, & Warren, 2006, p. 412). Initially, skipgrams are seen as a better means as they can handle constituency variation. Guthrie et al. (2009) suggest that they provide “a much fuller model of language with little loss” (p. 45). However, some instances of word associations may still be missed in skipgram searches. Besides, the size of skipgrams is limited to trigrams (or 3-word skipgrams) with up to four skips (Wilks, 2005, 2008; Guthrie et al., 2009). More importantly, they cannot handle instances realised in the form of ‘interest rate’ and ‘rate of interest’, i.e. positional variation (Cheng et al., 2009).

Due to the limitations of n-gram and skipgram searches, researchers from the Research Centre for Professional Communication in English of the Hong Kong Polytechnic University have developed a corpus linguistic programme to address the challenge (Warren, 2009a).

1.1 ConcGram and concgrams

ConcGram 1.0 (Greaves, 2009) is a corpus linguistic program designed to uncover the co-occurrence of words fully and in an automated way. The products of ConcGram 1.0 are termed ‘concgrams’, defined as comprising all of the permutations of the association of two or more words, irrespective of whether the words occur in different sequence relative to one another (i.e. positional variation, AB and BA) or when one or more words drop between the co-occurring words (i.e. constituency variation AB and ACB) (Cheng et al., 2006). The concgrams that are found by the software are all instances of word co-occurrences. Since not all of them are necessarily meaningfully associated, it is useful and necessary for users to open up the concordances with the concgram search function “to distinguish between ‘co-occurring’ words (i.e. concgrams) and ‘associated’ words (i.e. phraseology)” (Warren, 2009a, p. 3). In other words, while the list of concgrams extracted using ConcGram 1.0 are “objective, automatically generated data” (Warren, 2009a, p. 3),
the determination of meaningfully associated instances is subjectively based on the interpretation of and parameters set by the users or researchers.

As the primary function of ConcGram is to perform fully automated concgram search and extraction from a text or a corpus, using ConcGram in such an unfettered mode allows users to conduct truly “corpus-driven” studies (Tognini-Bonelli, 2001) without inputting any prior, pre-defined search command (Cheng et al., 2006). Cheng et al. (2006) state that this fully automated capability of the phraseological search engine further increases the likelihood that researchers discover new co-selections or patterns of language use. It is also possible that the users nominate a word or words to search as a concgram search query (Cheng et al., 2006). Another innovative feature of the software is that all of the instances of a concgram are displayed in one set of concordance lines (Cheng et al., 2009), meaning that instances of all variations, including n-grams, positional variation and constituency variation, are displayed in a single set of concordance lines for analysis. Such a reader-friendly design of the display of concgram concordances makes it manageable for the user (Warren, 2009a). The functions of ConcGram 1.0 hence enables researchers to provide a more extensive and authentic description of the pattern, use and meanings of language.

The phraseological search engine ConcGram and concgramming findings have been discussed in a number of papers (see, for example, Cheng et al., 2006; Cheng et al., 2009; Cheng, 2012). These papers outline the methodology and highlight the functions of ConcGram. By comparing n-grams and skipgrams, Cheng et al. (2006) describe the functions and features of ConcGram 1.0, and discuss the potential contribution of studying conegrams to identify phraseologies. In spite of the relatively short period of development, ConcGram has been used as a corpus analytical tool in various studies across different registers or genres, for example, local and overseas newspapers, magazines, engineering brochures, etc. (see, for example, Cava & Venuti, 2008; Milizia & Spinzi, 2008; Cheng & Lam, 2010, 2012; Warren, 2010; Cava, 2010; Sun, 2010; Cheng, 2009, 2011). Cheng and Lam (2010), for example, used ConcGram 1.0 to generate two-word conegrams from the two ‘human rights’ corpora of newspaper reports collected in pre and post 1997 in Hong Kong. They discuss whether there are any changes in the ways human rights are represented in pre and post colonial periods. Cava and Venuti (2008) investigate the linguistic choices that authors use in journal article abstracts to position themselves
within the discourse community and positively evaluate their work at the same time. Using *ConcGram* to generate concordances of some associated words identified in the corpus, Cava and Venuti (2008) examine the recurrent lexical patterns and find “a significant co-occurrence with terms specifically related to the presence of the researcher” (ibid., p.154). In a recent publication on a study of media discourse, Cheng and Lam (2012) examine the difference in western and Chinese perceptions of Hong Kong before and after the handover of Hong Kong to China in 1997 by comparing the corpora of overseas newspapers and magazines and local newspapers between the two periods of 1996-1998 and 2006-2008. The analysis was done based on the co-selection of words in the corpora, specifically the two-word concgram *political/Hong Kong* situated in different corpora.

Other studies put greater emphasis on the value of concgramming on pedagogy and language learning (see, for example, Cheng, 2007, 2010; Greaves & Warren, 2007; Warren, 2009b, 2011). For example, Greaves and Warren (2007) introduce the use of *ConcGram* as a computer driven methodology to the teaching and learning of phraseology. The methodology is outlined in the paper with examples, and possible interactive and collaborative teaching and learning activities are discussed. The contributions of ‘concgramming’ in language teaching and learning are suggested in the paper both in terms of raising teachers’ and learners’ “critical awareness of the nature and role of phraseology in the English language” (ibid., p. 304) and encouraging them to play the role of language researchers. Warren (2009b) studies the phraseologies and their role in the realisation of intertextuality of discourse flows in e-mails collected from the workplace by extracting concgrams from the professional e-mail corpus. This paper explores the implications of applying the notion of phraseology and combining it with intertextuality in the learning and teaching of English for professional communication.

2. The present study

The aims of the present study are three-fold: (1) to describe and discuss five concgrams to illustrate different forms of phraseological variations, i.e. n-grams, constituency variation and positional variation, (2) to demonstrate the analytical procedure of concgrams to show how they can be analysed from raw data to a
description about meaning language use with human interpretation, and (3) to empower teachers and researchers with the relevant knowledge and skills in order to achieve a fuller understanding of phraseology.

As a part of the on-going larger scale project investigating the phraseologies specific to engineering and financial services registers, the phraseologies selected for this paper come from two Hong Kong profession-specific corpora which were compiled by the Research Centre for Professional Communication in English of the Hong Kong Polytechnic University http://rcpce.engl.polyu.edu.hk/. The Hong Kong Engineering Corpus (HKEC) consists of 9,224,384 words and the Hong Kong Financial Services Corpus (HKFSC) 7,341,937 words. The two corpora are large collections of texts collected from the engineering and financial services sectors respectively in Hong Kong, with the help of professional associations. Examples of genres in the HKEC are code of practice, guide, ordinance, project summary, and so on, and some of those in the HKFSC are annual report, investment product description, prospectus, speech, etc.. Figure 1 shows the front page of the HKEC website with the built-in ConcGram online version.
Exploring phraseological variations by concgramming

The corpora were first searched by ConcGram 1.0 to generate a list of all the unique words, i.e. type, and then a list of two-word concagrams. The two-word concagram list was generated with each word in the unique word list acting as an origin for the search of all its co-occurring words within a default concordance string of 50 characters on each side of the centred word. The built-in exclusion list function was used so that the 50 most frequently occurring grammatical words in the BNC were excluded in the generation of the two-word concagram list, so as to obtain a list with lexical words. The next step was to generate the concordance for each two-word concagram for study. The phraseologies analysed and discussed in this paper were originally two-word concagrams from the two profession-specific corpora. Three concagrams were chosen from the HKFSC and two from the HKFSC (see Table 1). They are lexically rich and were selected to illustrate, as far as possible, the full range of patterns of phraseological variation.
Table 1. Five two-word concgrams selected for analysis

<table>
<thead>
<tr>
<th>Two-word concgrams</th>
<th>Corpus</th>
<th>Frequency of word co-occurrence</th>
<th>Percentage of word co-occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>fair / value(s)</td>
<td>HKFSC</td>
<td>4,058</td>
<td>0.055%</td>
</tr>
<tr>
<td>management / risk(s)</td>
<td>HKFSC</td>
<td>1,581</td>
<td>0.022%</td>
</tr>
<tr>
<td>interest / rate(s)</td>
<td>HKFSC</td>
<td>4,293</td>
<td>0.058%</td>
</tr>
<tr>
<td>energy / saving(s)</td>
<td>HKEC</td>
<td>1,290</td>
<td>0.014%</td>
</tr>
<tr>
<td>energy / use</td>
<td>HKEC</td>
<td>1,616</td>
<td>0.018%</td>
</tr>
</tbody>
</table>

The concordance for each two-word concgram was analysed manually to first identify the degree of word associations, followed by identifying the pattern of phraseological variations. Table 2 describes the frequencies of association between the two words.

Table 2. Frequency of word association

<table>
<thead>
<tr>
<th>Two-word concgrams</th>
<th>Corpus</th>
<th>Frequency of word association</th>
<th>Percentage of word association</th>
</tr>
</thead>
<tbody>
<tr>
<td>fair / value(s)</td>
<td>HKFSC</td>
<td>3,608</td>
<td>0.049%</td>
</tr>
<tr>
<td>management / risk(s)</td>
<td>HKFSC</td>
<td>1,263</td>
<td>0.017%</td>
</tr>
<tr>
<td>interest / rate(s)</td>
<td>HKFSC</td>
<td>3,466</td>
<td>0.058%</td>
</tr>
<tr>
<td>energy / saving(s)</td>
<td>HKEC</td>
<td>1,290</td>
<td>0.014%</td>
</tr>
<tr>
<td>energy / use</td>
<td>HKEC</td>
<td>1,616</td>
<td>0.018%</td>
</tr>
</tbody>
</table>

3. Analysis of phraseological variations

In this section, the five phraseologies will be discussed in terms of the patterns of phraseological variations they display. The first one is fair/value(s). This phraseology (N=3,608) is found among the ten most frequent ones in the HKFSC. It displays only limited patterns of variations. Almost all of the instances are n-grams (3,581 times, 99.3%) forming a noun phrase, with fair acting as an attributive adjective and value(s) as the head noun. Since the n-gram ‘fair value(s)’ has the highest frequency, it can be regarded as the canonical form of the phraseology (Cheng et al., 2009).

1 January 2005. Previously, the change in the fair value of investment properties was recognised in
2 investment properties, as permitted by HKAS 40. Fair values are determined by independent professional
costs. At each balance sheet date the fair value is remeasured, with any resultant gain or

Figure 2. Examples of fair/value(s) as n-grams

The remaining 27 instances denote the constituency variation, with some examples shown in Figure 3. In all of these instances, only one word drops between fair and value(s) and it is typically ‘market’ (24 times). The other 3 instances are ‘asset’, ‘trading’ and ‘carrying’. In such cases, fair and value(s) together with the intervening word form a larger noun phrase. The constituency variation highly adheres to the canonical form, in terms of their “meaning, syntactic entity, and frequency of occurrence” (Cheng et al., 2009, p. 243).

1 committed and forecast transactions with a net fair asset value of $8 million (2006 : $1 million
2 for on the balance sheet based on their fair market values as at the Listing Date. (3)
3 and liabilities arising from the acquisition: Fair Carrying Value Amount HK$M HK$M Fixed assets

Figure 3. Examples of fair/value(s) displaying constituency variation

The phraseology management/risk(s) includes the inflected forms, risk and risks. It occurs 1,263 times in the HKFSC. This phraseology is found to display positional variation. The first positional variant is management…risk(s) (82 times, 6.5%). All of these instances are non-contiguous with intervening words (see Figure 4 for examples). Typically, management is followed by the preposition ‘of’. The patterns of the intervening words are: (1) management + of + risk(s), (2) management + of + + determiner + risk(s), (3) management + of + noun + risk(s), and (4) management + of + adjective + risk(s). Pattern (1) and (2) are used to specify the type of management that is related to risk(s). In pattern (3) and (4) where a noun or an adjective is found intervening, it functions as a modifier of the word risk(s). More frequent modifiers are ‘credit’, ‘market’, ‘financial’, and ‘operational’, which function to delimit the type of risk(s) being managed and therefore provide further specifications of the type of management.

1 procedures in place for the identification and management of risks are adequate. The Audit Committee
2 is mandated to provide high level centralised management of credit risk for HSBC worldwide. Group
3 Directors reviews and approves policy for the management of the strategic risk. The Board has

Figure 4. Examples of management…risk(s) displaying constituency variation

The other positional variant risk(s)…management (1,184 times, 93.5%) denotes
both n-grams (1,176 times) and constituency variation. Both of the words are nouns and occur contiguously to form a compound noun. Indeed, the n-gram ‘risk management’ expresses the same meaning of the positional variation ‘management of risk(s)’ which is to specify the type of management. In other instances, the n-gram is a part of a bigger compound unit, for example, ‘risk management functions’ and ‘risk management strategy’ in lines 2 and 3 in Figure 5.

Instances displaying constituency variation in the sequence of risk(s)⋯management are infrequent (5 times, 0.4%). The patterns of these five instances are (1) risk + noun + conjunction + management, (2) risk + conjunction + noun + management, and (3) risk + noun + management. Pattern (1) occurs three times, for example, ‘risk control and management’ and ‘risk monitoring and management’ (lines 2 and 3, Figure 6). The word ‘monitoring’ in line 3 is a gerund transformed from a verb, but it functions as a noun, and thus also included in pattern (1). These instances can be expanded as ‘risk control and risk management’ and ‘risk monitoring and risk management’. In such cases, risk is directly modifying management, and therefore they express the same meaning as the n-gram ‘risk management’. The other two patterns only occur once in the corpus.

The canonical form of the phraseology is the n-gram risk(s) management. The other two variations adhere to the canonical form with a small “degree of turbulence” (Cheng et al., 2009, p. 243).

The above analyses show that fair/value(s) and management/risk(s) exhibit a limited number of variations, whereas the three other phraseologies to be discussed display a complete range of realisations of all of the possible phraseological
Interest/rate(s) is one of the most frequent phraseologies in the HKFSC (3,466 times), with two inflected forms of rate, i.e., rate and rates. The phraseology displays both positional variations, interest''rate(s) and rate(s)''interest. However, the frequency proportions of these two variants are distinctly different. In each of the positional variants, both the occurrences of n-grams and constituency variations are observed.

Regarding the first positional variant interest''rate(s) (3,391 times, 97.8%), 3,215 instances occur as n-grams (Figure 7), forming a compound noun which expresses the meaning of a certain amount or percentage related to interest. There is often an adjective preceding the n-gram to modify interest rate(s), for example, ‘variable’, ‘effective’, ‘market’, ‘fixed’, ‘floating’, ‘commercial’, ‘low’, and so on. The high frequency of this n-gram pattern makes it the canonical form of the phraseology.

1. bank deposits and time deposits carry variable interest rates, ranging from 3.100% to 4.825% (2005: 2. component of convertible bonds. The effective interest rate of the liability component is 4.05%. In 3. year-on-year, benefiting from higher market interest rates as well as the increase in higher

When the two words are non-contiguous, four typical patterns with intervening words are observed: (1) interest + preposition + rate(s), (2) interest + preposition + adjective + rate(s), (3) interest + preposition + determiner + adjective + rate(s), and (4) interest + verb + preposition + determiner + rate(s). The intervening preposition is predominantly ‘at’. The word rate(s) is modified by an adjective, such as ‘prime’, ‘fixed’, ‘bank’, ‘variable’, ‘floating’, and ‘effective’. Some of the modifiers are found to be the same as those that co-occur with n-grams. The determiner may be the indefinite article ‘a’ or the definite article ‘the’. In pattern 4, interest + verb + preposition + determiner + rate(s), the verb is typically ‘bear’ and it appears in its present participle form ‘bearing’, for example, ‘interest bearing at a rate’. It is also observed that when interest and rate(s) exhibit constituency variations, rate(s) is often followed by the preposition ‘of’ and then a numeral in percentage form, for example, lines 1 and 2 in Figure 8. This pattern is more specific to instances exhibiting constituency variation than those as n-grams. In both
the n-gram and constituency variation pattern, interest and rate(s) co-occur to denote the meaning of an amount or a level of the interest, but such a level is expressed explicitly in the constituency variation with more frequent collocation of a numeral.

1 a PRC State-owned bank, is unsecured, bears interest at a rate of 5.58% per annum and is repayable
2 will mature on 12 August 2004 and will carry interest at the rate of 2.20% per annum payable
3 by the PRC government. Cash at banks earns interest at floating rates based on daily bank deposit

Figure 8. Examples of interest…rate(s) displaying constituency variation

The difference in the pattern and the meanings between the n-gram and the constituency variation of interest…rate(s)… can be a good example demonstrating the different language use even using the same phraseology. While the n-gram ‘interest rate(s)’ tends to refer to a general concept or item of the interest, the constituency variation has a tendency to express a particular level or amount of interest.

The other positional variant rate(s)…interest occurs only 75 times. Only 4 instances are n-grams (see Figure 9). After examining the concordance lines of these instances and considering the low frequency, it is suggested that the n-gram ‘rate interest’ is rarely used in financial services texts and even when it is used, there is a modifier preceding it, for example ‘floating rate interest’ in line 1, or it is part of a larger n-gram, such as ‘fixed rate interest income’ in line 2 or ‘floating rate interest bearing financial assets’ in line 3. In other words, unlike the n-gram in the other positional variation, ‘interest rate(s)’, in which the two words form a compound noun and can stand alone by their meaning, ‘rate interest’ tends not to stand alone. The word rate combining with its preceding adjective modifies the word interest as in line 1, or, as in the case of lines 2 and 3, it combines with the adjective (e.g. ‘fixed rate’) to modify a compound noun formed by interest and another noun (e.g. ‘interest income’), and together they form a larger compound noun unit (i.e. ‘fixed rate interest income’).

1 these swap contracts, we pay floating rate interest and receive fixed rate interest payments. We
2 investment tenor with a high yield fixed rate interest income,” said Frank Turley, Head of Retail
3 in the table are the Group’s floating rate interest bearing financial assets and financial

Figure 9. Examples of rate(s)…interest as n-grams
The remaining instances of rate(s)…interest display constituency variation. Figure 10 shows some examples. One pattern is found, i.e., rate(s) + of + interest. Whether the word rate appears in its singular or plural inflected form, the intervening word is ‘of’. This pattern typically has a modifier preceding the word rate(s), such as ‘annual’ (line 1), ‘market’ (line 2), ‘valuation’, ‘higher’, and so on. Besides this, as illustrated in line 3, there is a single occurrence of ‘the rate applicable to interest’. This is not regarded as a second pattern, as “in corpus studies, it is recurrent instances that are significant” (Cheng et al., 2009, p. 245).

1 than equity shares or land, that annual rate of interest which, if used to calculate the present
2 approach using the prevailing market rates of interest available to the Company for financial
3 Derive from the PRC, the rate applicable to interest, rental, licence fees and other income by

Figure 10. Examples of rate(s)…interest displaying constituency variation

Overall, all four patterns of phraseological variation are exhibited in the phraseology interest/rate(s). They can be listed in the following order with the canonical form at the top and the pattern with the highest level of turbulence in meaning at last:

1. interest rate(s) (3,215 times)
2. interest * rate(s) (176 times)
3. rate(s) * interest (71 times)
4. rate interest (4 times)

The two constituency variations can be further broken down to more configurations based on the number of intervening words, represented by asterisks each representing an intervening word (see, Cheng et al., 2009). Since the aim of this paper is to describe and illustrate different types of phraseological variations by studying concgrams, all non-contiguous instances of a positional variant are grouped as one configuration for the sake of simplicity and demonstration.

This phraseology interest/rate(s) demonstrates a complete set of patterns of phraseological variations. The canonical form is ‘interest rate(s)’. The two words interest and rate(s) in the other configurations constitute a “textual object” (Sinclair & Mauranen, 2006, p. 149), meaning that the two words form a single linguistic
entity (ibid.), or constitute a “textual incident” (ibid., p. 154) which are formed by two or more textual objects (ibid.). Thus, the other three patterns are regarded as variants to the canonical form.

Although the complete set of phraseological variations are displayed, one of the patterns, ‘rate interest’, is found to have higher turbulence in its meaning, when compared to the canonical form ‘interest rate(s)’. Two phraseologies from the HKEC will be discussed below to demonstrate not only the complete patterns of phraseological variations but also how all the patterns share the meaning.

The phraseology energy/saving(s) (1,043 times) in the HKEC contains inflected forms of saving, i.e. saving and savings. In the case of the first positional variant energy … saving(s), 930 of the 966 instances are n-grams (Figure 11). Syntactically, when the two words occur as n-grams, they function either as a compound noun where energy serves as a noun adjunct modifying saving(s) (lines 1 and 2), or as a compound modifier with energy and saving modifying another noun (line 3). The nouns modified by ‘energy saving’ are typically ‘measures’, ‘feature’, ‘technologies’, ‘opportunities’, etc.

1  glazing will lead to largely different amount of energy saving per unit area of glazing used in
2  of inattention and scarce capital to realise energy savings that will pay for itself in the long
3  faces potential for loss of revenue through energy saving measures. Yet even the most optimistic

**Figure 11.** Examples of energy…saving(s) as n-grams

The two words display constituency variation in the same sequence for 36 times. The typical forms identified are (1) energy + noun + saving, (2) energy + conjunction + noun + saving, and (3) energy + noun + conjunction + saving. The three examples in Figure 12 realise these three patterns respectively. In line 1, ‘cost’ is the intervening word. In such case, energy is not directly modifying saving, but forms with the intervening noun as a larger unit. Patterns (2) and (3) have a conjunction and they can be reformed to show a clearer syntactic relationship between energy and saving. For example, line 2 can be rewritten as ‘energy saving and cost saving potentials’, or line 3 can be rewritten as ‘energy usage and energy saving’. In such cases, energy and saving have the same relationship as that when they denote as n-grams, and thus having the same functions of the n-gram ‘energy saving’, i.e. either a compound noun or a compound modifier.
1 enhancement and monitoring and sharing of the energy cost saving between the building owner and the
2 analysis has provided useful insights into the energy and cost saving potentials and their financial
3 Lighting application as a key for efficient energy usage and saving as well as being environmental

Figure 12. Examples of energy·saving(s) displaying constituency variation

The other positional variant saving(s)·energy occurs 77 times (7.4%). About one-third of these are n-grams (25 times) (Figure 13). The typical pattern of these instances is noun + preposition + saving energy. The word saving is the gerund form, so it compounds with energy to form a noun phrase. While it shares a similar meaning of the n-gram in the other positional variation ‘energy saving’, the typical pattern of ‘saving energy’ suggests that ‘saving energy’ is used in the texts to refer to a particular aspect related to saving energy. The aspects are represented by a noun, for example, ‘benefits’, ‘tips’, and ‘purpose’.

1 be longer. Apart from the benefits of saving energy, District Cooling Scheme can eliminate the need
2 24% This booklet provides tips for saving energy at home.* Some simple energy saving tips: 1.
3 nowadays in Hong Kong for the purpose of saving energy in lighting. The operating frequency of

Figure 13. Examples of saving(s)·energy as n-grams

Two-thirds of the instances of this positional variant display constituency variations. Typical patterns include (1) saving(s) + preposition + energy, (2) saving + adjective + energy, and (3) saving(s) + preposition + (determiner) + adjective + energy. In pattern (1), two prepositions, ‘in’ or ‘of’, that affect meaning are found. When ‘in’ is used, energy is not used alone but acts as a noun adjunct modifying another noun, for example, ‘saving in energy consumption’ (line 1). Whereas when ‘of’ is used, energy stands alone, such as ‘savings of energy’ (line 2). The intervening adjective in pattern (2) modifies the word energy to specify the particular type of energy being saved, such as ‘saving electrical energy’. Line 3 is an example of pattern (3), where the adjective ‘overall’ modifies ‘energy consumption’.

1 cost payback period and annual saving in energy consumption were clearly listed for
2 Energy Management Opportunities where savings of energy and money can be made. Energy Audits Our
3 show that significant saving in the overall energy consumption of the chilling system can be

Figure 14. Examples of saving(s)·energy displaying constituency variation

The four patterns of the phraseological variations of energy/saving(s) are listed
below:

1. energy saving(s) (930 times)
2. saving energy (25 times)
3. saving(s) * energy (52 times)
4. energy * saving(s) (36 times)

The predominant occurrence of the n-gram ‘energy saving(s)’ makes it the canonical form of the phraseology. The other patterns of variations basically adhere to the canonical form as they share much of the same meaning, i.e. saving energy or saving some energy-related items such as ‘energy cost’, ‘energy consumption’, and ‘energy bills’.

Another example from the HKEC is energy/use (1,118 times), with 807 instances (72%) denoting the positional variant energy…use, and 311 instances (28%) displaying the other positional variation use…energy.

Of the 807 instances of the energy…use, 598 (74.1%) are an n-gram ‘energy use’, with use acting as a noun (Figure 15). Thus the two words energy and use combine to form a compound noun. Some typical collocates to the left of the n-gram are ‘annual’, ‘air-conditioning’, and ‘operating’.

There are 209 instances denoting the constituency variation. Of these instances, 179 (85.6%) are ‘energy end-use’ (line 1, Figure 16), with use being part of the hyphenated compound ‘end-use’, and energy combined with ‘end-use’ to be a larger n-gram or compound noun. Since it is a hyphenated compound noun, it is arguable as to whether the word ‘end’ should be called an intervening word. While ‘energy end-use’ has a similar meaning with the n-gram ‘energy use’, ‘energy end-use’ focuses on the ultimate consumption or application of the energy. This difference is indicated by the word ‘end’.

1. the load factors (LFi) in order that the annual energy use calculated from the summation process
2. by the user before the program can predict the energy use and to perform LCA and LCC calculations for
3. purpose). Furthermore, calculation of operating energy use of all or individual services installations

Figure 15. Examples of energy…use as n-grams

There are 209 instances denoting the constituency variation. Of these instances, 179 (85.6%) are ‘energy end-use’ (line 1, Figure 16), with use being part of the hyphenated compound ‘end-use’, and energy combined with ‘end-use’ to be a larger n-gram or compound noun. Since it is a hyphenated compound noun, it is arguable as to whether the word ‘end’ should be called an intervening word. While ‘energy end-use’ has a similar meaning with the n-gram ‘energy use’, ‘energy end-use’ focuses on the ultimate consumption or application of the energy. This difference is indicated by the word ‘end’.

1. on their respective net calorific values. The energy end-use data set will be updated regularly and
benchmarks for building performance, covering energy and water use, indoor environmental conditions.

Figure 16. Examples of energy···use displaying constituency variation

Other patterns of the constituency variation include (1) energy + conjunction + noun + use, (2) energy + noun + conjunction + use, and (3) energy + preposition + modifier + use. In line 2, ‘energy and water use’, an example of pattern (1), can be expanded as ‘energy use and water use’. Pattern (2), for example ‘energy production and use’, is similar to pattern (1) in the sense that it can also be expanded as ‘energy production and energy use’. Thus in these two patterns, energy directly modifies use constituting a ‘textual object’ (Sinclair and Mauranen 2006, p. 149). A textual objects is defined as including “the main traditional word classes, nouns, adjectives, verbs and adverbs, either on their own or as heads of phrases; a textual object is a construct that must combine with another in order to be deployed in a communicative act” (Sinclair and Mauranen 2006, p. 149, pp. 154-55). In pattern (3), the intervening preposition is typically ‘for’ and the modifiers are ‘on-site’ (line 3) and ‘wide-scale local’, which function to provide details of the use of the energy.

19 of the 311 (6.1%) instances of the other positional variant use···energy are used as an n-gram ‘use energy’, with use used as a verb 16 times (see Figure 17). There is typically an infinitive ‘to’ preceding the verb use. In the other instances, use is combined with end to form a hyphenated compound, i.e. ‘end-use’. The word energy is used as a noun in 5 instances, for example ‘to use energy’ (lines 1 and 3, Figure 17). In these cases, energy and use co-select as an n-gram ‘use energy’ that shares a similar meaning as another n-gram ‘energy use’. Line 2 shows that energy is used together with an adjective, for example, ‘efficient’ and ‘saving’, to modify a noun phrase, for example, ‘to use energy efficient lighting installation’.

1 helped customers better understand how to use energy more wisely and contribute to a greener world.
2 ISO14000 environmental management scheme to use energy efficient lighting installations in their
3 in the short term is for community to use energy more efficiently. As responsible individuals,

Figure 17. Examples of use···energy as an n-gram

In the 292 instances of non-contiguous use···energy, the main patterns found are (1) use + of + energy, (2) use + of + adjective + energy, and (3) use +
adjective/adverb + energy. The concordance lines in Figure 18 are examples of these patterns. For the two patterns with the preposition ‘of’, the word use functions as a noun. The intervening adjective in patterns (2) and (3) is to make the type of energy in use specific. In addition to ‘renewable’ (lines 2 and 3), other adjectives found include ‘solar’, ‘wind’, ‘clean’, ‘green’, and ‘lighting’. Adverbs are typically ‘more’ and ‘less’.

1 investments in the production and the use of energy from sustainable sources. Before the upcoming
2 require the power companies to use renewable energy 3 and potentially lead to wider use of renewable energy in the future. As Hong Kong’s first wind power

Figure 18. Examples of use---energy displaying constituency variation

The instances of this phraseology realise the complete patterns of phraseological variations and each of them shares the same meaning as that of the canonical form ‘energy use’, i.e., to use of the energy power or specific types of energy power, with little turbulence. All the patterns of phraseological variations are n-grams, constituency variation, and the positional variation.

4. Conclusion

This paper has described the concgramming analytical procedures and discussed the linguistic realisations of different phraseological variations with five examples from two profession-specific corpora. Using fair/value(s), management/risk(s), interest/rate(s) from the HKFSC, and energy/saving(s) and energy/use from the HKEC, the paper has illustrated all of the possible phraseological variations, namely n-grams (contiguous words), constituency variations (non-contiguous words), and positional variations (words occurring in a different sequence relative to one another) that are specific to the examples of two-word concgrams (Cheng et al., 2009).

The discussion of the examples shows that some phraseologies convey the same meaning irrespective of the variations displayed, whereas for some phraseologies, different positional or constituency variations may express different meanings. This provides insights and suggestions for pedagogical and research values of studying phraseological variations. The paper suggests that language teachers of general English or specialised English introduce the concept of phraseological variations in
order to provide a better and more specific description of the patterns and meanings of collocation, and thus a fuller description of language use. Language learners are able to learn the language independently by exploring the phraseological profile of the language. Moreover, this paper also has implications for researchers of phraseological investigations. Concgrams provide researchers useful raw data towards quantifying the extent of phraseologies in a corpus. The analytical procedures of the phraseologies described in this paper have demonstrated how the patterns of phraseological variations can be derived from the raw data of concgrams with human interpretation. The fully automated capability of the software ConcGram enables researchers to conduct corpus-driven studies in which the researchers do not have any specific words or phrases in mind. Thus the probability of finding new phraseologies of patterns of language use is potentially much higher (Cheng et al., 2006).

References


Byrd, P., & Coxhead, A. 2010. On the other hand: Lexical bundles in academic writing and in the teaching of EAP. University of Sydney Papers in TESOL 5: 31-64.


Cava, A. M., & Venuti, M. 2008. The good me or the bad me? Identity and evaluation in research article abstracts. Linguistica e Filologia 27: 139-156.


Cheng, W. 2010. Hong Kong Engineering Corpus: Empowering professionals-in-training to learn the language of their profession. In M. C. Campoy-Cubillo, B. Bellés-Fortuño, & M. L. Gea-Valor (Eds.), Corpus-Based Approaches to English Language Teaching (pp. 67-78). London and New York: Continuum.


Exploring phraseological variations by concgramming

Guthrie, D., Guthrie, L., & Wilks, Y. 2009. What is a “full statistical model” of a language and are there short cuts to it? In D. Hlaváčková, A. Horáček, K. Osolsobě, & P. Rychly (Eds.), After half a century of Slavonic (pp. 45-56). Brno: Tribun EU.
Warren, M. 2009b. The phraseology of intertextuality in English for professional


---

**Winnie Cheng**  
Department of English  
The Hong Kong Polytechnic University  
E-mail: winnie.cheng@polyu.edu.hk

**Maggie Leung**  
Department of English  
The Hong Kong Polytechnic University  
E-mail: maggie-sn.leung@polyu.edu.hk

Received: 2012. 10. 23  
Revised: 2012. 12. 17  
Accepted: 2012. 12. 18