

A Corpus-based Study of the Structures and Functions of Academic Chunks in the Discipline of Engineering

CHEN Xin^a, XIAO Gengsheng^b

School of Foreign Languages, University of South China, Hengyang, 421001, China

^aemail: englishow@126.com, ^bemail: xgstudy@126.com

Keywords: academic chunk; corpus-based study; engineering discipline

Abstract. Chunks are commonly used in academic written register, and there are quite remarkable differences in chunk features between disciplines. Based on a corpus investigation into 4-word chunks in research articles of engineering discipline, the present study has found that researchers in this field most frequently use the chunks of structures passive followed by prepositional phrase and noun phrase with *of*-phrase fragment. Besides, due to the empirical and experimental nature of engineering studies, researchers use research-oriented chunks most prevalently, text-oriented chunks the second, participant-oriented chunks the fewest. The findings on the structures and functions of academic chunks in the present study can provide pedagogical implications for academic writing in engineering field.

Introduction

Chunks are recurrent multi-word expressions, regardless of their idiomaticity, and regardless of their structural status, which can be regarded as extended collocations: bundles of words that show a statistical tendency to commonly go together in natural discourse. [1] Chunks are so prevalent that they play a very vital role in linguistic production and language learning. Whether a language has been acquired or not, to some extent, depends on the degree to which the learner can proficiently understand and produce the chunks required for linguistic input and output. Because chunks can be stored and retrieved as a whole, the generation of which does not require the analysis of grammatical rules, the mastery of chunk knowledge can greatly determine the accuracy, fluency, and idiomaticity of the learner's language production. [2]

Owing to the importance of chunks in languages, researchers have studied chunks from a wide range of perspectives. Firstly, scholars conducted some studies on how chunks are used differently in different registers and great register differences have been identified in terms of high-frequency chunks. In conversation, common chunks have been found to include sequences such as *do you want me to, I said to him, going to be a, I don't know what*, while in academic writing the common chunks are remarkably different from frequently-used conversation chunks, for typical academic chunks include sequences such as *in the case of the, there was no significant, it should be noted that*. Furthermore, researchers also have paid attention to second/foreign language learners' acquisition of chunks. For example, it has been found that Chinese learners of English tend to be incapable to produce formally diversified and functionally appropriate chunks and they cling to literally translated chunks from Chinese and overuse the self-centered *I think* type of chunks which downgrades their discourse interaction to an ineffective and impolite mode.

In recent years, studies have been blossoming in the fields of English for Specific Purposes (ESP) and English for Academic Purposes (EAP), in which scholars gradually focus their attention on the relationship between knowledge of chunks and fluency of discourse in academic domain. It has been found that knowledge of chunks is indispensable to academic writing, and chunks are frequently used in research articles. For the past decade, researchers have shifted their attention to disciplinary variation of chunks in academic writing, which has been gradually becoming the focus for studies in the fields of EAP and ESP. However, the characteristics of chunks still remain unknown in many disciplines. Thus, the present study aims to investigate the forms and functions of chunks in the discipline of engineering.

Corpus Description and Chunk Retrieval

The language data for the present study consist of the written texts of 120 research articles in the field of engineering, energy and technology, which represent the discipline of engineering. All the texts are from the well-renowned international electronic database for academic journal publications *Elsevier*, in which we select 6 influential journals with relatively higher impact factors in engineering discipline, namely, *International Journal of Plasticity*, *Renewable & Sustainable Energy Reviews*, *Journal of Power Sources*, *Bioresource Technology*, *Solar Energy Materials and Solar Cells*, *International Journal of Hydrogen Energy*, with 20 articles from each. Moreover, we removed the noisy information by deleting those parts including the titles, abstracts, key words, acknowledgements, references, appendices, pictures, tables, figures, and charts in the academic articles, and kept the cleaned texts to build English Academic Article Corpus of Engineering Discipline (EACED), with corpus capacity of 688, 120 tokens.

Because 4-word chunks are more common than 5-word chunks, and meanwhile, compared with 3-word chunks, 4-word chunks can better demonstrate their structure and function, the present study only takes 4-word chunks as the object of research. Antconc is used as a corpus tool for the retrieval and statistical analysis of 4-word chunks according to the following procedures. Firstly, open all the corpus files of EACED in Antconc, and then enter the tab of Clusters. Secondly, in Search Term, select N-Grams with a tick. Then, because 4-word chunks are under investigation in present study, N-Gram Size is fixed as 4 by setting both the minimum size and maximum size as 4. After that, the minimum N-Gram frequency is set as 20, and the generation of chunk list is sorted by frequency. Through the automatic generation of 4-word chunks, careful manual examination, and elimination of those non-language units and those chunks existing in no more than 5 files, the present study altogether obtained 187 types of chunks, with frequency of chunks amounting to 3997 and tokens 15, 988, which accounts for 2.3 percent of the tokens in the whole corpus.

It is very significant to explore the frequently-used chunks in a certain field, which is a very important part of language features. Hyland (2008) conducted a survey of chunks and created chunk lists of several disciplines, and concluded that there was significant disciplinary variation in terms of most frequent 4-word chunks.[3] However, the discipline of engineering, energy and technology has not been investigated, and the chunk list in this discipline still remains unsettled. By retrieving the chunks from EACED, this research produced a list of the following 20 chunks as the most frequent 4-word chunks in engineering discipline. According to frequency, the top 20 frequently-used 4-word academic chunks in the field of engineering, energy, and technology are *on the other hand*, *in the case of*, *it can be seen*, *can be used to*, *as a function of*, *as shown in figure*, *is shown in figure*, *is based on the*, *at the same time*, *with respect to the*, *it is found that*, *the effect of the*, *the size of the*, *in the case the*, *a function of the*, *in terms of the*, *is one of the*, *in the form of*, *the results of the*, *it is necessary to*.

Structural Analysis of Chunks in Engineering Discipline

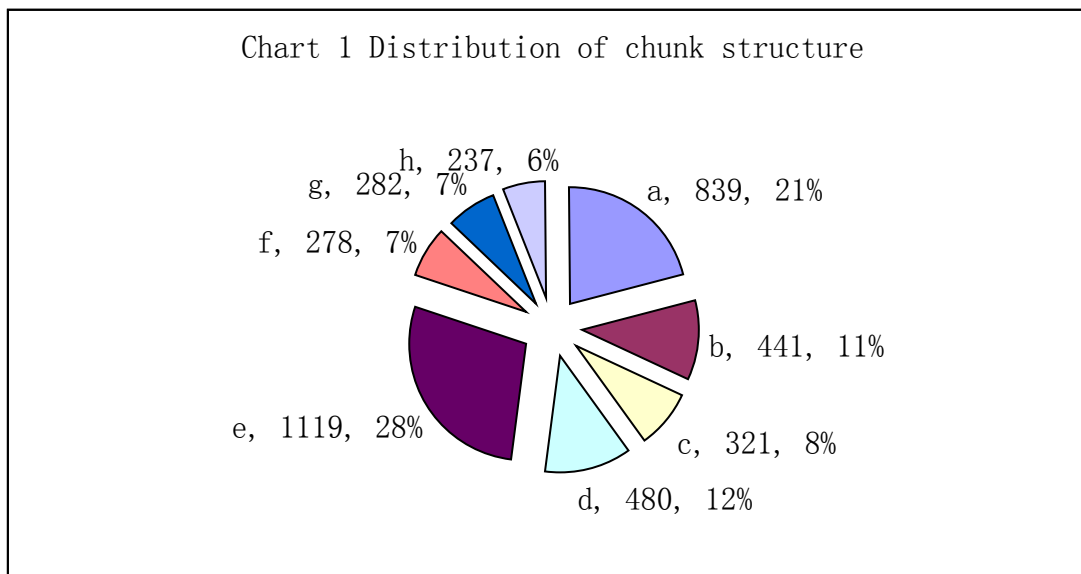
According to theoretical framework of chunk structure put forward by Biber et al (1999), 4-word chunks take the following language forms. [4]

Table 1 Most common patterns of 4-word academic chunks (Biber et al., 1999: 997-1025)

Structure	Examples
a. Noun phrase + of	the end of the, the nature of the, a large number of
b. Other noun phrases	the fact that the, one of the most, the extent to which
c. Prepositional phrase + of	at the end of, as a result of, on the basis of, in the context of
d. Other prepositional phrases	on the other hand, in the present study, with respect to the
e. Passive + prep phrase fragment	is shown in figure, is defined as the, can be found in
f. Anticipatory it + verb/adj	it is important to, it is possible that, it should be noted
g. Be + noun/adjectival phrase	is the same as, is a matter of, is due to the, be the result of
h. Others	as shown in figure, should be noted that, is likely to be

Based on the structural classification of academic chunks, we divide all the chunks (with type number of 187

and frequency of 3997) into 8 categories. Moreover, the present study conducted statistical computation of the frequencies of each structure type chunks. The distribution of 4-word chunks in terms of structural is shown in Chart 1.



According to Chart 1, passive + prepositional phrase is the most prevalent structure of academic chunks in engineering discipline, and the frequency of this structure type is 1119, taking up 28% of all chunks. Chunks of this structure aim to establish a locative and logical relation. To be specific, writers tend to use passive chunks followed by prepositional phrase fragments to guide readers through the text (1) or to identify the basis for an assertion in an argument (2).

(1) The conductivity of system 3B measured as a function of LitrIF concentration *is shown in Table S4 and Fig. 5. (Solar Energy Materials & Solar Cells)*

(2) This methodology *is based on the* flux transfer approach, used to model the distribution of light energy in round skylights of different proportions of height and width, wall reflectance and transmittance of the glass. *(Renewable and Sustainable Energy Reviews)*

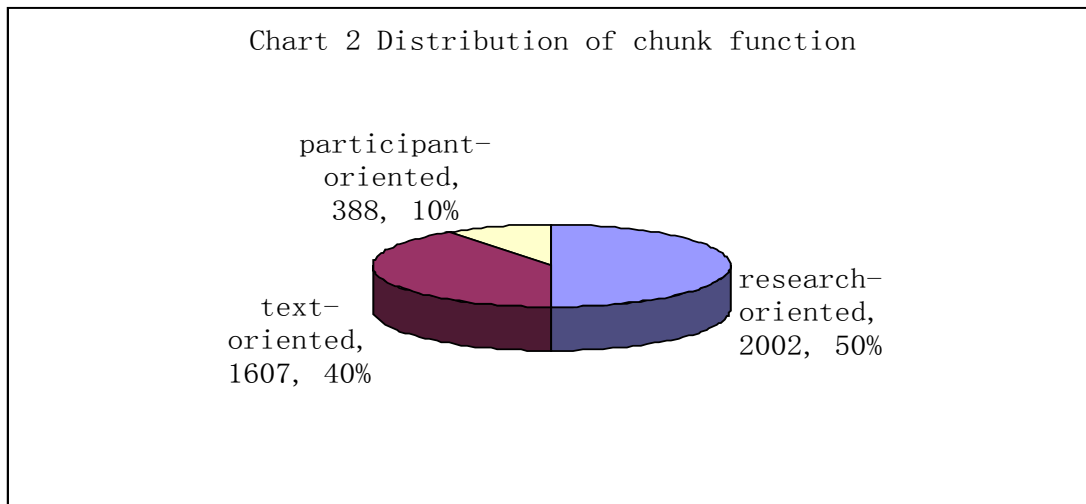
Then, it is obvious that the noun phrase with *of*-phrase fragment is the second most common structure of 4-word chunks in this discipline, which accounts for 21% of all forms. Chunks of this structure convey several meanings and are often used to express quantity, quality, existence, place or size.

(3) This relative TEEI index assumes that the isolated turbine and the turbines in the array benefit equally from the economies of scale of producing *a large number of* turbines at the same time. *(Renewable and Sustainable Energy Reviews)*

(4) As a consequence, *the size of the* microstructure is generally no longer negligible with respect to the specimen dimensions and size effects occur modifying the mechanical properties. *(International Journal of Plasticity)*

Functional Analysis of Chunks in Engineering Discipline

According to theoretical framework of chunk function put forward by Hyland (2008), 4-word chunks can be classified into three categories, namely, research-oriented chunks, text-oriented chunks and participant-oriented chunks. This standard of chunk categorization is mainly concerned with the primary discourse function that a chunk serves in the academic article. Research-oriented chunks are mainly used to help writers to structure their activities and experiences of the real world, e.g. to indicate location, procedure, quantification and topic etc. Text-oriented chunks are mainly concerned with the organization of the text and its meaning as a message or argument, functioning as transitional signals, resultative signals, structuring signals, and framing signals. Participant-oriented chunks are focused on the writer or reader of the text, including stance features and engagement features.



From Chart 2, it can be seen that research-oriented chunks are the most common, which account for half of all chunks in the corpus. The density of research-oriented chunks in the field of engineering, energy, and technology, is mainly due to the fact that in the discipline of engineering, academic articles have to clearly describe various aspects of research, e.g. research objects, contexts, models, equipment, materials, procedures and so on. The following examples are typical research-oriented chunks.

(5) The comparison result P_e is used to manage *the operation of the* fuel cell and affects the system response based on the energy storage media used. (*International Journal of Hydrogen Energy*)

(6) The force required to expand *the surface of the* droplet is less than that required to penetrate through a constricting pore throat and so all other interfaces remain static until the droplet is removed. (*International Journal of Hydrogen Energy*)

Besides, text-oriented chunks are also indispensable in academic writing of engineering domain. It can be seen that chunks of this function occupy 40% of all chunks. After analyzing the frequently-used chunks in this field, we can see that there are quite a large number of text-oriented chunks in the research articles, which serve the function of organizing the discourse. To be specific, by using some text-oriented chunks, e.g. *on the other hand, in contrast to the*, the writer can establish additive or contrastive links between elements; such text-oriented chunks as *it was found that, these results suggest that* aim to mark inferential or causative relations between elements; other text-oriented chunks are used to function as structuring signals (e.g. *in the present study, as shown in figure*) and framing signals (e.g. *in the case of, on the basis of*). The following two examples are typical text-oriented chunks.

(7) *In addition to the* economic/rural development arguments for community ownership, community ownership may also reduce opposition to developments, and possibly stimulate behavior change towards lower carbon use. (*Renewable and Sustainable Energy Reviews*)

(8) As a consequence, the size of the microstructure is generally no longer negligible *with respect to the* specimen dimensions and size effects occur modifying the mechanical properties. (*International Journal of Plasticity*)

Eventually, despite the relatively lower percentage (10%) of participant-oriented chunks in engineering discipline, its role still can not be neglected. Although academic articles are remarkably different from daily conversation, they still can clearly reflect the interaction between the writer and the readers. [5] In research articles of engineering discipline, we can see many participant-oriented chunks, which serve two functions, i.e. to convey the writer's attitude and evaluations (e.g. *may be due to, are likely to be*) and address readers directly (e.g. *we can see that, as can be seen*). The following two examples in the contexts can vividly demonstrate the interactive function of participant-oriented chunks.

(9) First, *it is possible that* co-ownership would reduce the opposition of some people within the recipient community to on-shore wind. (*Renewable and Sustainable Energy Reviews*)

(10) *It should be noted*, that at each time step (15 min), the power generated by the primary sources (wind turbine and solar-PV), P_{Ren} , is calculated and compared to load demand. (*International Journal of Hydrogen Energy*)

Conclusion

Based on a corpus study of 4-word chunks in the research articles of engineering, the present research has found that chunks are widely used in academic writing of engineering discipline, which take 8 different structural forms, with passive + prepositional phrase as the most frequent one, and the noun phrase with *of*-phrase fragment the second. Moreover, academic chunks can be classified into three groups according to discourse function, research-oriented, text-oriented, and participant-oriented. In the field of engineering, research-oriented chunks are the most prevalent due to the empirical or experimental nature of engineering studies, and meanwhile, text-oriented chunks are also quite significant in this field owing to the importance of organizing text in academic writing, and eventually, with relatively fewer participant-oriented chunks in research articles of engineering, they play a vital role in establishing interaction between the writer and the readers.

The corpus-based analysis of academic chunks of engineering discipline has clearly revealed the structural and functional features of chunks in this field. Owing to the disciplinary variation of chunks in different fields, it is imperative that researchers should acquire chunk knowledge in a certain domain. [6] Therefore, the present study can provide pedagogical implications for EAP and ESP teaching and academic writing especially for scholars and students working in the field of engineering.

Acknowledgement

This article is sponsored by the Projects Supported by Hunan Provincial Social Science Foundation (Grant No.14YBA330 & 14YBA331), Key Discipline of Foreign language and Literature of USC, and Key Base of Corpus Translation Studies of USC.

References

- [1] Cortes, V. Lexical bundles in published and student disciplinary writing: Examples from history and biology [J]. *English for Specific Purposes*, 2004, 23(4): 397-423.
- [2] Zhang, Y. A study on the correlation between L2 learners' ability to identify prefabricated chunks and their L2 proficiency [J]. *Journal of the Foreign Language World*, 2008, (3): 62-66.
- [3] Hyland, K. As can be seen: Lexical bundles and disciplinary variation [J]. *English for Specific Purposes*, 2008, 27(1): 4-21.
- [4] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. *Longman Grammar of Spoken and Written English* [M]. Harlow: Pearson, 1999.
- [5] Hyland, K. Stance and engagement: A model of interaction in academic discourse [J]. *Discourse Studies*, 2005, 7(2): 173–191.
- [6] Wang, F. & Wang, H. A corpus-based comparative study of academic lexical bundles in linguistics and engineering research articles [J]. *Journal of the Foreign Language World*, 2015, (2): 16-24.