THE JOURNAL OF TEACHING ENGLISH FOR SPECIFIC AND ACADEMIC PURPOSES
Vol. 9, Nº 1, 2021, pp. 71–84

UDC: 811.111'255.4:371.3 https://doi.org/10.22190/JTESAP2101071D

Original scientific paper

ASSESSING LEARNERS' ACADEMIC PHRASEOLOGY IN THE DIGITAL AGE: A CORPUS-INFORMED APPROACH TO ESP TEXTS

Andreea Dincă, Mădălina Chitez

West University of Timișoara, Romania

Abstract. *In the field of English for Specific Purposes (ESP), as in any other type of interlanguage, phraseology contributes significantly to successful academic writing (Biber and Barbieri 2007). For particular learner varieties, such as Romanian English, few studies have examined formulaic sequences (Hyland 2008), mainly focusing on lexico-grammatical patterns (Chitez 2012, 2014). The proposed paper investigates the use of phraseology in Romanian students' academic papers, written during their ESP courses, by adopting a double contrastive perspective: first, we contrast texts produced in two different disciplines (Literature Studies and Information Technology), and second, we compare academic phraseology in learner language with native speaker phraseology. For the analysis we have compiled two corpora (ESP-LIT and ESP-IT), each consisting of 40 texts representing a discipline specific didactic genre - essay. For reference we used the Academic Phrasebank (Davis and Morley 2018). The aim is to find out whether the use of academic formulaic expressions differs according to the discipline and the extent to which students integrate expert academic phrases into their writing. The methodology can be replicated for different language learning settings.*

Key words: *phraseology in ESP, discipline-specific phraseology, corpus-based contrastive phraseology, ESP in Romania*

1. INTRODUCTION

Using English in educational and professional settings is a reality that everybody acknowledges and embraces (Nelson et al 2020). From early school years until later in life, learning and using English either in teaching scenarios or in everyday situations have become usual global citizen's activities. Educational institutions, in particular, have been challenged to provide group-adapted solutions to the growing need of mastering English language skills. English for Specific Purpose (ESP) with its multiple extensions or sub-disciplines (e.g. English for Academic Purposes - EAP) is one of these solutions. As a "learner needs-based approach" (Belcher 2009, 3), the field of ESP incorporates versatile teaching methods that can address an array of language related concerns, such as the "macro- (rhetorical, whole-text) and micro- (lexicogrammatical) level characteristics of the written and spoken genres" (ibid. 4), or any other applied linguistics aspect useful to the English language learners.

Submitted September 30th, 2020, accepted for publication November 9th, 2020
Corresponding author: Andreea Dinca. West University of Timișoara, Romania | E-mail: andreea.dinca@e-uvt.ro

At the lexicogrammatical level, phraseology plays an essential role considering the fact that mastery of academic phases supports ESP learners in their effort to become expert academic writers rather than native-like writers (Römer and Arbor 2009). Academic phrases have been investigated and compiled by numerous scholars either in their realization as multiword units (Szudarski 2018, 75), as lexical bundles extracted with the support of n-gram approaches (e.g. Hyland 2008), or as phrase phrases, i.e. p-frames (Golparvar and Barabadi 2020). Even if academic phrase lists are not easily compiled, nor can be always integrated into "teaching practice" (Granger 2017, 23), their importance for teaching and learning processes cannot be disputed (Biber and Barbieri 2007).

In this study, we scrutinize the use of phraseological units in ESP writing from a double contrastive perspective, using corpus linguistics approaches: on the one hand, our aim is to identify salient differences, if any, between two disciplines, and on the other, to establish whether the use of such phrases in the analysed L2 writing is different from the use of phraseology in English L1 writing.

2. ESP IN CONTEXT

3.1. Varieties of English

As already mentioned, ESP is only one of the varieties of English learners are exposed to during teaching practice (Belcher 2017). Several typologies of English as a foreign language are present in pre-university, university curricula and further education contexts: (a) English as a Foreign Language (EFL) is the umbrella term for processes involving the learning of English as a non-native language by speakers of other languages. EFL implements learner-centered methods, where learners "are given a meaningful role in pedagogic decision making by being treated as active and autonomous players" (Nosratinia and Zaker 2014, 1). The domain closest to EFL in point of general linguistic competence building is (b) English as a Lingua Franca (ELF). What distinguishes them is "the nature of speakers' goals: EFL is considered successful when it converges to a target model, ELF when it is mutually intelligible" (Hülmbauer 2009, 328). (c) English Language Teaching (ELT), is the teaching counterpart of EFL. ELT focuses on the professional aspects of the field, numerous studies emphasizing the English teachers' professional identity (TPI), which is shaped by the process of teaching English (Pennington 2014). (d) Teaching English to Speakers of Other Languages (TESOL) is also a teacher-oriented branch which "was traditionally a discipline led by teachers" (Rose 2019, 898). Criticism has been raised that, in recent years, there is "wave of new theory surrounding teaching, (e.g. translanguaging, ELF-aware pedagogy) but many of these ideas are generated by researchers, and are yet to be accompanied by a matching volume of teacher input on how new perspectives can improve their language teaching practices" (ibid.).

Furthermore, the domain of ESP itself can be subdivided into several other branches that represent rather specialized language domains:

> *There are, and no doubt will be, as many types of ESP as there are specific learner needs and target communities that learners wish to thrive in. Perhaps the best known of these (especially among language educators who are themselves most often situated in academia) is EAP, or English for Academic Purposes, tailored to the needs of learners at various, usually higher, educational levels (see Hyland, 2006, for an excellent overview of EAP issues and practices). Less well known (to many academics) and potentially more diversified, given the breadth and variety of*

> *the worlds of work, is English for Occupational Purposes (EOP). The fastest growing branches of EOP are those associated with professions that are themselves constantly expanding and generating offshoots, such as English for Business Purposes (EBP); English for Legal Purposes (ELP); and English for Medical Purposes (EMP). There are also numerous other less well known but equally intriguing varieties of EOP, such as English for Air Traffic Controllers, English for Tourist Guides, English for Horse Breeders, and English for Brewers.* (Belcher 2009, 2).

Our study reflects phenomena of higher education ESP practice which are relevant for the ESP and TESOL communities, if we refer to practical aspects of academic writing examples in context. It also aims at describing EFL language features characterizing the Romanian undergraduate learners.

3.2. ESP in Romania

Ever since the Romanian education made the shift from communism-imposed foreign language teaching (i.e. primarily Russian as a Foreign Language) towards English as a Foreign Language, which happened in the first decade after the fall of the dictatorship in 1989, school and university curricula have been constantly enriched with English language lessons and courses. English is prioritized in all educational cycles and students study it not only as a compulsory subject but also during additional optional classes. As a result, English is quite frequently, one of the usual assessment disciplines in high school graduation exams (i.e. Romanian *Bacalaureat*) but also at the university. At the same time, little has been done in the direction of creating research-based teaching materials targeting the Romanian student learners. In particular, academic writing issues have been largely ignored.

> *Even though Romanian higher education has adhered to the Bologna Process and by law No. 288/2004 students are expected to write theses to graduate from each of the three university cycles, academic writing teaching in Romania is not guided by educational policy and writing support is provided according to each university's internal policies.* (Bercuci and Chitez 2019, 736)

The use of digital analyses including corpora to assess ESP phenomena in the Romanian user groups has been scarcely attempted. A recent study by Ene and Sparks (2020) presents some of the most significant research contributions in the area of EFL writing, in Romania, in general, without a particular focus on corpus linguistics methods. Several other recent studies by Bercuci and Chitez (2019) and Chitez and Bercuci (2019, 2020), indicate that, while ESP students can use corpora as authentic language materials to consult and improve their academic writing, ESP teachers can create learner corpora either for salient feature identification or student self-correction.

3.3. Phraseology in ESP: digital analysis outcomes

There is a growing body of literature that recognizes the importance of phraseology in mastering a foreign language. The majority of these studies use large sets of data and digital corpora that can be analysed and assessed using digital tools (see section 4.1.2.). For example, Hyland (2008a, 2008b) shows that multiword units play a crucial role in language learning and fluent linguistic production (p. 4). In addition, Paquot (2017) argues that word combinations of various types have a major role in areas like "language acquisition, processing, fluency, idiomaticity and change language acquisition" (p. 122). In the context of academic writing, previous research has demonstrated the importance

and usefulness of phraseological units (e.g. Chen and Baker, 2010; Simpson-Vlach and Ellis, 2010; Ädel and Erman 2011, Paquot and Granger, 2012). It has been argued that multiword units are an essential component of writing in specific fields of study or registers. Furthermore, mastering key phrases/multiword units shows the degree of communicative competence of a certain member in that particular field of study (Hyland 2008, p. 5). Therefore, in order to navigate within a certain academic register and discipline, one should be able to use typical recurrent word combinations successfully. In an attempt to identify the most used phraseological units in academic writing, scholars have created lists of salient formulaic sequences that appear in academic writing, such as: Academic Formulas List (AFL, (Simpson-Vlach and Ellis, 2010) and Academic Collocation List / ACL (Ackermann and Chen, 2013).

4. ANALYSIS

4.1. Data, tools and methods

4.1.1. Data

For the purpose of this study, we have compiled two learner corpora, ESP-IT and ESP-LIT, containing student academic writing from the West University of Timisoara, Romania. They are a sub-set of the Romanian Corpus of Academic Genres (ROGER[1]), a bilingual (i.e. Romanian-English) comparable corpus currently under construction. The configuration of the corpora for this study is:

(a) Corpus of English for Specific Purpose in the Information Technology Discipline / ESP-IT – it amounts to 63,842 tokens and consists of 40 texts written during their ESP classes by undergraduate students (1st and 2nd year), enrolled in the Informatics faculty. The texts included in the corpus are papers that students write regularly and represent across-the-curriculum didactic genres such as essay and scientific paper.

(b) Corpus of English for Specific Purpose in the Literature Studies / ESP-LIT - it amounts 67,529 tokens and consists of 40 texts written by postgraduate students of Literature Studies (1st and 2nd year). The texts that make up the corpus are discipline-specific essays.

ESP-IT and ESP-LIT are largely comparable, with one characteristic that distinguishes them: The Information Technology corpus consists of papers written by undergraduate students, whereas the Literature Studies corpus contains postgraduate level papers. However, considering that one of the objectives of the study is to identify salient phraseology features that characterise student writing in different disciplines, a different language proficiency level (undergraduate versus graduate) might support saliency detection (Granger and Bestgen 2014).

4.1.2. Tools

For the corpus-based analysis, we used tools such as the Lancsbox (Brezina, Weill-Tessier and McEnery 2020) package, and the programming language, Python (Van Rossum and Drake 2009). Lancsbox was used for concordancing (using the *KWIC* function) and for generating 4 and 5-Ggram lists (using the *N-gram* function) from each of the two learner corpora. Python was used to: (1) compare the two disciplines based on their 4 and

[1] More information at: https://roger.projects.uvt.ro/.

5-Gram lists and (2) to identify the extent to which students use expert phrases in their writing.

4.1.3. Methods

Corpus-based contrastive analysis

As mentioned before, the primary aim of the present study is to use digital methods (i.e. corpus linguistics) adapted for linguistic research in order to find out whether the use of academic formulaic expressions differs according to the discipline. In order to do this, we compared discipline-specific corpora and extracted phraseology shared by both of them.

The following steps were taken in order to conduct the analysis:

- First, the phraseological units were extracted using the *lexical bundle* (LB) approach: LBs are multiword units defined as "frequently recurrent strings of uninterrupted word-forms" (Hyland, 2008, p. 5). They are extracted automatically from the data set using the *N-gram* function of corpus concordance tool packages. With the help of Lancsbox's *N-gram* function we extracted the most frequent 4- and 5-grams from the two learner corpora. The cut-off frequency threshold was set at 0.25 times per ten thousand words. The dispersion criterion was also addressed, an N-gram had to occur in at least 2 texts. After extraction, the N-gram lists were saved into machine-readable files (.csv).
- Then, we used a self-developed (see below) Python programme to compare the ESP-IT 4 and 5-Gram lists with their corresponding N-gram lists from the ESP-LIT corpus and extract the common phraseological units.

Python programming for phrase list comparison

Acknowledging the importance of programming skills in research, the current paper presents two cases of using Python for linguistic research, performed with novice Python programming skills. We decided to implement individualized approaches to handling our data as there were no already available tools to perform the necessary analyses. We were supported, during the process, by an Information Technology engineer who improved our solutions when necessary.

Our first analysis aimed at extracting and comparing N-gram lists from ESP-IT and ESP-LIT with the aim of discovering the common phrases in the two corpora. In order to do this, we needed to extract 4 and 5-Grams from the two corpora and compare the correspondent lists two by two. The function of extracting the N-grams, taking into consideration relative frequency and dispersion rate, is already provided by packages for language data analysis, such as Lancsbox (Brezina, Weill-Tessier and McEnery 2020), so it was not necessary to write a Python program to perform this task. However, for the comparison of the lists, there was no tool available, and a procedure like this done manually would have been a very tedious task. Therefore, we tried to find an alternative to the manual analysis, by writing a Python program. Several steps were undertaken: First, the N-gram lists that also contained the raw and normalized frequency were saved in .csv files. Then, with the help of a *for loop* we searched for phrases in ESP-IT that also existed in ESP-LIT and if their relative frequency was higher than 2. In this way, we identified the N-grams shared by both corpora. We stored our results in a json format file because this type of file is suitable for data storage, being easily buildable and readable. The output JSON file is displayed in Figure 1.

```
{
   "5-gram":"when it comes to the",
   "IT_frequency":4.0,
   "IT_relative_freq":0.63,
   "LIT_frequency":5.0,
   "LIT_relative_freq":0.74
},
{
   "5-gram":"of this paper is to",
   "IT_frequency":3.0,
   "IT_relative_freq":0.47,
   "LIT_frequency":3.0,
   "LIT_relative_freq":0.45
}
```


Fig.1 Output file of n-gram comparison

In the second type of analysis, we wanted to find out if expert phrases were to be found in the student corpora. We compiled an expert phrase list that contained 36 academic phrases specific to short paper introductions, using the Manchester Academic Phrasebank (Davis and Morley 2018). Our idea was to try to find the strings of words contained in every phrase of the expert list in the two learner corpora. The first step we took was to set up a rule by which the program would identify words in the data. So, we established that Python would consider a word the text which has an empty space before and after it. However, we had a problem with the words that were followed by a punctuation mark, because even if the word was similar to one in the expert list, it was not taken into consideration if it had the punctuation mark just after it: the program was modified in order to omit the punctuation marks.

One challenge that this program raised was that, during the initial stage of the analysis, we got numerous irrelevant results. A learner phrase was considered compatible with the expert phrase if it displayed only one or two words contained by the expert phrases searched, such as “this”, “paper” or “which” or the program picked up words that were used in the sentences, but which were not used to form the searched phrases. We tried several solutions to improve this. First, we modified the program in order to make sure that when a certain phrase was chosen, the words occurred close enough to each other in order to actually form the phrase searched. Then, we chose to display only the sentences that contained at least 60 % of the expert phrase searched (e.g. at least 4 out of 7 words).

```
"fileName":"ESPLIT1006.txt",
      "data":[
         {
            "sentenceSearched":"In this paper, I argue that",
            "context":"  In this paper, I will try to look at the impact t
hat immigration has made in the American society, especially in the nowada
ys political and social context, which is well known to present a rather h
ostile attitude towards immigrants \u2013 be they legal or illegal ",
            "precision":"83.33333333333334%",
            "listFoundWords":[
               "In",
               "this",
               "paper,",
               "I",
               "that"


"sentenceSearched":"The aim of this paper is to",
            "context":"  Given the issues that the keyword proposed to ana
lysis provides, the aim of the following paper is to highlight the ambival
ence of the term, as it is presented in a series of reliable sources which
 are meant to portray different views regarding the subject of immigration
 ",
            "precision":"71.42857142857143%",
            "listFoundWords":[
               "aim",
               "of",
               "paper",
               "is",
               "to
```


Fig. 2 Output JSON file expert phrase analysis

We did this also because we were aware of the fact that students might use variations of the expert phrases. Finally, we added a list of stop words (with the purpose to not be taken into consideration by the program) that contained words like *to*, *which*, *that*, *in*. The results were stored in a JSON file (Figure 2). We plan on improving this program because the results still needed considerable manual checking: for example, the text highlighting was done manually.

4.2. Analysis and results

The first part of the analysis investigated whether the use of academic formulaic expressions in the academic papers written in English L2 (i.e. ESP) by the Romanian students is similar in the two disciplines, Information Technology and Literature Studies. With the help of the Python programme, we compared the 4 and 5-gram lists from the two disciplines. In this way, we identified the 4 and 5-grams shared by both disciplinary discourses (displayed in Table 1) as well as the discipline-specific ones (displayed in Table 2).

4.2.1. Common N-grams in ESP-IT and ESP-LIT

As Table 1 indicates, the 4 and 5-grams which occur in both disciplines are context-independent discourse-organizing structures. The most frequent 5-grams that are shared by the two groups are "*due to the fact that"* and "*one of the most important"*. The 4-grams shared by the two discipline-specific ESP learner corpora are more numerous than

the 5-grams, the two corpora sharing fifteen common 4-grams, among which *"on the other hand", "as well as the"*, or *"as a result of"*.

Table 1 4- and 5-Grams shared by ESP-IT and ESP-LIT

5-Gram	ESP-IT freq.	ESP-IT rel. freq.[2]	LIT freq.	ESP-LIT rel. freq.
due to the fact that	6	0.94	7	1.04
one of the most important	4	0.63	4	0.59
when it comes to the	4	0.63	5	0.74
of this paper is to	3	0.47	3	0.45
4-Gram	ESP-IT freq.	ESP-IT rel. freq.	LIT freq.	ESP-LIT rel. freq.
one of the most	16	2.51	25	3.71
when it comes to	15	2.35	18	2.67
is one of the	14	2.2	13	1.93
on the other hand	10	1.57	7	1.04
due to the fact	6	0.94	8	1.19
the fact that it	5	0.78	4	0.59
as well as the	5	0.78	11	1.63
in this case the	5	0.78	3	0.45
the fact that the	4	0.63	11	1.63
we can say that	4	0.63	4	0.59
as a result of	4	0.63	5	0.74
it comes to the	4	0.63	5	0.74
is based on the	3	0.47	3	0.45
for the first time	3	0.47	4	0.59
take a look at	3	0.47	3	0.45

4.2.2. Discipline specific N-grams in ESP-IT and ESP-LIT

Our results show that discipline specific phrases are prevalent in both corpora: they are both topic related and discourse-organizing structures. The phrases found in ESP-IT refer to approaches specific to the hard sciences, where the authors are rather concerned about presenting information about the procedures they use (e.g. *"for the running time is"*) and display results *(*e.g. *"table has the following structure")*.

The texts from the Literature Studies corpus, on the other hand, are rich in argumentative structures (e.g. *"it is safe to say"*) and topic-related phrases (*"prose in the 20th century"*). A selection of the most used discipline specific phrases in ESP-IT and ESP-LIT can be observed in Table 2.

[2] Relative frequency, normalized per 10k words (pttw)

Table 2 ESP-IT Discipline specific N-grams

Topic related	Rel freq.[3]	Discourse-organizing	Rel freq.
ESP-IT		ESP-IT	
5-Grams		*5-Grams*	
the array is already sorted	1.1	is one of the most	1.1
the creation of personal computing	0.63	to be the most inefficient	0.79
in ascending or descending order	0.63	we can see that the	0.63
when it comes to sorting	0.63	tasks for efficient use of	0.47
to kill a mockingbird is	0.74	in the context of the	0.60
the beginning of the novel	0.60	and the way in which	0.60
the united states of America	0.60	of the ways in which	0.60
in the early twentieth century	0.45	is one of the reasons	0.60
4-Grams		*4-Grams*	
the array is already	1.57	in this paper we	1.41
the number of elements	1.1	to be the most	0.94
the running time of	1.1	we can see the	0.94
the time complexity of	1.1	for the purpose of	0.47
in the American society	1.19	in the case of	2.23
in the novel is	0.89	the way in which	2.09
the status of the	0.89	in the face of	2.09
in favor of the	0.59	is the fact that	1.49

The results point to another interesting trend as well: Figures 3 and 4 display clear difference between the two disciplines in the use of content-related and discourse-organizing N-grams. It looks like IT students are more inclined to use content-related phrases, whereas literature students show a clear higher interest in using discourse-organizing structures that help them support their claims.

Fig. 3 Dispersion of 5-Grams in ESP-IT and ESP-LIT

[3] Relative frequency, normalized per 10k words (pttw)

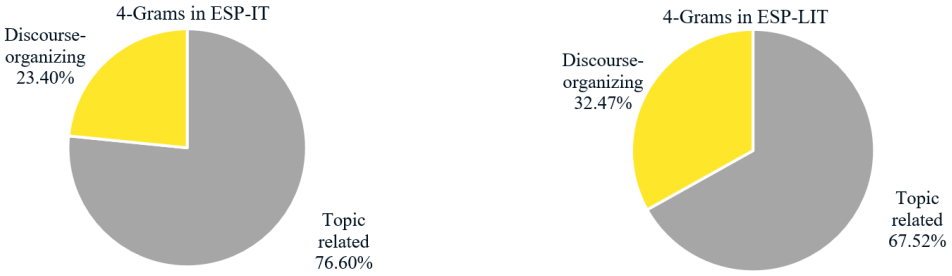

Fig. 4 Dispersion of 4-Grams in ESP-IT and ESP-LIT

4.2.3. Expert phraseology in ESP student writing

Our analysis also revealed that students show a certain degree of familiarity with using variations of expert phrasal elements. However, it could be noticed that they use variations of a limited number of phraseological constructions.

Similarities in the use of expert phrases in ESP-IT and ESP-LIT

A formulaic phraseological construction that proved to be frequent in both disciplines, is the structure that contains *This paper* in initial position, followed by various words such as *aims, illustrates, contains*, etc., as can be seen in Table 3. It occurs in 5/40 texts in ESP-IT and in 3/40 texts in ESP-LIT. Students also prefer variations of the expert phrases that best fit their context. A higher than two words matching rate between the expert list and the learner corpora was found in only two instances: *"**This paper** also **aims** to collect"* and *"**This paper presents** a theoretical introduction"* (ESP-IT). The other constructions used are "*This paper"*, followed by other words than the ones contained in the Academic Phrasebank (Morley 2014).

Table 3 Similarities in ESP-IT and ESP-LIT

Expert Phrase	ESP-IT	Occ.	ESP-LIT	Occ.
This paper*	*makes* a comparison between *presents* a theoretical introduction and practical comparison *is about* the time performance of *contains* the description of *also aims to* collect	5	*illustrates* the ways in which the term *illustrate* how the term *will focus* on the *focuses on* discussing	4
This research paper*	*makes* a comparison between	1	-	0

Differences in the usage of expert phrases in ESP-IT and ESP-LIT

Interestingly, there are also differences in the ways in which the students from the two analysed corpora use the expert formulaic phrases displayed in Table 4. Structures that contain "*In this paper/essay*" in initial position are used by students to explain what they

will do in their papers. The difference in usage, however, lies in the way they use pronouns in their writing. Information Technology students use the first person plural, *we*, even if their paper is single-authored, whereas the Literature Studies ones prefer the first person singular *I*.

Another difference that can be observed is the way in which IT and LIT students choose to state the aim of their paper. The students in ESP-IT used more varied words such as *"the goal/purpose/ scope of this paper",* whereas the students in ESP-LIT seem to prefer the "*the aim of this paper*". However, as the two learner corpora are fairly small, a larger dataset could provide a better understanding of this trend. Other phraseologies not included on the expert list: "The structure of the paper is as following:"; "The structure of this paper is as follows".

Table 4 Differences in ESP-IT and ESP-LIT

	ESP-IT	Occ.	ESP-LIT	Occ.
	use of **"we"** versus **"I"**			
In this paper *	***we** are going to compare* ***we** investigate the performance of* ***we** will use the presentation of the* ***we** will implement the mentioned*	4	***I** will focus on the current challenges* ***I** will try to look at the impact that immigration*	2
In my paper*	-		***I** am going to focus on*	1
In this essay*	-		***I** will argue that*	1
	use of **"goal"** versus **"aim"**			
The * of * paper	***The goal of this paper** is to compare* ***The purpose of this paper** is to go a little in-depth on* ***The scope of this paper** is to analyse and compare*	3	*Therefore, **the aim** of this paper is to show that* *Given the context, **the aim** of the following paper is to highlight* ***The purpose of this paper*** is to study how	3

5. Discussion and Conclusions

Although the two sets of data (i.e. ESP discipline-specific corpora) capture linguistic phenomena that are representative for a certain learner population at a specific time during their English learning process, they efficiently support digital analyses resulting in frequency-based and automatic retrieval examples. The two approaches that we implemented, the n-gram contrastive analysis and Python-programmed algorithm for the automatic comparisons, have offered us an overview upon phrase use phenomena in the two learner groups.

The typology of the phrases analysed with the help of corpora is dependent on the phrase extraction method. That is to say that, when standard n-gram analysis is performed, the co-occurring units are not all particularly relevant for academic writing conclusions. That is why we decided to divide the n-grams into two groups: discourse-organizing n-grams versus topic-

related n-grams. By doing that, several salient features of academic phrase use in the two disciplines could be identified: first, the IT students used considerably more content-related phrases (both 4- and 5-grams) than discourse organizing phrases compared to the Literature Studies students. Quite interestingly, the author-role phraseology used by the IT students indicates preferences towards the use of the 1st person plural "we" rather than the 1st person singular "I", like the philology students use in their texts. Considering that the corpus is composed of the same genre, namely essay, it can be asserted that is indeed a discipline-specific phraseology characteristic. At the same time, there are evident similarities in the way students of both disciplines use academic phrases, such as the "*due to the fact that"* and "*one of the most important*. They seem to be the most frequent phrases in both corpora, which highlight linguistic choices that can be explained either through pedagogical practice (i.e. same type of academic phrase-use guides and training) or through the teddy bear phenomenon (Hasselgren 2007) manifesting in certain interlanguage groups.

The second type of analysis, where we developed a Python programme to compare learner phrases with expert phrases extracted from the Academic Phrasebank (Davis and Morley 2018) revealed that the learners use academic phrases that are present in the native-speaker list (e.g. *this paper + presents*/*illustrates*/*makes a comparison*), but, on the other, the degree of lexical variation is quite low: both groups of students (Informatics and Philology) use only a few of the academic phrases that are present in the Academic Phrasebank. Another distinctive feature seems to be the selection of particular lexical components of discourse-organizing multiword units: while students in Informatics use a more varied range of nouns forming introductory phrases (e.g. the goal/aim/scope), in literary essays lexical variation within this type of phrases was lower and quite repetitive (e.g. preference for "aim" as support noun).

The results of the study can be exploited pedagogically by ESP teachers in two different configurations. One of them would be to understand which ESP learner group needs guided assistance for a particular language learning component (e.g. Chitez 2017). For example, it seems that IT students need better practice in academic conventions (i.e. discourse organizing phrases) than philology students. Or, teachers could choose to replicate this approach, build their own learner corpora and conduct corpus analyses in order to identify which aspects of language need further assistance. In this way, ESP research-based teaching (Lehtonen 2018) can be supported and best-practice examples shared and improved.

REFERENCES

Ädel, Annelie. *Metadiscourse in L1 and L2 English*. Amsterdam: Benjamins, 2006.

Belcher, Diane. "What ESP Is and Can Be: An introduction." In *English for Specific Purposes in Theory and Practice*, edited by Diane Belcher, 1- 20. Ann Arbor, MI: University of Michigan Press, 2009.

Belcher, Diane. "Recent developments in ESP theory and research: Enhancing critical reflection and learner autonomy through technology and other means". In *Synergies of English for specific purposes and language learning technologies*, edited by Nadezna Stojković, M, Tošić, and V. Nejković, 2-19. Newcastle upon Tyne, UK: Cambridge Scholars, 2017.

Bercuci, Loredana, and Madalina Chitez (2019). "A corpus analysis of argumentative structures in ESP writing". *International Online Journal of Education and Teaching* (IOJET) 6, no. 4 (2019): 733-747.

Biber, Douglas, and Federica Barbieri. "Lexical bundles in university spoken and written registers." *English for specific purposes* 26, no. 3 (2007): 263-286. https://doi.org/10.1016/j.esp.2006.08.003.

Brezina, Vaclav., Pierre Weill-Tessier and Anthony McEnery (2020). *#LancsBox* (version 5.x). http://corpora.lancs.ac.uk/lancsbox, 2020

Chen, Yu-Hua and Paul Baker. "Lexical Bundles in L1 And L2 Academic Writing." *Language Learning & Technology* 14, no. 2 (June 2010): 30-49. http://dx.doi.org/10125/44213

Chitez, Madalina, and Loredana Bercuci. „Calibrating digital method integration into ESP courses according to disciplinary settings". *New Trends and Issues Proceedings on Humanities and Social Sciences*, 7, no. 1 (2020), 20-29.

Chitez, Madalina, and Loredana Bercuci. „Data-driven learning in ESP university settings in Romania: multiple corpus consultation approaches for academic writing support". In *CALL and complexity – short papers from EUROCALL 2019*, edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley and Sylvie Thouësny, 75-81. Research-publishing.net, 2019.

Chitez, Madalina. "Corpus-informed vocabulary teaching: a case study on energy discourse in German". In *Doing Applied Linguistics. Enabling Trandisciplinary Communication*, edited by Daniel Perrin and Ulla Kleinberger, 238-243. Berlin: de Gruyter, 2017.

Chitez, Madalina. *Learner corpus profiles: the case of Romanian learner English*. Linguistic Insights Series (Series Editor: Maurizio Gotti). Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien: Peter Lang, 2016.

Chitez, Madalina. "Lexical frequency profile applications on learner corpora: a Romanian learner English explorative analysis." In *New trends and methodologies in applied English language research II: Studies in variation, meaning and learning. Linguistic Insights Series,* Vol. 145 (Series Editor: Maurizio Gotti), edited by Paula Rodríguez-Puente, David Tizón-Couto, Beatriz Tizón-Couto and Iria Pastor-Gómez: 15-36. Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien: Peter Lang, 2012.

Davis, Mary, and John Morley. "Facilitating learning about academic phraseology: teaching activities for student writers." *Journal of Learning Development in Higher Education*. Special Edition (2018): 1- 17.

Ene, Estela, and Sydney Sparks. "EFL Writing in Romania. Reflections on Present and Future". In *Education beyond Crisis. Challenges and Directions in a Multicultural World*, edited by Daniela Roxana Andron and Gabriela Gruber, 217–230. Brill | Sense, 2020.

Golparvar, Seyyed Ehsan, and Elyas Barabadi. "Key phrase frames in the discussion section of research articles of higher education". *Lingua: International review of general linguistics* 236 (March 2020): 102804. https://doi.org/10.1016/j.lingua.2020.102804.

Granger, Sylviane. "Academic phraseology a key ingredient in successful L2 academic literacy". In *Academic Language in a Nordic Setting - Linguistic and Educational Perspectives*, edited by Ruth Vatvedt Fjeld, Kristin Hagen, Birgit Henriksen, Sofie Johansson, Sussi Olsen and Julia Prentice. Oslo Studies in Language 9, No. 3 (2017): 9-27.

Granger, Sylviane, and Yves Bestgen. "The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study." *International Review of Applied Linguistics in Language Teaching* 52, no. 3 (2014): 229-252. https://doi.org/10.1515/iral-2014-0011.

Hasselgren, Angela. "Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary". *International Journal of Applied Linguistics*, 4, no. 2 (2007), 237-258.

Hyland, Ken. "As can be seen: Lexical bundles and disciplinary variation". *English for Specific Purposes* 27, no. 1 (2008a): 4-21. doi: 10.1016/j.esp.2007.06.0017.

Hyland, Ken. "Academic clusters: text patterning in published and postgraduate writing" *International Journal of Applied Linguistics* 18, no. 1 (2008b): 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x.

Hyland, Ken. „English for academic purposes: An advanced resource book". London: Routledge, 2006.

Hülmbauer, Cornelia. „'We don't take the right way. We just take the way that we think you will understand' - The shifting relationship of correctness and effectiveness in ELF communication". In *English as a lingua franca: studies and findings*, edited by Mauranen, Anna and Ranta, Elina, 323-34. Newcastle upon Tyne: Cambridge Scholars Press, 2009.

Lehtonen, Tuula. "Practitioner Research as a Way of Understanding my Work: Making Sense of Graduates' Language Use". In *Key Issues in English for Specific Purposes in Higher Education*, edited by Yasemin Kirkgöz and Kenan Dikilitaş, 129-140. Cham: Springer, 2018.

Nelson, Cecil L., Zoya G. Proshina, and Daniel R. Davis. *The Handbook of World Englishes*. Hoboken, NJ: Wiley Blackwell, 2020.

Nosratinia, Mania, and Zaker, Alireza. „Metacognitive attributes and liberated progress: The association among second language learners' critical thinking, creativity, and autonomy". *SAGE Open* 4, no. 3 (2014): 1-10. doi: 10.1177/2158244014547178

Paquot, Magali. "The phraseological dimension in interlanguage complexity research." *Second Language Research* 35, no. 1 (2019): 121-145. https://doi.org/10.1177/0267658317694221.

Paquot, Magali, and Sylviane Granger. "Formulaic language in learner corpora." *Annual Review of Applied Linguistics* 32, no. 32 (2012): 130-149. http://doi.org/10.1017/S0267190512000098.

Pennington, Martha, and Barbara Hoekje. "Framing English language teaching". *System* 46 (2014): 163-175. https://doi.org/10.1016/j.system.2014.08.005.

Römer, Ute, and Ann Arbor. "English in Academia: Does Nativeness Matter?" *Anglistik: International Journal of English Studies* 20, no. 2 (September 2009): 89-100.

Rose, Heath. Dismantling the Ivory Tower in TESOL: A Renewed Call for Teaching-Informed Research. TESOL Quarterly 53, no. 3 (September 2019), 895-905.

Simpson-Vlach, Rita, and Nick C Ellis. "An academic formulas list: New methods in phraseology research." *Applied linguistics* 31, no. 4 (2010): 487-512. https://doi.org/10.1093/applin/amp058.

Szudarski, Paweł. *Corpus linguistics for vocabulary. A guide for research*. London and New York: Routledge, 2018.

Van Rossum, Guido and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.