IDENTIFICATION OF THE MOST COMMON PHRASEOLOGICAL UNITS IN THE ENGLISH LANGUAGE IN ACADEMIC TEXTS: CONTRIBUTIONS COMING FROM CORPORA

Ergasheva Dilafruz Abdusalomovna
SamSIFL, The faculty of English philology and translation studies, trainee teacher.

ABSTRACT.
Academic-scientific phraseological units in the English language play a key role in the communication of/to experts, once they reproduce frequent and expected expressions in varied disciplines. This is paper aims at identifying and analyzing the 100 non-specialized academic-scientific phraseological units in the English language, present in eight major fields of knowledge. The theoretical background referred to Phraseology and Corpus Linguistics. Regarding methodology, an academic corpus was compiled with more than 120 million words. The most relevant units are 'the case of', 'as a result of' and 'at the end of'. The list with the most functional phraseological units in the English language might provide a valuable pedagogical linguistic reference for the study of the academic genre.

Keywords: specialized communication, frequent expressions, linguistic-textual processing, statistical analysis.

Introduction

Given the relevance of the theme and the insertion of researches in international scenario in the English language, noted by the increase of the number of versions or texts written by language researchers, it is necessary to study and provide supporting material to researchers, teachers, students, and translators.

This way, our investigation aims to bring awareness concerning academic linguistic production, and highlight that, besides terminological and lexical common units, the academic text is also constituted of phraseological units, which potentially give authenticity to the scientific article in English, fulfilling the target reader's expectations – thus, having common vocabulary, terminological vocabulary and typical structures of academic-scientific communication.

Literature and arts. Regarding the research outlined, we will present the theoretical foundation based mostly on Phraseology and Corpus Linguistics. In this article I will try to describe the methodological procedures linked to the collection and processing of the corpus, as well as the statistical calculations used. In the third section, we will analyze the obtained data and the results that we have found. In the fourth section, we will discuss the final considerations.

Theoretical foundation. The theoretical foundation resorts to Phraseology and Corpus Linguistics. Ellis (5; pp.1-13) explains that from the 1950s on, structural patterns started to be called 'constructions' or 'phraseologisms'. In comparison to the past century, a considerable amount of studies has been developed in the Phraseology purview, with contributions in a prominent position coming from studies affiliated with cognition, description, acquisition, teaching of native and foreign language, and also with Terminology, as phraseological units also occur in specialized texts.

In order to differ the theoretical line from the object of study, we will adopt the term Phraseology - with an initial capital letter - to refer to the discipline which studies the phraseological units and also the term phraseology - with an initial lower case letter - to allude to the object of study of this discipline, the phraseological units. In this context, the object of study tackled here may be discussed under varied names, depending on the theoretical strands: multi-word expressions, statistical phrases, chains, formulaic sequences, multi-word units, grouping, combinations of recurring words, lexical package, n-grams etc.

When dealing with phraseology in the academic context, Coxhead (3; pp.149-155) stresses the importance of typical words in the academic communication as well as stresses the importance of the groups to which these words belong. The author believes that we are still in a preliminary phase regarding the understanding of formulaic sequences nature in the academic context.

Coxhead (4; pp.160-164) affirm that 4-word sequences are the most explored in written language studies. We will analyze 4-word sequences in written academic communication that commonly occur in academic scientific texts, but do not have the status of a specialized phraseological unit. For instance, the sequence constituted by 4 words 'scientific name of the animal' is likely to occur in the Biology field, whereas, in the Mathematics field, its occurrence may be rare or even null. Our purpose is to identify the units that may be, at the same time, common in several areas of expertise. Therefore, the phraseological unit focused here is not identified a priori as pertaining to a particular area of knowledge.

Taking into account the relevance of academic Phraseology knowledge and mastery, we agree with Hunston (7; pp.101-112), when she states that knowing phraseological units implies a linguistic competence relating to linguistics resources used in texts of specific learning contexts. Furthermore, it is assumed a degree of the subject or the themes knowledge covered by these texts, because they constitute, with the terms, transmitting units of specialized knowledge. For that matter, Biber (3; pp.178-189) points that about 21% of his academic corpus consisted of phraseological units. This is finding leads to believe that the theme in question has an important role in academic texts, making them relevant for reception and production. In the academic field, Paquot (6; pp.101-119) brings to light the English as a foreign language students' ability of producing phraseological units, identifying an excessive use of structures in their native language. The author suggests that more teaching materials should be published so as to incorporate information based on language observation in scientific texts, for example. Biber, Conrad and Reppen (4; pp.17-119) believe that pattern descriptions in different linguistic genres identify characteristics in the genres, through similarities or differences. However, there is a need of a large-sized corpus to detect important events.

Besides the possibility of language study through isolated words, there is another path to be followed along with phraseological units. In relation to the language field, the addition of these items in pedagogical materials can bring benefits to learners, based on empiric data collected in actual usage. Since the use of any idioms by its users provide an infinite generation of data, it can be inferred that such amount of data can be analyzed according to researchers' interests in many areas, including Corpus Linguistics.

In a new linguistic construct, not only does it offer a set of computerized techniques to the verification of traditional phenomena connected, for example, to the lexicon, semantics or syntax; on the contrary, the analysis of a corpus can revel facts about a language which may have never been thought. This way, such line of thought proves to have an exploratory orientation –being not only a new methodology of language studying, but also a new way of research.

Sinclair (8; pp.79-83) argues that the observation of real language turns out to be a safe way of describing a language, providing the study of a range of patterns, many of which unexpected. Without

empiric data, the task of indicating if certain linguistic phenomenon happens naturally in a context becomes hampered.

By taking into consideration what was stated, the approximation between Phraseology and Corpus Linguistics to study academic texts provides other paths to linguistic exploration, adding empiric approachto the theoretical reflection.

With the framework of Phraseology and Corpus Linguistics, we identified and analyzed the academic- scientific phraseological units in English, shared by eight major areas of knowledge (Exact and Earth Sciences, Biological Sciences, Engineering, Health Sciences; Agrarian Sciences, Applied Social Sciences, Linguistic, Literature and Arts). As we sought to demonstrate, the simple absolute frequency of the phraseological units would not suffice to investigate the most common ones. The mere observation of the normalized frequenciesin the different corpora would not be enough to state that a phraseological unit has a uniform occurrenceamong different areas.

References:

1.Biber, D., Conrad, S., & Reppen, R. (2004). Register variation and English for specific purposes. In D. Biber, S. Conrad, & R. Reppen. Corpus Linguistics: investigating language structure and use (pp. 135-171). Cambridge, MA: Cambridge University Press.

2.Biber, D., Conrad, S., & Reppen, R. (2004). Register variation and English for specific purposes. In D. Biber, S. Conrad, & R. Reppen. Corpus Linguistics: investigating language structure and use (pp 103-106). Cambridge, MA: Cambridge University Press.

3.Biber, D., Conrad, S., & Reppen, R. (2004). Register variation and English for specific purposes. In D. Biber, S. Conrad, & R. Reppen. Corpus Linguistics: investigating language structure and use (pp 178-189). Cambridge, MA: Cambridge University Press.

4.Biber, D., Conrad, S., & Reppen, R. (2004). Register variation and English for specific purposes. In D. Biber, S. Conrad, & R. Reppen. Corpus Linguistics: investigating language structure and use (pp 107-119). Cambridge, MA: Cambridge University Press.

5.Coxhead, A. (2008). Phraseology and English for academic purposes: challenges and opportunities. In F. Meunier, &S. Granger (Ed.), Phraseology in foreign language learning and teaching (p. 149-156). Philadelphia, PA: John Benjamins Publishing Company.

6.Coxhead, A. (2008). Phraseology and English for academic purposes: challenges and opportunities. In F. Meunier, &S. Granger (Ed.), Phraseology in foreign language learning and teaching (p. 160-164). Philadelphia, PA: John Benjamins Publishing Company.

7.Ellis, N. (2008). Phraseology: the periphery and the heart of language. In F. Meunier, & S. Granger (Ed.), Phraseology in foreign language learning and teaching (p. 1-13). Philadelphia, PA: John Benjamins Publishing Company.

8.Hunston, S. (2010). Corpora and language teaching: issues of language. In S. Hunston. Corpora in applied linguistics (p. 137-169). Cambridge, MA: Cambridge University Press.

9.Hunston, S. (2010). Corpora and language teaching: issues of language. In S. Hunston. Corpora in applied linguistics (p. 101-112). Cambridge, MA: Cambridge University Press.

10.Sinclair, J. (2004a). How to build a corpus. In M. Wynne (Ed.), Developing linguistic corpora: a guide to good practice (p. 79-83). Oxford, UK: Oxbow Books.

11.Sinclair, J. (2004b). Trust the text: language, corpus and discourse. New York, NY: Routledge.