

RESEARCH ARTICLE

Corpus of Academic Learner English (CALE): A new corpus at the intersection of corpus linguistics and English for academic purposes

Aysel Şahin Kızıl

İzmir Bakırçay University, Turkey / Contact: aysel.sahinkizil@bakircay.edu.tr 

Abstract

Offering insights into varieties of English language, corpus compilation practices and corpus-based research have great potential to enhance the field of English for Academic Purposes (EAP). Among different types of corpora, learner corpora seem particularly promising as they document different aspects of learner language that might otherwise go unnoticed. Acknowledgement of potential contribution of learner corpora to EAP has led up to scientific efforts to compile learner corpora for academic purposes. This paper is the description of such an effort. The aim of this study is twofold: first, to highlight the possible contributions of language corpora to EAP; and second, to introduce the Turkish subcomponent of the Corpus of Academic Learner English (CALE-TR), as a new English for Academic Purposes corpus, currently being compiled for the investigation of learner academic writing.

© 2020 The Literacy Trek & the Authors – Published by The Literacy Trek

APA Citation

Şahin Kızıl, A. (2020). Corpus of Academic Learner English (CALE): A new corpus at the intersection of corpus linguistics and English for academic purposes. *The Literacy Trek*, 6(2), 41-54. <https://doi.org/10.47216/literacytrek.791664>

Keywords

English for academic purposes; learner corpora; EAP; Corpus of Academic Learner English; CALE-TR

Submission date

08.09.2020

Acceptance date

08.12.2020

Introduction

Being a relatively new but highly vigorous area of corpus linguistics, learner corpora research has been a major focus of interest over the past few decades. Broadly defined as electronic collections of foreign or second language learners' language samples in accordance with a set of predefined criteria, learner corpora prove to be promising in many areas of language-related studies. Collected in many different settings so far, learner corpora have largely been gathered within the context of foreign language learning (Granger & Paquot, 2011). One such world-wide known learner corpus is the International Corpus of Learner English (ICLE). The initiation of

ICLE project has inspired a number of academic attempts in corpus compilation within different settings along with its opening up new venues for corpus ideas.

One such inspiration is the idea of compiling and exploiting language corpora to enrich the field of English for Academic Purposes (EAP) (Flowerdew, 2014). Though relatively little steps have been taken to back up EAP with learner corpora, a quick look at the relevant literature and instructional materials reveals that EAP practitioners and material designers mainly rely on native corpora. Learner corpora are hardly exploited for EAP related activities, which is regarded as a shortfall (Charles, 2013; Gilquin, Granger, & Paquot, 2007).

Two major causes could be proposed to explain the relative neglect of exploiting learner corpora in EAP studies. One could be the lack of awareness among the researchers and EAP practitioners regarding the potentials of learner corpora for the field. Improving the field of EAP through learner corpora research does not have a long history only dating back to 1990s (Flowerdew, 2014). This might cause a delay in developing insights into the possible contributions of learner corpora to the field. The second reason for not using learner corpora commonly in EAP studies could be linked to the scarcity of learner corpora specially designed for EAP field (Charles, 2013; Flowerdew, 2014). Despite the little steps taken so far in terms of collecting corpora of academic learner English (e.g. the Varieties of English for Specific Purposes dAtabase (VESPA), the Corpus of Academic Learner English (CALE)), the field is still in need of further initiatives for EAP specific learner corpus projects (Charles, 2013).

Given this need, CALE as a corpus project targeting the intersection of corpus linguistics and EAP was started by the University of Bremen, Germany in 2014. Since then, it has been expanding and seems quite promising with its special focus on lexico-grammatical variation in the language of advance learners majoring in linguistics and applied linguistics.

The aim of this paper is twofold: first, to highlight the potential contributions of language corpora to EAP; second, to introduce the subcomponent of CALE, as a new English for Academic Purposes corpus, currently being compiled for the investigation of learner academic writing. This paper is structured as follows: Section

2 presents an overview of English for Academic Purposes (EAP) by touching upon its historical development along with its current situation. Section 3 depicts the impacts of corpora on the field of EAP by highlighting the learner corpora in particular. Section 4 introduces CALE as a newly emerging corpus of academic English along with a small-scale corpus analysis demonstrating potential of CALE for investigating advance level learner English. Section 5 concludes the article.

Literature Review

English for Academic Purposes

Situated within the broader field of English for Specific Purposes (ESP), English for Academic Purposes (EAP) is considered as one of the two branches along with English for Occupational Purposes. A broader definition usually offered in the literature for EAP is teaching English with the purpose of providing learners with guidance in their academic study or research in English (Harwood & Petric 2011). Hyland (2006) regards it as a “theoretically grounded and research informed practice” (p.1) since nature of language, learning and teaching all have immediate effects on EAP.

The study of English in academic context dates back to 1960s when Britain started to face with an increasing demand from international students for study purposes. In 1960s, Britain also had to respond to the rapid growth of English as a common language for business and science in many countries (Hamp-Lyons 2011). While the second context gives way to the development of ESP, the first context paved the way for the advent of EAP. This parallel outset justifies the view of EAP as a sub-set of ESP. Earlier steps taken within the field of EAP were quite similar to ESP; however, EAP has differentiated in the course of time.

Initial practices regarding EAP mainly focused on register analysis, which was pertaining to the comparison of scientific English with general English to find out frequently occurring grammatical and lexical items in each register (Harwood & Petric, 2011). This comparison stemmed from the idea that such analyses could inform the syllabus and textbook designs. Some of the first EAP books, therefore,

benefited to a great extent from the findings of register analysis. 1970s witnessed a shift of attention towards linguistic analysis of patterns above the sentence or utterance level (Swales, 1988). Discourse patterns, textual structure and rhetorical functions were the focus among the EAP practitioners. This shift of focus led to the advent of authentic language use within EAP. In parallel with the developments in ELT, communicative skills gained significance in EAP as well. From the beginning of 1980s, greater significance was attached to the skills learners needed to perform effectively in academic contexts. This brought specific skills (e.g. reading, listening) to the foreground, which, in turn affected the design of EAP materials (Swales, 1988). In 1980s, need analysis became one of the central concepts in EAP due to the diversity of the profile of EAP students and their learning conditions (Hamp-Lyons, 2001). In an overview of EAP, Carkin (2005) states that “needs assessment of the diverse learners in EAP underlies syllabus design, materials development, text selection, learning goals and tasks, and, ultimately, evaluation of students and course or program success” (p. 87). Depending on the context, approaches to need analysis and relevant data collection techniques vary greatly. Questionnaires, interviews, focus group studies, observation of lectures and activities are among the main tools for data collection for EAP need analysis (Gilbert, 2013). Findings from such techniques have produced valuable insights for the growth of EAP.

Together with the contributions of aforementioned developments, EAP currently has become more mature. Hamp-Lyons (2011) evaluates the current situation of EAP as follows:

A pattern seems to be becoming established whereby the usual teaching solutions in English first language contexts within universities, where (as is usually the case these days) a large proportion of students, particularly graduate students, are international students and non-native speakers of English, are for English language teachers to start from EAP solutions based on formal or informal needs analysis”.

Today, EAP has secured its position as a discipline in language education, and EAP instruction takes place in a range of contexts. One of them is the higher education settings in English-speaking countries. It is also popular in settings where English is official language and medium of instruction. EAP courses are also offered in university settings where university subjects are partly or wholly taught in English.

Finally, EAP instruction has a significant place in settings where English is considered to be important for study and publications (Harwood & Petric, 2011).

Prevalence of EAP courses have led the practitioners from different contexts to embark on a quest for informing and analysing academic language. One valuable strand in this direction is the exploitation of corpus linguistics for EAP (Flowerdew, 2014; Hamp-Lyons, 2001). Following section highlights major contributions of corpus linguistics on EAP with an emphasis on learner corpora.

Language Corpora and EAP

Gilquin, Granger and Paquot (2007) identify two significant contributions of corpus linguistics to EAP by stating that corpus-based analysis of academic language provides “detailed descriptions of its distinctive linguistic features, and more specifically its highly specific phraseology, and careful analyses of linguistic variability across academic genres and disciplines” (p.320).

The first contribution stems from the very nature of language samples compiled through a set of strict design criteria. This makes the language corpora distinctive in being representative of a language or a variety for linguistic research (Granger, 2004; Sinclair, 2005) Linguistic searches through the language corpora are realized in the form of a set of analysis ranging from frequency counts to multidimensional analysis (Nesi, 2016). Each way of searching sheds light to different aspect of academic language. Frequency counts enable researchers to produce academic wordlists not only in the form of simple itemisations of every word but also through “complex comparisons within and between corpora, and lists of terms restricted to specific domains” (Nesi, 2016 p.211). Some of the best known yields of this type of corpus analysis are Academic Wordlist (AWL) by Coxhead (2000), Academic Keyword List (AKL) by Paquot (2010) and The Academic Vocabulary List (AVL) by Gardner and Davies (2014). Concordancing, another method of corpus query, helps to find out the co-text of a linguistic feature (e.g. preferred co-occurrences and grammatical structures) (Gilquin et al., 2007). This type of analysis has also been oriented to the disciplinary variations. Studies by Cortes (2004) and Hyland (2008) are of this kind. Finally, corpus-based studies employing Multidimensional Analysis (MDA) with its focus on group of texts representing

registers or language variations have contributed greatly to the field of EAP through revealing “variation in the clustering patterns of linguistic features across genres, disciplines, and levels of study” (Nesi, 2016). Biber and Gray (2010), for instance, analysed a corpus of academic research articles published in various disciplines and report significant findings regarding the discourse style of academic writing.

Most of the studies conducted so far have relied on corpora including expert academic writings. Although there is no doubt that corpora of expert writing have informed the field of EAP greatly, recent EAP literature calls for the need to integrate learner corpora into EAP studies (Coxhead, 2011). Acknowledging the significance of analysis of learner writings for EAP, Paquot (2007) states:

Descriptions of EAP words should then be refined on the basis of a careful analysis of learner corpus data that will highlight learners’ attested ‘difficulties’ in using these words (cf. Flowerdew 1998; Granger 2004a). Learner corpora are the only type of data that can reveal learners’ typical ‘difficulties’ in terms of underuse, overuse, misuse and idiosyncratic use of EAP words or multi-word sequences (p. 127).

Similar thoughts are also worded in Gilquin et al., (2007) claiming that learner corpora have the potential to draw a precise picture of learner academic language, which, in turn, could guide the practitioners in the field of EAP.

Insights into the importance of learner corpora for EAP activities have led a number of researchers to initiate projects to compile academic language samples produced by the language learners. Table 1 presents the major academic learner corpora around the world.

Table 1. Major learner corpora around the world

Corpus	Project Name	Language Background of the learners	Size in words	Medium	Learners/ Proficiency levels	Text/Task Type
BAWE	British Academic Written English		6.5 million words	Written	undergraduate s; postgraduates	Essay, narrative recount, proposal, research reports
MICUSP	Michigan Corpus of Upper-Level Student Papers	Multiple	2.6 million	Written	undergraduate s; postgraduates	Proposal Research paper Argumentative essay Report Critique

VESPA	The Varieties of English for Specific Purposes dAtabase	Multiple	220.000 Expanding	Written	undergraduate s; postgraduates	Research papers Reports Dissertations
CroLTeC	Croatian Learner Text Corpus		1 million	Written	From A2 to C1	Literary essays, book reviews
ACAW	The Aachen Corpus of Academic Writing	German	240.000 Expanding	Written	Advanced	Academic research writing
ALEC	The Advanced Learner English Corpus	Swedish	1.3 million	Written	Advanced	Essays in linguistics and literature
The BATMAT Corpus		Swedish Finnish	3 million	Written	Advanced	BA dissertations MA dissertations
BELLCE	Belarussian Learner Corpus of English	Russian Belarussian	58 texts	Written	High intermediate Advanced	Essays
CEEAUS	The Corpus of English Essays Written by Asian University Students	Multiple L1	200.000	Written	Advanced	Essays
CAWE	The Chinese Academic Written English corpus	Chinese	400.000	Written	Advanced	Dissertations in linguistics and applied linguistics
CHALC	The Cologne-Hanover Advanced Learner Corpus	German	210.000	Written	Advanced	Term papers and essays
ICLE	The International Corpus of Learner English	Multiple	3 million	Written	Advanced	Essays
JEFL	The Japanese English as a Foreign Language Learner Corpus	Japanese	700.000	Written	From beginning to advanced	Essays
CALE	The Corpus of Academic Learner English	Multiple	Expanding	Written	Advanced	Review Summary Research paper Abstract Proposal

Although Table 1 is far from being a comprehensive list covering all learner corpora of academic English, it could still make a base for understanding the increasing role of learner language for EAP studies. One advantage learner corpora offer to EAP is that their being adequately large gives the researchers wider empirical basis than available before (Gilquin et al., 2007). Additionally, learner corpora enable researchers to quantify learner data through a number of automated methods and tools. This lets the researchers to develop better insights into the idiosyncratic aspects of learner language. One of the method of analysis used with learner corpora is Contrastive Interlanguage Analysis (CIA) (Granger, 1996), which is attributed a key role in determining learner-specific features of language. It involves the comparison of learner language with native language (L2 vs L1) and the comparison of different varieties of learner languages (L1 vs L1) (Gilquin et al., 2007).

Both types of comparisons have produced fruitful results regarding the properties of learners' academic language in terms of syntactic, phraseological, stylistic and pragmatic aspects. It has been suggested that language learners experience problems of frequency, register, positioning and semantics in their academic language (Cotos, 2014; Gilquin et al., 2007; Gilquin & Paquot, 2008). Being at the heart of corpus studies, frequency related investigations have shown what is typical in learner English. To illustrate, Granger (1998) and Hinkel (2002) found that learners underuse academic vocabulary in their essays and their writings are characterized by features not typical of written discourse. Investigating stance-related properties of learner academic writings, Chen and Baker (2010) have concluded that "the tendency to hedge less and instead adopt an overstating tone seems to be universal for learners from different L1 backgrounds" (p. 43). Based on such findings, there is a consensus on the potentials of learner corpora for EAP as worded by Flowerdew (2001) who states that "insights gleaned from learner corpora... to complement those from expert corpora for syllabus and materials design" (p.364).

In order to get sound inferences from any learner corpora, one common suggestion is that it should be compiled in accordance with a set of strict, pre-identified criteria, and it should be representative as much as possible. One such learner corpora within EAP context is Corpus of Academic Learner English (CALE), which is further described in the following section.

Corpus of Academic Learner English (CALE)

CALE is a project led by University of Bremen, Germany, within the context of “Lexico-grammatical variation in advanced learner varieties” funded from January 2014 to January 2017 by the Central Research Development Fund of the University of Bremen, Germany, one of eleven “Universities of Excellence” in Germany that compete internationally in top-level research (Callies, 2015). Describing the learner language represented in the corpus as “advanced learner varieties” (ALVs), CALE specifies four main purposes. First, it aims to create a database for detailed quantitative and qualitative description of advanced learner varieties. Second, it aims to produce detailed case studies to investigate lexico-grammatical variation. Additionally, it aims to develop a corpus-driven, text-centered method based on linguistic criteria for the assessment of writing proficiency in the academic register. Final purpose of CALE is to contribute to the field of EAP through applying the findings to teaching academic writing at advanced levels.

Being still in progress, CALE comprises of six sub-corpora: CALE-GR collected from German EFL learners, CALE-LIT from Lithuanian EFL learners, CALE-POL from Polish EFL learners, CALE-RUS from Russian EFL learners, CALE-CHI from Chinese EFL learners and CALE-TR from Turkish EFL learners.

CALE-TR started to be compiled in 2015 by the researchers from Firat University and Çukurova University, Turkey, and is still in progress. The corpus project was funded by Firat University Scientific Research Projects Unit¹ from 2015 to 2017. In gathering CALE-TR, the same design principles and conventions are being followed as specified in CALE guidelines. These guidelines are also followed by the other international partners compiling other components of CALE. Based on these common design principles considered strictly in compilation of the sub-corpora of CALE, it is possible for the researchers to identify properties of academic learner English in a given first language (L1) background. Additionally, researchers have the opportunity to compare academic interlanguage of a given group with the other groups from different L1 background. This type of cross-linguistic comparison enables corpus linguists to determine universal and L1-related properties of academic interlanguage (Kilimci, 2014). The comparable native corpora of CALE-TR are

Michigan Corpus of Upper-Level Student Papers (MICUSP; Römer & O'Donnel, 2011) or the British Academic Written English corpus (BAWE; (Alsop & Nesi, 2009; Callies, 2015). Comparing learner academic English with native academic English has the potential to identify the non-native features in academic interlanguage.

Design Criteria for CALE-TR

CALE-TR is planned to comprise discipline and genre-specific texts. It is, therefore, regarded as an EAP corpus (Callies & Zaytseva, 2013). It includes seven academic text types written by the EFL students doing their MA and PhD studies. Figure 1 illustrates the text types making up CALE-TR.



Figure 1. Major text types in CALE-TR

The texts are discipline-specific as the topics are related to the fields of Linguistics and Applied Linguistics (Callies, 2015). In order to ensure that collected texts are in similar form, the textual and linguistic features for each text types were predefined by following the characteristics identified for academic prose (Biber & Conrad, 2009). Accordingly, each text is defined in terms of its general purpose, its specific purpose, author's skills and stance. Structural features, length and functional features were also predetermined for each text type and included in the guidelines which were given to the students before asking them to submit their writings. Following is an example illustrating the features of reading report.

Table 2. Profile for the “reading report” in CALE-TR

Purpose	Features
<p>General purpose: Informational-to inform and explain</p> <p>Specific purpose: (1) To understand the structure, content and line of argumentation of a scholarly research paper. (2) To summarize a text coherently</p> <p>Skills: The author demonstrates ability to capture general line of argumentation in a scholarly research and present the main points, hypotheses and methodology in an effective way.</p> <p>Stance: Author’s evaluation/absent</p>	<p>Structural properties: structured into covert sections such as topic, research questions, hypotheses etc.</p> <p>Length: 600 words</p> <p>Functional Properties: making it clear to an outsider what the paper is about.</p>

Each text submitted by the students is accompanied with two metadata files; one is a questionnaire on foreign language background of the student writer and the other is the questionnaire about the text with informed consent form. Each student writer is required to fill out the metadata files while submitting their texts to CALE-TR. This enables researchers to conduct detailed comparative analyses on the data provided by the students.

To summarize, it is clear that CALE has a great potential in informing the field of learner corpus research with its distinguishing focus on advanced academic language used by learners. With this potential, CALE and its sub-components could be exploited to investigate a wide range of aspects of learner academic writing. Exploring, for instance, rhetorical structures, collocational patterns and lexicogrammatical variations in general in learner English could inform the instructional practices and lead up to better “teaching approaches, manuals and syllabuses” (Vyatkina & Boulton, 2017 p.6). In the words of Flowerdew (2015) “English is increasingly used as a lingua franca in business and academia; comparative studies on the language of proficient L2 users and L2 learners would thus be helpful” (p.481). CALE-TR could make a great source for the academic language of L2 learners.

Conclusion

This paper is an attempt to describe the intersection of language corpora and EAP with its special focus on CALE-TR as a new source of learner corpus. The corpus introduced here, CALE-TR, has potential uses in researching advance level English. It also provides a fruitful avenue for contrastive analysis of advance level written English due to its strictly defined and followed design criteria. The CALE-TR is not its final form yet. However, when the corpus compilation is complete and made available, it is thought that it could provide a sound basis for the researchers and practitioners in the field of EAP and SLA.

Notes on the contributor

Aysel Şahin Kızıl (Ph.D.) works as an Associate Professor at İzmir Bakırçay University. She holds an M.A degree in Applied Linguistics and a PhD in English Language Teaching. She has taught linguistics and applied linguistics at the undergraduate level, and theories of second language acquisition at the graduate level. Her current research interests include corpus linguistics and the use of ICT tools in language education.

Endnote

ⁱ CALE-TR as a research project (Project no: ISBF.15.01) is funded by the Commission of Scientific Research Projects, Fırat University. Data compilation and annotation process is still in progress.

References

- Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71–83.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20.
- Callies, M. (2015). Corpus of Academic Learner English (CALE). Retrieved December 24, 2017, from <https://blogs.uni-bremen.de/cale/corpus-design/>

- Callies, M., & Zaytseva, E. (2013). The Corpus of Academic Learner English (CALE). *Dutch Journal of Applied Linguistics*, 2(1), 126–132.
- Carkin, S. (2005). English for academic purposes. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 85–99). Mahwah, NJ: Lawrence Erlbaum Associates.
- Charles, M. (2013). English for academic purposes. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 137–153). Oxford: Wiley.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423.
- Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2), 202–224.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355–362.
- Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small Corpus Studies and ELT* (pp. 363–369). Amsterdam: Benjamins.
- Flowerdew, L. (2014). Learner corpus research in EAP : Some key issues and future pathways. *English Language and Linguistics*, 20(2), 43–60.
- Flowerdew, L. (2015). Learner corpora and language for academic and specific purposes. In G. Gilquin, S. Granger, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 465–484).
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327.
- Gilbert, J. (2013). English for academic purposes. In G. Motteram (Ed.), *Innovations in learning technologies for English language teaching* (pp. 117–144). London: British Council.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319–335.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41–61.
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast: Text based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.
- Granger, S. (Ed.). (1998). *Learner English on computer*. Edinburgh: Addison Wesley Longman Limited.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Connor & T. A. Upton (Eds.), *Applied Corpus Linguistics. A Multidimensional Perspective* (pp. 123–146). Amsterdam: Rodopi.
- Granger, S., & Paquot, M. (2011). Language for specific purposes learner corpora. In U. Connor & C. A. Chapelle (Eds.), *Language for Specific Purposes. The Encyclopedia of Applied Linguistics* (pp. 1–5). Oxford: Blackwell- Wiley.

- Hamp-Lyons, L. (2001). English for academic purposes. In R. Carter & D. Nunan (Eds.), *The Cambridge Guide to Teaching English to Speakers of Other Languages* (pp. 126–130).
- Harwood, N., & Petric, B. (2011). English for academic purposes. In J. Simpson (Ed.), *The Routledge Handbook of Applied Linguistics* (pp. 243–258). London: Routledge.
- Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah: Lawrence Erlbaum.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. New York: Routledge.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21.
- Kilimci, A. (2014). LINDSEI-TR: A new spoken corpus of advanced learners of English. *International Journal of Social Science and Education*, 4(2), 401–410.
- Nesi, H. (2016). Corpus studies in EAP. In K. Hyland & Philip Shaw (Eds.), *The Routledge Handbook of English for Academic Purposes* (pp. 206–217). New York: Routledge.
- Paquot, M. (2007). Corpora and ICT in language studies : PALC 2005. In J. Walinski, K. Kredens, & S. Gozdz-Roszkowski (Eds.), *Corpora and ICT in language studies* (pp. 127–140). Berlin: Peter Lang
- Paquot, M. (2010). *Academic vocabulary in learner writing from extraction to analysis*. London: Continuum.
- Römer, U., & O'Donnel, 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.
- Sinclair, J. (2005). Corpus and text: Basic principles. In Martin Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books.
- Swales, J. M. (1988). *Episodes in ESP*. Hemel Hempstead: Prentice Hall International.
- Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, 21(3), 1–8.