



**Списание за наука**

**„Ново знание“**

ISSN 2367-4598 (Online)

ISSN 1314-5703 (Print)

*Академично издателство „Талант“*

*Висше училище по агробизнес и развитие на  
регионите - Пловдив*

**New Knowledge**

**Journal of Science**

ISSN 2367-4598 (Online)

ISSN 1314-5703 (Print)

*Academic Publishing House „Talent“*

*University of Agribusiness and Rural Development  
Bulgaria*

<http://science.uard.bg>

## **A CORPUS-BASED STUDY OF LEXICAL ITEMS IN BULGARIAN AND ENGLISH ECONOMICS RESEARCH ARTICLES**

**Ivaylo Dagnev**

*Medical University –Plovdiv, Medical College, Bulgaria*

**Abstract:** The article offers a comparative analysis of the lexical means used in Bulgarian and English research articles with an economic focus. The research is based on a broad theoretical foundation, using a genre and terminological toolkit, alongside a linguistic analysis of lexical items. Methodologically, the research is based on two separate corpora of articles in Bulgarian and English language articles of almost equal number of words. The most frequent lexical units are established, and an analysis of lexically-related groups - collocations, coligations and lexical bundles is also offered. The importance of these lexical devices in the context of textual organization and for the terminological system of the genre of the economics research article is emphasized.

**Key words:** economics research articles, comparative study, Bulgarian, English, lexical items

# КОРПУСНО ИЗСЛЕДВАНЕ НА ЛЕКСИКАЛНИ ЕДИНИЦИ В БЪЛГАРО И АНГЛОЕЗИЧНИ НАУЧНО- ИЗСЛЕДОВАТЕЛСКИ СТАТИИ В ОБЛАСТТА НА ИКОНОМИКАТА

Ивайло Дагнев

*Медицински университет – Пловдив, Медицински колеж*

**Резюме:** Статията предлага съпоставителен анализ на лексикалните средства използвани в български и англоезични научно-изследователски статии с икономическа насоченост. Изследването се базира на широка теоретична основа, използваща жанров и терминологичен инструментариум, наред с езиков анализ на лексикалните единици. В методологически план, разработката се базира на два отделни корпуса от статии в български и англоезични статии с приблизително еднакъв брой думи. Установяват се най-честотните лексикални единици, като се предлага и анализ на лексикално-свързаните групи – колокации, колигации и лексикални снопове. Изтъква се значението на тези лексикални средства в контекста на текстуалната организация и за терминологичната система на жанровата форма на икономическата научно-изследователска статия.

**Ключови думи:** научно-изследователски статии, икономика, сравнително езикознание, български, английски, лексикални единици

## I. INTRODUCTION

### 1. The concept of genre

The perspective of genre leads to the view of a text as a member of a category: a culturally recognized artifact, a grouping of texts according to some conventionally recognized criteria, or according to specific, culturally defined goals. Genres are categories established by consensus within a culture and may change over time when the circumstances change. category: a culturally recognized artifact, a grouping of texts according to some conventionally recognized criteria, or according to specific, culturally defined goals. Genres are categories established by consensus within a culture and may change over time when the circumstances change. The main definition that Swales (1990:58) gives to genre is: "... a set of communicative events, participants sharing common communicative goals, recognizable by the expert members of the discursive community."

### 1.1. Genre analysis and the language of science

In the 1980s, many scholars such as Miller (1984) and Bhatia (1987) began researching genre analysis, in an effort to solve the problems of written scientific discourse, but it was Swales' book, "Genre Analysis" (Swales 1990), which put the field on a firm scientific footing. As a result of Swales' work, genre analysis has become an important component in the development of language skills, mainly because the recognition and understanding of genre characteristics help to create predictable expectations about the context in which a text is being realized. Through this type of analysis, texts themselves can be segmented into specific elements that make the compositional structure and the discourse markers that constitute them recognizable.

The frame Swales (1990) offers for studying genre comprises three phases. The first phase concerns the discourse community. Every discussion of genre analysis must be considered through the lens of this community. The second, inextricably linked to the former are the conventional expectations associated with genre. In terms of research articles (RAs), this parameter indicates the degrees of identification, contextualization and representation of any economics problem or hypothesis, methods of research, results and explanations, and the probable solution to the

problem. Thirdly, according to Swales (1990: 58), the third phase implies possible pedagogical benefits, as language learners need a contextual and situational orientation defined by genre characteristics. This is of particular importance to studying a foreign language for specific or academic purposes, as the acquisition of language skills is thus identified by mastering certain genre forms.

In purely linguistic terms, researchers describe the basic characteristics of the language of science as accuracy, clarity, logical train of thought, objectivity and informativity. The language of science also presupposes consistency and argumentation of the exposition, standardization and a relative authorial distance from the explicated object of study. The scientific text then is a system of strict laws in which text typology is built from the bottom up - the publication of higher information levels is built up as a publication of lower-level publications (e.g. the sum of thematically related articles is a prerequisite for their unification in a monograph, etc.) (Staneva 2001: 370).

## **1.2. Genre forms in academic writing. General outline**

In terms of academic genres, the main genre forms can be subdivided according to their length:

- short-length genre forms: annotation, abstract, short review, report, presentation;
- medium length genre forms: article, scientific message, lecture;
- long-length genre forms: monograph, dissertation, study, reference book.

The main feature of all these genre forms is formal standardization, which is directly related to the cultural and scientific traditions of the respective country, as well as to the intrinsic relationships associated with a scientific field. Over the last decades, globalization has evolved into "global" criteria in terms of the structural and lexical and grammatical requirements of most genres in response to the creation of global databases with scholarly metrics such as Scopus, Web of Science, Google Scholar and others.

## **1.3. The subgenre of research article**

The article is the academic genre heavyweight; whose structure serves as the mainframe of all other genre forms. The former could be either a research article or a review article. Swales (1990) first focuses on the ideational characteristics of the RA - he introduces the notion of rhetorical move. A rhetorical move indicates "a change in information flow" (Swales 1990: 140-141) - in his opinion, the main function of a research article is to convince the reader of the significance of the presentation of the scientific product through publication. Swales (ibid.) gives an example with the "Introduction" section of the RA, pointing out that it is made up of formal and structural components (moves) that underlie the rhetorical function of the scientific text. For example, the rhetorical *Move: Establishing a Territory* aims to delineate and identify a scientific study in a particular context; *the Move: Indicating a Gap* seeks to establish uncharted territory in the relevant field of science, thus underscoring the significance of the study.

### **1.3.1. Principal characteristics of academic and economics research articles**

As mentioned above, scientific articles are divided mainly into two groups: original (or research) and overview. Original scientific articles are called primary literature, while review papers are part of the secondary scientific literature, summarizing the results of published research, analyzing books, or advancing science. There are marked differences between the two types of article, although according to Davis (2015), the review article has a similar structure to the research one. However, in our opinion, the structure of the review article, is much more *laissez faire* and does not always conform to the generally accepted and standardized IMRD structure of the research variant.

With regard to the business domain, there are essentially two sub kinds of economics research papers: empirical papers, which run data through a model (a series of mathematical equations); and theoretical papers, which begin with a model based on certain premises and then prove that certain outcomes will ensue. These two kinds of papers reflect what the discipline considers to be legitimate economics. Both kinds of papers typically have abstracts and both are written in sections with subheadings. The format of sections and their content have become fairly standard over the years. For empirical papers, the sections (with some variation) include: Introduction; [Literature Review]; Data; Methodology/Model; Results; Conclusion. For theoretical papers, there will be an Introduction and a section presenting the basic model. Then there will be usually subsections and further sections discussing various scenarios that can be acted out as the basic model is extended. Theoretical papers abound in math, and usually the proofs are put in an appendix.

## 2. Corpus analysis in the frame of comparative linguistics

The focus on comparative linguistics, as a thorough linguistic analysis of academic genres, is inconceivable without probably the best mechanism contemporary linguistics has to offer - corpus analysis. The main objectives of such an analysis are highlighting those linguistic features that can be conducive to a better understanding of academic discourse and consequently leading to an improved written production.

As Johansson (2003: 5) notes, one of the most serious problems before comparative analysis is equivalence, in other words, how to know what to compare after different languages conceptualize notions and their lexical representations to a large extent differently. One method, gaining ever increasing popularity over the past two decades, is corpus research, which is also possible thanks to the improvement of computers technologies. Its advantages are undeniable, since the statistical processing of large texts and lexical arrays gives representativeness that goes beyond boundaries of intuition and subjectivity.

Strictly speaking, the definition of the term "corpus" is complex and does not necessarily involve computer processing, as the very idea of orientation of linguistic research towards pragmatics and descriptiveness reflects much wider trends in the social development of societies, such as, for example, the processes of globalization, mass communication, acceleration of dynamism in communication etc. According to Biber (1993: 245), "the corpus is, in principle, an organized body of texts in their natural environment". For Sinclair and Moranen (2006: 23):

"The corpus is a collection of texts in electronic form, selected according to certain external criteria to represent linguistic diversity as a source for linguistic analysis."

When we look at other definitions of corpora, we come across several common traits shared by numerous authors (Garside, Leach, Sampson, 1987; Sinclair, 1996; Stefanovic, Grice, 2006):

- a) the corpus is composed of language material (written text, speech fragments, etc.);
- b) this material is selected and organized according to explicit linguistic criteria and is authentic;
- c) the material serves as a sample of a more comprehensive linguistic expression;
- (d) the material is usually electronically processed;
- e) serves for linguistic analysis.

In terms of data collection principles, there are two main approaches: **the corpus-based approach** and **the empirical corpus approach** (corpus-driven approach). The first approach is the so-called top-down, i.e. the textual material is organized on the basis of certain theoretical principles, while the second method relies on linguistic modeling based on already extracted empirical material (Tognini-Bonelli, 2001). In reality, both methods are not so different (McEnery, Jiao, Tono, 2006: 9):

"It appears that the empirical corpus approach is not so different from corpus-based – while the latter is assumed to isolate theory from the data or standardize it to fit the theory, the first filters

the data by ostensibly randomized samples, although there are no guarantees that the corpus has not been selectively examined to non-conforming (to the theoretical model - b. m.) data should be avoided."

If we take a closer look at the empirical corpus model, we will see that even if information is extracted on the basis of already received data, the very perspective of observation and interpretation necessitates a certain type of inference which is dependent on preliminary concepts.

### **3. Vocabulary in economics research articles**

The vocabulary characterizing economics research articles can be viewed in the light of English for academic purposes (EAP) which is a rich and fast moving area of endeavor, growing rapidly beyond the historical areas of English as a first language context. Academic vocabulary can be seen as a layer of vocabulary that occurs across a range of academic subject areas, meaning users would encounter this vocabulary in biology or chemistry, linguistics or law. There is argument about whether there is a core of academic vocabulary (Hyland & Tse 2007), for example, because the occurrence of specific words in a particular area does not necessarily mean that these items are used in the same way or with precisely the same meaning in different subject areas.

On the other hand, the subject area of economics and business is too large to be encompassed in a single study. There are so many subspheres of applied economics that analyzing in detail its terminological vocabulary is an extremely tall order. Moreover, in practical terms, specialized vocabulary in a given subject area is often the least problem of a researcher, so what may really count as a true investigative analysis are those terms that cut across all areas, or the great majority of the latter, the so-called **metaterms** (Mavrodieva, Tisheva 2014:123), which could be divided into two classes:

a) **General metaterms**, such as: administrative, aspect, authority, classification, competence, consumption, distribution, empirical, globalization, indicator, methodology, parameter, qualitative, scenario, technological, validity, etc.

b) **Specific (to economy) metaterms**, such as: auction, audit, bankruptcy, benchmark, commodity, entrepreneur, recession, turnover, unemployment, etc.

### **4. The terminological angle of academic writing**

Precision and accuracy are maybe the two terms that best encapsulate the idea of scientific style when dealing with high abstract level texts and it is terminology that is deemed as the most immediately salient feature of specialized communication. Terminology in science dictates the need for such precision. Today, science depends heavily upon highly specialized and ever expanding technical vocabulary, so inaccurate rendering of even a few terms can mar interpretation significantly. Moreover, the coining of new terms by researchers is ongoing as a measure of scientific advancement, involving new discoveries and development of new sub disciplines. What can be applied from Cabre's theory of terminology (Cabre, 1999) are these criteria regarding the linguistic aspect:

a) Terms are lexical units and can participate in lexical and syntactic structures, as well as being limited in terms of lexical combinatorics based on the principles limiting all lexical units;

b) Formally, they can coincide with lexemes from the general discourse;

c) The meaning of the terms is discrete, ie. clearly delineated within the specialized area; Standardization is a key concept in terminology, which involves two aspects:

a) unifying and fixing each referent, and

b) unifying and standardizing its designation (Sager 2001: 255).

Terms differ from words in that the former are endowed with a special form of reference, they refer to "a discrete conceptual entity, constituting the knowledge space of a particular subject field" (Sager 2001: 261). Thereby, to mark the specificity of reference, agreements (in the discourse communities) are bound to be reached over the precise meaning and linguistic forms by means of the processes of regularization, harmonization and standardization (ibid). Hence, special

subjects can be differentiated from general knowledge by the nature of reference, which subsumes the use of additional concepts. The difference, as Sager (ibid) mentions, is one of degrees, rather than one of two extremes. As far as terms are concerned, behind each one there should ideally stand a clearly defined concept, systematically related to other concepts that make up the knowledge structure of a text or a discourse area in question. The common lack of polysemy within one specialized domain accounts for the context independency of terms. In terminology, the polysemic nature of language is limited; terms are the result of convention, they are formed on the basis of an agreement among specialists in a given field, and are motivated by the intention to facilitate communication in the domain (Sager 1990: 56-57).

### 5. Lexical features characterizing ERA

In terms of lexical features, we have to discern between **single lexical items** and **multiword lexical items** – the latter being **collocations**, **colligation** and **lexical bundles**. According to the Oxford dictionary, a collocation represents “the habitual juxtaposition of a particular word with another word or words with a frequency greater than chance”. A colligation, in its turn, is a set of words arranged in the way they function in a syntactic structure--i.e., forming a syntactic pattern, eg.

*Example: I allowed him to leave.*

The grammatical structure is **to allow s.o. to do sth.** So, the main difference between collocation and colligation is that collocation is a grouping of two or more words that are commonly used together in English, whereas a colligation is a grouping of words based on certain grammatical patterns. The basic idea of colligation is that just as a lexical item may be primed to co-occur with another lexical item, so also it may be primed to occur in or with a particular grammatical function. Alternatively, it may be primed to avoid appearance in or co-occurrence with a particular grammatical function (Hoey, 2005:43).

According to Hoey (2005: 43) colligations actually encompass three distinct aspects of distributional attraction between linguistic items: (i) the relationship between a lexical item and a grammatical context (e.g. [consequence + BE + subordinate clause]; Hoey, 2005: 57–58), (ii) the relationship between a lexical item and a particular syntactic function in which the item can be used (e.g. consequence is often used as part of a complement; Hoey, 2005: 44–48), and (iii) the relationship between a lexical item and the position in a phrase, clause, sentence, text or discourse where the item can be used (e.g. consequence is often used as part of the theme in a sentence; Hoey 2005: 49–52). Thus, Hoey (2005) uses colligation as a cover term which encompasses both grammatical patterns and patterns of information structure associated with a lexical item.

Traditionally, collocation has been considered as a relationship between only two lexical elements (Jones and Sinclair, 1974; Sinclair, 1987). Therefore, the frequent co-occurrence of more than two words has often been treated as a separate phenomenon (however, see Kjellmer 1984 and Smadja 1993). Frequently occurring multiword combinations have been labeled in different ways in the linguistic literature, e.g. *lexical bundles* (Biber & Conrad, 2009), clusters (Kenny, 2000: 99) and *multi-word strings* (Mauranen, 2000: 120). In some cases, it has been argued that these multiword combinations are reducible to binary collocations. However, there is evidence to suggest that not all three-item-collocations can be split up into two basic constituents and that multiword collocations might, in fact, be quite common (see the discussion in Siepmann 2005: 417).

Many linguistic studies of academic discourse in the past few decades have focused on the description of multi-word expressions (eg. Nattinger and DeCarrico, 1992; Lewis, 1993; Schmitt, 2004), challenging the traditional dichotomy between grammar and vocabulary. These studies are part of a long line of research that goes back to Firth (1951), Altenberg (1998), Sinclair (1991), suggesting that language in use relies to a great extent on prefabricated word patterns (“how are you”, “as it were”), and that the ability to use them appropriately is a significant component in

successful language acquisition. Following Altenberg (1993, 1998), who used a frequency-driven, fixed word approach to identify multi-word sequences, Biber et al. identified lexical bundles as specific type of multi-word expressions, and defined them as “recurring sequence of three or four words” (Biber et al., 1998: p. 990). Biber’s lexical bundle approach has been applied in many subsequent corpus studies in academic writing.

Lexical bundles have several specific characteristics that distinguish them from the other multi-word sequences, such as idioms, collocations and colligations. By definition, they are extremely common, not idiomatic in meaning, and not perceptually salient (Biber and Barbieri, 2007, p. 269). They are clusters of 3-6 contiguous words identified in electronic corpora exclusively on the basis of their frequency in a certain register, may be parts of clauses (eg. “I don’t want to”) or phrases (eg. “in the case of”), and usually do not represent complete structural and semantic units. Biber et al. (1999) suggest that 15% of the bundles in conversation are structurally complete, while only 5% in the academic prose are complete structural units. The bundles in the academic prose are “phrasal rather than clausal” (Biber et al., 2004, p. 382), with 70% of the bundles consisting of noun phrase expression (eg. “the nature of the”), or a sequence that bridges across two prepositional phrases (eg. “as a result of”). Hyland (2009) states that most bundles in academic writing are parts of noun and prepositional phrases, and reports different patterns across disciplines: social sciences (business studies and applied linguistics) employ more bundles beginning with a prepositional phrase, while science and engineering texts use more passive bundles. Jalali et al. (2015) found that the largest structural category of lexical bundles in medical research articles was prepositional phrases. Cortes (2004) also reports that the majority of the bundles used in academic history writing in English and Spanish are prepositional phrases.

**6. Aim of the research:** The paper strives to analyze in a comparative fashion specific vocabulary (metaterms) in Bulgarian and English language economics research articles.

## II. METHODS

With regard to scope, the paper centers on **economics research articles** as these are the mainframe of academic production in the field of economics. The study employs both corpus-based approach and empirical corpus approach methodologies as its purpose is conducting a detailed study of the selected language material. On the other hand, due to our desire for verifiability, we restricted ourselves to sources of open access or those that can be extracted on the basis of institutional subscription. Regarding languages, there are two corpora – one is composed of articles, published in Bulgarian journals, in online version (with the abbreviation BLC - 37 articles), and the other from articles in English language journals (abbreviated ELC - 16 articles). Our guiding principle has been the number of words, not that of articles, as the ones in English are more than double in size. As far as the BLC is concerned, we have extracted articles from “Ikonomika 21”, “Biznes upravljenie”, “Dialog”, “Ikonomicheski I sotsialni alternative”, “Novo znanie”, etc. All the articles were published over the last two years. The ELC have been compiled with articles available at ScienceDirect: “Journal of Development Economics”, “Economics: The Open-Access” and “Research Papers in Economics and Finance” and also published in above-mentioned period. The areas of research of both groups of articles, viz. BLC and ELC is quite varied – they encompass both highly theoretical ones, heavily reliant on mathematics and statistics and ones delving into practical economic applications of various business spheres.

For the sake of analysis, we have abstained from building very large corpora. Nevertheless, the total count of words in all corpora is 394594 words, which is a formidable number and accounts for excellent representativeness by this criterion. The total number of words in BLC is 192819, while in ELC the number is 201 775. All articles were initially scanned into PDF files and then into word processing files for ease. For data processing, we used *WordSmith*, version 6.0 (Scott 2012), freely available, employing all three of its features: Concordance, KeyWord and WordList, which allow the monitoring of word contexts and repetitive lexical constructions. To this end, the

three corpora were merged into three respective files. Separately, we also used a second word processing program, *AntConc*, due to the fact that different programs sometimes give ambiguous figures for the number of words, phrases, etc., especially when referring to “unpopular” word processing software in lesser-used in science languages like Bulgarian.

### III. RESULTS AND DISCUSSION

As said in the Introduction, we have carried out textual analyses of both corpora in terms of **single word terms** and **multiword terms**, viz. collocations, colligations and lexical bundles. With regard to single word items, we present the lemma not the specific word form.

**Single word items** with the highest number of tokens in BLC are:

- nouns - *predlagane*<sup>12</sup>, *potrebiteli*, *razvitie*, *produkti*, *danni*, *biznes*, *process*, *faktura*, *resursi*, *nachin*, *upotreba*, *informatsia*, *infrastruktura*, *tehnologii*, *otnoshenie*, *upravlenie*, *riteil*, *oborudvane*, *zaemi*, *iztochnik*, *onlain*, *izpolzvane*, *targovia*;
- adj – *okolna*<sup>3</sup>, *ikonomicheski*, *bitovi*, *osnoven*, *evropeiski*, *targovski*, *elektronen*, *strategicheski*, *opredelen*, *planiran*;
- verbs<sup>4</sup> – *tryabva*, *imam*, *izpolzvam*, *pozvolyavam*, *reshavam*, *vazdeistvam*, *osiguryavam*;
- other<sup>5</sup> – *na*, *kato ot*, *pri*, *varhu*, *koeto*, *drugi*, *kakto...taka*.

**Single word items** with the highest number of tokens in ELC are:

- nouns – *demand*, *index*, *percent*, *level*, *season*, *rebate*, *treatment*, *effect*, *discount*, *production*, *increase* (n), *price*, *product*, *purchase*, *input*, *expenditures*, *result*, *impact*, *policy*, *land*, *offer*, *cost*, *extension*, *subsidy*, *response*, *payout*, *investment*, *survey*, *preference*, *information*, *loss*;
- adj – *average* (adj), *increasing*, *based*, *significant*, *purchased*, *statistical*;
- verbs – *be*, *may*, *increase*(v), *have*, *find*, *cover*, *take*, *can*, *offer*, *consume*, *cost* (v), *result*(v), *ensure*, *respond*.

All presented single word items in both corpora fall within the category of two hundred most frequently used words according to AntConc (4.1.0.).

In terms of collocational patterns, the most prominent ones in BLC are of the type *adj + noun*:

- *vazhna rolya*<sup>6</sup>, *narastvasht natisk*, *choveshki resursi*, *okolna sreda*, *kraini potrebiteli*, *visoka vazvrashhtaemost*, *zhiznen tsikal*, *mestni danatsi*, *visoki /niski dohodi*

With the *noun + (prep) (adj) noun* pattern (post modification - in Bulg. - *nesaglasuvano opredelenie*)

- *tegloto na otpadatsite*<sup>7</sup>, *konsumatsia na energiya*, *infrastrukturata na...*, *rolya v prinosa*, *prinos za razvitie*, *problemite na okolnata sreda*, *protsepa na predlagane*, *vazmozhnosti za razvitie*

With the combination *noun + verb*:

- *igrae rolya*<sup>8</sup>, *postavyam tseli*, *izgrazhdam strategiya*, *(lihvenite) protsenti se povishavat*

<sup>1</sup> All words in Cyrillic are transcribed. Their translations are provided in footnotes.

<sup>2</sup> Supply, consumers, development, products, data, business, process, invoice, resources, means, use, information, infrastructure, technologies, relation, management, retail, equipment, debts, source, online, application, trade

<sup>3</sup> environmental, economic, domestic, main, European, commercial, digital (online), strategic, definite, planned

<sup>4</sup> must (have to), have, use, allow, decide, influence, ensure

<sup>5</sup> of (on, about, etc), such as, from (by), upon (over), which, other, as...as

<sup>6</sup> important role, increasing pressure, human resources, environment, end users, high return, life cycle, local taxes, high/low revenues

<sup>7</sup> the weight of waste, energy consumption, infrastructure of..., contribution role, contribution to development, environmental problems, supply process, opportunities for development

<sup>8</sup> play a role, set a goal, build a strategy, interest rates are rising

In the ELC the most frequent ones are also of the type *adj + noun* (premodifier + head) or three- element multiword phrases:

- (adj + noun) - *risk aversion, insurance demand, insurance product, index insurance, extension office, significant difference, treatment effect, purchasing power/coverage/insurance, demand response, percent increase, insurance contract, developing countries, insurance product, baseline survey;*

- (three-element multiword items) – *statistically significant result, farm level characteristics, resource constrained farmers, insurance coverage period, traditional theoretical predictions, demand response curves, average treatment effects.*

With the combinations *noun + verb* or *verb + noun*:

- *estimates are, inflation is rising, reject the bill, purchase the insurance, results suggest*

With multi-unit structure such as (*det*) + *premodifier* + (*prep*) + *premodifier* + ... + *head*:

- *type of dry spell, nature of the incentive, presence of basis risk, indemnity based crop insurance*

We have found a number of colligations in both corpora:

- in BLC<sup>9</sup>: *(neshto) e kritichno vazhno, natiskat za plashtane narastva, vyzmozhnostite (za razvitie) se povishavat;*

- in ELC: *demand is (initially) increasing, insurance makes them (worse), dry spell occurs...(adv), spending..... to increase, sth no longer hold(s).*

In terms of lexical bundles, after the exclusion of those containing proper nouns and context-based bundles (e.g. the Ministry of Finance, in the European Union), a total of 112 different bundles were included in the analysis in the BLC. With regard to ELC, the number of lexical bundles is higher, as we have identified a total of 127 different items.

Table 1. Most frequent lexical bundles in BLC and ELC

Bundles in BLC <sup>10</sup>	Frequency (in tokens)	Bundles in ELC	Frequency (in tokens)
1. <i>ot druga strana</i>	46	1. <i>on the other hand</i>	44
2. <i>po tozi nachin se</i>	27	2. <i>as a result of</i>	37
3. <i>po odnoshenie na</i>	25	3. <i>as well as the</i>	33
4. <i>v zavisimost ot</i>	22	4. <i>at the same time</i>	27
5. <i>v rezultat na</i>	19	5. <i>on the basis of</i>	22
6. <i>s pomoshhta na</i>	19	6. <i>a result of the</i>	18
7. <i>ot znachenie za</i>	17	7. <i>the total number of</i>	18
8. <i>ot тази гледна точка</i>	15	8. <i>in the process of</i>	16
9. <i>v razvitiето na</i>	11	9. <i>the size of the</i>	14
10. <i>po sashtnostta si</i>	11	10. <i>the level of</i>	14
11. <i>za opredelyane na</i>	10	11. <i>the development of the</i>	11
12. <i>e edin ot</i>	10	12. <i>in the case of</i>	11
13. <i>ot edna strana</i>	9	13. <i>is one of the</i>	10

The findings from this research have demonstrated that bundles are “important building blocks of discourse associated with basic communicative functions” (Biber et al., 2004, p. 400), that “different registers rely on different sets of lexical bundles” (Conrad & Biber, 2005, p. 69), and that grammatical structure of lexical bundles is a distinct characteristic of registers (Biber et

<sup>9</sup> sth is of critical importance, the pressure for paying off is increasing, the opportunities for development are increasing

<sup>10</sup> on the other hand, in this way..., with respect to, according to, as a result of, with help of, of importance to, from this point of view, in the development of, in its essence, to determine the..., is one of

al., 1998). Lexical bundles play also a significant role in text and discourse construction as most of them, including the most frequently used in the corpora, serve as discourse markers. Discourse markers (DMs) are the intersection points in a text, performing rhetorical functions, marking the boundaries of steps and moves in RA, signifying a change in the information flow, signaling alterations in an author's stance. They are a subcategory of grammatical cohesion which join clauses and sentences together to form longer discourse (Richards & Schmidt, 2002). From what we have extracted in the corpora, we can say that both Bulgarian and English papers rely heavily on DMs, as some of the most frequently used lexical bundles, such as *ot druga strana/on the other hand, po otnoshenie na, ot тази gledna tochka, on the basis of, at the same time* etc. serve as DMs in terms of textual organization.

We have deliberately studied colligations and lexical bundles under a separate cover because we want to emphasize the crucial part these two categories play in what it means to know a language. As far as the frequency of occurrence of all three types of multiword expressions is concerned, both BLC and ELC can be said to be rich in formulaic language, which suggests a high level of regularization and standardization in both respective RA subgenres.

#### IV. CONCLUSION

In spite of its short volume and limited scope, the research provides a relatively detailed analysis in a comparative fashion of specific vocabulary (metaterms) in Bulgarian and English language economics research articles. It can be stated that these lexical elements are vital in constructing the rhetoric function of the studied articles in the corpora. The research also provides some quantitative data as far as lexical bundles are concerned. A possible follow-up investigation may look into the larger structures of text organization on the level of morpho-syntax, such as Tense, Aspect, syntactic structures, information flow (theme and rheme patterns) and on the plane of discourse – discourse markers, hedging, stance verbs, shell words, etc.

#### REFERENCES

1. Мавродиева, И., Тишева, Й., (2014). Академично писане за докторанти и постдокторанти. Университетско издателство „Св. Кл. Охридски“, София. 207 с. (Mavrodiieva, I., Tisheva, Y. Academic writing for PhD students and postdocs. University Publishing House "St. Cl. Ohridski", Sofia. 207 p.)
2. Станева, Хр. (2001) Стилистика на българския книжовен език. Абагар. В. Търново. (Staneva, Hr. Stilistika na bulgarskiya knizhoven ezik (Stylistics of the Bulgarian literary language. Abagar. V.Turnovo.)
3. Altenberg, B. (1993). Recurrent Verb- Complement Constructions in the LondonLund Corpus. In J. Aarts, P. de Haan and N. Oostdijk (eds), English Language Corpora: Design, Analysis and Exploitation (227–45). Amsterdam: Rodopi.
4. Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations'. In A. P. Cowie (ed.), Phraseology: Theory, Analysis, and Applications (101–22). Oxford: Oxford University Press.
5. Bhatia, V. (1987). Language of the law. Language Teaching, vol. 20, no. 4, pp. 227-234.
6. Biber, D. (1993). Representativeness in corpus design. Literary and Linguistic Computing, vol. 8, pp. 243-257.
7. Biber, D. and Barbieri, F. (2007). Lexical bundles in university spoken and written registers. English for Specific Purposes 26(3):263-286. DOI: 10.1016/j.esp.2006.08.003
8. Biber D., Conrad, S. (2009). Register, Genre, and Style. Cambridge. Cambridge University Press.
9. Biber, D., Conrad, S. and Cortes, V. (2004). If you look at...: Lexical Bundles in University Teaching and Textbooks. Applied Linguistics 25/3: 371-405. Oxford University Press.

10. Biber, D., Conrad, S., Reppen, R. (1998). *Corpus Linguistics, Investigating Language Structure and Use*, Cambridge. Cambridge University Press
11. Cabre Castellvi, M.T. (1999). *Terminology, Theory, Methods, Applications*. Translated by Janet Ann DeCesaris. Amsterdam: John Benjamins.
12. Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*. DOI: 10.1515/9783484604674.56
13. Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23(4):397-423. DOI: 10.1016/j.esp.2003.12.001
14. Davis, R. (2015). A genre analysis of medical research articles. PhD Thesis, University of Glasgow, p. 293. <http://theses.gla.ac.uk/6724/>
15. Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford.
16. Garside, R., Leach, G., and Sampson, G. (1987). *The Computational Analysis of English: A Corpus-Based Approach*. Longman, Inc., New York.
17. Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Psychology Press, p.202.
18. Hyland, K. (2009). *Academic Discourse: English in a Global Context*. London: Continuum.
19. Hyland, K., Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied linguistics*, 25(2), pp.156-177.
20. Jalali, Z.S., Moini, M.R., Arani, M.A. (2015). Structural and Functional Analysis of Lexical Bundles in Medical Research Articles: A Corpus-Based Study. *International Journal of Information Science and Management*. Vol. 13, No. 1, 2015, 51-69
21. Johansson, S. (2003). Contrastive linguistics and corpora. In S., Granger, J. Lerot, & S. Petch-Tyson, *Corpus-based approaches to contrastive linguistics and translation studies*, Amsterdam: Rodopi, pp. 31-44.
22. Jones, S. & Sinclair, J. (1974). English lexical collocations - A study in computational linguistics. *Linguistics*, pp. 15-61.
23. Kenny, D. (2000). Translators at play: exploitations of collocational norms in German-English translation. In *Working with German corpora*, Edited by: Dodd, B. 143–160. Birmingham: University of Birmingham Press.
24. Kjellmer, G. (1984). Why Great: greatly but not Big: \*Bigly? *Studia Linguistica* 38(1):1 - 19
25. Lewis, M. (1993). *The lexical approach*. Hove, UK: Language Teaching Publications.
26. Mauranen, A. (2000). Strange Strings in Translated Language: A Study on Corpora. In M. Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects* Manchester: St. Jerome Publishing, 119 -141
27. McEnery T., R. Xiao, Y. Tono. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. John Benjamins. Published online: 08 December 2006 <https://doi.org/10.1075/ijcl.11.4.09won>
28. Miller, C. R. (1984). Genre as Social Action. *Quarterly Journal of Speech* 70:151-167
29. Nattinger, J., DeCarrico, J. (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press, 1992. Pp. xvi 218.
30. Richards, J.C. & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics* (third ed.), Pearson Education, London.
31. Sager, J. (2001). Terminology. Standardization. In *Routledge Encyclopedia of Translation Studies*, ed. M. Baker. Routledge, London.
32. Sager, J. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
33. Schmitt, N. (2004). *Formulaic Sequences Acquisition, processing and Use*. Introduction. John Benjamins.

34. Siepman, D. (2005). Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography*, Volume 18, Issue 4, pp. 409–443, <https://doi.org/10.1093/ijl/eci042>
35. Sinclair, J. (1987). The Nature of the Evidence in J. Sinclair (ed.) *Looking Up: An Account of the COBUILD Project in Lexical Computing*. Collins ELT: 150–159
36. Sinclair, J. (1996). The search for units of meaning. *Textus* 9/1: 75-106. (Reprinted in *Sinclair 2004*: 24-48.)
37. Sinclair, J. (1991). *Corpus Concordance, Collocation*. Oxford University Press
38. Sinclair, J.M., Mauranen, A. (2006). *Linear Unit Grammar*. John Benjamins.
39. Smadja, Fr. (1993). Retrieving Collocations from Text: Xtract. *Computational Linguistics*, 19(1):143–178.
40. Stefanowitsch, A., & S.T. Gries (2006). *Corpus-based approaches to metaphor and metonymy*. Berlin: Mouton de Gruyter.
41. Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: CUP.
42. Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins

#### REFERENCE DICTIONARIES

Oxford dictionary, (<https://www.lexico.com/definition/collocation>)