A systematic analysis of a two-word concgram in Nepalese policy documents: A corpus-driven approach

Madhu Neupane Bastola,
Tribhuvan University, Nepal

Abstract

Corpus linguistics can inform language teaching in various aspects from syllabus designing to creating exercises based on the real use of language. However, its use in language teaching is still rare. In the context of Nepal, though corpus linguistics forms a part of the University Curriculum in English Education, the students are rarely offered a practical experience of corpus analysis. The same is the case with teacher training courses. This paper followed an analytical procedure for identifying phraseological variation within a two-word 'concgram' that is a set of co-occurring words. In this paper, a two-word concgram, make/effort, is analyzed to identify concgram configurations, the most frequently used form, and its meaning by using concordance lines. Lastly, the paper presents the implications of corpus analysis in English language teaching.

Keywords: Corpus, ConcGram, concgrams, phraseology, meaning shift units

Introduction

In a general sense, a corpus refers to a collection of texts. More specifically, a corpus is a collection of naturally occurring machine-readab-le discourse put together for linguistic research (Adolphs, 2006; McEnery & Hardie, 2012) and the area of linguistics that studies a corpus is corpus linguistics. Corpus linguistics has been informed by empiricism as its philosophical base for the study of language (Adolphs, 2006; Aijmer, 2009; McEnery & Hardie, 2012; Sinclair & Carter, 2004; Stubbs, 2005, 2007) for making sense of naturally occurring language data (Tognini-Bonelli, 2001). It has been a fertile field of research because it has a broader application in different areas such as language teaching, professional communication, and academic writing, to name but three. Based on the nature and purpose of research, corpus research has been divided into three types: corpus-based, corpus-driven, and data-driven (Adolphs, 2006; McEnery & Hardie, 2012; Rayson, 2008; Tognini-Bonelli, 2001)

Corpus-based research is mostly top-down because it usually explores the relationship between theory and data. According to Tognini-Bonelli (2001), researchers following this approach usually start with a predefined model of language or hypotheses about language and "analyse the corpus through these categories and sieve the data accordingly" (p.66). As the researchers are interested in how well their theories are accounted for by data, a theory is a priori to corpus-based research. This deductive methodology can sometimes cause a problem when the data does not fit to theory because of the inherent variability in language. Such problem may lead the researchers either to modify the theory or to insulate the data that fits the theory (Tognini-Bonelli, 2001). Nevertheless, corpus-based studies are common in second language acquisition because findings derived from such studies can be helpful in teaching-learning processes (Aijmer, 2009).

Unlike in corpus-based approach, in a corpus-driven approach, a corpus "is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system. The theoretical statements are fully consistent with and reflect directly, the evidence provided by the corpus" (Tognini-Bonelli, 2001, p. 84). The researchers use their knowledge to make sense from the data. This approach is inductive and takes a holistic stance in describing the system of language by using uncontaminated (i.e., not pre-tagged) raw data (Sinclair & Carter, 2004). Because of its holistic nature, corpus-driven research gives high priority to the representativeness of corpus to derive valid conclusions and emphasizes a continuous update of a corpus (Sinclair & Carter, 2004; Tognini-Bonelli, 2001). However, there is no fixed size for a sample to be representative. The appropriateness of the size depends on the claims that the researchers want to make.

In addition to corpus-based and corpus-driven approaches, Rayson (2008) proposes a data-driven approach, which, he claims, combines both corpus-based and corpus-driven research approaches. Unlike a corpus-based approach, the data-driven approach does not start with research questions instead starts with building a corpus. Then it goes on to annotate the corpus, retrieve data from the corpus, devise a research model or questions, retrieve data, and interpret the data for confirming the hypothesis. Though this model is similar to the corpus-driven method as it starts with building a corpus, it uses part of speech tagging commonly used in corpus-based studies. This study followed a corpus-driven approach to explore a corpus of Nepalese policy documents.

Phraseology and meaning shift units (MSUs)

Phraseology is the study of the structure of phraseological units referred differently as phrseologism (Gries, 2008), multiword sequences such as "lexical phrases, formulas, routines, fixed expressions, prefabricated patterns (or prefabs), and lexical items" (Biber, Conard, & Cortes, 2004) or meaning shift units (MSUs) (Cheng, Greaves, Sinclair,

and Warren, 2009). According to Cheng et al. (2009), phraseology is used "in an inclusive sense to describe the tendency for words to be co-selected by speakers and writers to achieve meanings" (p. 236). Sinclair (2004) refers to this tendency as the 'idiom principle' because the choice of one word greatly constraints the other choices available.

Different terms (e.g., n-grams, multiword expressions, skip grams, phrase frames, phraseologism, or MSUs) are used to refer to phraseological units. Among them, phraseologism and MSUs are more convincingly defined. A phraseologism refers to:

> the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is more significant than expected by chance. (Gries, 2008, p. 6)

This definition focuses on six different aspects in defining a phraseologism: the nature of the element (i.e., a lexical item or a grammatical unit), number of elements (two or more), frequency of occurrence (more than that can occur by chance), distance (contiguous or non-contiguous), flexibility (fixed or variable), and semantics (functioning as a semantic unit in terms of meaning). From this perspective, *in spite of* and *of course* are examples phraseologisms (Gries, 2008).

Sinclair (2004) used the term MSU to describe meaningful word associations. The basic idea behind using the term MSU is that "each new combination of words generates a shift in meaning, even if this is only relatively subtle, compared with other possible combinations involving one or more of the words" (Cheng et al., 2009, p. 237). In each collection of words, there is a shift in meaning. For example, in the unit *hard work* the meanings of *hard* and *work* have shifted to give a new sense. The meaning of *hard* in *hard work* differs from the meaning of *hard* in *hard surface*. The meaning of *hard* in these examples shifts as it combines with different words. According to Cheng et al. (2009), an MSU is composed of five categories of co-selection: the core (the element that is invariably present), semantic prosody (negative or positive orientation of meaning), semantic preference (restriction to regular co-occurrence), collocation (content words that usually co-occur with the core), and colligation (grammatical categories that usually co-occur with the core). Out of these five categories of co-selection, the core and semantic prosody are obligatory while others are optional. The analysis of a concgram *make effort* illustrates all these aspects of co-selection.

Implicitly subsumed in the definition of an MSU are the criteria of distance between elements (i.e., contiguous or non-contiguous), their flexibility or positional variations (e.g., *hard work* and *work hard*), and their frequency of occurrence in a corpus. Therefore, the definition of an MSU is more comprehensive as it incorporates all the defining aspects of a phraseologism discussed above and more. Therefore, this paper uses the

concept of MSU in analyzing word associations and uses ConcGram© (Greaves, 2009) program to identify such associations fully automatically.

ConcGram and concgrams

Examining the association between words has been an important area of study (Greaves & Warren, 2007) yet a difficult task (Cheng, Greaves, & Warren, 2006) in corpus linguistics. Previous research on extracting word association has focused on n-grams (i.e., contiguous words that constitute a phrase such as *a lot of people*) and skipgrams (i.e., non-contiguous word associations that are four words apart-e.g. *play a much less important role*). However, neither n-grams nor skipgrams can extract positional variations in word associations such as *role play* and *play a role* (Cheng et al., 2006). A new computer-mediated software ConcGram© (Greaves, 2009) has overcome this problem as it can automatically identify all the potential configurations of between 2 and 5 words including positional variations (e.g. *work hard* and *hard work*) and constituency variations (i.e., *work hard* and *work very hard*) (Cheng et al., 2009, 2006; Greaves & Warren, 2007). Such associations between words with the constituency and/or positional variations are called concgrams. In other words, concgrams are "co-occurring words irrespective of whether or not they are contiguous, and irrespective of whether or not they are in the same sequential order" (Cheng et al., 2009, p. 237). As in other corpus research software such as WordSmith © (Scott, 2012), concordance lines are used in ConcGram. Interestingly though, ConcGram highlights all the associated words of a concgram in each concordance line as shown in the following examples:

> the concerned corporation will be required to **make** efforts to control its cost of production or
> the situation Teachers were found to have **made** some **efforts** to address the language problems
> concrete steps in that direction. The plan will **make** all possible **efforts** in this regard.   The
> limited availability of sources and means have **made** our planned development **efforts** more
> training local leaders and social workers to **make** them more responsive to development **efforts**.
> economic policies, it becomes a national duty to **make** available some returns of development **efforts**
> inflow of information or conservation **efforts**made in the international context and to mobilize
> is, therefore, essential that further **efforts** be **made** to ensure a multi-purpose development and use of
> been the case in the past, then **efforts** will be **made** to promote highly essential industries from among
> plans in future. Hence, **efforts** are to be **made** further for extending various educational
> will be strengthened. **Efforts** will also be **made** to coordinate the research and development
> education. (3) **Efforts** should be continued to **make** Sanskrit education timely, socially beneficial
> **efforts** for industrial development have been **made** in preceding years, even with the mentioned

There can be two-, three-, four-, or five-word concgrams. Identification of concgrams helps us to better understand idiomatic nature of language by observing the co-selection pattern in a spoken or written discourse. The search for concgrams can be automatic or user nominated. ConcGram is corpus-driven as it extracts the concgrams fully automatically (refer to Cheng et al., 2006, Cheng et al., 2009 or Greaves & Warren, 2007 for more information about the process). The ConcGram software identifies all the co-occurrences of words that comprise a concgram. The researchers need to identify the associated ones and discard others because some of such co-occurrences shown

in concordance lines may not be associated (Cheng et al., 2009). The co-occurrences of *make* and *effort* in the following examples are not associated:

> **effort**s, the country has not been able to **make**any significant progress in raising domestic
> **effort**s to their sustainable development and to **make** them capable by their participation. Organisation

This paper analyzes one of the most frequent two-word concgrams *make effort* in a Nepalese Policy Corpus (NPC hereafter) of about two million words and identifies all possible concgram configurations and configuration functions.

The present study

This research aimed to examine one of the most frequently occurring congrams (i.e. *make effort*) with positional variation in the NPC. For this purpose, first, the corpus of Nepalese Policy documents consisting of 11 periodic plans (1956-2016), 9 educational plans (1956-2016), 3 constitutions (1990-2015), and two language policy documents (1990-2016) was created by combining the soft copies of these documents available various websites. The corpus has around 2 million words. Second, from the NPC, the lists of unique words and two-word congrams excluding function words were generated. One of the most frequently occurring two-word concgrams (i.e., *make effort*) with both constituency variation (e.g., *make efforts*, *make* great *efforts*) and positional variation (e.g., *make effort* and *efforts made*) was chosen for further analysis informed by Cheng et al.'s (2009) framework. Third, concordance lines for the selected concgram were generated by using ConcGram© (Greves, 2009) because of its robust nature in extracting phraseological units that have positional variations as well constituency variations (Cheng et al., 2009; Warren, 2009). Wild card and word/prefix functions were used in the search process to identify all inflectional variations of *make* (i.e., *make*, *made*, *making*, *makes*) and *effort* (i.e., *effort* and *efforts*) (see Greves, 2009 for more information about the use of wild clad and word/prefix function). The concordance generated 459 co-occurrences of *make effort* in the corpus. Out of them, 22 co-occurrences were excluded from the analysis because they did not belong to the same MSU that is *make effort*. Nine of the co-occurrences were generated because of the use of wild card and word/prefix functions used in the search process:

> location. The NASA is a tool that decision **make**rs can use to target **effort**s and resources for
> are delayed because of shortage of construction **mate**rials, and duplication of **effort** occurs. Limited

In thirteen of them, two words (i.e., *make* and *effort*) were not associated:

> for their gradual utilisation will be **made**. Other Means of Transport ·**Effort**s to link the
> improved. Institutional reforms will be **made** to **make** co-operative movement effective. **Effort**s will

Therefore, altogether there were 437 instances (*make effort*- 45 and *effort made*- 392) of the concgram *make effort* in the NPC. The MSUs and the canonical concgram configuration and functions were identified keeping two factors into consideration: whether or not the concgram units belonged to the same MSU and how frequently

they occurred in the corpora. The canonical function related to the canonical form and degrees of turbulence (i.e., variation in a canonical form) were also identified. WordSmith© (Scott, 2012) was used to examine collocation and colligation of the word *effort*.

Finally, to better understand the use of the concgram, the findings obtained from the NPC were compared with other reference corpora, Brown corpus, Frieburg-Brown Corpus, London-Oslo/Bergen Corpus, and Frieburg-London-Oslo/Bergen Corpus. The reference corpora are briefly described in this section.

Brown corpus (BC): Brown corpus is 1-million-word corpus of standard American English complied by W.N. Francis and H. Kucera at Brown University in 1961. The corpus was named after Brown University. The texts for the corpus were sampled from 15 different text categories. The corpus contains 500 texts of just over 2000 words. (https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html)

Frieburg-Brown (Frown)Corpus (FC): Frown corpus was compiled by Christian Mair. It matches the Brown corpus in its design but was compiled to represent American Standard English of the early 1990s. It was named after University of Freiburg, Germany. (http://clu.uni.no/icame/manuals/FROWN/INDEX.HTM)

London-Oslo/Bergen Corpus (LOB corpus): The LOB corpus was prepared with the cooperation between the University of Lancaster, the University of Oslo, and the Norwegian Computing Centre for the Humanities at Bergen. It is like Brown corpus in design and sample (i.e., fifteen different text categories and 500 printed texts of just over 2000 words). However, unlike Brown corpus, it represents standard British written English of the 1960s (http://clu.uni.no/icame/lob/lob-dir.htm).

Frieburg-London-Oslo/Bergen (FLOB) Corpus: This corpus is also a member of brown family. Designed by Christian Mair in 1991, the corpus matches the well-known Brown and LOB corpora and represents standard written British English of the early 1990s (http://clu.uni.no/icame/manuals/FLOB/INDEX.HTM).

Hong Kong Financial Service Corpora (HKFSC): HKFSC is a 7-million-word corpus of collection of text from Hong Kong financial service. It was developed by the Research Centre for Professional Communication in English of The Hong Kong Polytechnic University (http://rcpce.engl.polyu.edu.hk/HKFSC/).

Results and discussion

Analysis of the first positional variation of the concgram (i.e., make effort)

Among the 437 instances of the concgram, there were 45 instances of the first positional variation (i.e., *make effort)*. The different configurations of the variation are presented in Table 1.

Table 1. The configurations of the first positional variation in the NPC

S.N.	make/effort	Frequency
1.	make effort	11
2.	make*effort	22
3.	make**effort	11
4.	make***effort	1

*Note: The number of * refers to the number of words between make and effort.*

The data shows that *make * effort* (with one word in between) had the highest frequency of occurrence (22 times). This most frequently occurring form is the canonical form. The canonical form has either determiner (7 times) or a modifier (13 times) as the intervening word. The second most frequent configurations were *make effort* (11 times) without any intervening word and *make**effort* (11 times) with two intervening words. The followings are some representative examples of the different configurations of *make effort* in the NPC:

market and production provision and to **makeeffort**s to move ahead   to cope international
The University and its colleges should **make** no **effort** to mould high school curricula to
to this, the National Planning Commission has **made** an **effort** to incorporate the suggestions received
essential for the programmes of all sectors to **make** a concerted **effort** to alleviate poverty. Hence,
 limited availability of sources and means have **made** our planned development **effort**s more

The configurations of *make effort* obtained from the NPC were compared with those obtained from the reference corpora described in the previous section. The results are shown in Table 2.

Table 2. The comparisons of the configurations of the first positional variation

make effort	make effort	make* effort	make** effort	make*** effort
BC		2	3	1
FC		7	3	
LOBC		6	2	1
FLOBC		7	2	
HKFSC	3	38	13	4
NPC	11	22	11	1

*Note: The number of * refers to the number of words between make and effort.*

The comparison also shows that *make*effort* is the canonical configuration with its highest frequency of occurrence. Unlike in the NPC, the configuration *make effort* without any intervening word was nonexistent in reference corpora except in HKFSC. Table 3 presents the configuration functions of the first positional variation.

Table 3. The configuration functions of make/effort in the NPC

S.N.	make effort	Frequency
1.	Verb + object	11
2.	Verb + det. /mod. + object	22
3.	Verb + det. /mod. + mod. + object	11
4.	Verb + det. + mod. + mod. + object	1

In the configurations outlined in Table 3, *make* functions as a verb and *effort (s)* as an object. In eleven of these cases, verb and object are contiguous and, in 22 cases, there was a single intervening word (either a determiner or a modifier). In rest of the cases, there were two or three intervening words. In some cases, the intervening words were determiners and modifiers (e.g., *a concerted effort*) or two modifiers (e.g., *poverty alleviation efforts*). The words like *special*, *coordinated*, *concerted*, and *possible* modified the word *effort*. The configuration function with either one determiner or one modifier was the most frequent one.

The reference corpora were examined to compare the configuration functions of the first positional variation as identified in the NPC. The comparison confirmed the findings obtained from the NPC (see Table 4).

Table 4. The configuration functions of make/effort in the reference corpora

	Verb + object	Verb + det. /mod. + object	Verb + det. /mod. + mod. + object	Other
BC		8	6	1
FC		7	3	
LOBC		6	3	1
FLOBC		7	2	
HKFSC	3	38	13	4
NPC	11	22	11	1

The followings are some representative examples from the reference corpora:

Verb + object

    two broad guidelines, we   shall continue to **makeeffort**s in the following tasks: (a)

    our infrastructural development, we should **makeeffort**s in the related planning and

Verb + det. /mod. + object

> to standards and expectations, or who **made** exaggerated **effort**s to achieve these goals.
> second he turned his eyes. He seemed to **make** no **effort** to speak or even smile. Presently
> and in face of falling prices, we have **made** strenuous **effort**s to rationalise our
> and his Foreign Minister **made** great **effort**s to curb Tirpitz and smooth British

Verb + det. /mod. + mod. + object

> over from Georgi Zaroubin, he **made** a determined **effort** to change this idea.
> and saucepan drop from his hands,**make**s diverse convulsive **effort**s to rise, and finding

Other

> Only four towns indicated that they **made** any more than a normal **effort** to list property
> loving him; from the tension, the weariness that **made** even breathing an **effort**.
> We in the HKMA have taken the initiative and **made** a start in this **effort**, and have got quite a lot
> competitive market. At the same time, we have **made** significant progress in our **effort**s to improve

As observed in the NPC, *Verb + det. /mod. + object* was the canonical configuration function in all the reference corpora.

Analysis of the second positional variation of the concgram (i.e., effort(s) made)

In comparison to *make effort*, *effort made* was significantly more frequent in the NPC corpus as there were 392 stances of this positional variation. Table 5 presents the configurations of this variation.

Table 5. The configurations of effort made in the NPC

S.N.	make effort	Frequency
1.	effort made	36
2.	effort* made	44
3.	effort ** made	256
4.	effort*** made	39
5.	effort**** made	5
6.	effort***** made	3
7.	effort******made	3

Table 6 shows the comparison of textual configurations of the second positional variation of the concgram in the NPC corpus with those in the reference corpora.

development. In spite of the **effort**s **made** in the past to ensure people's participation of
has been overlooked and not much **effort**is **made** towards involving the users in construction is
 not adequate. 8. Adequate **effort**is not **made** for the development and extension of appropriate
to achieve this goal, all the **effort**s to be **made** by the government in economic, political,
determination is imperative. Big **effort** must be **made** on domestic front for mobilization of internal,
This shows that every possible **effort** has to be **made** at the government level for the mobilisation of
No coordinated **effort** has so far been **made** to see that the prescribed syllabus is adhered
education. (3) **Effort**s should be continued to **make** Sanskrit education timely, socially beneficial
ineffective, **effort**s in that direction will be **made** afresh with a change of emphasis. The emphasis
of the **effort**s and achievements in the country **made** in this short duration, some thought provoking
efforts for industrial development have been **made** in preceding years, even with the mentioned

Table 6. The comparison of the configuration of effort made

	effort made	effort* made	effort ** made	effort *** made	effort (** ** or more) made	
BC	1	9	5	2	3	20
FC		2				2
LOBC	1	4	4	2		12
FLOBC	2	6	2		1	11
HKFSC	10	17	26	2	8	63
NPC	36	44	256	39	11	392

The comparison shows that *effort**made* (with two intervening words) is the canonical configuration form. However, the frequency of this configuration is noticeably high in the NPC and HKFSC than in other reference corpora. It seems that the form *effort**made* is used frequently in policy documents. However, further research is required to verify this claim. Table 7 presents the configuration functions of the second positional variation.

Table 7. The configuration functions of second positional variation in the NPC

S.N.	make effort	Frequency
1	Noun (PrepP) + past participle	36
2	Noun + to infinitive	8
3	Sub + is/am/are/was/were + past participle	32
4	Sub + Aux + (not) + (adv.) +be + past participle	333
5	Sub + aux + to be + past participle	11

What follows is the brief description of the configuration functions described in Table 7.

Noun (post modifier) + past participle [*efforts (of the government) made*]: In this case, *made* is a participle adjective post modifying the noun *effort*. If we examine further, the past participle form is an example of a *reduced clause* (e.g., *efforts -that are- made*). As can be seen in the second example, there can also be a prepositional phrase in between the oun and the participle form.

first attempt towards organizational **effort**s **made** in the planned development of tourism, in Nepal.

the many **effort**s at privatization **made** in the past were unsuccessful. After the

Noun + to infinitive (*effort to make*): In this case, *to-infinitive* form functions as the post modifier of the noun *effort (s)*.

them as part of the ceaseless **effort** to **make** educational administration more efficient and

and political context, and lack of **effort** to **make** it timely. Lack of comprehensiveness in teacher

be reviewed from time to time in an **effort** to **make** them capable of fulfilling an increasing needs

Situation   (a) Culture **Effort**s to **make** local agencies and users' groups participate in

Sub + Aux (is/are/was/were) + (not) + past participle: This is one of the frequent configuration functions in the NPC corpora. In this case, the word *effort* functions as the subject of the passive clause. Among different forms of *be* verbs, *were* was the most frequent one (were-25 times, was-3 times, is- 3 times, are- 3 times). As is evident in the examples below, such forms can be negative, however, they were less frequent than the positive ones.

order to achieve this objective, an **effort** is **made** to present the distribution of the languages in

of current capacities alone. An **effort** was **made** to estimate the resource requirements on a

of overstaffing continues. Unless **effort**s are **made** to install cost-effective administrative

Some institutional and procedural **effort**s were **made** during the Eighth Plan in order to **make**

is not adequate. 8. Adequate **effort** is not **made** for the development and extension of appropriate

on the scale, unless, timely **effort**s are not **made** to raise them tip. Priority should be given to

noticed that sufficient **effort**s were not **made** in the area of agricultural prices and market

that **effort**s to increase production should be **made** all t he more extensive in view of population

Sub + modal/be/have + (adv.) + (not) + be/been/being + past participle: This is the most frequent configuration functions of the second positional variation (i.e. *effort made*). Table 8 presents the variations within this configuration function.

Table 8: Sub + modal/be/have + (adv.) + (not) + be/been/being + past participle

S.N.	Form	Frequency
1.	Sub (post modifier) + will + (adv.) + be + past participle	176
2.	Sub (post modifier) + has/have/had + (adv.) + (not) + been + past participle	55
3.	Sub (post modifier) + shall/should + be + past participle	21
4.	Sub (post modifier) + is/are + being + past participle	12
5.	Sub (post modifier) + is/are/have + to be + past participle	11
6.	Sub (post modifier) + must be + past participle	4
7.	Sub (post modifier) + would be + past participle	1

The followings are some examples of the different configuration functions presented in Table 8.

> wise basis. 9. Especial **effort**s will be **made** to implement programs, with the cooperation of the
>
> will be prohibited. **Effort** will also be **made** to **make** the people aware of the disastrous
>
> tax system, 4.Meaningful **effort**s have not been **made** to ensure adequate returns to the increasing
>
> addition to these, some **effort**s have also been **made** in the construction of suspension bridge, wooden
>
> been pointed out that further **effort**s should be **made** in the coming plans for boosting the general
>
> sources. To solve this problem **effort** is being **made** to tap the ground water resources. After
>
> potential embodied in them. **Effort**s are being **made** to fulfill such commitments from the government
>
> determination is imperative. Big **effort** must be **made** on domestic front for mobilization of internal

*Sub + aux + to be + past participle (*some representative examples):

> of Internal resources Great **effort**s have to be **made** for the mobilisation of additional resources to
>
> been realised that more **effort**s have to be **made** towards achieving targets of woman and child
>
> In view of the situation, **effort**s are to be **made** henceforth for the proper arrangement in

The configuration functions of *effort made* identified in the NPC were compared with such functions in the reference corpora (see Table 9).

Table 9. The comparison of the configuration function of effort/make

Form	BNC	BC	FC	LOBC	FLOBC	KHFSC	NPC
Noun + past participle	5	1		1	2	10	36
Noun (PrepP)+ to infinitive	5	2		2	2	11	8
Sub + is/am/are/was/were + (adv.) + past participle	15	9	5	4	7	7	37
Sub + Aux + (not) + (adv.) +be + past participle						23	333
Sub + aux + to be + past participle	3	1	1	1	1		11
Sub + finite clause						9	

The comparison also shows that the same configuration function *Sub (PreP/and noun/ infinitive clause) + Aux + (not) + (adv.) + be+ past participle* was the canonical configuration function. However, the use of this function only in the NPC and HKFC seem to suggest the frequent use of this configuration function in policy documents. Followings are some representatives examples taken from reference corpora:

Noun+ past participle

immensely encouraged by the splendid **effort made**, particularly by our own British

it may have. And I am pleased that the **effort**s **made** by Hong Kong as a whole have been duly

Noun + (PrepP) + to + verb

new Brahms Second is a major **effort** to **make** a record that sounds like a real orchestra rather)

efforts of the Home Office and the police to **make** the public more aware of the need to

community support for all of our **effort**s to **make** Hong Kong a   better place to live. This is an

Sub + Aux (is/are/was/were) + (not) + past participle

training programs for those skills. No **effort** is **made** in the same studies to present

practical plans and strenuous **effort**s are **made** to keep them in the active fellowship.

it is not surprising that many **effort**s were **made** to give it concrete expression. A number

Some time ago an **effort** was **made** to sell more yoghurt in greater

Sub + modal/be/have + (adv.) + (not) + be/been/being + past participle

sewer maintenance division said **effort**s will be **made** Sunday to clear a stoppage in a sewer

Within each chapter an **effort** has been **made** to group together those crystals with similar

(5)_ Unsuccessful **effort**s have been **made** to replace high mileage allowances with state

in the extensive **effort**s President Kennedy has **made** to enlist solid bipartisan support for

side every conceivable **effort** had been **made** to arrive at agreement on a procedure for the

where a major Soviet **effort** is being **made**. But the issue is also a basic

disagreements occur every **effort** should be **made** to resolve the issue and to avoid dissention.

with systemic consequences. So **effort**s must be   **made** to strengthen global prudential and regulatory

Sub + finite clause

years of premium payments. The **effort**s you **make** now and over the next few years will assure you

and sponsors for your superb **effort**s that **make** this event a huge success. I look forward to

organisations. Your hard work and **effort**s have **made** this event a great success. And, of course, my

is the result of the tremendous **effort**s we have **made** to maintain and, where practicable, improve our

our shared view of the **effort**s that must be **made** by businesses, governments and consumers, to

to the painstaking **effort**s which have been **made** over the years to develop and agree operational

of the **effort**s that our community has **made** since the Asian financial crisis. The United

I applaud the **effort**s of all those who helped **make** this important agreement possible, in

for the extra **effort** many of you have had to **make** to come here today. It's a privilege and a

for the **effort**s and the contributions they have **made** in this momentous year. Guo Shuqing

As indicated in Table 10, the concgram *make effort* seems to be common in discourse related to policy and plans because the normalized frequency of its occurrence is noticeably high in the NPC and HKFC.

Table 10. Normalization of frequency of make/effort in different corpora

	BC	FC	LOBC	FLOBC	KHFSC	NPC
Number of words in corpora	1 million	1 million	1 million	1 million	7 million	2 million
Concgrams per 10,000 words	0.35	0.13	0.21	0.2	0.82	2.18

Collocation, sematic preference, and semantic prosody of make/effort

Collocations: Table 11 shows the adjectives, nouns and verbs that frequently collocated with *effort(s)* in the corpora used in this research.

Table 11. Collocates of effort(s)

Adjectives				Nouns	Verbs
active	cooperative	greater	possible	alleviation	initiate
addition	coordinated	increased	present	background	intensify
additional	coordinating	institutional	private	community	made
alleviation	decentralization	integrated	reasonable	conservation	make
collaborative	determined	necessary	recent	coordination	making
collective	developmental	joint	repeated	decontamination	meet
community	effective	limited	serious	defense	plan
concentrated	established	local	significant	development	
concerted	extensive	marketing	special	education	
considerable	further	national	strenuous	future	
consistent	future	particular	sustained	government	
continued	genuine	past	timely	initiative	
continuing	governmental	persistent	tremendous	management	
continuous	great	planned	targeted	mobilization	

Semantic preference*:* Semantic preference refers to the semantic field that the collocational words belong to. In this case, the word *effort* has semantic preference to verbs of making or changing (e.g., *make, develop, plan, increase*); adjectives of difficulty (e.g., *special, coordinated, continued*), and noun with institutional sense (e.g., *development, government*). The word *effort* was in common use in the plan and policy documents when concrete implementation strategies to achieve the desired goals were lacking.

Semantic prosody: Semantic prosody refers to negative or positive orientation of meaning of an MSU. The construction *make effort* has a nuance of negative semantic prosody as it has the meaning of doing something good to change or improve the existing situation when the plan for concrete action is lacking. See the examples below:

> much more data is needed. **Effort**s will be **made** together as much data as possible.
> Plan: Increase in Exports **Effort**s will be **made** to increase exports by 100 percent within the
> on imported materials, continued **effort**s will be **made** to replace imports by domestic production by
> and wasteful expenditures. **Effort**s will be **made** to change all sectors of the country's economy

The detailed reading of the text where the above sentences were used showed that there were no specific targets set and no implementation plans *to replace imports by domestic production*, *increase exports by 100 percent* and *change all sectors of the economy*. Similarly, words like *special*, *earnest*, *rational*, *continued*, *necessary*, *coordinated*, *additional*, and *possible* that collocated with *effort* carried the sense that achieving the desired goal was not easy. The predominant use of *effort* in a passive structure with no responsible agent makes the lack of action apparent.

Conclusion

This study has examined the phraseology of a two-word concgram. The analysis shows that *make/effort* constitutes an MSU. The second positional variation (i.e. *effort made*) is more common than the first positional variation (i.e. *make effort*). The MSU has the semantic preference for verbs of change (e.g., *make*, *develop*, *change*), noun having organizational sense (e.g., *nation*, *government*, *institution*) and adjectives associated with difficulty and time duration (e.g., *special*, *coordinated*, *continued*). It has negative semantic prosody of *difficulty* and seems to be mostly used in passive constructions without agents. As there are no implementation plans to achieve the goals and the agency that makes efforts is usually obfuscated, the responsible ones

are difficult to identify and/or held accountable. Therefore, plans for making efforts do not necessarily make a difference.

Implications of the study

Corpus linguistics has multiple benefits in language teaching (Ebeling, 2009). However, its use in language teaching is still rare and "teachers are still unwilling to or lack the skill to use corpora as an aid to get new insights into English" (Aijmer, 2009, p. 1). In the context of Nepal, though corpus linguistics forms a part of the University Curriculum in English, students are rarely offered hands-on experiences (Granath, 2009). Highlighting its importance, Granath (2009) calls corpus *a native-speaker consultant* and considers corpus even more informative than a dictionary or a grammar book. Corpus can also be used for designing teaching materials and exploring texts produced by second language learners. Granath (2009) suggests that "different types of corpus exercises can profitably be made part of regular foreign language courses" (p. 48).

Corpus informed teaching can introduce learners with real language rather than made-up examples. Teachers can use corpora to create teaching materials, demonstrate variation in grammar, and show contextual variations in meaning. Corpora can help teachers present a contrastive analysis of lexical items in different domains, teach the meaning of multiword units, and use corpus as a reliable tool to answer students' questions related to the use of language. Alternatively, teachers can ask students to draw their own conclusions based on corpus data (Adolphs, 2006; Ebeling, 2009; Granath, 2009; McEnery & Hardie, 2012; Römer, 2009). According to Adolphs (2006), three main implications of corpus analysis in language teaching are syllabus design, material development, and data-driven learning (Adolphs, 2006). For this reason, it is better to introduce corpus analysis in pre- or in-service teacher training and teacher education courses so that teachers could help themselves and their students to discover the use of language and acquire knowledge and skills (Ebeling, 2009; Granath, 2009; Römer, 2009). In this regard, the analysis of concgram can help the learners to better understand the association between words, common phenomena in the English language.

References

Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies*. London: Routledge.

Aijmer, K. (Ed.). (2009). *Corpora and language teaching*. Amsterdam: John Benjamins.

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

Cheng, W. (2009). Income/interest/net: Using internal criteria to determine the aboutness of a text. In K. Aijmer (Ed.), *Corpora in language teaching* (pp. 157–178). Amsterdam: John Benjamins.

Cheng, W., Greaves, C., Sinclair, J. M., & Warren, M. (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, *30*(2), 236–252.

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

Ebeling, S. O. (2009). Oslo Interactive English: Corpus-driven exercises on the Web. In *K. Aijmer (Ed.), Corpora and Language Teaching* (pp. 67–82). Amsterdam: John Benjamins Publishing Company.

Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47–66). Amsterdam: John Benjamins Publishing Company.

Greaves, C. (2009). ConcGram [a phraseological search engine]. Amsterdam: John Benjamins.

Greaves, C., & Warren, M. (2007). Concgramming: A computer driven approach to learning the phraseology of English. *ReCALL*, *19*(3), 287–306.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger and F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 3–26). Amsterdam: John Benjamins.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519–549.

Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 83–100). Amsterdam: John Benjamins Publishing Company.

Scott, M. (2012). WordSmith Tools version 6. Lexical Analysis Software.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education / Mike Scott and Christopher Tribble*. Amsterdam: John Benjamins.

Sinclair, J. M., & Carter, R. (2004). *Trust the text: Language, corpus and discourse*. New York, N.Y: Routledge.

Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature*, *14*(1), 5–24.

Stubbs, M. (2007). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, *3*, 166–193.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Warren, M. (2009). Why concgram? In C. Creaves, *Concgram 1.0: A phraseological search engine* (pp. 1–11). Amsterdam: John Benjamins.