UDK 808.3

Alimardonova Fotima Primkulovna
TerSU, Master's Department
Faculty of Linguistics (English),
1st year student

DEVELOPING A CONCEPTUAL MODEL OF CLASSIFYING PHRASEOLOGICAL UNITS: CLASSIFICATION PRACTICES AND THEORETICAL FRAMEWORK

Abstract: This paper analyzes existing literature of lexical bundle classification practices and proposes a conceptual model of lexical phrase typology based on the findings. First, the study focuses on the literature review about multi-word unit classification, considering different aspects of lexical unit structure and their functions in texts. Second, the study attempts to devise a theoretical model which can be widely used to make lists of phrases no matter what field of specialization is encountered. We hope that this model enables educators and researchers from different fields to classify effectively different phrases in term of structure, function and content connection.

Keywords: classification, structure, function, practice, lexical bundle, framework, target discourse, criteria.

Introduction

Lexical phrases can be identified by using a variety of terms such as fixed expressions, multi-word units, lexical bundles, routines and prefabricated patterns. Biber et. al (2002) considered them as text building blocks, based on frequency - driven order of words. Evidently because these text builders can support smooth flow of cohesive devises which constitute the overall text cohesion. This innovative language level which can be found with multiple names (like we mentioned above) in many literature resources includes numerous advantages in itself. This can be clear with the fact that students from different fields can facilitate effectively their vocabulary acquisition and store a huge number of words in their mental lexicon

when they use so called lexical bundles in spoken and written production which hold practical importance. It is also known that effective writing requires students to gain fuller understanding of word usage in context. Sinclair at al. (2007) stressed that lexis is responsible for the organization and patterning of language. There is so much to say about the benefits that come from lexical bundle usage. However, the classification of these phraseological units becomes one of the key issues that we need to handle because of a number of reasons. It must be the first and foremost reason that classification enables us to maintain the confidentiality, ease of access, and integrity of lexical resources in our language system. With clear classification, students can find themselves much easily and fast study multiword-units in context. There is an enormous amount of research and studies devoted to typology of lexical phrases around the world. Below, in the section relevant to these resources, common practices about lexical phrase classification are identified in short review. Later, our article also tries to create a useful framework for teachers and researchers to classify the lists of specific phrases in both ESP and EAP contexts.

Existing Literature review

Having more obvious view of lexical phrase classification, a number of researchers such us Cortes (2004), Hyland (2008), and Grabowski (2013) came to the conclusion that functional classification is more applied to four- word lexical phrase typology possibly because of its easier specification. Both structural and functional classifications were included in the framework designed by Biber et al. (1999). In terms of structural classification, lexical phrases are divided into **noun phrases (the way in which), prepositional phrases (about the nature of), and verb phrases (is based on)**. In the case of functional classification, Biber et al. (2004) distinguished the following three categories, considering them as signals to discourse fluency.

Stance expressions link authorial presence and author's certainty with presented information by using impersonal and personal lexical units such as *"it is possible"* and *"I want you to"* respectively.

Discourse organizers support and hold the cohesion of texts by expressing the relationships between previous and the following textual information. Typical examples include *on the one hand, to look at the*.

Referential expressions connect directly to physical or abstract objects or more obvious context of texts in order to identify those objects and support topic continuity. As an example, *in terms of the* is a typical example to say.

Olga Mudraya (2006) illustrated instructional activities in order to help engineering students acquire language prefabs (or multiword units or phraseological units) by using lexical and corpus approaches. In this regard, we can see that her **semantic** research organized lexical bundles within the corpus data into three main categories: **technical, nontechnical and sub-technical units.** Olga's study focused on concordance data in which students were required to analyze syntactic patterns in general and technical senses.

Breeze (2013) classified lexical bundles into two main types, **content and non-content** phrases. To further categorize, the content lexical bundles are subdivided into **abstract concepts, agents, documents and dates**.

In efforts to identify different lexical characteristics of specific discourse communities, researchers found the following types of lexical bundles from functional and structural points of view.

Cortes (2004) identified the potential differences that in history discipline texts, two major types of lexical phrases are noun phrases and prepositional phrases while biology texts includes wider range of structural lexical phrase types and specifically ***it+be+adjective*** **or** ***verb+complement*** **phrases** which emphasize disciplinary pragmatic hedging.

Hyland (2008) determined the common difference across various branches of science that the existence of **research-oriented lexical phrases** focused on the communication of empirical methods and showed the tendency towards the noun phrases (*the performance of the*). He also revealed **text-oriented bundles type** in

humanities discourse because of the strong emphasis of fluent argumentation in these disciplines.

Another type of lexical bundles which is worthy to note here is participant-oriented bundles which focus on the author of the text such as *it should be noted.*

Veronika Tomakova (2016) retrieved lexical bundles in legal texts and categorized them from the perspectives of functional and structural view. In the discourse of legal English community, it is confirmed that noun and prepositional phrases occupy dominant position in legal English genre. In her study, Veronika used the methodology of Breeze (2013) and made some necessary changes like time and actions subdivisions of content phrases.

From what have been outline above, one can conclude that different disciplines have similarities and differences in their target discourse in terms of lexical phrase classification. That is why there should be a common method that is applicable to any field of specification for identifying their bundle types.

In this regard, a number of studies have offered their frameworks as we see below.

Methods and materials

To inform our classification model, we analyzed common practices and methods relevant to lexical bundle classification processes from different fields and identified common core characteristics of these processes. One of the typical practice by Philippa Otto (2020) is that although his work which is based on use of three - part methodology in the example of civil engineering is solely devoted to choose **specialized vocabulary**, we can extract the relevant practice how the list of **content words in target discourse** can be made and narrowed with surrounding context. This practice is “key word analysis and filters” which enables us generate a list of words in contexts that occurred more frequently in the specialized corpus rather than in general English corpus. This technique compares the frequency of each word in target corpus with its frequency in general English corpus. Then, it determines a list of words with

unusual high frequency which mean they play a particular role or function in a specialized discourse. With proper concordance software (**AntCorc.tool**) which is to analyze the corpus, the list of keywords is ranked from higher to lower according to their keyness value specified by log likelihood ratio (6.63 limit or higher). This statistic is intended to show the most representative phrases of language in target corpus. In this ranking, all word forms are treated as individual words analyzed separately without lemmatizing. After that, a set of filters is applied to reduce and refine the lists.

Another main practice to select **content specific bundles** is made by Dougal Graham (2014) which further informed how to determine common lexical phrases based on four main criteria frequency, occurrence in multiple texts, corpus-specificity and by co-occurrence. He also proposed six categories for determining marked lexical phrases. These are (1) marked part of speech, (2) marked word form (3) non-prototypical word meaning (4) marked collocation (5) non-literal phrase meaning (6) specialized syntax.

To identify **technical/non-technical/sub-technical** aspects of the given lexical bundles, a lexical frequency instructional model by Olga Mudraya (2006) is highly recommended since it can clarify how word frequency analysis and word list organization can be implemented.

More importantly, the analytical framework designed by Biber et al. (1999) to identify structurally and functionally classified lexical phrases highly informed our classification model because it helps us determine three main categories of lexical bundles (**Stance expressions**, **Discourse organizers, and Referential expressions)** from functional perspective and find structurally classified bundles (**noun, prepositional and verb phrases)** as well.

The model represented in this paper is designed to classify lexical bundles in term of structure, function, content specificity (or technical level). This model includes two sections: **bottom-up register analysis** which focuses on frequently

occurring pervasive lexico-grammatical features (in our case lexical buddle) in target texts because of their well-suited functionality and **practice –based classification** which uses the existing methods of classifying lexical phrases (above mentioned typological methods) and special criteria relevant to discipline nature.

Bottom - up register analysis is the starting point of classifying target lexical bundles since the very first thing when beginning special corpus analysis is to automatic segmentation of all the texts in specialized corpus and identify various types of lexico grammatical features and examine their patterns. In the case, the research focus should be related to examine co-occurrences of multiword units (lexical bundles/phrases/collocations and so on). Based on Philippa Otto's practice, "key word analysis and filters", this step is to try to generate a list of words in contexts that occurred more frequently in the specialized corpus rather than in general English corpus. With available and practical concordance tools, it is easy to examine frequency lists and concordance lines to identify lexical multi word patterns.

Practice –based classification is the core stage of classifying the target lexical phrases. This process goes with analyzing lexical phrase units in terms of structure, function and content. For content based classification, Olga Mudraya and Dougal Graham's proposals will be taken in to practice while Biber's analytical framework enable us categorize multi-word units based on structural and functional criteria.

Results and Discussion

Using a set of existing methods and top-down transfer analysis to classify lexical phrases is a model with several benefits. First, because it provides clear and concise guidelines for instructors and researchers with which they can generate a list of reliable and well-classified multi-word units. Secondly, it has the potential applications for different fields of specification so that overall framework of this model allows specific purpose instructors analyze their target community's discourse. Thirdly, using the bottom-up register analysis approach enables teachers to induce lexical patterns with more clear focus rather than top-down approach does.

Unique to this study was that the model identified common core practices of effectively classifying representative pieces of target discourses from different fields as an important guidance for mastery of effective lexical bundle classification. Lexical bundles classified within this model amply fulfill the expectation of providing clear view of specific purpose lexical phrase units since it takes the account of essential aspects of lexical bundles into consideration. As noted earlier in this article, the list of lexical bundles representing the target domain of written and spoken academic discourse can serve reference materials for ESP teachers in higher education settings.

In applying the model developed for this article, three main criteria on classification stood out related to the structure, function, and content connection when analyzing specific purpose texts. From the functional classification perspective, the behavior of lexical phrases within the text is to be identified while structure puts stress on grammatical aspects and order of lexical phrases. Content connection is related to specific aspects of target texts.

Generally speaking, researchers can find a clarified view of how to analyze and classify target lexical phrases for target discourse community when they rely on this model. That is to say that application of this model can promote the analytic skills to identify common types of lexical bundles from specified corpus.

Conclusion

With the view of the superior position of the existing practices of lexical bundle classification within students' target discourse situations and in line with bottom –up register analysis which defines the initial process of classifying multi-word units as the way of segmenting automatically the contents of the specialized corpus and examining the frequency and range of generated list of lexical bundles, the goal of this investigation was to revise the common practices and develop a conceptual model which serves educators and researchers to classify their multi-word units in their specialized language contexts for different fields of specialization.

The first of the two parts of the model sought to provide a common view on analyzing target register with segmentation process and computer – assisted tools and produce a list of the most common lexical phrases in target contexts with the objective of establishing a base for practice –based classification. On the other hand, the second part of this model, relying on the results of the first stage, was designed to provide proper uses of existing methods of classification and therefore draw attention to those language bundles which are specific to target discourse community. As implied earlier, because of its major role in classification, the second part of our model targeted only the structure, function and content specialization of lexical phrases as outline by the common practices.

The methods described in the second stage allow instructors even to provide students with clear guidelines to classify common lexical bundles in the classroom without distracting them to a huge number of bundle types. By using this model as a starting point in lexical bundle classification, instructors and researchers can take advantage of the strengths of existing methods and supplement their teaching materials with a list of specific phraseological units in classified forms.

Overall, the application of the generated model in this article imparts a clear and concise way to analysis and classification of linguistically marked lexical bundles, support the researchers by including hands-on practices which facilitate students' discourse processing and enrich their mental lexicon in a clear organized way.

References:

1. Ken Hyland (2008). "as can be seen: Lexical bundles and disciplinary variation" 27(1):4-21.

2. Olga Mudraya (2006). "Engineering English: A lexical frequency instructional model" pp.235-256.

3. Elena Cotos (2017). "Language for specific purposes and corpus-based pedagogy" pp1-23.

4. Philippa Otto (2020). “Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering” pp32-36.

5. Dougal Graham (2014). “The engineering Phrases List: Towards Teachable ESP Phrases”pp1-16.

6. Veronika Tomankova (2016). “Lexical Bundle in Legal texts corpora – Selection, classification and pedagogical implications” pp75 – 94.

7. Douglas Biber et.al (2002). “The Longman Grammar of Spoken and Written English” 34(4).

8. Ken Hyland (2012). “Bundles in Academic Discourse”pp. 150 - 169

9. Altenberg, B. (1993). Recurrent word combinations in spoken English. In Arcy, J. D. (Ed.), *Proceedings of the Fifth Nordic Association for English Studies Conference* (pp. 17–27).

10. Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14, 275–311.

11. Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263–286.

12. Biber, D., Conrad, S., & Cortes, V. (2003). Towards a taxonomy of lexical bundles in speech and writing. In Wilson, A., Rayson, P., & McEnery, T. (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* (pp. 71–92).

13. Cheng, W. (2007). Concgramming: A corpus-driven approach to learning the phraseology of discipline-specific texts. *CORELL: Computer Resources for Language Learning*, 1, 22–35