Phraseology in two Omani undergraduate Civil Engineering genres: *The N of (the) N* pattern

Priya Mathew
Centre for Academic Writing, Middle East College, Knowledge Oasis Muscat,
P.O. Box 79, PC 124, Al Rusayl, Sultanate of Oman

ABSTRACT
The study of phraseology with respect to continuous and discontinuous frames in academic writing has gathered increasing research attention over the past decade. Their prevalence in expert writing and the influence of discipline and genre on their frequency and type have led to studies that attempted to identify the most productive discontinuous frames. The aim of this study is to investigate the pattern *the N of (the) N*, a prolific pattern in expert academic writing, in two Omani corpora of undergraduate Civil Engineering genres, Case Studies (CS) and Methodology Recounts (MR). The two strands of inquiry involve 1) a comparison between the semantic noun categories of the first (N1) and second noun (N2) used in this pattern and; 2) the N1-N2 sequences in the pattern which realize specific discourse functions in these two genres. Strings belonging to this pattern were retrieved from the two corpora through the corpus interface, *Sketch Engine*. Findings indicate the prevalence of this frame in the two genres and genre influence on the choice of semantic noun categories. It was also found that the N1-N2 sequences in the pattern are used to realise distinct discourse functions in the two genres. This is one of the first corpus-based studies on university student writing in the Middle East and considering that English is the language of instruction and assessment in many of these countries, these findings have significant pedagogical implications. The implications are that EFL students in such *lingua franca* contexts can be supported by a more discipline-specific approach.

Keywords: Civil engineering; corpora; Omani undergraduates; p-frames

INTRODUCTION

Engineering has been one of the most sought after disciplines by students across universities in the world and it is no wonder that this discipline has attracted the attention of linguists belonging to various traditions. Engineering discourse has been examined in terms of genre structure (Artemeva & Freedman, 2008; Artemeva, 2009; Cheng, 2007; Conrad, 2014, 2017; Gardner, 2008; Kanoksilapatham, 2015, Nesi & Gardner, 2012), vocabulary (Durrant, 2014; Hsu, 2014; Mudraya, 2006; Ward, 2009), and specific linguistic devices (Gardner et al., 2018; Leedham, 2011; Leedham & Fernandez-Parra, 2017). The differences between novice writers and experts (Conrad, 2017; Hyland, 2008) as well as between students at various levels of proficiency have also been undertaken (Gardner et al., 2018; Staples et al., 2016). Some other studies have explored disciplinary and genre variation among Engineering sub-disciplines (Gilmore & Millar, 2018; Kanoksilapatham, 2015; Nekrasova-Beker, 2019; Rezoug & Vincent, 2018).

Several studies have focused on frequent multi-word sequences in engineering discourse.

Email: priya@mec.edu.om

Investigations of both continuous and discontinuous sequences have revealed that they vary across disciplines (Gardner & Xu, 2019; Hyland, 2008; Nekrasova-Beker & Becker, 2020), and levels of proficiency (Hyland, 2008). These studies have mainly focused on textbooks and research articles (Gilmore & Millar, 2018; Nekrasova-Beker & Becker, 2020; Wood & Appel, 2014) with a few exceptions. Rezoug and Vincent (2018) examined continuous sequences or Lexical Bundles (LBs) in Masters dissertations across four engineering disciplines and Hyland's (2008) LB study also included Masters dissertations and PhD theses in Electrical Engineering.

The question of the pedagogical utility of lists of continuous multi-word sequences (MWS) or Lexical Bundles has been raised time and again and various solutions have been proposed. However, before the question of pedagogical approaches to teaching multi-word sequences is considered, the relevance of bundles retrieved from genres that have little to do with the written tasks expected of students needs to be considered. There has also been limited research on multi-word sequences in undergraduate student writing in engineering disciplines, and no study has been undertaken, to the best of my knowledge, on discontinuous frames in undergraduate student writing, perhaps due to the occluded nature of these genres.

The present study attempts to investigate one particularly prevalent structure in academic writing, *NP + of phrase* (Biber et al. 1999; Cunningham 2017), in two important undergraduate Civil Engineering genres in a higher education institution in the Sultanate of Oman. The *NP + of* phrase construction was converted into the discontinuous frame, *the N of (the) N* and the semantic categories of the nouns used in this frame were examined across the two genres to identify their frequency and functions. The extension of the *NP + of phrase* to the second noun made these sequences more perceptually salient for pedagogical purposes as well as enabled the identification of the most frequent first noun and second noun combinations used to fulfil specific discourse functions associated with the two genres.

Continuous and discontinuous multi-word sequences

The seminal work of researchers such as Firth (1957), Bolinger (1976), and Pawley and Syder (1983) established the significance of phraseology in the English language. The latter pointed out the role of frequently repeated strings in facilitating fluency, making language natural and increasing processing speed while also highlighting the challenges that second language learners encounter in mastering this aspect of written and spoken discourse. A key development in this field was marked when Sinclair (1991) applied corpus linguistic methods to the study of phraseology. His idiom principle posits that expert users of the language have a stock of "semi pre-constructed phrases" (1991, p. 110) at their disposal. Pawley and Syder (1983) and Sinclair (1991) focus on the readily available fixed as well as slightly variable repeated multi-word sequences that are part of the lexicon of fluent and natural speakers and writers of the English language. Many approaches to the study of phraseology have been adopted since by researchers such as Construction Grammar (Goldberg, 1995), Pattern Grammar (Hunston & Francis, 2000), Lexical Priming (Hoey, 2005), and the psycholinguistic aspects of formulaic language (Wray & Perkins, 2002).

Both continuous and discontinuous multi-word sequences have been researched. There have been numerous corpus-driven studies on fixed continuous sequences commonly known as Lexical Bundles (LBs) starting with Altenberg (1998) who identified 470 frequent word sequences in the *London–Lund Corpus*. The Lexical Bundle approach involves the retrieval of fixed repeated strings of various lengths according to criteria based on frequency and dispersion. Biber et al.'s (1999) reported the LBs of various lengths found in conversation and academic prose. They examined the structure and discourse functions of these LBs and pointed out how many of these sequences were not perceptually salient. They were often found to function as a bridge between two structural units which later caused them to name LBs as "the building blocks of discourse" (Biber & Barbieri, 2007; Biber, 2006; Conrad & Biber, 2004; Chen & Baker, 2010). The main characteristics of LBs are that they are recurrent, fixed, non pre-defined, and not perceptually salient. Most of the LB studies since (Ädel & Erman, 2012; Chen & Baker, 2010; Rezoug & Vincent, 2018) have also examined the structure and functions of LBs drawn from various corpora based on frameworks developed by Biber et al. (1999) and Hyland (2008).

LB studies have reported that the situational variables of discipline, genre, and level of proficiency play a crucial role in the selection of LBs. Many of these studies have stressed the pedagogical implications of these findings for both L2 students (Ädel & Erman, 2012; Chen & Baker, 2010) as well as professionals (Pan et al., 2016). However, most studies of LBs have been on published research articles (RAs). Therefore, the suggestions for pedagogical utility have been mainly based on findings drawn from corpora of expert texts written in genres not attempted by students (Ädel & Erman, 2012; Chen & Baker, 2010). These recommendations may not be applicable because student genres are different from expert genres in their primary function (Hüttner, 2008). LB studies on occluded student genres such as undergraduate and postgraduate student assignments are quite limited.

Although limited research on discontinuous frames or p-frames has been undertaken, this approach is gathering increasing research interest. Renouf and Sinclair (1991) conducted one of the earliest studies of this type by examining the fillers in seven frames (e.g. *a + ? + of*) which they called "collocational frameworks". Marco's (2000) corpus-based study involved the identification of the most common frames in Medical research articles. He found that the most productive ones included the *NP + of* structure (e.g. *the number of, a variety of*). Some corpus-based studies involved *a priori* identification of p-frames with a variable slot. Biber's (2009) study of the phraseological patterns in lexical bundles derived from conversation and academic writing revealed that there are fundamental differences in the p-frames in writing and conversation. For example, the variable slot is usually a content word in the latter while in conversation, these sequences are fairly constant.

The corpus-driven method of identifying discontinuous sequences was an important methodological development (Cunningham, 2017; Gray & Biber, 2013; Nekrasova-Beker & Becker, 2020; Römer, 2009). For example, Römer's (2009) study adopted a corpus-driven approach to examine the 280 most frequent p-frames in a corpus of academic book reviews (e.g. *it * be interesting to* and *it would be * to*). Gray and Biber (2013) used custom-made programs to retrieve all the discontinuous frames from two corpora of conversation and written academic prose and show how a corpus-driven method of retrieval has a higher recall. These findings indicate that some significant frames may be missed out when an 'LB to frame' method (Biber, 2009; Grabowski, 2013) is adopted.

Another development in the study of discontinuous frames came with Fletcher's (2012) free program, *kfNgram,* which can be used to retrieve these sequences. A few studies have illustrated (Cunningham, 2017; Nekrasova-Beker & Becker, 2020) how this program can be employed to retrieve p-frames. Most studies of discontinuous sequences included examination of their structure and discourse functions (Gray & Biber, 2013; Nekrasova-Beker & Becker, 2020; Römer, 2009), as was typical in most LB studies (Hyland, 2008; Rezoug & Vincent, 2018). Grabowski's (2015) domain-specific framework (e.g. *reading-oriented, topic*) attempted to categorise the discontinuous frames in pharmaceutical genres. Similarly, Cunningham (2017) developed a framework to categorise the p-frames in Mathematical RAs (e.g. *aboutness, coherence*). These studies also reported on the extreme variability of the slot and their research-oriented function in academic writing.

Other work in phraseology has focussed on the association between the frequency of continuous and discontinuous sequences and their prevalence in genres (Ädel & Erman, 2012; Ebeling & Hasselgård, 2015; Hyland, 2012), differences between novice and expert writing (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004; De Cock, 1998; Nekrasova-Beker, 2019), disciplinary variation (Hyland, 2008) and variation across sub-disciplines (Grabowski, 2013; Nekrasova-Beker, 2019).

Phraseological investigations of discontinuous sequences in Engineering discourse are extremely limited indeed although some LB studies have been conducted (Gardner & Xu, 2019; Hyland, 2008; Rezoug & Vincent, 2018). A very recent study was conducted on engineering p-frames by Nekrasova-Beker and Becker (2020) who compared the structure, discourse functions, slot position, and other general characteristics of p-frames in corpora of published material used in five engineering disciplines at the undergraduate level. They reported that the majority of the frames were variable in terms of slot productivity across the sub-disciplines of Engineering they examined.

This review of studies on continuous and discontinuous sequences reveal that these phraseological phenomena are important for both novice writers as well as second language learners to master. Their study is especially important for second language learners in EFL contexts who are instructed and assessed in English. The academic success of these students depends on the quality of their written assignments, which should be characterized by frames and LBs that are the currency of their disciplinary discourse. However, most of the studies of this aspect of phraseology are based on corpora of research articles and textbooks and not on student work. Research on the phraseology of student assignments should therefore be the next major development in this area. It can also be seen that very few studies have focused on Engineering, a discipline which has been identified as a popular choice with students worldwide. It is evident that phraseological research has not caught up with other strands of research in Engineering discourse such as genre analysis (Artemeva, 2009; Nesi & Gardner, 2012), Multi-Dimensional Analysis which included Engineering texts (Gardner et al., 2018; Hardy & Römer, 2013) and disciplinary vocabulary (Durrant, 2014; Hsu, 2014; Mudraya, 2006; Ward, 2009), to name a few.

Clearly, the findings related to phraseological patterns in academic writing have pedagogical implications. The need for novice writers to acquire these sequences for entry into their respective discourse communities has been the most significant finding in all of these studies. Wray (2006, p. 593) points out, that "when we speak, we select particular turns of phrase that we perceive to be associated with certain values, styles and groups". However, the pedagogical utility of studies that have examined Lexical Bundles has been questioned (Biber &

Barbieri, 2007; Byrd & Coxhead, 2010; Liu, 2012) because of the challenges associated with teaching lists of word sequences which were most often not perceptually salient.

Some attempts have been made to make them more teachable. For example, Ädel and Erman (2012) tried to shortlist the most important and teachable LBs on the basis of their perceptual salience and the judgment of EAP instructors while Byrd and Coxhead (2010) suggested the use of longer LBs to increase their salience. Another solution could be to perhaps identify extremely productive frames which previous studies have found to comprise a significant proportion of academic texts such as the pattern *NP of*. This pattern has been found to be prevalent in academic prose in general (Biber et al., 1999, p. 365) and discipline specific texts in continuous (Gardner & Xu, 2019; Rezoug & Vincent, 2018) and discontinuous sequences (Nekrasova-Beker & Becker, 2020; Marco, 2000). Biber et al. (1999, p. 636) observe that many of the most common bundles in academic writing include an *NP + of phrase*, and that most of these express size and amount. It is thus logical to presume that some Engineering disciplines such as Civil Engineering which are applied and deal with concrete nouns, might contain a high proportion of this pattern. Gray and Biber (2013) also found that this pattern comprised 11 of the 15 most frequent p-frames in academic writing. Gardner and Xu (2019) corroborate these findings by reporting that their Engineering LBs primarily consisted of long nominals realized by the pattern "the x of the y, as in these examples: The case/ context/ development/ effect(s)/end/ number/ presence/ size/ sum / total number/ use of (the)".

In-depth analysis of pre-selected frames, including the semantic categories of the most frequently found parts of speech which constitute the fillers in p-frames is a fruitful line of inquiry. Most of the studies discussed above identify lists of frames and even though their structural and functional characteristics are discussed, the association between the semantic categories of the fillers comprising a restricted list of words to realize a particular discourse function is often not investigated. This is because even in studies where pre-selected frames are examined (Marco 2000) or those in which the most frequent frames generated through a corpus-driven retrieval method are examined, the focus is on more than two frames (Cunningham, 2017; Gray & Biber, 2013) thus affording no space for detailed examination of the components of each frame. The trend has been to look for generalisations so that frames can be categorized and grouped on the basis of their structure, discourse function or teachability.

This study proposes another thread of inquiry which involves in-depth analysis of a prevalent frame that have been identified in previous studies. Investigation of its use in a particular discipline and their variation in genres within the same discipline have the benefit of being directly useful for pedagogical interventions so that learners are not confronted with a bewildering array of discontinuous frames. Instead, learners can focus on a single productive frame and explore the types of fillers used to realise a range of discourse functions in the selected discipline and genre. We can identify, based on previous findings, the part of speech of the fillers in a particular frame; this can make this type of research specific and more narrowly focused, especially for second language learners who might already be burdened with the task of learning a wide range of vocabulary which might not necessarily be integral to the discourse of their community. For example, the variable slot in the frame, *the * of* is realized by a noun (e.g. *the development of, the use of* in Gray and Biber 2013; *the movement of the, the weight of the* in Nekrasova-Beker, 2019) and is often reported as the most frequent frame in Engineering (Gardner & Xu, 2019; Nekrasova-Beker, 2019; Rezoug & Vincent, 2018). To increase the perceptual salience of the frame as proposed by Byrd and Coxhead (2010) and also to explore the relationship between the first and second variable slots to express specific discourse functions, the frame can be extended to include the most frequent POS that occurs after *the * of*. A frequent class of words that occur in this slot are nouns (Gardner & Xu, 2019; Marco, 2000) sometimes preceded by a definite article, and this is usually because the first noun restricts the range of word types that can fit into this slot.

The present study which examines the p-frame, the *N of (the) N* in Case Studies and Methodology Recounts in Civil Engineering attempts to answer the following research questions:

1. What are the semantic noun categories of the first noun in the *N of (the) N* sequences and what discourse functions do they fulfil?
2. What are the most frequent first noun (N1) and second noun (N2) combinations in the two genres and what discourse functions do the N1-N2 combinations fulfil?

METHOD
The Omani corpus of academic writing-civil engineering
The corpus used in this study was the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE), comprising undergraduate assignments from two Civil Engineering programmes offered by a higher education institution in the Sultanate of Oman. This corpus was used because it is representative of Civil Engineering student assignments in this higher education institution and

also because not many studies of undergraduate writing in this discipline have been undertaken. Samples of all the discursive types of assignments which students attempted until Semester 7 were collected after obtaining students' consent. The assignments were then categorized into genres based on Nesi and Gardner's (2012) framework, assisted by interviews with subject lecturers, reading of the assignments themselves and other documentary sources such as assignment briefs and course information guides. The reading of the assignments themselves were thus triangulated with other contextual information to inform the classification process. Nesi and Gardner's (2012) taxonomy of undergraduate genres was based on the 6.5-million-word corpus of student writing in four UK universities; they identified 13 genre families in their corpus holdings. Out of the six main genres identified in OCAW-CE (Case Study; Methodology Recount; Explanation; Exercise; Site Investigation Report; Manual), the Case Study and Methodology Recount genres were identified as the most prevalent and significant based upon their distribution and frequency across the semesters as well as the views of subject lecturers. A total of 31 Case Studies comprising 105,806 words were collected while 63 Methodology Recount texts comprising 39,748 words were collected (see Table 1).

Table 1
The Case Study and Methodology Recount sub-corpora in the Omani Corpus of Academic Writing-Civil Engineering

Genre	No. of assignments	No. of words	Average no. of words	Semester
Case Study	31	105,806	3,413	1, 2, 4, 5, 6, 7
Methodology Recount	63	39,748	641	1, 2, 3, 4, 5, 7
Explanation	20	68,041	3,402	1, 2, 3, 4, 5, 7
Exercise	28	28,692	1025	5, 6, 7
Site Investigation Report	4	9,577	2,394	3
Manual	3	8,390	2,797	4
Total	149	260, 453		

The multi-word sequence *the N of (the) N* was chosen since the noun phrase followed by an *of-*phrase fragment has been identified in numerous studies (Biber et al. 1999; Cunningham 2017; Gardner & Xu 2019; Marco 2000) as a distinctive feature. Studies on discontinuous sequences or phrase frames (p-frame), frequent multi-word sequences with a variable slot, are either corpus-based as they are identified on the basis of the LBs retrieved (Biber 2009) or corpus-driven (Cunningham, 2017; Nekrasova-Beker & Becker, 2020, Römer, 2009). With the corpus-based method, however, not all the p-frames may be identified and in the present study too, we had to choose a method that would retrieve most instances of this pattern in the Case Study and Methodology Recount sub-corpora so that more accurate comparisons of the pattern in the two genres were made possible. Therefore, the strings realizing the *the N of (the) N* pattern were retrieved using the query language, CQL, from Sketch Engine. This method of retrieval is more efficient than the 'LB to frame' method (e.g. Biber 2009) which involves first retrieving LBs and then identifying patterns to study from amongst the LBs retrieved. It is thus likely that the CQL method in this study has enabled retrieval of most instances of this pattern from the two corpora.

The following query was used in *Sketch Engine* to retrieve instances of *the N of (the) N* pattern in the Case Study and Methodology Recount sub-corpora. To ignore case, the default attribute, *lc=word (lowercase)* was selected.

```
"the" [tag="N.*"] "of" [word="the"]? []{0,2} [tag="N.*"]
```

This query enabled the retrieval of the second noun even if it occurred up to two spaces after the second optional *the*. The inclusion of two optional words before the second noun was found to be optimal as instances of the pattern were captured both within the node group as well as close to it (see Table 2).

Table 2
Examples of N2 Beyond the Node Group

	Node Group	
(03) a portion of the previously mixed soil into	*the cup of the liquid limit*	***apparatus*** *at the point where the cup rests on the*
(04) back to front by using grooving tool. Rotate	*the handle of liquid limit*	*testing **device** to begin vibration and counting*

The recall rate was much higher when the optional two spaces were included in the query. For example, 2,141 instances were retrieved from the CS sub-corpus while only 1,070 lines were retrieved without the optional spaces. The precision was almost 100% although there were duplications which had to be manually eliminated as in examples (1) and (2). In concordance (01), *seawater* is considered the second noun (N2) in *the N of (the) N* pattern while in concordance (02), *desalination* is the second noun.

(01) *...system. Untraditional methods: including **the process of <u>seawater</u>** desalination and wastewater treatment to make*

(02) *...system. Untraditional methods: including **the process of seawater <u>desalination</u>** and wastewater treatment to make the water*

After the duplications were discarded, there were 1280 instances of the pattern in CS texts and 583 in the MR sub-corpus. The next step was to identify the first noun (N1) and the second noun (N2) in the concordance lines retrieved. Identifying N1 did not pose a problem as they always occurred as the second word in the node group but the N2s had to be identified carefully because there were up to two pre-modifiers before some of them. Examples (03) and (04) show N2s (bolded) occurring beyond the node group.

(03) *...a portion of the previously mixed soil into the cup of the liquid limit **apparatus** at the point where the cup rests on the...*

(04) *Rotate the handle of liquid limit testing **device** to begin vibration and counting...*

The noun types used in the *N of (the) N* pattern in the CS and MR sub-corpora were then categorized manually using Biber's (2006) semantic noun category framework. According to this framework, there are eight semantic noun categories: Abstract, Animate, Cognitive, Concrete, Group, Place, Quantity, and Technical/Concrete. Two additional categories, Gerundial Noun and Proper Noun were added based on an inductive and corpus-driven approach. Examples of nouns under the various categories in the framework from the CS and MR sub-corpora and Biber's T2K-SWAL corpus are provided in Table 3.

Table 3

Examples of Noun Categories in T2K-SWAL (Biber 2006) and the CS and MR Texts

Source	Noun categories used for classifying CS and MR texts (number of nouns in CS and MR assignments per million words)	Examples from CS and MR assignments
Biber (2006) (Main noun categories: *Abstract, Animate, Cognitive, Concrete, Group, Place, Quantity, Technical/Concrete; 17% overlap of specific nouns found in Biber's (2006) T2K-SWAL corpus with CS and MR assignments*).	1. Abstract **CS: 8071** **MR: 4000**	**CS:** *consumption, availability, plans phases,* **MR:** *results, removal*
	2. Animate **CS: 113.4** **MR: 0**	**CS:** *researchers, people, creators, designer, architect, student, mayor, residents*
	3. Cognitive **CS: 198** **MR: 377**	**CS:** *concept, idea, views, diagnosis* **MR:** *analysis, calculation,*
	4. Concrete **CS: 312** **MR: 956**	**CS:** *roof, train, dome,* **MR:** *cup, sand, water, device, hook*
	5. Group **CS: 141.7** **MR: 0**	**CS:** *institution, household, team, company, society, school, commission, family*
	6. Place **CS: 850** **MR: 1886**	**CS:** *country, region, south* **MR:** *surface, top, base, interior*
	7. Quantity **CS: 1521** **MR: 6365**	**MR:** *size, volume, ratio, thickness, height, mass, length, percentage*
	8. Technical/Concrete **CS: 226** **MR: 132**	**CS:** *bill, schedule, element, shade, radius, arc, records* **MR:** *data*
	9. Proper Nouns **CS: 434.7** **MR: 0**	**CS:** *Sultanate of Oman, the Kingdom of Bahrain, the Republic of Iraq*
	10. Gerund **CS: 132** **MR: 201**	**CS:** *painting, building, paving,* **MR:** *opening, lighting, settling*

There was only a 17% overlap between the nouns retrieved from the CS and MR sub-corpora and the T2K-SWAL corpus on the basis of which Biber's (2006) framework is based. This low

proportion of shared nouns can be attributed to differences in the composition and topic of these two corpora. The T2K-SWAL corpus contains a few engineering texts but mostly comprise spoken and written texts used in university contexts such as service encounters, textbooks and course packs. Student writing is also not a part of this corpus, unlike OCAW-CE.

The complete list of nouns found in T2K-SWAL guided us in the categorisation of the nouns in the two CE sub-corpora. Although the same nouns were often not found, some similar nouns such as *thesis* helped to categorise the noun *project* which was found in the corpus under the *Technical/Concrete* category. Sometimes, the same noun could fit under various categories. For example, the noun *solution* was sometimes used in the abstract sense and at other times, it referred to a concrete noun. While making these kinds of categorising decisions, the co-text was considered. Another challenge with adopting Biber's (2006) framework was that the categories were not mutually exclusive. For example, the nouns in the *Cognitive* and *Abstract* category often overlapped (e.g. *calculation, knowledge*). This was resolved by reading the co-text and determining the function of the word; if the function of the *Abstract* noun was to express some kind of mental process, it was categorised as *Cognitive*. Since the framework does not contain sub-categories, more fine-grained categorisation was not possible. For example, the Quantity nouns *age, century, week* were grouped with nouns like *temperature, heat* and *energy* which have standardized measures.

We decided to retain the categories provided in Biber's framework (2006) as a proliferation of categories would reduce reliability. However, two additional noun categories were added to the framework based on the concordance lines retrieved for this pattern: *Gerundial* and *Proper* Nouns. Huddleston and Pullum (2002) described 'Gerundial Nouns' as *–ing* verb forms which have the characteristics of nouns, such as functioning as subject or object in clauses or taking pre-modifiers. In Example (05), *paving* is a Gerundial Noun under this definition. These –ing forms were retrieved by *Sketch Engine* when they occurred as N1 in *the N of (the) N* pattern.

> (05) *distance: when there is a body adjacent to the* ***paving*** *of bridge pillar or a wall or shoulder supported mile or…*

The procedure described above for categorizing nouns highlights the need for manual analysis and filtering irrespective of examples from other corpora or the accuracy of automatic taggers.

Comparison of noun types across the Case Study and Methodology Recount Sub-corpora
To examine the N2 nouns and their relationship with N1, the concordance lines were manually categorized into N1-N2 sequences. The aim of this analysis was to find out if either genre preferred specific N1-N2 sequences within the pattern, and if they did, what function they realized.

The number of instances of each noun category found in N1 position in the two sub-corpora were counted and normed per 1000 words. Normalisation was necessary because of the differing word count of each text and for enabling comparisons of frequency across the two genres. The mean occurrences of each noun type was then calculated before applying the statistical z-test to find out if there was a statistically significant difference in the means of each noun category across the two sub-corpora.

There were 25 different N1-N2 combinations although not all of them were represented in both sub-corpora. Their occurrences per 1000 words were normed in order to standardize the frequency count process. Only the 10 sequences which were found in both CS and MR texts were chosen for analysis in this study and for the z-test as zeros can distort the results. The z-test was chosen over the t-test because 'the requirement for normal distributions is less important' (Blaikie, 2003, p. 194) and a sample size above 30 is considered 'large' by statistical standards. The confidence interval was set at 95%, which means that the probability of differences occurring due to chance is only 5% and below (Blaikie, 2003).

RESULTS AND DISCUSSION
The discourse of the CS and MR texts seems to be phrasally rather than clausally complex since the number of nouns is twice that of verbs. The pattern *the N of (the) N* constitutes 7% of the tokens in the MR texts and 6% of CS tokens. Considering that all instances of this pattern may not have been retrieved, the prevalence of this pattern in these two genres in CE is evident.

Noun types in CS and MR texts
Figure 1 illustrates the distribution of N1 types in the Case Study and Methodology Recount texts. It is immediately apparent that the two genres favour different semantic noun categories in this pattern. The CS texts use Abstract Nouns much more frequently while the MR texts favour Quantity, Place and Concrete nouns.

The raw frequencies of the ten noun types in the CS and MR sub-corpora were normed in each individual text. This method was followed since the word count of each text is different. This reduced the possibility of distortion of results due to the idiosyncrasies of individual writers or nature of the task types. Table 4 shows the results of the z-test that was applied to find out if there are significant differences in the mean of each noun category in the two sub-corpora.
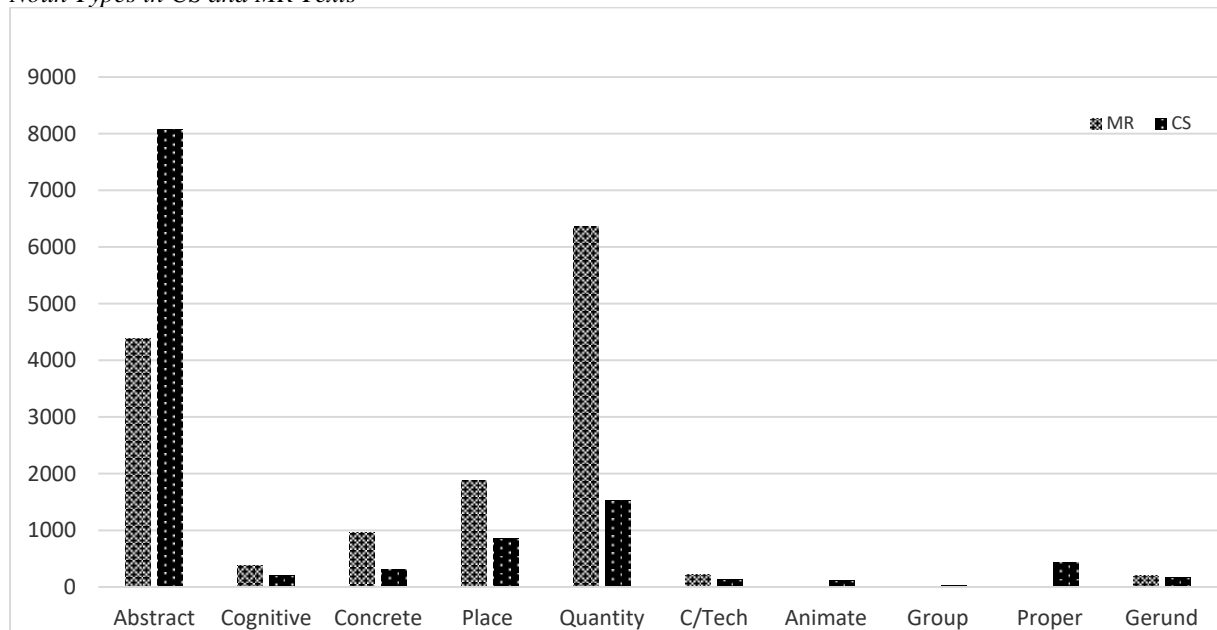
Figure 1
Noun Types in CS and MR Texts

Table 4
Results of Z-Test

S. No	N 1	Mean (MR)	Mean (CS)	sd (MR)	sd (CS)	Z Value	Sig.
1	Abstract	3.698	8.275	3.453	3.863	5.589	**
2	Cognitive	0.369	0.211	1.198	0.395	0.947	ns
3	Concrete	1.004	0.256	1.577	0.434	3.506	**
4	Place	2.020	0.811	3.601	0.702	2.568	*
5	Quantity	7.042	1.520	5.837	1.172	7.219	**
6	Concrete-Technical	0.191	0.177	0.562	0.365	0.146	ns
7	Animate	0.000	0.114	0.000	0.214	2.959	**
8	Group	0.000	0.024	0.000	0.094	1.437	ns
9	Proper	0.000	0.411	0.000	0.653	3.503	**
10	Gerund	0.183	0.166	0.621	0.342	0.180	ns

** Indicates significance at 1% level

* Indicates significance at 5% level

ns *indicates that the differences are not significant (p > 0.05)*

The normalised frequency of Abstract nouns in the CS sub-corpus (8071 pmw; e.g. *expectations, aspirations, dissemination*) is twice that of the MR texts (e.g. *design, change, delay*). There was a significant difference (p<0.05) in the means between the CS (M=3.86, SD=3.86) texts and MR (M=3.69, SD=3.45) texts; z =5.58. The Abstract nouns that are less frequent in general corpora occur with a high frequency in CS texts. For instance, the Abstract noun *sustainability* has a frequency of 52 pmw in the 15 billion- word English Web 2015 corpus while in the CS texts, the frequency is 845 wpm. Technical nouns were not frequent in both text types although there are a few exceptions in the MR texts (e.g. *elasticity, plasticity*). These findings support those of Biber (2006) who reported that more than 50% of the nouns in written university registers are Abstract. The fact that N1s belonging to the Abstract category in *the N of (the) N* pattern in

the MR and CS texts constitute 50% and 70% respectively bears closer scrutiny.

With regard to Quantity nouns, their normalised frequency is four times in the MR texts (6365 pmw) as compared to CS texts (1521 pmw). The z-score (z=7.21) reveals that there is a statistically significant difference in the means between MR (Mean=7.04, SD=5.83) and CS (M=1.52, SD=1.17) texts. Most of the Quantity nouns in the MR texts (bolded) are followed by a Concrete noun (underlined) in N2 position as can be seen in Example 06. In the CS texts, however, the Quantity nouns in N1 position are followed by other noun categories such as Place (e.g. *the size of the bedroom*), Proper nouns (e.g. *the area of Qatar*) and Abstract (e.g. *the volume of traffic*) nouns. These differences can be attributed to the different functions of these nouns in the two genre types; the CS texts make frequent references to the case (e.g. *a stretch of highway, a building*) and associated

descriptions while MR texts are limited to experiments and the concrete objects used in them.

> (06) *on it and I left the cylinder for 3 hours. Due to the **density** of the <u>sand</u> which is heavier than clay and silt, sand…*

The normalised frequency of Place nouns in the MR texts (1886 wpm) is more than double that found in CS texts (850 wpm). The difference in the means of Place nouns in the MR texts (M=2.02, SD=3.6) and the CS texts (M=0.81, SD=0.7) is statistically significant with z=2.56, p<.05. All the Quantity nouns in N1 position in the MR texts (with one exception; bolded in Examples 07 and 08) are followed by Concrete nouns (underlined) representing parts of laboratory equipment or samples used in experiments.

> (07) *the latest weight of M1, over the pit so that the **base** of the <u>cylinder</u> covers the pit concentrically.*
> (08) *density of the sand is determined by dividing the **mass** of <u>sand</u> in the calibrating cylinder with the **volume** of the <u>sand</u>…*

On the other hand, Place nouns in the CS texts primarily refer to the position of objects (e.g. *the **surface** of the <u>door jamb</u>; the **front** of the <u>tractor</u>*), to the location of cities and countries (e.g. *the South of Shanghai; the city of Mumbai*), direction (e.g. *south; northwest*) or to relatively large objects which is part of the case being studied (e.g. *base of the retaining wall*). The differences in the functions realised by these nouns in the two text types is mirrored in the type of nouns used both in N1 and N2 positions.

The results of the comparison of the normalised frequencies of Concrete nouns in the MR and CS sub-corpora are not surprising based on the trends observed for Place nouns; their frequency in MR texts (about 956 wpm) is three times that in CS texts (about 312 wpm). The Concrete nouns in N1 position in MR texts mainly refer to parts of laboratory equipment (e.g. *hook of the spring balance, nozzle of the Eureka can*) while in the CS texts, they refer to the material they are made of or large objects (e.g. *the bridges of glass, the sleepers of the railway track*). Although Biber (2006) notes that Concrete nouns are prevalent in Engineering textbooks, this noun type is less frequent in these two CE genres when compared to other noun types such as Abstract nouns.

Although only four types of Cognitive nouns are used in the MR texts (*calculation, determination, analysis, theory*) compared to the six types in the CS texts (*idea, concept, view, calculation, diagnosis, analysis*), their normalised frequency is higher in the MR texts (377 pmw vs. 198 pmw). The difference in the means of Cognitive nouns in the two sub-corpora was not statistically significant as p<.05. However, the difference in the functions realised by the Cognitive nouns in each

genre was evident. In the MR texts, Cognitive nouns in N1 position (bolded) were mainly used to refer to measurements or calculations of physical phenomena expressed by N2 (underlined) (e.g. *the **determination** of the <u>particle size distribution</u>, the **calculation** of the <u>seepage rate</u>.* In the CS texts, on the other hand, the Cognitive nouns in N1 are followed by N2s representing concepts or people (e.g. *the concept of sustainability, the views of others*). These differences can be explained by the nature of the tasks attempted in MR and CS texts.

There was no significant difference in the means of *Concrete-Technical* nouns between the MR texts (M=.19, SD=.56) and the CS texts (M=.17, SD=.36) with z=.14, p>.05 although the normalised frequency was higher in the MR texts (226 wpm) compared to the CS texts (132 wpm). The N1s (bolded) refer to technical concepts in both MR (e.g. the **axis** of the thrust) and CS texts (e.g. *the **bill** of quantities*), supporting Biber's (2006) findings that this noun category in the T2K-SWAL corpus express technical concepts which are only accessible to disciplinary insiders.

The normalised frequency of Gerundial nouns in the MR texts (201 wpm) is higher compared to that of CS texts (161 wpm). However, the z-test showed that this difference is not significant (MR texts (M=.18, SD=.62) and CS texts (M=.16, SD=.34) with z=.18, p>.05). In both genres, they refer to common processes in Civil Engineering such as *paving, painting* and *lighting* as well as general words such as *monitoring* and *opening*.

Animate, Group and Proper Nouns are not used as N1s in *the N of (the) N* pattern in MR texts. There is no statistically significant difference in the means of four nouns categories in the CS and MR texts: Cognitive, Concrete/Technical, Group and Gerund. This can be accounted for by the nature of the MR texts which are restricted to the recounting of experiments while their presence in CS texts (bolded) can be explained by focus on specific case studies and associated descriptions involving the names of places, the people undertaking the activities (Example 10) and the groups involved in them (Example 11).

> (10) *The building design still needs to be approved by the **mayor** of Wuhan, but the construction could start as soon as …*
> (11) *HCT Green-Nest is the first eco house design competition that the **Research Council of Oman** has sponsored.*

A comparison of the percentages of these various noun categories in N1 position in the two sub-corpora underline the different discourse functions realised by them. The higher proportion of Abstract nouns in the CS sub-corpus can be explained by their focus on descriptions, explanation of theories, reviewing the literature, and providing recommendations. The higher percentage of Quantity and Concrete nouns in the MR texts can be

explained by the nature of the task which entails reference to parts of equipment (e.g. *cup, shuttle, cone*) and samples used in labs (e.g. *rock, soil, sand*).

Noun-Noun sequences in *the N of (the) N* patterning in MR and CS texts
There were 25 shared Noun-Noun sequences in the MR and CS texts. The z-test applied after the counts of sequences were normalized in each text revealed that the difference in means in six of these categories are statistically significant: Abstract-Abstract, Place-Concrete, Quantity-Concrete, Abstract-Technical/Concrete, Concrete-Concrete

and Quantity-Quantity; see Table 5). Examples of such noun-noun sequences in *the N of (the) N* pattern in the MR and CS texts are provided in Table 6.

The distribution of these noun-noun sequences by percentage in the MR and CS texts are provided in Figure 2. These numbers mirror the trends in N1 frequency with sequences containing Abstract Nouns being the most frequent in CS texts and those containing Quantity, Place and Concrete nouns being favoured by MR texts. As discussed earlier, the preference for specific noun sequences by the two genres can be accounted for by the tasks they undertake.

Figure 2
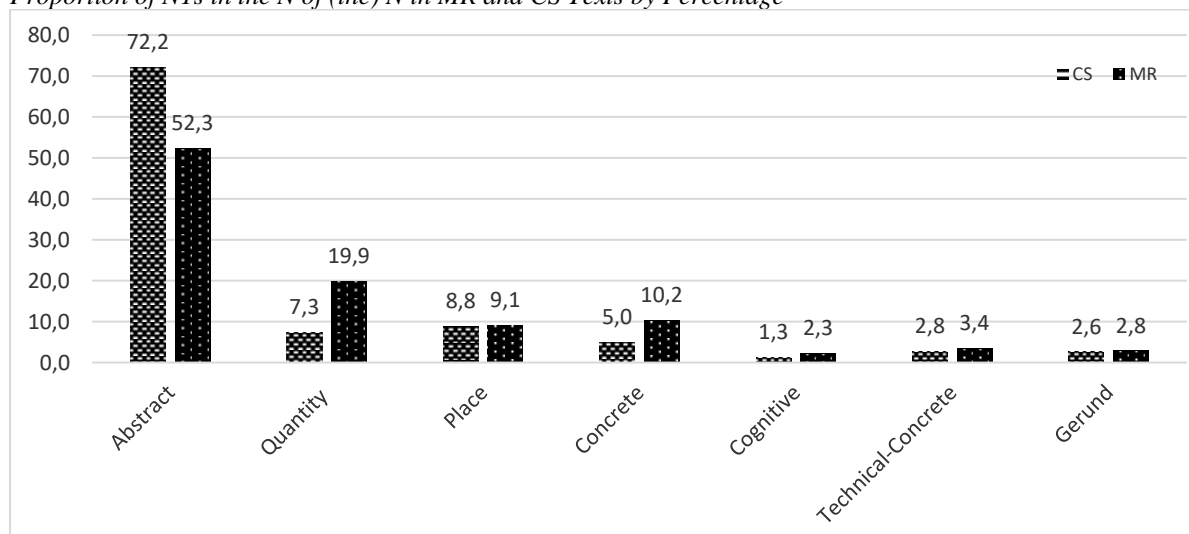*Proportion of N1s in the N of (the) N in MR and CS Texts by Percentage*

Table 5
Noun-Noun sequences in MR and CS Texts

	Mean (MR)	Mean (CS)	sd (MR)	sd (CS)	z value	significance
Abstract-Abstract	3.834	0.926	2.145	1.521	6.760	**
Place-Concrete	0.316	1.504	0.422	2.821	3.268	**
Quantity-Concrete	0.533	5.812	0.500	5.564	7.469	**
Abstract-Tech/Concrete	0.802	0.131	0.764	0.505	4.433	**
Concrete-Concrete	0.175	0.941	0.357	1.566	3.692	**
Quantity-Quantity	0.161	0.474	0.289	1.142	2.046	*

** Indicates significance at 1% level
* Indicates significance at 5% level
ns *indicates that the differences are not significant (p > 0.05)*

These findings regarding N1 noun categories and noun-noun sequences in the two genres can be interpreted in the light of a recent study by Gardner et al. (2018) on the two types of writing complexity they found in student assignments: *Information Density* characterized by Abstract nouns, nominalisations and attributive adjectives prevalent in advanced writing in the Social Sciences and *Compressed Procedural Density* which is realized by the use of Concrete and Quantity nouns, among other linguistic features, favoured by assignments in the Hard Sciences. The Case Studies in this corpus,

similar to the Engineering CS assignments in BAWE, tended to be business-oriented and less technical and therefore their preference for Abstract nouns, which are sometimes nominalizations, used to realise the first type of complexity in student writing, *Information Density*, is not surprising. The prevalence of Concrete and Quantity nouns in the MR texts found by Gardner et al. (2018) to contribute to *Compressed Procedural Density* can be attributed to the highly technical nature of these texts.

Table 6
Examples of Noun-Noun Sequences in MR and CS Texts

	MR	CS
Abstract-Abstract	*the **principles** of <u>mechanics</u>; the **consistency** of the <u>compaction</u>*	*the **objective** of the <u>project</u>; the **quality** of the <u>air</u>*
Place--Concrete	*the **base** of the <u>dam</u>; the **outside** of the <u>bottle</u>*	*the **interior** of the <u>building</u>; the **surface** of the <u>wall</u>*
Quantity-Concrete	*the **mass** of an <u>object</u>; the **volume** of <u>water</u>*	*the **width** of the <u>lane</u>; the **amount** of <u>refrigerant</u>*
Abstract-Technical/Concrete	*the **type** of the <u>graphs</u>; the **types** of <u>project</u>*	*the **life** of the <u>project</u>; the **need** for artificial <u>light</u>*
Concrete-Concrete	*the **cone** of the <u>cylinder</u>; the **hook** of the spring <u>balance</u>*	*the **sleepers** of railway <u>track</u>; the **roof** of Malaysian <u>houses</u>*
Quantity-Quantity	*the **percentage** of the total <u>weight</u>; the **ratio** of the total <u>mass</u>*	*the **period** of <u>time</u>; the **ratio** of the <u>weight</u>*

CONCLUSION

In order to address the research questions of the study, all instances of *the N of (the) N* pattern were retrieved from the Case Study and Methodology Recount sub-corpora in the Omani Corpus of Academic Writing (OCAW) in Civil Engineering. The semantic categories of the nouns as well as the N1-N2 semantic noun category combinations in *the N of (the) N* pattern were examined in order to find out the similarities and differences between the two genres in the use of this pattern. Additionally, each N1-N2 noun category sequence was mapped with its discourse function.

Significant differences were found in the N1s preferred by the two text types. CS texts preferred Abstract nouns while MR texts favoured Quantity, Place and Concrete nouns. Statistically significant differences were also detected in the N1-N2 combinations between the two genres. Frames with Abstract nouns in N1 or N2 positions (Abstract-Abstract and Abstract-Technical/Concrete) were favoured by the CS texts, reflecting the tasks students attempt in this genre such as the definition of concepts and review of the literature. MR texts, on the other hand, prefer N1-N2 sequences that include Quantity, Concrete, and Place nouns (Quantity-Quantity, Concrete-Concrete, Place-Concrete, Quantity-Concrete), which are used to measure and quantify objects or samples used in experiments. The pedagogical implications of these findings are that Omani students can be made conscious of the semantic noun categories used in this pattern and how an N1-N2 sequence fulfils a particular communicative function in a genre. These findings will also have relevance for students in similar *lingua franca* contexts after consideration of situational variables such as the first language of students, requirements of academic departments, and institutional and national policies on assessment and curricula.

REFERENCES

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In P. Cowie (Ed.), *A phraseology: Theory analysis and applications* (pp. 101-122). Oxford University Press.

Artemeva, N. (2009). Stories of becoming: A study of novice engineers learning genres of their profession'. In C. Bazerman, A. Bonini, & D. Figueiredo (Ed.), *Genre in a Changing world. perspectives on writing* (pp. 158-178). The WAC Clearinghouse and Parlor Press.

Artemeva, N., & Freedman, A. (2008). *Rhetorical genre studies and beyond*. Inkshed Publications.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics, 14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written register. *English for Specific Purposes, 26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finnegan, E. (1999). *The Longman grammar of spoken and written English*. Longman.

Blaikie, N. (2003). *Analyzing quantitative data: From description to explanation*. Sage.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, *1*, 1–14.

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*(5), 31-64.

Cheng, A. (2007). Transferring generic features and re-contextualizing genre awareness: Understanding writing performance in the ESP genre-based literacy framework. *English for Specific Purposes, 26*, 287–307. https://doi.org/10.1016/j.esp.2006.12.002

Conrad, S. (2014). Expanding multi-dimensional analysis with qualitative research techniques. *Multi-dimensional Analysis, 25*, 273-95. https://doi.org/10.1075/scl.60.09con

Conrad, S. (2017). A comparison of practitioner and student writing in civil engineering. *Journal of Engineering Education, 106*(2), 191-217. https://doi.org/10.1002/jee.20161

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica, 20*, 56-71. https://doi.org/10.1515/9783484604674.56

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30-49. http://dx.doi.org/10125/44213.

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes, 25*, 71-83. https://doi.org/10.1016/j.jeap.2016.11.005

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics 3*(1), 59–80. https://doi.org/10.1075/ijcl.3.1.04dec

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics, 35*(3), 328-356. https://doi.org/10.1093/applin/amt016

Ebeling, S. O., & Hasselgård, H. (2015). Learners' and native speakers' use of recurrent word-combinations across disciplines. *Bergen Language and Linguistics Studies, 6*, 87-106. https://doi.org/10.15845/bells.v6i0

Firth, J. R. (1957). *Papers in linguistics*. Oxford University Press.

Fletcher, W. H. (2012). *kfNgram*. Information & Help. Retrieved from https://www.kwicfinder.com/kfNgram/kfNgramHelp.html/

Gardner, S. (2008). Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description: An illustration through university history and engineering assignments. In E. H. Steiner & N. Neumann, S. Saarbrücken (Eds.), *ESFLCW 2007: Data and interpretation in linguistic analysis* (pp. 1-34). Universität des Saarlandes.

Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics, 10*, 1-30. https://doi.org/10.1093/applin/amy005

Gardner, S., & Xu, X. (2019). Engineering registers in the 21st century: SFL perspectives on online publications. *Language, Context and Text, 1*(1), 65-101. https://doi.org/10.1075/langct.00004.gar

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, *51*, 1-17. https://doi.org/10.1016/j.esp.2018.02.002

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Grabowski, L. (2013). Register variation across English pharmaceutical texts: A corpus-driven study of keywords, lexical bundles and phrase frames in patient information leaflets and summaries of product characteristics. *Procedia-Social and Behavioral Sciences*, *95*, 391-401. https://doi.org/10.1016/j.sbspro.2013.10.661

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1), 109-136. https://doi.org/10.1075/ijcl.18.1.08gra

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora, 8*(2), 183–207. https://doi.org/10.3366/cor.2013.0040

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge. https://doi.org/10.4324/9780203327630

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes, 33*, 54-65. https://doi.org/10.1016/j.esp.2013.07.001

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

Hüttner, J. (2008). The genre(s) of student writing: Developing writing models. *International Journal of Applied Linguistics, 18*(2), 146-165. https://doi.org/10.1111/j.1473-4192.2008.00200.x

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Huddleston, R., & Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes, 37*, 74–86. https://doi.org/10.1016/j.esp.2014.06.008

Leedham, M. E. (2011). *A corpus-driven study of features of Chinese students' undergraduate writing in UK universities* [online PhD thesis]. The Open University. http://oro.open.ac.uk/29228/1297/Leedham_Thesis.pdf

Leedham, M., & Fernández-Parra, M. (2017). Recounting and reflecting: The use of first person pronouns in Chinese, Greek and British students' assignments in engineering. *Journal of English for Academic Purposes, 26*, 66-77. https://doi.org/10.1016/j.jeap.2017.02.001

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, *31*(1), 25-35. https://doi.org/10.1016/j.esp.2011.07.002

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

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes, 25*(2), 235-256. https://doi.org.10.1016/j.esp.2005.05.002

Nekrasova-Beker, T. M. (2019). Discipline-specific use of language patterns in engineering: A comparison of published pedagogical materials. *Journal of English for Academic Purposes*, *41*, 1-12. https://doi.org/10.1016/j.jeap.2019.100774

Nekrasova-Beker, T., & Becker, A. (2020). The use of lexical patterns in engineering. *Advances in Corpus-based Research on Academic Writing: Effects of discipline, register, and writer expertise*, *95*, 228-254.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge University Press.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes, 21*, 60-71. https://doi.org/10.1016/j.jeap.2015.11.003

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory. In J. Richards & R. Schmidt (Ed.), *Language and communication* (pp. 191-226). Longman.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Ed.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 128-143). Longman.

Rezoug, F., & Vincent, B. (2018). Exploring lexical bundles in the Algerian Corpus of Engineering. *Arab Journal of Applied Linguistics*, *3* (1), 47-77.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, *7*(1), 140-162. https://doi.org/10.1075/arcl.7.06rom

Sinclair, J. M. (1991). *Corpus, concordance, collocation.* Oxford University Press.

Staples, S., Egbert, J., Biber, D., & Bethany, G. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149-183. https://doi.org/10.1177/0741088316631527

Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, *15*, 1-13. https://doi.org/10.1016/j.jeap.2014.03.002

Ward, J. (2009). A basic engineering word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*, 170–182. https://doi.org/10.1016/j.esp.2009.04.001

Wray, A. (2006). Formulaic language. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 590–597). Elsevier.

Wray, A., & Perkins, M. (2002). The functions of formulaic language: An integrated model. *Language and Communication*, *20*(1), 1-28. https://doi.org/10.1016/S0271-5309(99)00015-4