

Corpus Linguistics Meets Academic Writing: Examples of Applications in the Romanian EFL Context



Mădălina Chitez

Abstract Corpus-based academic writing studies have been increasingly used to verify hypotheses regarding processes of university writing and learning. In the Romanian context, research in the areas of academic writing and corpus linguistics has been relatively scarce. Academic writing in Romanian is not explicitly taught, whereas academic writing in English is part of the curricula of a major or minor in English. The Romanian corpus linguistics field is mainly represented by the Romanian Academy Research Institute for Artificial Intelligence Institute (RACAI) whose activity consists of the creation of corpora to support natural language processing (NLP) investigations. There are only few learner and specialized corpora available for research. In the present chapter, the Romanian Corpus of Learner English (RoCLE) is used in order to exemplify the manner in which corpora can be used in academic writing classes. Three topics have been selected for exemplification: contrastive linguistics, academic phraseology, and move analysis. For each topic, a brief description of the theoretical background with relevance for the Romanian context is given, followed by examples of corpus-based analyses extracted from RoCLE. Based on the same examples, pedagogical recommendations indicate possible directions of corpus use in teaching academic writing.

Keywords Corpus linguistics · Academic writing · Academic writing teaching · Genre-based corpus research · Romanian corpora · EFL

M. Chitez (✉)

Faculty of Letters, History and Theology, Department of Modern Languages and Literatures,
West University of Timișoara, Timișoara, Timiș, Romania

e-mail: madalina.chitez@e-uvt.ro

© Springer Nature Switzerland AG 2018

M. Chitez et al. (eds.), *University Writing in Central and Eastern Europe:*

Tradition, Transition, and Innovation, Multilingual Education 29,

https://doi.org/10.1007/978-3-319-95198-0_10

1 Introduction

Since almost all activities at the university are connected in some way to academic writing, the amount of information related to student writing is vast. New methodologies, from such areas as corpus linguistics, writing research, or academic literacy, can shed light on a range of areas relevant to university writing teaching and learning: Which are the typical genres of a certain discipline? How and what do students learn with the help of a certain genre? Which linguistic features shape the profile of a text, and to what extent are they identifiable? How do the rhetorical features that characterize academic writing differ from one discipline to another?

Academic writing teaching and corpus linguistics are increasingly joining forces in interdisciplinary approaches that analyze, evaluate, and optimize student writing (Aull 2015; Biber et al. 2007; Flowerdew 2005; Flowerdew and Forest 2015; Gotti and Giannoni 2014; Hyland 2009; Nesi and Gardner 2012; Römer and O'Donnell 2011; Swales 2004; Upton and Connor 2001). In the Romanian university context, such approaches have never been used extensively, let alone integrated.

2 The Context: Academic Writing in Romania

Since the Fall of Communism, Romania has gone through a long process of transition from the Stalinist norms in education to new developments in higher education that were developed in Western countries (see also chapters “[Academic writing at Babeş-Bolyai University. A Case Study](#)”, “[Institutional Writing Support in Romania: Setting Up a Writing Center at the West University of Timișoara](#)”, and “[Perceptions About “Good Writing” and “Writing Competences” in Romanian Academic Writing Practices: A Questionnaire Study](#)”). In 25 years of continuous reforms in higher education, connected with the rise in number of disciplines and specializations (Chitez 2014, pp. 21–23) and the growing importance of English, the necessity of adaptation to new writing requirements emerged:

For cultures such as Italy and Romania, it is more a problem of introducing such new genres that interfere less with traditional ones but reach acceptance and provide a certain degree of comfort for all actors. Genre awareness and a deeper understanding of what genres accomplish in education are factors that supposedly play a crucial role for creating new teaching directives in the future. (Chitez and Kruse 2012, p. 175)

At Romanian universities, neither academic genres nor academic writing are taught explicitly. Faculty do not address, in general, issues of genre writing or writing process challenges due to a widely accepted view that learning to write is something specific to elementary and secondary education (ibid.). In fact, the majority of informative materials on written genres in the Romanian educational context refer to three primary categories: creative writing (e.g., *compunerea*, “composition”), argumentative writing (e.g., *comentariul literar*, “literary commentary”), and formal writing (e.g., *scrisoare*, “letter”; ibid., pp. 172–173). At the university level, it is the disciplinary setting that influences the academic writing performance: Students in

the humanities, especially those studying foreign languages, have a greater chance of being exposed to structured genre-use and production training than students in the engineering disciplines, for instance. Almost all foreign-language departments offer practical modules to students where they learn how to write common genres such as stories, journal articles, and formal letters. Typically, such courses take place in foreign languages but not in Romanian, so that it may be a greater challenge for the Romanian students to handle academic texts in their mother tongue than in a foreign language.

Only in the recent past have academic writing research and support initiatives in Romania been launched. The SNF-SCOPES project LIDHUM (see chapter “[Studying and Developing Local Writing Cultures: An Institutional Partnership Project Supporting Transition in Eastern Europe’s Higher Education](#)”), coordinated by Otto Kruse and Mădălina Chitez, was one of the few initiatives in the field. The statistical analysis performed after the implementation of the EUWRIT survey (Chitez et al. 2015a) indicated that the repertoire of educational genres is similar in Eastern European countries but some genres are culture specific: for example, the seminar paper in Romania (Bekar et al. 2015, p. 130). Another project, OPEN RES,¹ was conducted at the Babeş-Bolyai University of Cluj and has resulted in several publications on academic writing training. Scattered publications on academic writing in Romania cannot easily be accessed (e.g., Andronescu 1997; Pavlenko and Bojan 2014) or need to be reinforced by extensive empirical evidence (e.g., Frăţila (Pungă) 2006).

3 Romanian Text Corpora

In Romania, corpus linguistics has been given very little attention compared to what it has received in other international research contexts, especially North America and Western Europe. It should be emphasized that by *corpus linguistics* we are referring to the research discipline that uses linguistic evidence extracted from electronic linguistic databases in order to conduct hypothesis-driven or explorative linguistic research. The alternative discipline, *computational linguistics*, shares with the field of corpus linguistics the key element of linguistic database construction, but its primary aim is to develop techniques of natural language processing, so its focus lies on information technology methods rather than on linguistic theory. Certainly, the two disciplines can overlap to a lesser or greater extent according to the underlying research questions in a project or a case study investigation.

Several research projects have resulted in compilations of corpora, which have also contributed to relevant linguistic analyses for the Romanian context.

¹Open School for Academic Self-improvement. Research, Academic Writing and Career Management PN-II-PCCA-2011-3.1-0682 212/2.07.2012 grant funded by UEFISCDI—www.uefiscdi.gov.ro

One of the few larger collections of student texts is the Romanian Corpus of Learner English (RoCLE) databank (see Chitez 2014). This corpus complies with the general collection norms of the ICLE corpora (Granger et al. 2009), thus including specific genres (argumentative essays and/or literature essays) written by native speakers of Romanian in English. Informants are students having English as major or minor at their university and being enrolled in their third or fourth year of study. There are 352 texts in the corpus, which consist of 201,551 words. The corpus is used for the description of the salient features of student academic written discourse in English as a foreign language (more details below).

Lately, other small-scale corpora in the area of applied linguistics have been constructed that reflect the current use of language (either native or foreign language). For example, Herteg and Popescu (2013) proposed the use of comparable corpora consisting of English and Romanian newspaper business texts compiled by students (15,000 words compiled by each student in each language) for the extraction of business collocations in both languages.

On the other hand, Romanian corpus-based computational linguistic projects have been successfully conducted for several years now. The institution consistently active in corpus collection processes is the Romanian Academy Research Institute for Artificial Intelligence “Mihai Drăgănescu” (RACAI) led by Academician Dan Tufis, who has gained national and international recognition in the field (see Macoveiciuc and Kilgarriff 2010). Table 1 below almost exclusively includes (with the exception of RoWaC) RACAI corpora.

4 At the Confluence of Academic Writing and Corpus Linguistics: Three Examples

4.1 *Linguistic Fields*

Contrastive Linguistics

There is a consistent body of corpus-based contrastive research on academic discourse in numerous languages (d’Angelo 2012; Fløttum et al. 2006; Johansson 2007; Mauranen 1993, 1994; Siepmann 2005). Many such studies make use of parallel corpora to investigate translation challenges (e.g., Mikhailov and Cooper 2016). By aligning the source and target texts, translators (or language learners) can better understand the mechanism and options of linguistic equivalence. In fact, “corpora have perhaps strengthened the trend away from word-equivalence to phrasal equivalence” (Krishnamurthy 2006, p. 253), which makes them also interesting to the academic writing contrastive field, given the importance of rhetorical appropriateness in genre use. Often followed by “further research with monolingual corpora in both languages” (Mauranen 2002, p. 182), translation-related research can either take the form of a genuine contrastive study or turn into an in-depth analysis of salient linguistic phenomena (Ebeling et al. 2013). Other corpus-based

Table 1 Romanian corpora

Corpus	Type of collection	Details about the corpus	Corpus size
CoRoLa ^a	Original collection	Reference corpus written and spoken language; foreseen function styles: scientific, belletristic, journalistic, juridic, administrative, and memorialistic; Romanian	Ongoing
RoWaC ^b	Original collection	Web texts retrieved using web crawling; BootCaT, a newspaper archive (www.adevarul.ro); and a site for copyright-free books (www.biblioteca-online.ro); Romanian	50 million words
ROCO ^c	Original collection	Series of various registers of corpora: news, literature, and legislation; Romanian	35 million words
Ro-Wordnet (RoWN) ^d	Part of EuroWordNet via BalkaNet	Parallel corpora; Romanian and English	43,765 synsets
Romanian FrameNet ^e	Part of FrameNet 1.1 corpus	Parallel corpora; Romanian and English	1094 sentences
Romanian TimeBank ^f /Timex	Translated from TimeBank 1.1	Parallel corpora; Romanian and English	186 news articles, with 72,000 words
RoSemCor ^g	Part of SemCor	Parallel corpora; Romanian, English and Italian	12 articles from SemCor
Acquis Communautaire ^h	Part of multilingual Acquis Communautaire Corpus (AC)	Parallel corpora; Romanian and English	12,000 Romanian documents and 6256 parallel English-Romanian documents; 16 million words.
Romanian MULTEXT-East Corpus ⁱ	Part of EU Copernicus project MULTEXT-East (ME) Corpus	Parallel and comparable corpora for six Central and Eastern European languages; both text and speech corpora; the text corpus includes the translations of the Novel "1984" by George Orwell	100,000 words per corpus

^aSee Barbu-Mititelu et al. (2014)^bSee Macoveiciuc and Kilgarriff (2010)^cSee Tufis and Irimia (2006)^dSee Tufis et al. (2006)^eSee Ion and Barbu-Mititelu (2005) and Trandabat and Husarciuc (2008)^fSee Macoveiciuc and Kilgarriff (2010)^gSee Cristea and Forascu (2006)^hSee *ibid.*ⁱSee Dimitrova et al. (1998)

studies make use of independent comparable corpora to look at particular linguistic features, such as quantity approximation in De Cock and Goossens' study (2013). In general, areas that benefit the most from corpus-based contrastive analyses seem to be bilingual and monolingual lexicography (Granger and Lefer 2013, p. 158).

There has, however, been little theoretical contrastive research on academic writing in Romanian versus other languages (see Chitoran 2013, for English-Romanian contrastive analyses). A few contrastive remarks have been offered by Chitez's (2014) corpus-based study on several grammatical topics (articles, genitive, prepositions).

Academic Phraseology

Formulaic language, or phraseology (Granger and Meunier 2008; Stubbs 2001; Wray 2008), has long centered round the concept of lexicogrammar. As McEnery and Gabrielatos (2006, p. 41) point out, some linguists (Halliday 1991, 1992) prefer the term *lexicogrammar* because it is quite difficult to separate lexis from grammar, as they appear to be “the same thing seen by different observers” (Halliday 1992, p. 62). However, researchers tend to position their theories closer to one of the two ends of the lexico-grammatical continuum: lexical (Stubbs 2001; Halliday 1992) or grammatical (Sinclair 1966, 2004). In time, linguists have generally agreed that the notions of collocation (Biber et al. 1999, 2009; Ellis 1996; Firth 1957; Goldberg 2006; Römer 2009; Stefanowitsch and Gries 2003), lexical bundle (Biber et al. 1999, 2004; Biber 2006; Biber and Barbieri 2007), chunks (Wray 2008), and n-gram (Jarvis et al. 2012; Jarvis and Paquot 2012) are the key elements in lexicogrammatical approaches. In corpus linguistics, collocations and phrases are often interchangeable. Moreover, multiple studies have also indicated a certain degree of correlation between the users' language competence and the phraseology profile of their discourse (Granger and Bestgen 2014; Laufer and Waldman 2011; Levitzky-Aviad and Laufer 2013; Nattinger and Decarrico 1992; Nesselhauf 2003). Academic writing research has taken on this awareness (see Bondi 2014; Charles et al. 2009) and integrated it into applied linguistics studies focusing on the compilation of the Academic Word List (Coxhead 2000; Nation 2001), the Academic Collocation List (Ackermann and Chen 2013; Durrant 2009; Laufer and Waldman 2011; Simpson-Vlach and Ellis 2010), and the Academic Phrasebank (Chitez et al. 2015b; Morley 2005).

Move Analysis

In genre research, certain structural patterns can be found repeatedly. These structures may be called conventional and are present as such in many instructional manuals and guides, but they also have functional meaning in organizing the discourse. Many studies (see also chapter “[Research Articles as a Means of Communicating Science: Polish and Global Conventions](#)”) have followed the research line of Swales (2004), who defined a sequence of moves and steps in which authors of research

articles position themselves within a research field by first “establishing a territory” then defining a “niche” that they then, in a third step, “occupy.” Variations of this Creating a Research Space (CARS) model have been detected in research articles from many different cultures. The highly formalized evaluation model for move structures within the whole research article (not only the introduction) from Kanoksilapatham (2005), which follows the introduction, methods, results, and discussion (IMRAD) structure of the research article, can serve as a model for the coding of complex move text structures.

Bondi (2009), for example, followed a rather open comparison of Italian and English historical and economic discourse searching for differences in genre characteristics. She found large variations not only between disciplines and languages, but also between different approaches within the disciplines of the same language. She also analyzed statements of purpose in historical and business discourse. Such statements are often in fixed phraseological patterns (“in this paper,” “this paper is,” “of this paper,” “purpose of this,” “this paper examines,” etc.) and are therefore easily accessible in searches.

In Romanian academic genres, the open approach seems more appropriate since we cannot assume that Romanian students follow the standard IMRAD sequence in their research-related academic genres, given the fact that the typical Romanian university genres are much more variable and less conventionalized than research articles. However, we should be aware that the basic linguistic features of academic genres are identifiable.

4.2 *How Can Corpora Be Used? Exemplifications from the Romanian Corpus of Learner English*

Data

The examples in the present study are extracted from the RoCLE (see Chitez 2014) (Table 2).

The proportion of argumentative essays is around 75% of the total number of texts while literature essays make up 25% of the databank. In this way, some research questions can be addressed concerning the rhetorical patterns in either register settings (formality level in argumentation) or disciplinary settings (literature studies).

Example 1: Romanian-English Collocation Pattern Transfer

In her study, Chitez (ibid.) has identified several grammatical areas with potential for language and academic writing teaching: articles, genitives, and prepositions. For example, the use of the prepositions *in*, *on*, and *to* in collocation patterns are some of the cases with a great Romanian-English interference risk (Table 3):

Table 2 Topic distribution in RoCLE

Main topics (argumentative)	Number of texts	Other topics (argumentative)	Number of texts	Topics (literature)	Number of texts
Pollution	52	Animals	3	Moby Dick, Oscar Wilde's The Picture of Dorian Gray, Prometheus Unbound, Wuthering Heights, My favourite book, My favourite novel, The Romantic Spirit in British Literature, N. Hawthorne, generalities relying on short fiction with examples of your own, etc.	85
Crime	33	Child abuse	3		
Feminism	25	Europe	3		
Press	23	School	3		
Equality	17	Television	3		
Abortion	11	Advertising	2		
Teenagers	9	Future	2		
Orphanages	8	Society	2		
Miracles	8	Nature	2		
Technology	8	Universities	2		
Euthanasia	7	Other (1x)	27		
Homosexuality	7				
Money	7				
Total texts	215		52		

Table 3 Romanian versus English prepositional collocation patterns in RoCLE

	Interference	RoCLE example	Interference	RoCLE example
In	Grammatical transfer from Romanian into English: <i>la</i> ↔ <i>in</i>	<i>[continue in] all level</i> <i>[join in] the suffering</i> <i>[put in] (danger)</i>	Phraseology transfer from Romanian into English	<i>in generally</i> <i>in special</i> <i>in consideration</i>
On	Grammatical transfer from Romanian into English: <i>pe / de</i> ↔ <i>on</i>	<i>pay [attention on] the</i> <i>[concerned on]</i>	Phraseology transfer from Romanian into English	<i>on the same time</i> <i>think on</i> <i>(sit) on contrary sides</i>
To	Grammatical transfer from Romanian into English: <i>la / de</i> ↔ <i>to</i>	<i>[are going to] shopping</i> <i>aims [to a vision]</i>	Phraseology transfer from Romanian into English	<i>point to</i> <i>show to</i> <i>think to</i> <i>accorded to</i> <i>conducive to</i>

What academic writing experts can learn from the analysis of preposition collocation lists is the fact that Romanian students might use academic writing phraseology incorrectly (e.g., “in special”) mainly because they translate phraseology from Romanian (e.g., Ro: *în special*).

Another example is the use of the collocation pattern created by the demonstrative pronoun *this* together with a singular common noun. It has been shown (Chitez 2014, p. 116) that Romanian students excessively use expressions such as *this situation*, *this way*, *this kind*, and *this movement*. Considering that demonstrative ana-

phors are relevant markers of scientific writing (Lundquist 2007), this phenomenon can be also further investigated and exploited for the benefit of the Romanian students writing in English.

Example 2: Genre-Related Academic Phraseology

Two types of texts are included in RoCLE: argumentative and literature essays. It is interesting to look at the types of phrases that appear frequently in such genres (Fig. 1).

Surprisingly, students use the first person pronoun almost as frequently in argumentative essays (0.439% of the total number of words in RoCLE-ARG) as in literature essays (0.272% of the total number of words in RoCLE-LIT). The patterns most frequently encountered in argumentation are: *I think, I know, I believe, I want, I consider, and I agree*. In literature essays, students use past tense constructions such as *I read, I found, I used, or I loved*, which convey a rather narrative style to the literature text.

Example 3: Move-Analysis Indicators in RoCLE Essays

In order to be able to assess rhetorical moves in academic writing, a standard genre rhetoric move patterning has to be defined. In the case of argumentative essays, we will consider one of the structural units proposed by literature: situation, problem,

N	Concordance	Set	Tag	Word #
109	beautiful period in every people life. I think the adolescence is unique,			410
110	everyone takes part consciously or not. I think that pollution is a sort of slow			14
111	. I do don have the same opinion. I agree that they are special, but not			122
112	, on my entire family, was a big one. I was crying all the time, begging for a			497
113	a life. People have no right to do so. I agree in this particular matter with			510
114	not expect any extra money, but since I gave it to him/her, he / she could not			681
115	person, like a normal one? Sincerely, I don't know if I can look to that men			214
116	I read in newspapers or I hear everyday that a crime happened			7
117	, or "what can I do about it, since I am so unimportant". What is even			104
118	not expect any extra money, but since I gave it to him/her, he / she could not			682
119	Sincerely, I'm impartial with homosexual people,			3
120	harmony. I wake up as I was sleeping, I realize that my mother feels good and			731
121	even inslave them, and beat them. So I have no respect for them, the white			330
122	life. Will you be able to sleep? I don't think so. A murder changes			271
123	dye. If I were put in this situation, I would not know what to do. First of			197
124	last solution on a desperate situation. I wonder if death, generally speaking is			39
125	a good thing for women to do it or not. I remembered that my mother told me			31
126	In order to support my thesis statement I will give some arguments which I			110

Fig. 1 The use of the first person pronoun *I* in collocation patterns in argumentative essays

solution, and evaluation (see Tirkkonen-Condit 1985). The multiword phrases associated with each of the moves can be searched for and extracted from the texts. In the case of the evaluation procedures, we checked the use of *conclusion/conclude* or similar expressions and noticed the following patterns:

- (a) In most argumentative essays, students use a concluding phrase to mark the evaluation move in their text, with frequent expressions being built around markers such as [*to conclude*]/[*in conclusion*]/[*concluding*], [*thus,*], [*so,*], [*all (in all)*], and [*these facts*], or in rare constructions such as [*to cap it all*] and [*taking everything into account*].
- (E1) Prove your independence to yourself and face today or tomorrow without cell phone and you will feel differently. <ICLE-RO-AIC-0002.2>
- (E2) If we remain silent and passive it will become one. But our immediate reaction and vivid interest will struggle against it and it will set forth an example and model for the others to follow. <ICLE-RO-AIC-0003.1>
- (E3) We should meditate on this problem, we should understand that if there are people who would give anything to spend yet another moment with their loved ones, why, and how could you decide to put an end to what God gave us: Life! <ICLE-RO-AIC-0008.1>
- (b) If the markers are not present, the alternative rhetorical phenomenon is the “recommendation,” either in the form of second person addressing or as a collective “we” formula (see E1–E3).

5 Pedagogical Recommendations

There are multiple ways by which the data extracted from corpora can be introduced into teaching scenarios. For the specific task of supporting student academic writing, a corpus can be implemented as follows:

Corpora Can Facilitate Induced-Learning Writing-Related Tasks students can be given the task of analyzing databanks in either one language (Romanian) or in contrast (e.g., Romanian versus English L2; English L1 versus English L2) in order to identify salient features of academic writing use (see example of an exercise in contrastive phraseology in English and Romanian below)

Please use the corpus databank of Romanian Learner English (RoCLE corpus) in order to extract academic phraseology containing the following keywords: *author, paper, intend, follow, important*. Do the same with the corpus databank of Native Speaker English (LOCNESS corpus). Compare lists of phrases and identify interferences and/or transfer from native language Romanian into English.

Corpora as control for reference instruments students can be given the task of analyzing databanks in English L1 and English L2 in order to check vocabulary/phraseology listed in dictionary entries or academic phrase banks.²

²Possible sources for analyzing academic phraseology: <http://www.phrasebank.manchester.ac.uk/>

Corpora as Support for Academic Writing Tools students can have access to a specialized corpus within an electronic academic writing tool (see Chitez et al. 2015b) that can give them immediate support with linguistics problems encountered during the actual writing process.

6 Conclusions

We have used the RoCLE to exemplify the conception of corpus-based exercises to be completed in an academic writing class. A corpus of EFL texts containing specific types of genres (e.g., argumentative and literature essays) can be valuable to language tutors dealing with these text types.

As has been shown in this viewpoint paper, if we look at several highly used prepositions, there are numerous cases of collocation-pattern transfers from native Romanian into English. The analysis can be replicated for other grammatical elements or units. The RoCLE can also be employed for genre-specific analyses, such as the typology and use of phraseology in argumentative versus literature essays. In our case analysis, we showed that authorship (use of personal pronoun *I*) is rendered through genre-specific expressions. In the third example, we looked at the evaluation move in the sub-corpus of argumentative essays (RoCLE-ARG). Results showed frequent constructions (e.g., *in conclusion*) or multi-word Romanian borrowings (e.g., *taking everything into account*) and the tendency to replace overt concluding with recommending strategies.

The final recommendations on how to use corpora to improve students' academic writing include free corpus consultation to increase awareness of target-language phraseology and vocabulary and comparison between corpora and dictionaries for lexical accuracy. A further application of corpora for academic writing, even if it is beyond the scope of the present article, would be to integrate corpora into academic writing tools.

It is essential to highlight that students can benefit from the use of corpora simply by learning to access, use, or create them. It is the teachers' task to guide the learning process towards areas that are relevant for the academic writing field, such as the ones exemplified in this paper (contrastive collocation patterns, genre-specific phraseology, and rhetoric move constructions). Thus, the applicability of corpora is linked to the teacher's creativity and openness to new methodologies in the classroom. In the Romanian EFL context, such approaches would be quite innovative, triggering or resulting in meaningful research in both corpus linguistics and academic writing or leading to unexplored but effective teaching strategies.

References

- Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL)—A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12, 235–247.
- Andronescu, S. C. (1997). *Tehnica scrierii academice*. Bucuresti: Editura Fundatiei “Romania de Maine”.
- Aull, L. (2015). *First-year university writing: A Corpus-based study with implications for pedagogy*. London: Palgrave Macmillan.
- Barbu-Mititelu, V., Irimia, E., & Tufiş, D. (2014) CoRoLa the reference corpus of contemporary Romanian language. In *Proceedings of the ninth international conference on language resources and evaluation (LREC'14)*. Reykjavik, Iceland, 2014, 1235–1239.
- Bekar, M., Doroholschi, C. I., Kruse, O., & Yakhontova, T. (2015). Educational genres in Eastern Europe: A comparison of the genres in the humanities departments of three countries of three different universities in three different countries. *Journal of Academic Writing*, 5(1), 119–132.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263–286.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. New York: Longman.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405.
- Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.
- Biber, D., Kim, Y.-J., & Tracy-Ventura, N. (2009). A Corpus-driven approach to comparative phraseology: Lexical bundles in English, Spanish, and Korean. *Japanese/Korean Linguistics*, 17, 75–94.
- Bondi, M. (2009). *Polyphony in academic discourse. Cross-linguistic and cross-cultural perspectives on academic discourse*. Amsterdam: John Benjamins Publishing Company.
- Bondi, M. (2014). Integrating corpus and genre approaches: Phraseology and voice across EAP genres. In M. Gotti & D. S. Giannoni (Eds.), *Corpus analysis for descriptive and pedagogical purposes* (pp. 43–62). Bern: Peter Lang.
- Charles, M., Pecori, D., & Hunston, S. (Eds.). (2009). *Academic writing: At the Interface of Corpus and discourse*. London: Continuum.
- Chitez, M. (2014). *Learner corpus profiles: The case of Romanian learner English*. In M. Gotti (Series Ed.), *Linguistic insights series* (Vol. 173). Bern/Berlin/Bruxelles/Frankfurt am Main/New York/Oxford/Wien: Peter Lang.
- Chitez, M., & Kruse, O. (2012). Writing cultures and genres across Europe. In M. Castelló & C. Donahue (Eds.), *University writing: Selves and texts in academic societies* (pp. 33–51). Bingley: Emerald Group.
- Chitez, M., Kruse, O., & Castelló, M. (2015a). *The European writing survey (EUWRIT): Background, structure, implementation, and some results* (Working papers in applied linguistics 9). Winterthur: ZHAW Zurich University of Applied Sciences. online at: <https://digitalcollection.zhaw.ch/handle/11475/1016>
- Chitez, M., Rapp, C., & Kruse, O. (2015b). Corpus-supported academic writing: how can technology help? In F. Helm, L. Bradley, & S. Thouésny (Eds.), *Critical CALL proceedings of the 2015 EUROCALL conference, Padova, Italy*. Dublin, Ireland: Research-publishing.net.
- Chitoran, D. (Ed.). (2013). *Contrastive studies in English and Romanian*. Bucharest: Bucharest University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Cristea, C., & Forascu, D. (2006). Linguistics resources and technologies for Romanian language. *Journal of Computer Science of Moldova*, 14(40), 33–73.

- D'Angelo, L. (2012). Identity conflicts in book reviews: A cross-disciplinary comparison. In G. Maurizio (Ed.), *Academic identity traits. A corpus-based investigation* (pp. 79–94). Bern: Peter Lang.
- De Cock, S., & Goossens. (2013). Quantity approximation in English and French business news reporting: More or less the same? In K. Aijmer & B. Altenberg (Eds.), *Advances in corpus-based contrastive linguistics: Studies in honour of Stig Johansson* (pp. 139–156). Amsterdam: John Benjamins.
- Dimitrova, L., Ide, N., Petkevic, V., Erjavec, T., Kaalep, H. J., & Tufiş, D. (1998). Multitext-east: Parallel and comparable corpora and lexicons for six central and eastern European languages. In C. Boitet & P. Whitelock (Eds.), *Proceedings of the joint 17th international conference on computational linguistics 36th annual meeting of the association for computational linguistics (COLING-ACL 1998)* (pp. 315–319). Montreal: Morgan Kaufmann Publishers.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28, 157–169.
- Ebeling, J., Ebeling, S. O., & Hasselgård, H. (2013). Using recurrent word combinations to explore crosslinguistic differences. In K. Aijmer & B. Altenberg (Eds.), *Advances in corpusbased contrastive linguistics: Studies in honour of Stig Johansson* (pp. 177–200). Amsterdam: John Benjamins.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–126.
- Firth, J. (1957). *Papers in linguistics, 1934–1951*. Oxford: Oxford University Press.
- Fløttum, K., Dahl, T., & Torodd, K. (2006). *Academic voices*. Amsterdam/Philadelphia: John Benjamins.
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24(3), 321–332.
- Flowerdew, J., & Forest, R. W. (2015). *Signalling nouns in academic English*. Cambridge: Cambridge University Press.
- Frăţila (Pungă), L. (2006). Collocations in ecological discourse. *B.A.S. British and American Studies*, 12, 211–227 Available at: https://www.academia.edu/14049718/Collocations_in_Ecological_Discourse.
- Goldberg, A. E. (2006). *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.
- Gotti, M., & Giannoni, D. S. (Eds.). (2014). *Corpus analysis for descriptive and pedagogical purposes: ESP perspectives* (Linguistic insights series, Vol. 200) (Series editor: Maurizio Gotti). Bern/Berlin/Bruxelles/Frankfurt am Main/New York/Oxford/Wien: Peter Lang.
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *IRAL*, 52(3), 229–252.
- Granger, S., & Lefer, M.-A. (2013). Enriching the phraseological coverage of high-frequency adverbs in English–French bilingual dictionaries. In K. Aijmer & B. Altenberg (Eds.), *Advances in corpusbased contrastive linguistics: Studies in honour of Stig Johansson* (pp. 157–176). Amsterdam: John Benjamins.
- Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International Corpus of Learner English (Version 2). Handbook and CD-Rom*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer & B. Altenberg (Eds.), *English Corpus linguistics: Studies in honour of Jan Svartvik* (pp. 30–43). London: Longman.
- Halliday, M. A. K. (1992). Language a system and language as instance: The corpus as a theoretical construct. In J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 61–77). Berlin: Mouton.

- Herteg, C., & Popescu, T. (2013). Developing students' intercultural and collocational competence through analysing the business press. A corpus-based pedagogic experiment among Romanian MA students. *Procedia Social and Behavioral Sciences*, 83, 378–382.
- Hyland, K. (2009). Corpus informed discourse analysis: The case of academic engagement. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 110–128). London: Continuum International Publishing Group.
- Ion, R., & Barbu-Mititelu, V. (2005). Towards ROMANCE FrameNet. The translation task. In *Proceedings of international workshop ROMANCE FrameNet, EUROLAN 2005*, Cluj-Napoca, România.
- Jarvis, S., & Paquot, M. (2012). Exploring the role of n-grams in L1 identification. In S. Jarvis & S. A. Crossley (Eds.), *Approaching transfer through text classification: Explorations in the detection-based approach* (pp. 71–105). Bristol: Multilingual Matters.
- Jarvis, S., Bestgen, Y., Crossley, S. A., Granger, S., Paquot, M., Thewissen, J., & McNamara, D. (2012). The comparative and combined contributions of n-grams, Coh-Metrix indices, and error types in the L1 classification of learner texts. In S. Jarvis & S. A. Crossley (Eds.), *Approaching transfer through text classification: Explorations in the detection-based approach* (p. 154177). Bristol: Multilingual Matters.
- Johansson, S. (2007). *Seeing through multilingual corpora. On the use of corpora in contrastive studies*. Amsterdam: John Benjamins.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3), 269–292.
- Krishnamurthy, R. (2006). Corpus Lexicography. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 250–254). Oxford: Elsevier.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647–672.
- Levitzky-Aviad, T., & Laufer, B. (2013). Lexical properties in the writing of foreign language learners over eight years of study: Single words and collocations. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (EUROSLA Monographs 2, pp. 127–148). Eurosla Monographs Series 2. European Second Language Association.
- Lundquist, L. (2007). Academic discourse as social control and system(s), seen through the use of demonstrative noun phrases in French scientific texts. In K. Fløttum (Ed.), *Language and discipline perspectives on academic discourse* (pp. 219–242). Newcastle: Cambridge Scholars Press.
- Macoveiciuc, M., & Kilgarriff, A. (2010). The RoWaC corpus and Romanian word sketches. In D. Tufis & C. Forascu (Eds.), *Multilinguality and interoperability in language processing with emphasis on Romanian*. Bucharest: Romanian Academy Publishing House.
- Mauranen, A. (1993). *Cultural differences in academic rhetoric: A text-linguistic study*. Frankfurt am Main: Peter Lang.
- Mauranen, A. (1994). Two discourse worlds: Study genres in Britain and Finland. *Finlance*, 13, 1–40.
- Mauranen, A. (2002). Will “translationese” ruin a contrastive study? *Languages in Contrast*, 2(2), 161–186.
- McEnery, T., & Gabriellatos, C. (2006). English corpus linguistics. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 33–71). Oxford: Blackwell.
- Mikhailov, M., & Cooper, R. (2016). *Corpus linguistics for translation and contrastive studies. A guide for research*. Oxford/New York: Routledge.
- Morley, J. (2005). *Academic Phrasebank: An online writing resource*, The University of Manchester. Available from <http://www.phrasebank.man.ac.uk>
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. New York: Oxford University Press.

- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.
- Pavlenko, S., & Bojan, C. (2014). Eight reasons to attend an academic writing course. The impact of academic writing courses on participants. *Studia Universitatis Babeş-Bolyai, Philologia*, 59(3), 209–223.
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 83–98). Philadelphia: John Benjamins.
- Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of upper-level student papers (MICUSP). *Corpora*, 6(2), 159–177.
- Siepmann, D. (2005). *Discourse markers across languages: A contrastive study of second level discourse markers in native and non-native text with implications for general and pedagogic lexicography* (Routledge advances in corpus linguistics). London/New York: Routledge.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512.
- Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 410–430). Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.
- Stefanowitsch, A., & Gries, S. T. (2003). Collocations: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- Swales, J. M. (2004). *Research genres. Explorations and applications*. Cambridge: Cambridge University Press.
- Tirkkonen-Condit, S. (1985). *Argumentative text structure and translation*. Jyväskylä: University of Jyväskylä.
- Trandabat, D., & Husarciuc, M. (2008). Romanian semantic role resource. In *Proceedings of the sixth international language resources and evaluation (LREC'08)*, Marrakech, Morocco, May, 28–30, 2008, 2806–2810.
- Tufis D., & Irimia, E. (2006). RoCo-news: A hand validated journalistic corpus of Romanian. In *Proceedings of the fifth international conference on language resources and evaluation (LREC-2006)*, Genoa, Italy, 2006 (pp. 869–872).
- Tufis, D., Barbu-Mititelu, V., Bozianu, L., & Mihaila, C. (2006). Romanian WordNet: New developments and applications. In P. Sojka, K.-S. Choi, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the 3rd conference of the Global WordNet Association* (pp. 337–344). Jeju: Masaryk University in Brno.
- Upton, T. A., & Connor, U. (2001). Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre. *English for Specific Purposes*, 20(4), 313–329.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries* (Oxford applied linguistics). Oxford: Oxford University Press.