English for Specific Purposes 58 (2020) 155–169

Contents lists available at ScienceDirect

English for Specific Purposes

journal homepage: http://ees.elsevier.com/esp/default.asp

The creation and evaluation of a grammar pattern list for the most frequent academic verbs

Hong Ma[a,*], Manman Qian[b]

[a] Department of Lingistics, Zhejiang University, Hangzhou 310058, Zhejiang Province, China
[b] Program of Applied Linguistics and Technology, Iowa State University, Ames, IA, USA

ARTICLE INFO

Article history:
Available online 10 March 2020

Keywords:
Frequency list
Computational linguistics
Grammar patterns
Academic discourse

1. Introduction

The frequency of academic words is essential to vocabulary pedagogy in English for academic purposes, because it informs instructional decisions concerning which words are to be prioritized given limited in-class time and independent study time (Coxhead, 2000). Driven by the pedagogical needs of accessing words frequently used in academic writing, many corpus linguists have developed academic word lists by adopting different methods and procedures. The most widely cited lists include the Academic Word List (AWL) (Coxhead, 2000) and the Academic Vocabulary List (AVL) (Gardner & Davies, 2014). In constructing the AWL, words were selected based on word families, a family being defined as “a stem (headword) plus all closely related affixed forms” (Coxhead, 2000, p. 216), because learners are likely to comprehend the affixed forms if they have knowledge of the stem (Bauer & Nation, 1993). Gardner and Davies, 2014, nevertheless, found this word selection criterion problematic for not taking grammatical parts of speech into consideration. According to Gardner and Davies, 2014, one word form is likely to be used as different grammatical parts of speech, and grammatical parts of speech are associated with meaning differences. For example, the word form “report” means “an official document” or “telling people something

* Corresponding author.
E-mail addresses: 0016183@zju.edu.cn (H. Ma), mqian@iastate.edu (M. Qian).

https://doi.org/10.1016/j.esp.2020.01.002

happened" when used as a noun or a verb respectively. Gardner and Davies, 2014 thereafter created the AVL by incorporating grammatical parts of speech and treated the same word form with different grammatical parts of speech as separate entities.

Since researchers in vocabulary acquisition and corpus linguistics recognized formulaic expressions as constituting an indispensable dimension of vocabulary knowledge (Milton, 2009; Read, 2000), lists focusing on different aspects of formulaic expressions in academic texts have been compiled to complement the aforementioned classic vocabulary lists (AWL and AVL). The Academic Formulas List (Simpson-Vlach & Ellis, 2010) and the Academic Collocation List (Ackermann & Chen, 2013) are two typical examples of lists of formulaic expressions that appear frequently in academic writing.

Following this trend of developing frequency lists of formulaic expressions, the current work proposes a method for developing a frequency list of grammar patterns, a type of formulaic expression that has not yet received sufficient attention from the field of language research and pedagogy in SLA (Römer, O'Donnell, & Ellis, 2014). In this project, we narrowed our scope to extracting grammar patterns of the 115 verbs that ranked among the 500 most frequent academic words in the AVL. Despite the narrowed focus, the extraction methodology adopted in the project can be replicated and used to extract grammar patterns for content words (including nouns, verbs, adjectives and adverbs) in any given corpus.

The ground-breaking work presented in the Collins COBUILD English dictionary (1995), which encompasses the grammar patterns extracted from the Bank of English by Hunston, Francis, Manning, and Mason, served as an important resource for the development of the current study's verb pattern list. This important publication, *Collins COBUILD English dictionary,* focusing primarily on the generic patterns (e.g. be V-ed about, V n -ing, and V n -ed), provided no information concerning the frequency ranking of the patterns of specific words (e.g. be concerned about, increase from, and develop out of) (See the *Collins Cobuild English dictionary* for detailed explanations of the symbols used for representing grammar patterns). In this project, we took the existing resource a step further and generated frequency information for the grammar patterns of specific words. We believe that the complementary frequency information will faciliate more efficient use of grammar patterns in language pedagogy, especially the specific audience of EAP language learners.

In this study, we also propose a method for extracting grammar patterns published in the *Collins COBUILD English dictionary* using a rule-based computational approach. Our methodology is replicable for similar future endeavors and is likely to add to the body of literature on methods for word list development.

1.1. *Centrality of formulaic expressions and academic vocabulary*

Linguists have come to understand that language consists of not only individual words, but also formulaic expressions as semantic units (Jaen, 2007; Martinez & Schmitt, 2012; Qian & Schedl, 2004). The centrality of formulaic expressions for meaning-making was established by the articulation of the idiom principle, which states that "a language user has available to him or her a larger number of semi-preconstructed phrases that constitute single choices, even though they might appear to the analyzable into segments" (Sinclair, 1991, p.110). The establishment of the centrality of formulaic expressions has also informed and enriched the instruction of academic vocabulary. A range of empirical strategies has been utilized to develop frequency lists for different types of formulaic expressions in academic writing (Ackermann & Chen, 2013; Simpson-Vlach & Ellis, 2010).

The procedure adopted in developing the Academic Collocation List (Ackermann & Chen, 2013) inspired the current research. Ackermann and Chen (2013) began with extracting a list of content words that occur more than five times per million words in the target corpus. A collocation program was then employed to extract collocations of the identified content words. Similarly in this study, we first identified the frequent content words (only verbs in our study) in a target corpus and then extracted grammar patterns for the identified content words using self-developed scripts.

1.2. *The need for a frequency list of grammar patterns*

Despite the centrality of formulaic expressions identified in the English language, the number of frequency lists for formulaic expressions falls short. For instance, even though grammar patterns are important to language pedagogy, a frequency list for grammar pattens is still lacking.

Grammar patterns are defined as "all the words and structures which are regularly associated with the word and which contribute to its meaning" (Hunston & Francis, 1999, p.37). Specific examples of grammar patterns include: 1) "Appear" and "manage" can be followed by a to-infinitive but not a present participle; 2) "Finish" and "suggest" can be followed by a present participle but not a to-infinitive; 3) The meanings of the verbs such as "forget", "remember", "stop", and "try" vary depending on whether they are followed by to-infinitives or present participles (Hunston, Francis, & Manning, 1997). Grammar patterns are important to language pedagogy, since they permeate each individual sentence (Hunston & Francis, 1998; Römer, O'Donnell, & Ellis, 2014) and are fundamental to language learners' accuracy and fluency (Hunston et al., 1997). Novice or nonnative writers' use of grammar patterns, however, tends to differ greatly from expert or native writers' pattern use (Römer, 2009; Römer & Berger, 2019; Römer, Skalicky, & Ellis, 2017). It is not uncommon for nonnative students to produce sentences with incorrect grammar patterns such as "Bear in mind that the nonsmokers *consist about* 65% of the Polish society." (This sentence was extracted from the International Corpus of Learner English).

Despite the importance of grammar patterns and language learners' difficulties in using grammar patterns, grammar patterns have not received sufficient attention in language pedagogy (Römer et al., 2014). To master a new language, language learners need to develop sensitivity to experts' use of formulaic expressions (Hyland, 2008). Considering the variability of language use across different registers (Hyland, 2008), this study adds to the frequency lists of formulaic expressions in academic texts by creating a grammar pattern frequency list for the most frequent 115 academic verbs in the AVL. Specific to students learning English for academic purposes (EAP), knowledge of pattern use in academic discourse undergirds their efficiency in comprehension and production when they are involved in real life tasks including attending seminars, reading articles, giving presentations, writing research papers, and engaging in other academic communication-based activities. To promote and guide the practical teaching of grammar patterns appearing frequently in academic texts, a grammar pattern list for the most frequent academic words is indispensable in providing useful information as to which grammar patterns should be prioritized for teaching and learning.

2. Materials

This section describes the materials used in the project, including 1) the corpus from which the frequent academic verb patterns were extracted, and 2) the previously published pattern list, on which the extraction of verb patterns was based. Additionally, a corpus the researchers compiled to evaluate the accuracy of the pattern extraction system is described.

2.1. Corpus used for pattern extraction

The 2.4 million-word corpus used for extracting patterns contains textbooks and published academic journal articles selected from five major academic disciplines (health science, life science, physical science, social science and engineering). The rationale of including both registers is that textbooks serve as the major source of academic language for university students (Biber, Conrad, & Cortes, 2004; Lei & Liu, 2016; Ward, 2007), while published journal articles represent the discourse community that university students are apprenticed to (Durrant, 2016; Gardner & Davies, 2014; Liu, 2012). Table 1 presents the composition of the current corpus.

The size of this corpus, even though slightly smaller than those adopted by previous researchers in corpus linguistis (e.g. Durrant, 2016; Hyland, 2008; Lei & Liu, 2016; Simpson-Vlach & Ellis, 2010), passed the threshold of a million tokens that Sinclair (1991) postulated as sufficient to yield useful in-depth linguistic information.

2.2. Previously presented grammar patterns

The grammar patterns of English verbs were derived from the Collins Cobuild English dictionary, where the patterns are presented in a word-oriented manner, with each verb linked to all the patterns the verb can be used in. The grammar patterns in this publication were extracted from the Bank of English through a corpus-driven approach, a way of investigating language by generating hypotheses based on observations of a large, principled collection of naturally occurring texts (written or spoken) stored electronically (Hunston & Francis, 1999). The Bank of English is considered a sufficient representation of the entire English language, as this corpus contains over 250 million words collected from various sources in English speaking countries such as newspapers, magazines, books, and daily conversations. However, the frequency of these grammar patterns was not published, leaving English teachers and material designers not sufficiently supported in terms of which patterns to prioritize in practical teaching and material development. Our research intended to address such deficiencies by deriving a frequency list for grammar patterns of verbs frequently appearing in the academic register.

2.3. Corpus for evaluation

To evaluate the performance of the pattern extraction system, we compiled a corpus of manageable size to test the accuracy of the system. The test corpus consisted of 10 research articles in applied linguistics totaling 77,085 words. Linguistics articles were selected because their academic genre resembled the target database used for extracting word lists and the genre is familiar to the human raters, extensively trained applied linguists, whose manual anlaysis was compared with the output of the pattern extraction program. Choosing a genre familiar to the raters further provided potential to ensure the accuracy and efficiency of their manual analysis.

Table 1
Composition of the Corpus.

Academic disciplines	Textbooks # of words	Published journal articles # of words	Total # of words
Health science	357478	109683	467161
Life science	260287	165534	425821
Physical science	286071	267935	554006
Social science	263698	255896	519594
Engineering	276245	221414	497659
Total # of words	1443779	1020462	2464241

3. Procedures

3.1. Identification of patterns of frequent verbs

To identify the most frequent academic verb patterns, a pattern extraction system was developed by drawing upon the new AVL, *Collins Cobuild English dictionary,* and integrating tenets of natural language processing (NLP). The developmental process involved three key steps. First, the 115 most frequent academic verbs were obtained from the top 500 words/lemmas, "words with a common stem, related by inflection only, and coming from the same part of speech", in the AVL (Gardner & Davies, 2014). Then, all possible patterns of each selected verb were collected by manually searching each entry for the verb in question in the *Collins COBUILD English dictionary*. Finally, the NLP approach was adopted to extract all the identified patterns for the 115 most frequent academic verbs from our corpus. Because in this study we were interested in the patterns of specific verbs (e.g. be concerned about, increase from, and develop out of) rather than the generic patterns (e.g. be V-ed about, V n -ing, and V n -ed), we focused the extraction solely on those grammar patterns containing the target verbs. We hereafter term the patterns of specific verbs as "verb-pattern combinations".

The NLP-based pattern extraction system consisted of two sequential modules: activation of Stanford CoreNLP using the Analyzer originally used for developing a web-based automated writing evaluation tool, CyWrite (Chukharev-Hudilainen & Saricaoglu, 2014) and rule-based pattern extraction using *Prolog,* which is a logical programming language commonly adopted for textual data analysis (Bramer, 2013). Stanford CoreNLP, a language parser, automatically (1) splits input text into sentences and tokens, (2) labels each token with part-of-speech (POS) tags (shown in Figure 1), (3) parses each sentence, outputs syntax trees (shown in Figure 1), (4) identifies Stanford Typed Dependencies (shown in Figure 1), (5) produces a hierarchical and textual description of the syntactic structure of each sentence (shown in Figure 2) and (6) saves all the textual descriptions in an Extensible Markup Language (XML) file readily analyzable for the rule-based extraction module (Chukharev-Hudilainen & Saricaoglu, 2014; De Marneffe, MacCarney & Manning, 2006). Figure 1 illustrates the syntax tree of the example sentence "Native students are provided with mentors." (The target verb-pattern combination is "provide n with n.") This tree also includes POS tags above the tokens and Stanford Typed Dependencies below the tokens.

The syntactic structure of the example sentence is also produced in the form of a textual description (shown in Figure 2.) and saved in an XML document, which is analyzed afterwards in the rule-based extration module.

The rule-based extraction module was programed using *Prolog*. To extract the patterns, the *Prolog* code was first written to specify the structure for each pattern by describing the syntax trees produced in the Cywrite Analyzer module incorporating both POS tags and Stanford Typed Dependencies. One of the strings of the *Prolog* code used to extract the verb-pattern combination "provide n with n" (as an example) is presented in Figure 3.

When the patterns described using Prolog matched the syntactic structures of sentences parsed in the CyWrite Analyzer module, these sentences were extracted with the specific patterns labeled. As shown in the example output (Figure 4), the

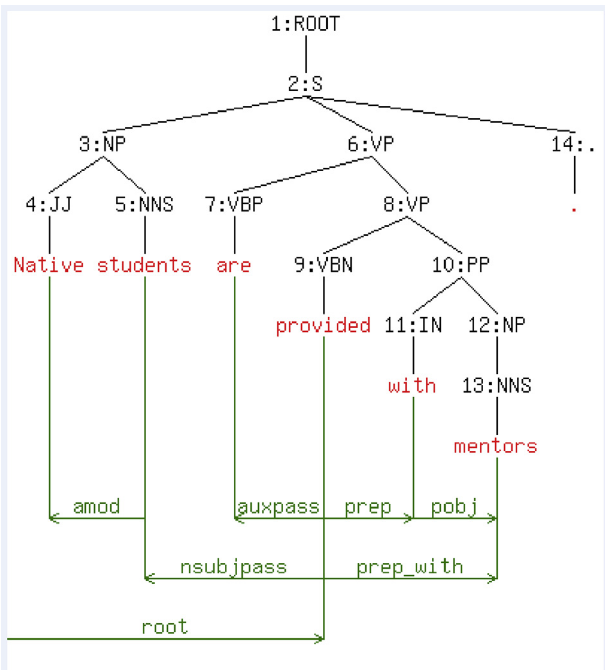

Figure 1. The syntactic tree of the example sentence.

```
<parse>(ROOT
  (S
  (NP (JJ Native) (NNS students))
  (VP (VBP are)
  (VP (VBN provided)
  (PP (IN with)
  (NP (NNS mentors)))))
  (. .)))
```

Figure 2. The textual description of the syntactic tree.

```
'bevedwith(X) :-
type(verb,X),dep(auxpass,X,_),token(with,W),precedes(X,W),
(dep(prep_with,X,_);dep(prepc_with,X,_);immed_precedes(X,W)).',
```

Figure 3. The prolog code for the passive form of "V n with n".

pattern "provide n with n" was extracted from the example sentence "Native students are provided with mentors" with "provide" as the target verb. All patterns were numbered for convenient tracking. For example, the pattern "V n with n" was numbered as "166_1" in our list. "Feature Counts" indicates the total number of the target patterns extracted from a sentence. Here for pattern "V n with n", "Feature Counts" equals 1, indicating that one target verb-pattern combination "provide n with n" was extracted from the example sentence.

3.2. Evaluation of the pattern extraction system

This section elaborates on our evaluation procedure in which output by the extraction system and by human raters were compared to calculate the accuracy performance of the pattern extraction system.

3.2.1. Obtaining output from the extraction system

The current pattern extraction system was applied to the test corpus to extract all coded verb-pattern combinations in text. The sentences containing the extracted verb-pattern combinations were subsequently retrieved from the test corpus and compared with the results from human analyses.

3.2.2. Obtaining output from human raters

Several methodological steps were taken to obtain the analysis by human raters. First, the self-compiled corpus was imported into Antconc, a concordance software program. After a search of the target verb lemmas in Antconc, all the sentences containing the target lemmas were then automatically extracted and listed as concordance lines in Antconc. As shown in Figure 5, a search of the target lemma "provide" yielded all the sentences that contained this lemma. Using Antconc to compile sentences with the target verb lemmas increased working efficiency by drawing the raters' attention exclusively to the sentences containing the target lemma, so that the raters did not have to read every single sentence in the corpus to identify the target grammar patterns.

The two human raters then manually analyzed each concordance line and independently tagged the patterns of the target lemma. The interrater reliability was 96.3%. Results from the two raters' individual analyses were then compared with differences negotiatiated to establish an agreed version. Next, the patterns identified by human raters were compared against the output from the extraction system. The next section explains how the data were analyzed to evaluate the performance of the pattern extraction system.

4. Data analysis

Precision and recall, measures widely used to evaluate NLP systems (Melamed, Green, & Turian, 2003), were used to assess the performance of the current pattern extraction system. When comparing a set of "candidate" items Y to a set of "reference" items X, precision equals the intersection of the two sets divided by the candidate set $\left(\frac{|X \cap Y|}{|Y|}\right)$. While recall equals the intersection of the two sets divided by the reference set $\left(\frac{|X \cap Y|}{|X|}\right)$ (Melamed et al., 2003). Specific to this project, candidate items are items that were retrieved by the pattern extraction system, and items identified by human raters are the reference items. The intersection refers to items simultaneously identified by both the pattern extraction system and human raters. Precision reflects the percentage of extracted patterns that are accurate, while recall indicates the percentage of manually identified patterns that are successfully extracted by using the rule-based approach.

```
Native students are provided with mentors

[pattern166_1]
        X = provided (node9)
[Feature Counts]
        pattern166_1: 1
```

Figure 4. The extraction of the target verb-pattern combination.

5. Results & discussion

The results of this project can be broadly grouped into two parts: 1) precision and recall of the pattern extraction system; and 2) a pattern frequency list extracted for the 115 most frequent academic verbs.

5.1. Precision and recall

The precision of the pattern extraction system was 95.2%, and the recall was 82.0%. These results suggest that 95.2% of the retrieved verb-pattern combinations were accurate, and the majority (82.0%) of the verb-pattern combinations existing in the evaluation corpus were successfully recognized and retrieved by the pattern extraction system.

5.2. The pattern list

All the verb-pattern combinations for the 115 most frequent verbs in the AVL list were extracted. A Perl script was written to sort the verb-patterns combinations based on frequency. Appendix A and Appendix B present the pattern list of the 115 most frequent academic verbs in a verb-oriented order and a pattern-oriented order respectively. Since the two different versions of the list are too long to present in the main body of this article, we present a small part of both versions to illustrate the possible use of the list. As exemplified in Table 2, showing part of the pattern list for the 115 most frequent academic verbs in a verb-oriented order, grammar patterns for each academic verb are ranked based on their frequency of occurrences in a descending order. To associate our list to the AVL, we rank the verbs based on their frequency order in the AVL.

This list of verb patterns is pedagogically useful in several ways. First, provided with the list, English language teachers and material designers would be able to make more informed decisions as to which patterns of a specific academic verb should be introduced earlier than others. To date, implicit learning realized through data-driven learning has been the most widely adopted method for teaching formulaic expressions (Charles, 2007; Yoon & Hirvela, 2004). However, inductive learning of an aspect of formulaic expressions can be problematic (Vannestal & Lindquist, 2007), since it is difficult for students to extrapolate the tendencies of language use given conflicting examples the students may encounter (Flowerdew, 2009; Hunston & Francis, 2000) and the daunting number of concordance lines required to perform the extrapolation (Coxhead, 2008). Guided by the current list, explicit teaching of academic verb patterns can be more conveniently and systematically enacted. As shown in Table 2, the verb “report” can be used with nine patterns in an academic register. EAP students are not likely to identify and learn all nine patterns independently, even though they are expected to use academic verbs with flexibility. Hence, drawing upon this list, teachers or material designers can manipulate students’ exposure to patterns of “report” by introducing its more frequent patterns first and then its less frequent patterns. With a clear goal of determining which specific patterns of an acadmic verb to search and explore, students can also work more efficiently to retrieve authentic sentences with the target patterns of a specific verb and learn practical uses of the lexis.

Table 3 presents part of the pattern list for the 115 most frequent academic verbs in a pattern-oriented order. Verbs frequently used with each grammar pattern are listed based on frequency of the specific verb-pattern combinations in a descending order.

The frequency list in a pattern-oriented order carries pedagogical value and enables English language teachers to prioritize verb-pattern combinations of higher frequency in their teaching process, since each generic pattern can be associated with numerous English verbs. Access to this current list of grammar patterns could also facilitate language learners’ independent study of grammar patterns in an order that takes the patterns’ frequency of occurrence and use into consideration. Take the pattern “V to n” for example. Eleven verbs among the most frequent 115 academic verbs can be used with the pattern “V to n”. It is, therefore, more efficient for students to learn more frequently used verb-pattern combinations, such as “contribute to n”, “relate to n”, and “refer to n”, and then extend their efforts to less frequently used combinations.

Consistent with previous research that adopted usage-based approach to the study of grammar patterns, this current project also identified Zipfian distribution of grammar patterns, with only a few verb-pattern combinations accounting for the lion’s share of each pattern (Goldberg, Casenhiser, & Sethuraman, 2004; Römer et al., 2014). Usage-based studies have emphasized the importance of frequency in learning grammar patterns, and showed that Zipfian distributional properties help make grammar patterns learnable (Römer et al., 2014), since the highly frequent verb-pattern combinations serve as prototypes with which other semantically related verb-pattern combinations can be associated (Goldberg et al., 2004).

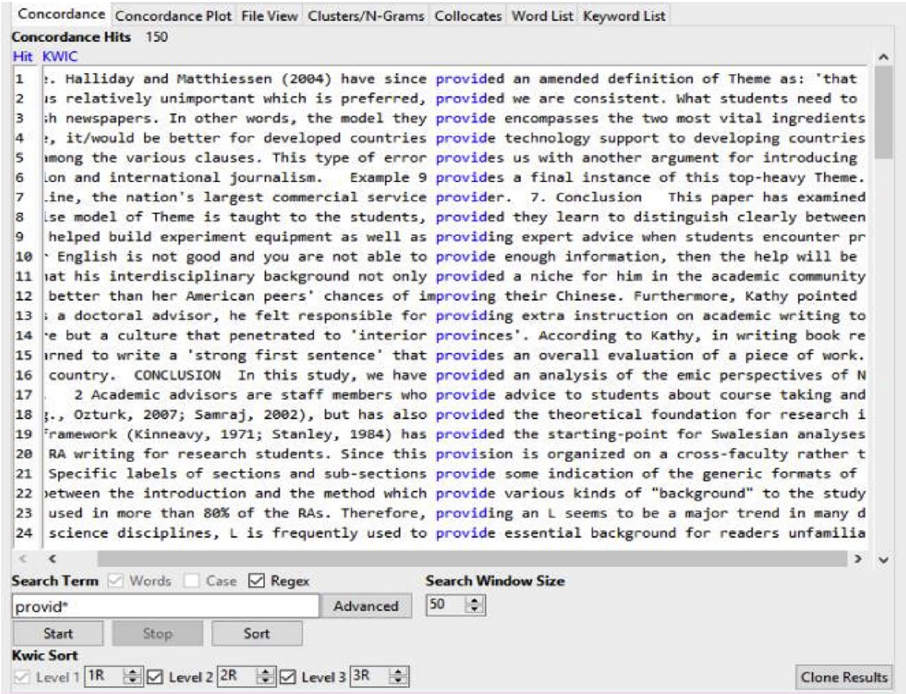

Figure 5. The concordance lines with the target verb "provide".

Table 2
Part of the Pattern List in a Verb-oriented Order.

Verbs	Patterns	Frequency
require	V n	996 (including 385 passive forms)
	V n to-inf	379 (including 225 passive forms)
	V that	88
	be Ved of	5
report	V n	341 (including 253 passive forms)
	V n to n	175 (including 159 passive forms)
	be Ved to-inf	141
	V that	50
	be Ved as	36 (including 7 active forms)
	V to n	32
	be Ved for	24
	It be Ved that	19
	V on n	10

Table 3
Part of the Pattern List in a Pattern-oriented Order.

V	occur (1526), increase (1259), exist (808), form (233), arise (195), differ (183), develop (170), emerge (158), indicate (134), reduce (101), perform (96), reflect (84), vary (81), improve (65), interpret (60), permit (43), lack (40), yield (23), conduct (18), publish (17), organize (16), locate (11), engage (10), result (4)	5335
V that	suggest (619), indicate (314), note (279), assume (278), demonstrate (226), ensure (160), argue (151), state (139), imply (118), conclude (112), reveal (92), require (88), observe (75), recognize (69), propose (57), predict (52), report (50), provide (36), determine (24), maintain (21), establish (21), illustrate (17), estimate (15), emphasize (13), emerge (12), reflect (4)	3042
V on n	depend (719), focus (347), rely (203), reflect (35), exist (14), report (10), improve (2)	1330
V to n	refer (397), contribute (234), relate (233), apply (219), extend (67), increase (34), occur (32), report (32), tend (18), yield (5), demonstrate (5)	1276

Note. The numbers in parentheses refer to the frequency of the verb-pattern combinations.

The idea of linking semantically associated verb-pattern combinations is also reflected in a pattern approach to grammar developed through a corpus-driven approach (Hunston & Francis, 1998). At the core of the pattern approach to grammar is the observation that words used with the same pattern tend to share similar meaning (see Sinclair, 1996 for details of the pattern approach to grammar). Researchers in favor of the pattern approach have speculated that this approach could promote students' efficiency of learning patterns by establishing the connection between individual word behaviors based on meaning, while they have long realized that “one of the questions that arises from considering the role of pattern in text is how frequent certain patterns are” (Hunston & Francis, 1998, p. 67). Lacking information on frequency of specific verb-pattern combinations, the published verb pattern teaching material, *Grammar Pattern 1: Verbs,* (Sinclair, 1996) ranked the base verbs alphabetically. For example, verbs with the pattern “V that” and sharing the meaning of “say” are listed alphabetically as follows: “accept, acknowledge, admit, advise, advocate, affirm, agree, allege, allow, announce, argue, ask...” (Sinclair, 1996, p. 98). Teaching patterns following an alphabetical order seems not to be pedagogically sound. The current frequency list of verb patterns will contribute to a more efficient pattern approach to grammar by narrowing teaching scope to the most frequent verb patterns and grouping the patterns based on meaning.

6. Conclusion

The present study successfully extracted a pattern list for the 115 most frequent verbs in academic English discourse using a self-built pattern extraction system with the help of Stanford CoreNLP and Cywrite Analyzer (Chukharev-Hudilainen & Saricaoglu, 2014). The development of this list contributes to the ongoing call for efforts to create frequency lists for formulaic expressions in academic discourse. The accuracy of the pattern extraction system measured through precision and recall is high, undergirding the usability of this modified list. However, it is worth bearing in mind that the results, obtained from the test corpus of only 77,085 words, do not confirm the total accuracy of the pattern extraction system when applied to the 2.4 million-word target corpus. It also should be noted that the list covers grammar patterns of only the 115 most frequent academic verbs. To benefit language learners in a wider context, the current list needs to be expanded to cover patterns of more content words (verb, nouns, adjectives and adverbs). Accordingly, the extraction method proposed in this study can be replicated to generate such lists by plugging in grammar patterns of frequent content words in the AVL.

The development of the current pattern extraction system can potentially advance studies on grammar patterns to a greater degree. Current work on grammar patterns compares native/novice writers’ and nonnative/expert writers’ use of the patterns, explores the differences in pattern use across different displines, and investigates the grammatical contexts of pattern use. However, these studies have focused on a limited number of grammar patterns, such as *that-clause* and *wh-clause* (Hunston, 2003), *it v-link ADJ that* and *it v-link ADJ to-inf* (Groom, 2005), nouns followed by a complement clause (Charles, 2007), and small sets of verb patterns, mostly “V preposition n” types (Römer & Berger, 2019; Römer et al., 2014, 2017). By replicating the current pattern extraction approach and applying it to corpora of varied registers and genres, studies on grammar patterns are likely to uncover practical uses of all grammar patterns.

Additionally, the current research may inspire other methods of extracting grammar patterns, which could be used to further improve the pattern extraction process and contribute to a more accurate list. Relying on Collins Cobuild English dictionary (1995) to specify possible grammar patterns, this rule-based programming approach may fail to extract patterns that were not identified previously. This issue could be successfully addressed by adopting machine learning-based techniques, which may well yield a fully automatic pattern extraction program.

Finally, the current list of verb patterns did not address the concern that a core list of formulaic expressions fails to consider disciplinary variability. Some scholars have challenged the practice of developing general academic word lists (or lists of formulaic expressions) (Ackermann & Chen, 2013; Hyland & Tse, 2007), because (1) words may associate with different meanings across disciplines, and (2) highly frequent academic words may also be frequent in general contexts (Gardner & Davies, 2014). Concerning the two criticisms, we side with Gardner and Davies, 2014 in arguing that generalized word lists cannot be discarded before a computer program is developed to accurately tag the distinct meanings of word forms, and that being included in a frequent general word list does not diminish the importance of a word (or a formulaic expression) that also frequently occurs in academic fields. Nevertheless, we cannot ignore the value of discipline-specific lists. Future lists of grammar patterns could be developed from corpora representing different disciplines to capture cross-disciplinary pattern uses. These discipline-specific grammar pattern lists may have the potential to uncover the frequency distribution of meaning across disciplines, since different patterns of the same word are generally associated with different meanings (Hunston & Francis, 1998). For example, when used with the pattern “V n *with* n”, the verb “provide” may refer to “provide something that someone needs or wants”, while the verb-pattern combination “provided that” is used “if you mean that the first thing will happen only if the second thing also happens” (see “The other banks are going to be very eager to help, provided that they see that he has a specific plan.” in Collins COBUILD English dictionary, p.1325). It can be expected that occurrences of frequency for these two patterns of “provide” are different across disciplines. Given the close connection between pattern and meaning, we anticipate that discipline-based lists of grammar patterns could contribute to the ongoing debate concerning the viability of a core list of academic word or formulaic expressions.

Acknowledgements

The authors appreciate the anonymous reviewers for their constructive comments on this paper. The authors would like to express sincere gratitude towards Dr. Chukharev-Hudilainen, Associate Professor, Program of Applied Linguistics & Technology, Iowa State University, for teaching us how to program and allowing us to use the system he developed (Cywrite Analyzer) in our research. We would also like to extend our appreciation to Yaqiao Zhou at iQIYI, Inc. for techinical support.

Appendix A. the Pattern List in a Verb-oriented Order

Provide	V n	1462 (including 135 passive forms)
	V n with n	139 (including 15 passive forms)
	V that	36
include	V n	480 (including 43 passive forms)
	be Ved in	106
develop	V n	968 (including 225 passive forms)
	V	170
	V out of n	19
	V from n	10
suggest	V that	619
	V n	138
	be Ved	106
require	V n	996 (including 385 passive forms)
	V n to-inf	379 (including 225 passive forms)
	V that	88
	be Ved of	5
report	V n	341 (including 253 passive forms)
	V n to n	175 (including 159 passive forms)
	be Ved to-inf	141
	V that	50
	be Ved as	36 (including 7 active forms)
	V to n	32
	be Ved for	24
	It be Ved that	19
	V on n	10
base	*be* Ved on	331
	V n on n	141
describe	V n	684 (including 211 passive forms)
	V n as	150 (including 98 passive forms)
	V wh	48
indicate	V n	491 (including 80 passive forms)
	V that	314
	V	134
	V wh	33
produce	V n	1028 (including 184 passive forms)
identify	V n	406 (including 58 passive forms)
	V n as	39 (including 11 passive forms)
	V n with n	16 (including 7 passive forms)
	V with n	4
support	V n	452 (including 66 passive forms)
increase	V	1259
	V n	1145 (including 192 passive forms)
	V from n	42
	V to n	34
	V by n	19
note	V n	335 (including 228 passive forms)
	V that	279
	V wh	7
represent	V n	799
	be Ved	177
	V n as	59 (including 22 passive forms)
determine	V n	1104 (including 371passive forms)
	V wh	157
	V that	24
	V to-inf	10
	V wh to-inf	7
occur	V	1526
	V to n	32
present	V n	189 (including 50 passive forms)
	V n in n	133 (including 85 passive forms)
	V n with n	34 (including 14 passive forms)
	V n to n	26 (including 8 passive forms)
	V n as	24 (including 9 passive forms)
reduce	V n	987 (including 224 passive forms)
	V	101
	V n to n	143 (including 75 passive forms)
involve	V n	561 (including 77 passive forms)

	V n in n	191 (including 164 passive forms)
	V ving	93
focus	V on n	347
	V n	78 (including 41passive forms)
relate	V to n	233
	V n to n	164 (including 109 passive forms)
	V n	44 (including 16 passive forms)
establish	V n	384 (including 148 passive forms)
	V that	21
	V n with n	15 (including 4 passive forms)
	V n as	11 (including 4 passive forms)
	It be Ved that	11
seek	V n	177 (including 23 passive forms)
	V to-inf	90
	V n from n	14 (including 2 passive forms)
	V n for n	8
compare	V n	314 (including 39 passive forms)
	V with n	289
	V n with n	74 (including 18 passive forms)
	V n to n	55 (including 9 passive forms)
argue	V that	151
	It be Ved that	23
	V for n	12
	V against n	9
	V over n	7
	V about n	2
state	V that	139
	V n	91 (including 41 passive forms)
examine	V n	386 (including 55 passive forms)
reflect	V n	159 (including 37 passive forms)
	V	84
	V on n	35
	V that	4
recognize	V n	221 (including 49 passive forms)
	V that	69
	V n as	53 (including 34 passive forms)
	be Ved by	23
maintain	V n	431 (including 87 passive forms)
	V that	21
	V n at n	14 (including 5 passive forms)
associate	V with n	519
	be Ved with n	368 (including 31 active forms)
design	V n	122 (including 33 passive forms)
	be Ved to-inf	65
	be Ved for	19
address	V n	264 (including 55 passive forms)
	V n to n	11 (including 5 passive forms)
define	V n	414 (including 164 passive forms)
	V n as	177 (including 109 passive forms)
	V wh	7
apply	V n	351 (including 150 passive forms)
	V n to n	296 (including 198 passive forms)
	V to n	219
	V for n	36
	V to-inf	23
contain	V n	1391 (including 52 passive forms)
form	V n	982 (including 191 passive forms)
	V	233
	V n into n	6 (including 4 passive forms)
	V into n	3
reveal	V n	213 (including 28 passive forms)
	V that	92
	be Ved as	7 (including 6 active forms)
	V wh	3
affect	V n	573 (including 149 passive forms)
achieve	V n	661 (including 263 passive forms)
conduct	V n	266 (including 125 passive forms)
	V	18
perform	V n	588 (including 337 passive forms)
	V	96
discuss	V n	429 (including 183 passive forms)
	V wh	19
exist	V	808

	V on n	14
improve	V n	477 (including 65 passive forms)
	V	65
	V on n	2
observe	V n	632 (including 444 passive forms)
	V that	75
	V n ving	14
demonstrate	V n	497 (including 121 passive forms)
	V that	226
	V wh	14
	V for n	7
	V to n	5
result	V in n	691
	V from n	220
	V	4
experience	V n	195 (including 26 passive forms)
control	V n	370 (including 114 passive forms)
measure	V n	580 (including 244 passive forms)
test	V n	206 (including 71 passive forms)
	be Ved for	34
tend	V to-inf	297
	V to n	18
	V n	7
refer	V to n	397
	V n to n	198 (including 161 passive forms)
obtain	V n	858 (including 370 passive forms)
contribute	V to n	234
	V n to n	29
	V n	28
assume	V n	320 (including 91 passive forms)
	V that	278
	be Ved to-inf	75
express	V n	451(including 215 passive forms)
	be Ved as	66 (including 16 active forms)
promote	V n	203 (including 16 passive forms)
	V n from n	3
participate	V in n	110
engage	V in n	95
	V n in n	29 (including 23 passive forms)
	V n	22 (including 8 passive forms)
	V with n	18
	V	10
publish	V n	138 (including 68 passive forms)
	V	17
encourage	V n to-inf	69 (including 24 passive forms)
	V n	23 (including 10 passive forms)
assess	V n	268 (including 68 passive forms)
	V wh	16
view	V n as	91 (including 52 passive forms)
	V n	62 (including 16 passive forms)
	V n in n	9
limit	V n	390 (including 199 passive forms)
	V n to n	119 (including 82 passive forms)
influence	V n	95 (including 29 passive forms)
emerge	V	158
	V from n	50
	V as	38
	V that	12
explore	V n	236 (including 51 passive forms)
	V n for n	18 (including 4 passive forms)
	V for n	2
generate	V n	415 (including 111 passive forms)
perceive	V n	81 (including 34 passive forms)
	V n as	44 (including 18 passive forms
ensure	V that	160
	V n	131 (including 4 passive forms)
select	V n	117 (including 52 passive forms)
	V n for n	37 (including 24 passive forms)
	V n from n	6 (including 4 passive forms)
emphasize	V n	53 (including 8 passive)
	V that	13
extend	V n	166 (including 31 passive forms)
	V n to n	70 (including 34 passive forms)

	V to n	67
	V from n to n	12
	V from n	12
	V beyond	11
evaluate	V n	314 (including 97 passive forms)
conclude	V that	112
	V n	39 (including 18 passive forms)
	V n from n	4 (including 1 passive forms)
consist	V of n	413
	V in n	29
adopt	V n	186 (including 48 passive forms)
	V n as	11 (including 2 passive forms)
depend	V on n	719
	V upon n	81
attempt	V to-inf	124
	V n	40 (including 15 passive forms)
predict	V n	365 (including 56 passive forms)
	V that	52
	V wh	28
employ	V n	143 (including 63 passive forms)
	V n to-inf	38 (including 25 passive forms)
	be Ved in	34 (including 6 active forms)
account	V for	186
	be Ved for	23
link	V n to n	146 (including 97 passive forms)
	V n	107(including 41 passive forms)
	V n with n	30 (including 15 passive forms)
analyze	V n	325 (including 129 passive forms)
range	V from n to n	55
	V between pln	10
enable	V n to-inf	139
	V n	117 (including 6 passive forms)
organize	V n	104 (including 52 passive forms)
	V	16
locate	V n	69 (including 24 passive forms)
	V	11
enhance	V n	323 (including 58 passive forms)
estimate	V n	309 (including 107 passive forms)
	V that	15
	V n at n	12 (including 8 passive forms)
	V wh	3
propose	V n	230 (including 119 passive forms)
	V that	57
	V to-inf	52
vary	V n	95 (including 34 passive forms)
	V	81
	V from n	63
construct	V n	141 (including 34 passive forms)
	V n from n	11 (including 9 passive forms)
rely	V on n	203
cite	V n	25 (including 12 passive forms)
	V n as	11 (including 8 passive forms)
lack	V n	116
	V	40
constitute	V n	176 (including 6 passive forms)
incorporate	V n	154 (including 68 passive forms)
	V n into n	60 (including 47 passive forms)
illustrate	V n	218 (including 61 passive forms)
	V wh	26
	V that	17
	V n with n	12
arise	V	195
	V from n	134
	V out of n	17
acquire	V n	143 (including 36 passive forms)
	V n from n	5 (including 2 passive forms)
characterize	V n	92 (including 11 passive forms)
	V n as	78 (including 12 passive forms)
	be Ved by	66
differ	V	183
	V from n	150
	V with n	5
review	V n	156 (including 50 passive forms)

	V for n	2
interpret	V n	106 (including 50 passive forms)
	V n as	87 (including 49 passive forms)
	V	60
display	V n	123 (including 23 passive forms)
	V n to n	9
derive	V n	153 (including 53 passive forms)
	V from n	135
	V n from n	131 (including 104 passive forms)
permit	V n to-inf	60 (including 12 passive forms)
	V n	55 (including 16 passive forms)
	V	43
	V n n	2
regard	V n as	175 (including 107 passive forms)
	V n with n	3
transform	V n	112 (including 21 passive forms)
	V n into n	49 (including 16 passive forms)
	V n from n	4 (including 2 passive forms)
imply	V n	135 (including 8 passive forms)
	V that	118
facilitate	V n	213 (including 16 passive forms)
yield	V n	214 (including 6 passive forms)
	V	23
	V to n	5
inform	V n	77 (including 30 passive forms)
	V n that	14 (including 7 passive forms)

Appendix B. The Pattern List in a Pattern-oriented Order

Patterns	Frequency	Total frequency
V n	contain (1339), provide (1327), increase (953), produce (844), represent (799), form (791), reduce (763), develop (743), determine (733), require (611), obtain (488), involve (484), describe (473), include (437), affect (424), improve (412), indicate (411), achieve (398), support (386), demonstrate (376), identify (348), maintain (344), measure (336), examine (331), predict (309), generate (304), compare (275), enhance (265), control (256), perform (251), define (250), discuss (246), establish (236), express (236), assume (229), evaluate (217), address (209), yield (208), estimate (202), apply (201), assess (200), facilitate (197), analyze (196), limit (191), observe (188), promote (187), reveal (185), explore (185), recognize (172), constitute (170), experience (169), illustrate (157), seek (154), present (149), conduct (141), suggest (138), adopt (138), test (135), extend (135), ensure (127), imply (127), reflect (122), lack (116), enable (111), propose (111), note (107), acquire (107), construct (107), review (106), display (100), derive (100), transform (91), design (89), report (88), incorporate (86), characterize (81), employ (80), publish (70), influence (66), link (66), select (65), vary (61), interpret (56), organize (52), state (50), perceive (47), inform (47), view (46), locate (45), emphasize (45), permit (39), focus (37), relate (28), contribute (28), attempt (25), conclude (21), engage (14), encourage (13), cite (13), tend (7)	24429
be Ved	observe (444), require (385), determine (371), obtain (370), perform (337), achieve (263), report (253), measure (244), note (228), develop (225), reduce (224), express (215), describe (211), limit (199), increase (192), form (191), produce (184), discuss (183), represent (177), define (164), apply (150), affect (149), establish (148), provide (135), analyze (129), conduct (125), demonstrate (121), propose (119), control (114), generate (111), estimate (107), suggest (106), evaluate (97), link (97), assume (91), maintain (87), indicate (80), involve (77), test (71), publish (68), assess (68), incorporate (68), support (66), improve (65), employ (63), illustrate (61), identify (58), enhance (58), predict (56), examine (55), address (55), derive (53), contain (52), select (52), organize (52), explore (51), present (50), review (50), interpret (50), recognize (49), adopt (48), include (43), focus (41), state (41), link (41), compare (39), reflect (37), acquire (36), perceive (34), vary (34), construct (34), design (33), extend (31), inform (30), influence (29), reveal (28), experience (26), locate (24), seek (23), display (23), transform (21), conclude (18), relate (16), promote (16), view (16), permit (16), facilitate (16), attempt (15), cite (12), characterize (11), encourage (10), emphasize (8), imply (8), engage (8), enable (6), constitute (6), yield (6), ensure (4)	9262
V	occur (1526), increase (1259), exist (808), form (233), arise (195), differ (183), develop (170), emerge (158), indicate (134), reduce (101), perform (96), reflect (84), vary (81), improve (65), interpret (60), permit (43), lack (40), yield (23), conduct (18), publish (17), organize (16), locate (11), engage (10), result (4)	5335
V that	suggest (619), indicate (314), note (279), assume (278), demonstrate (226), ensure (160), argue (151), state (139), imply (118), conclude (112), reveal (92), require (88), observe (75), recognize (69), propose (57), predict (52), report (50), provide (36), determine (24), maintain (21), establish (21), illustrate (17), estimate (15), emphasize (13), emerge (12), reflect (4)	3042

(continued on next page)

(*continued*)

Patterns	Frequency	Total frequency
V on n	depend (719), focus (347), rely (203), reflect (35), exist (14), report (10), improve (2)	1330
V to n	refer (397), contribute (234), relate (233), apply (219), extend (67), increase (34), occur (32), report (32), tend (18), yield (5), demonstrate (5)	1276
be Ved to	apply (198), refer (161), report (159), relate (109), link (97), limit (82), reduce (75), extend (34), compare (9), present (8), address (5)	937
V in n	result (691), participate (110), engage (95), consist (29)	925
V with n	associate (519), compare (289), engage (18), differ (5), identify (4)	835
V from n	result (220), differ (150), derive (135), arise (134), vary (63), emerge (50), increase (42), extend (12), develop (10)	816
be Ved as	define (109), regard (107), describe (98), view (52), express (50), interpret (49), recognize (34), report (29), represent (22), perceive (18), characterize (12), identify (11), present (9), cite (8), establish (7), adopt (2), reveal (1)	618
V to-inf	tend (297), attempt (124), seek (90), propose (52), apply (23), determine (10)	596
be Ved to-inf	require (225), report (141), assume (75), design (65), employ (25), encourage (24), permit (12)	567
V n to n	apply (98), reduce (68), relate (55), link (49), compare (46), refer (37), limit (37), extend (36), contribute (29), present (18), report (16), display (9), address (6)	504
V n as	regard (68), define (68), characterize (66), describe (52), view (39), interpret (38), represent (37), identify (28), perceive (26), recognize (19), express (16), present (15), adopt (9) , report (7), establish (7), reveal (6), cite (3)	504
V of n	consist (413)	413
be Ved with	associate (337), compare (18), provide (15), link (15), present (14), identify (7), establish (4)	410
be Ved in	involve (164), include (106), present (85), employ (28), engage (23)	406
V n to-inf	require (154), enable (139), permit (48), encourage (45), employ (13)	399
V wh	determine (157), describe (48), indicate (33), predict (28), illustrate (26), discuss (19), assess (16), demonstrate (14), define (7), note (7), estimate (3), reveal (3)	361
be Ved on n	base (331)	331
V n with n	provide (124), compare (56), associate (31), present (20), link (15), illustrate (12), establish (11), identify (9), regard (3)	281
V for n	account (186), apply (36), argue (12), demonstrate (7), review (2), explore (2)	245
be Ved for	test (34), select (24), report (24), account (23), design (19), explore (4)	128
be Ved from	derive (104), construct (9), select (4), seek (2), acquire (2), transform (2), conclude (1)	124
be Ved by	characterize (66), recognize (23), inform (7)	96
V n in n	present (48), involve (27), view (9), engage (6), employ (6)	96
V ving	involve (93)	93
V upon	depend (81)	81
be Ved into	incorporate (47), transform (16), form (4)	67
V from n to n	range (55), extend (12)	67
V n from n	derive (27), seek (12), acquire (3), conclude (3), promote (3), select (2), construct (2), transform (2)	54
It be Ved that	argue (23), report (19), establish (11)	53
V n into n	transform (33), incorporate (13), form (2)	48
V n on n	base (41)	41
V as	emerge (38)	38
V out of n	develop (19), arise (17)	36
V n for n	explore (14), select (13), seek (8)	35
V by n	increase (19)	19
V n ing	observe (14)	14
be Ved at	estimate (8), maintain (5)	13
V n at n	maintain (9), estimate (4)	13
V beyond	extend (11)	11
V between pln	range (10)	10
V against n	argue (9)	9
V n that	inform (7)	7
be Ved that	Inform (7)	7
V wh to-inf	determine (7)	7
V over n	argue (7)	7
be Ved of	require (5)	5
V into n	form (3)	3
V n n	permit (2)	2
V about n	argue (2)	2

References

Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL)-A corpus driven and expert-judged approach. *Journal of English for Academic Purposes, 12*, 235-247.

Bauer, L., & Nation, I. (1993). Word families. *International Journal of Lexicography, 6*, 253-279.

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

Bramer, M. (2013). *Logic programming with prolog*. New York, NY: Springer.

Charles, M. (2007). Argument or evidence? Disciplinary variation in the use of the noun that pattern in stance construction. *English for Specific Purposes, 26*(2), 203-218.

Chukharev-Hudilainen, E., & Saricaoglu, A. (2014). Causal discourse analyzer: Improving automated feedback on academic ESL writing. *Computer Assisted Language Learning*, 1-23.
Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.
Collins COBUILD dictionary (2nd ed.). (1995). London: HarperCollins.
Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier, & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 149-162). Amsterdam: John Benjamins.
De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC* (vol. 6, pp. 449-454)Genoa: ELRA.
Durrant, P. (2016). To what extent is the academic vocabulary list relevant to university student writing? *English for Specific Purposes, 43*, 49-61.
Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*(3), 393-417.
Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305-327.
Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15*, 289-316.
Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes, 4*, 257-277.
Hunston, S. (2003). Lexis, word form and complementation pattern: A corpus study. *Functions of Language, 10*(1), 31-60.
Hunston, S., & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics, 19*(1), 45-72.
Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
Hunston, S., & Francis, G. (2000). Pattern grammar: A course-driven approach to the lexical grammar of English. *Computational Linguistics, 27*(2), 318-320.
Hunston, S., Francis, S., & Manning, E. (1997). Grammar and vocabulary: Showing the connections. *ELT Journal, 51*(3), 208-216.
Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.
Hyland, K., & Tse, P. (2007). Is there an “Academic vocabulary”? *TESOL Quarterly, 41*, 235-253.
Jaen, M. M. (2007). A corpus-driven design of a test for assessing the ESL collocational competence of university students. *International Journal of English Studies, 7*(2), 127-147.
Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes, 22*, 42-53.
Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*, 25-35.
Martinez, R., & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics, 33*(3), 299-320.
Melamed, I. D., Green, R., & Turian, J. P. (2003). Precision and recall of machine translation. In *Companion volume of the proceedings of HLT-NAACL 2003 – short papers* (vol. 2, pp. 61-63). Edmonton, Canada, May—June. Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/N03-2021.
Milton, J. (2009). *Second language acquisition: Measuring second language vocabulary acquisition*. Clevedon: Multilingual Matters.
Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52.
Read, J. (2000). *Assessing vocabulary*. New York: Cambridge University Press.
Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*, 140-162.
Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition, 41*, 1089-1110. https://doi.org/10.1017/S0272263119000202.
Römer, U., O'Donnell, M. B., & Ellis, N. C. (2014). Second language learner knowledge of verb-argument constructions: Effects of language transfer and typology. *The Modern Language Journal, 98*, 952-975.
Römer, U., Skalicky, S., & Ellis, N. C. (2017). Verb-argument constructions in advanced L2 English learner production: Insights from corpora and verbal fluency tasks. *Corpus Linguistics and Linguistic Theory*, 1-27. https://doi.org/10.1515/cllt-2016-0055.
Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.
Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Sinclair, J. M. (Ed.). (1996). *Grammar patterns 1: Verbs*. London: HarperCollins Publishers.
Vannestal, M., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL, 19*(3), 329-350.
Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes, 6*, 18-35.
Yoon, H., & Hirvela, A. (2004). ESL student attitudes towards corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257-283.

Hong Ma is currently an assistant professor at Zhejiang University, China. She received her PhD in Applied Linguistics and Technology from Iowa State University. While at Iowa State, she led multiple research projects, which developed and evaluated a pattern grammar inspired computer-assisted vocabulary-learning tool (PAT GRAM).

Manman Qian received her PhD in Applied Linguistics and Technolgoy from Iowa State University. Her research interest and expertise revolve around overcoming challenges in second language pedagogy through computational approaches. She has led the development of *Linguatorium Auris*, a high-variability phonetic training system, and shared research results at multiple venues including *AAAL* and *Language Learning & Technology*.